Radial basis function network for ODEs: application to diabetes and insulin therapy models

DARIO ANDRINI

Index

- ANN (artificial neural network)→RBFN (radial basis function network)
- Simultaneous approximation of a function and its derivatives
- Solution of ordinary differential equations of arbitrary order
- Extension of RBF theory to vector valued case
- Application to biological models (β-cell cycle)

Artificial neural networks-architecture

- Fundamental units called NEURONS
  - Input units
  - Hidden units
  - Output units
- NEURONS are connected to each other ad connection points are called SYNAPSES
- Each neuron TOTAL INPUT signal is given by the weighted sum of all the input arriving at the neuron (synapses AMPLIFY or ATTENUATE the input signal)
- Each neuron OUTPUT signal is ruled by a function, called ACTIVATION FUNCTION, whose arguments are the input SIGNALS to the neuron

Artificial neural networks-architecture

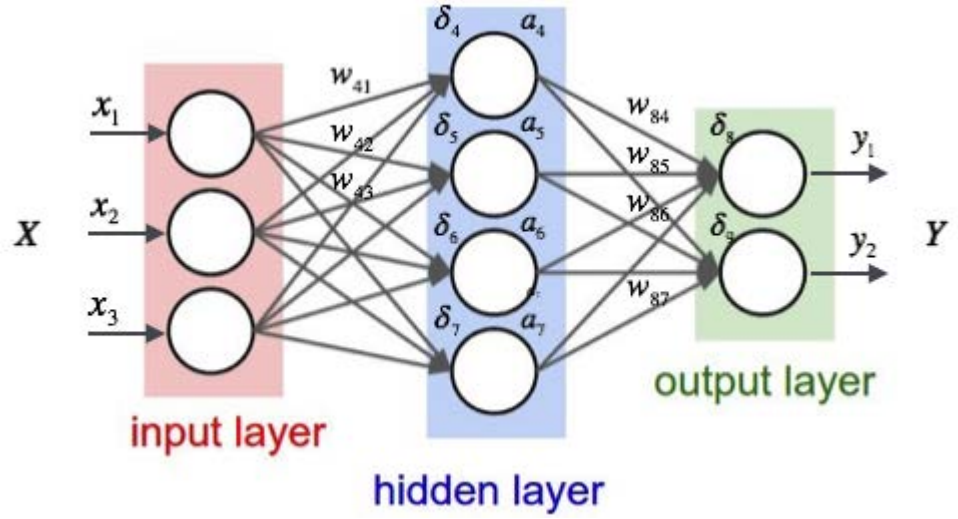
ACTIVATION FUNCTION (Φ):

$\boldsymbol{x} = \{x_1, x_2, \dots, x_N\}$ input vector

$\boldsymbol{y} = \{y_1, y_2, \dots, y_n\}$ output vector

$\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ threshold vector

ω_{ij} weights matrix

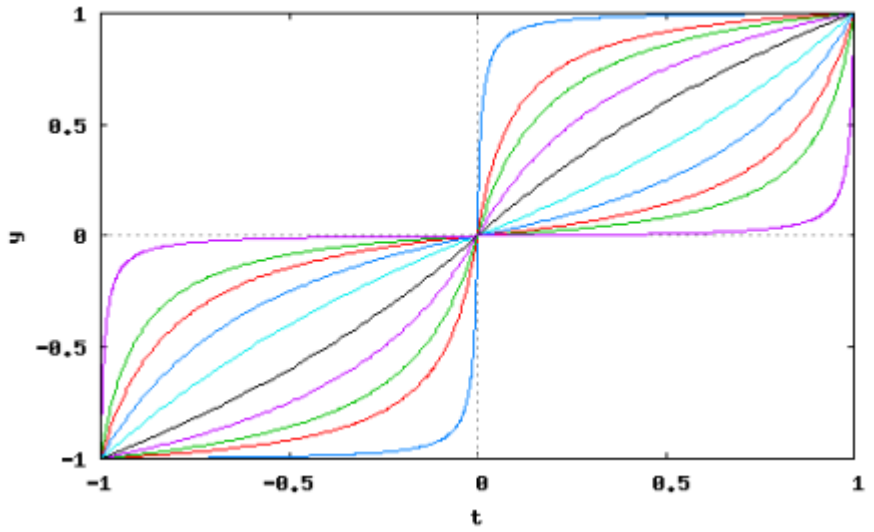


Response of the i-th neuron:

$$y_i = \Phi(\sum_{j=1}^{N} \omega_{ij} x_j - \theta_i)$$

Artificial neural networks-activation function

- Staircase function: $\Phi(A) = \begin{cases} 1 & if\ A > \theta \\ 0 & otherwise \end{cases}$
- Linear function: $\Phi(A) = kA$
- Sigmoid function: $\Phi(A) = \frac{1}{1+e^{-kA}}$



Artificial neural networks-learning phase

Learning is ruled by the choice of WEIGHTS:

- SUPERVISED LEARNING: according to the learning-samples, we choose the weights such that a measure of the distance between the output and the exact solution is minimized (→ an exact solution must exists and be known: THE LEARNING SAMPLES)

- UNSUPERVISED LEARNING: according to the learning-samples, we choose the weights such that certain desired conditions on the output are satisfied.

Artificial neural networks-main features

- ROBUSTNESS: it gives good results even in the case of noisy input or in the case some connections are damaged or destroyed
- FLEXBILITY: in the case of unsupervised learning there's no need to know the exact solution

Radial neural networks (RBFN)

- Only 1 hidden layer:
  - From INPUT layer to HIDDEN layer → radial activation function Φ
  - From HIDDEN layer to OUTPUT layer → linear transformation A

Radial neural networks (RBFN)

At the end, since the output layer is LINEAR the RBFN can be considered as a function s:

$$s(x): \mathbb{R}^m \to \mathbb{R}$$

$$s(x) = \sum_{j=1}^{N} w_j \varphi\left(\left\|x - \xi_j\right\|\right)$$

where:

N=number of neurons in the hidden layer

$\Phi := \varphi(\|\cdot\|_2)$

w_j=weight of neuron j in the linear output neurons

ξ_j=j-th centre

Radial neural networks (RBFN)

In the case of supervised learning the learning samples concretise in the imposition of the following conditions:

$$s(x_i) = d_i \qquad i = 1 \dots N$$

Given the set of N distinct points of data sites in $\mathbb{R}^m$:

$$\{x_i \in \mathbb{R}^m \mid i = 1 \dots N\}$$

Thus we are led to solve an interpolation problem → no need for a back propagation algorithm (and convergence issues related to it)

NB: if instead of N we have $i = 1 \dots d \geq N$ we are led to solve an APPROXIMATION PROBLEM

Simultaneous approximation of a function and its derivatives-DRBFN

DIRECT APPROACH:

$$f: \Omega \to \mathbb{R}^s$$

$$f(\tilde{x}) \approx \sum_{j=1}^{N} w_j \varphi(\|\tilde{x} - \xi_j\|)$$

$$\frac{\partial^k f}{\partial x_j \ldots \partial x_l}(\tilde{x}) \approx \sum_{i=1}^{N} w_i \frac{\partial^k \varphi}{\partial x_j \ldots \partial x_l}(\tilde{x})$$

This method is to sensitive to NOISE and don't give us a good approximation:

$$f(x) = \sum_k \hat{f}_k e^{ikx} \to f'(x) = \sum_k ik\hat{f}_k e^{ikx}$$

Simultaneous approximation of a function and its derivatives-IRBFN

INDIRECT APPROACH:

$$f^{(n)}(x) \approx \sum_{i=1}^{m} w_i \Phi_i(x)$$

$$f^{(n-1)}(x) = \int f^{n-1}(x)dx \approx \sum_{i=1}^{m} w_i \mathrm{H}_i^1(x) + C_1$$

$$\vdots$$

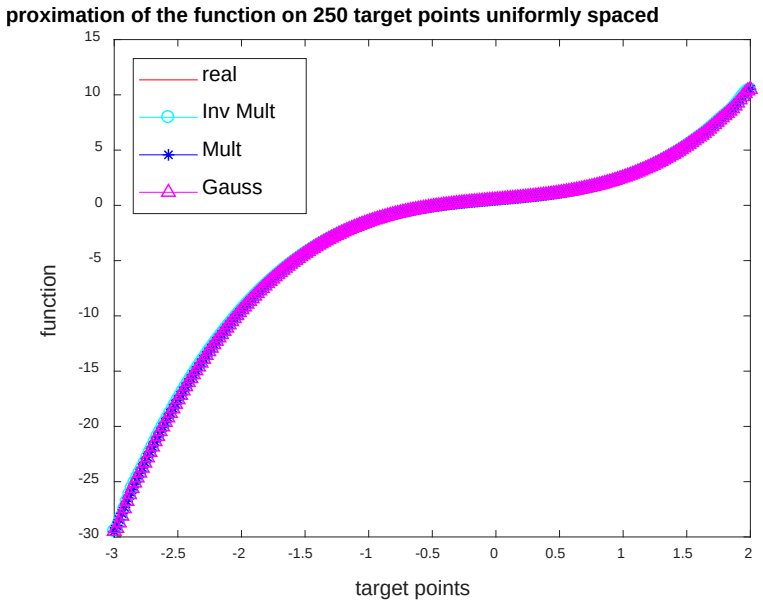
$$f(x) = \int f^{(1)}(x)dx \approx \sum_{i=1}^{m} w_i \mathrm{H}_i^{\mathrm{n}}(x) + C_1 x^{n-1} + C_2 x^{n-2} + \cdots + C_n$$

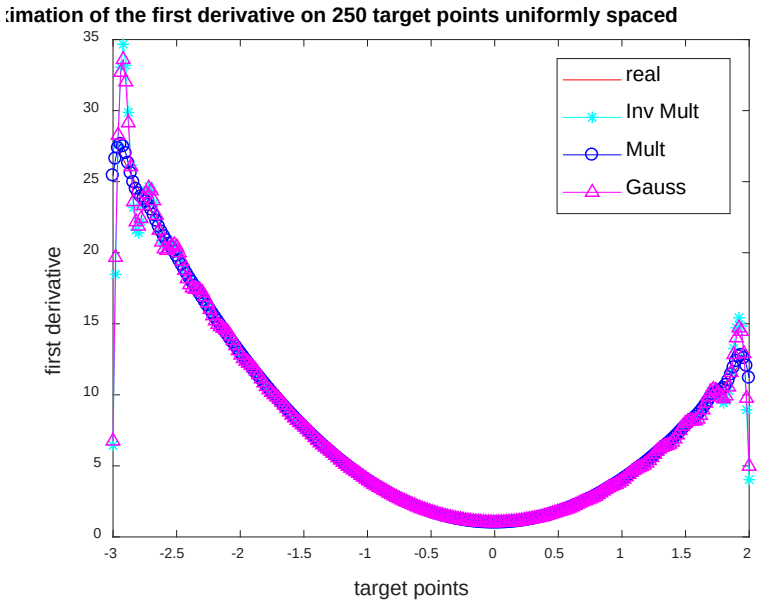
This leads to a non square system since we need to estimate the constantsC_i's. thus we need:

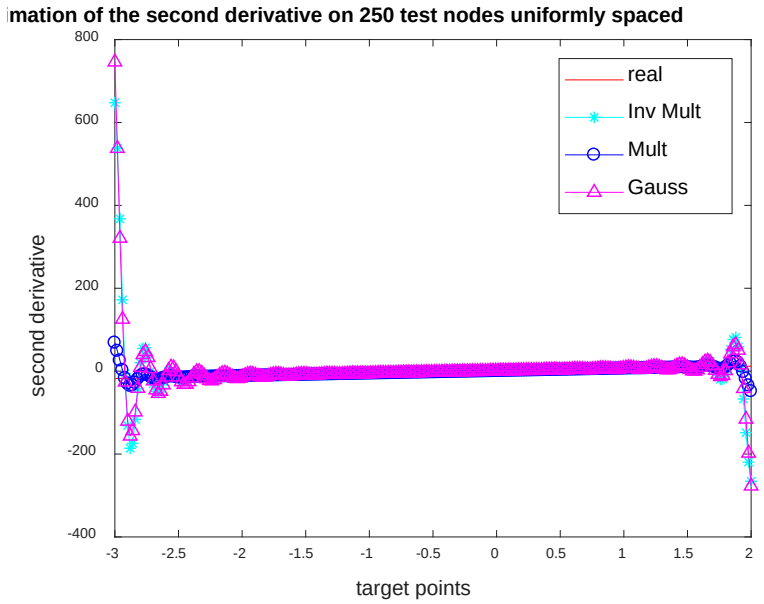
$$\#centers = \#datasites - n$$

Numerical experiments-DRBFN

$$y(x) = x^3 + x + 0.5$$

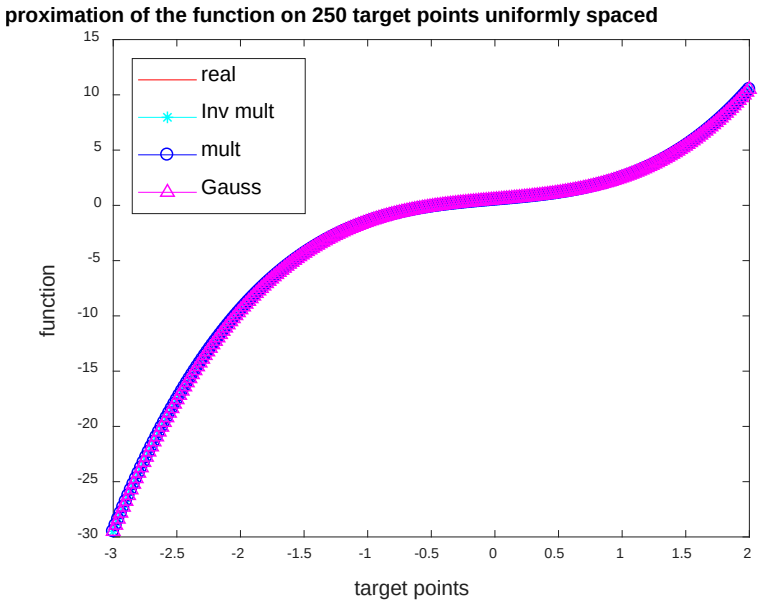


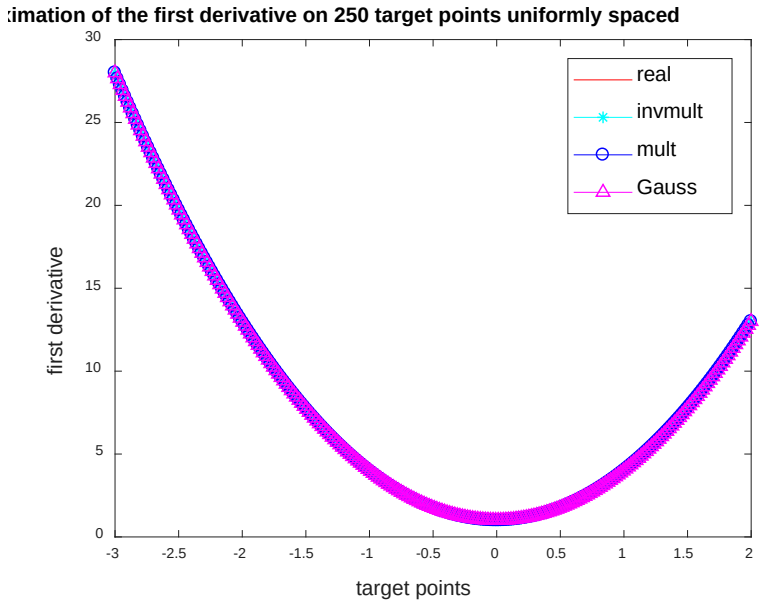


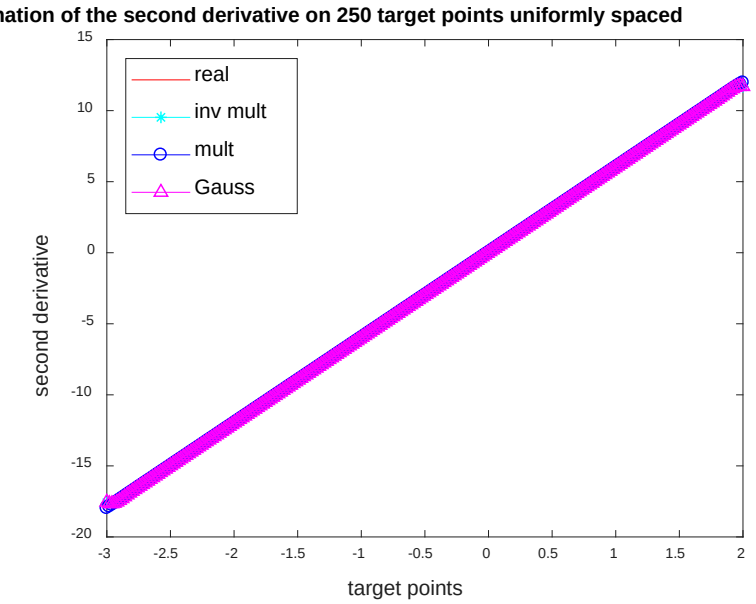


Approximation of derivatives up to order 2 with uniformly distributed points

Numerical experiments-IRBFN



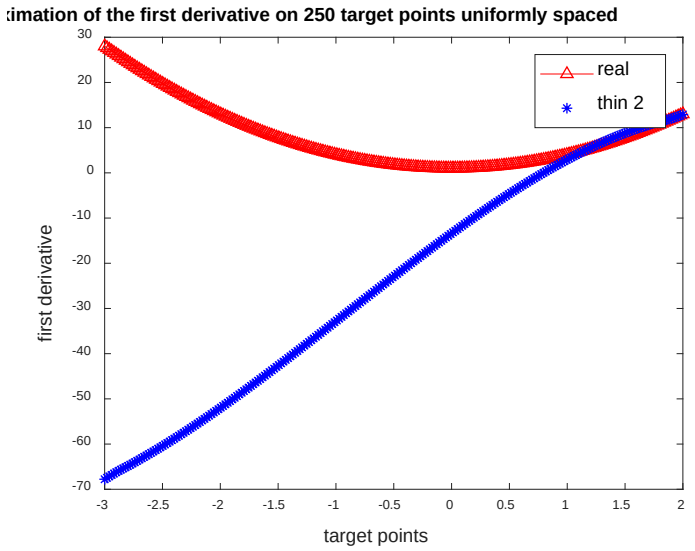


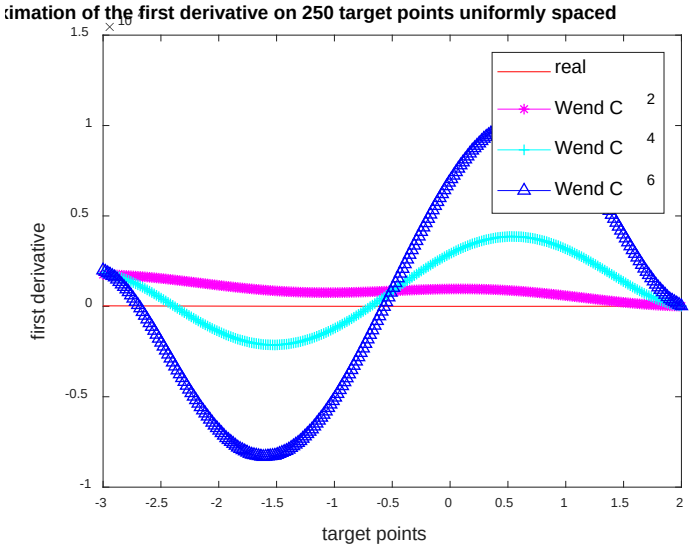


Approximation of derivatives up to order 2 with uniformly distributed points

Numerical experiments-IRBFN

Only trough the double application of the differential operator we obtain a function that is still in the Wendland native space. The same is true for also for thin plate splines. Thus only even derivatives can be well approximated





Localization properties

Localization: we want to approximate $f: D \subset \mathbb{R}^n \to \mathbb{R}$ with the approximant $s(x)$. If D_0 is any part of D then the contribution to $s(x)$ from $\{f(y): y \in D_0\}$ is small for values of x in D that are far from D_0

Example: $s(x) = \sum_{y \in \mathbb{Z}^n} f(y)\psi(x-y)$ and $\psi(x) = \sum_{k=1}^{m} \mu_k \varphi(\|x - x_k\|_2)$. We search in the space spanned by $\varphi(\|\cdot - x_k\|_2)$→ In 1-d we search for μ_k s.t $\psi(\cdot - y)$ are cardinal.

Minimum localization conditions(MLC) : $i) \int_{\mathbb{R}^n} |\psi(x)| dx < \infty$ and $ii) \int_{\mathbb{R}^n} \psi(x)dx = 1$

Example: for multiquadrics in the case $n = 1$ conditions on the coefficient μ_k exist s.t the MLC for each choice of the shape parameter (and these conditions are independent of the shape parameter)→MULTIQUADRICS HAVE LOWER ERROR

If $\varphi(r) \to 0$ as $r \to \infty$ if we guarantee $i)$ we have: $\int_{\mathbb{R}^n} \psi(x)dx = 0$→problem with constant functions

Network existence theorem

Theorem: suppose that a univariate function φ is analytic in $(-r, r)$, where $r > 0$ and for some integer $p \geq 1$,$\varphi^{(pj)}(0) \neq 0$ for $j \geq 0$. Then, if K is a compact subset of $\mathbb{R}^s$, for any $f \in \bar{C}^{(m1,\dots,m_q)}(K)$, and any $\varepsilon > 0$, there is a RBFN:

$$s(x) = \sum_{j=1}^{N} w_j \varphi(\lambda_j \|x - \xi_j\|)$$

For some suitable $\xi_j \in \mathbb{R}^s$, c_j and $\lambda_j \in \mathbb{R}$, where $j = 1 \dots N$, such that :

$$\left\|D^k f - D^k s\right\|_{L_\infty(K)} < \varepsilon$$

For any $k \in J(m1, \dots, m_q)$.

Observation: by density argument we can extend the result to L_p spaces (“UNIVERSAL APPROXIMANT”)

Solution of ordinary differential equations of arbitrary order

Let us consider the following example:

$$x^4 y^{(4)} + 4x^3 y^{(3)} + x^2(12 - x^2)y^{(2)} + 2x(x^2 - 12)y^{(1)} + 2(12 - x^2)y = 2x^5$$

In the interval $[1,11]$ with the boundary conditions:

$$\begin{cases} y(1) = A \\ y(11) = B \\ y'(1) = C \\ y'(11) = D \end{cases}$$

$A = 1 + e + \frac{1}{e}$, $\mathrm{B} = 1463 + 11(e^{11} + e^{-11})$, $\mathrm{C} = 2e$, $\mathrm{D} = -340 + 2e^{11}$

The exact solution is :

$$y(x) = x + x^2 + x^3 + xe^x + xe^{-x}$$

Solution of ordinary differential equations of arbitrary order

INDIRECT METHOD (for direct method is completely analogous):

With a little abuse of notation we write, substituting the expressions for the function and its derivatives in to the ODE:

$$x^4 \sum_{(i=1)}^{N} w_i \Phi_i(x) + 4x^3 \sum_{(i=1)}^{N+1} w_i \mathrm{H}_i^1(x) + x^2(12 - x^2) \sum_{(i=1)}^{N+2} w_i \mathrm{H}_i^2(x)$$

$$+2x(x^2 - 12) \sum_{(i=1)}^{N+3} w_i \mathrm{H}_i^3(x) + 2(12 - x^2) \sum_{(i=1)}^{N+4} w_i \mathrm{H}_i^4(x) = 2x^5$$

$$\begin{cases} \sum_{(i=1)}^{N} w_i \Phi_i(1) = A \\ \sum_{(i=1)}^{N} w_i \Phi_i\,(11) = B \\ \sum_{(i=1)}^{N+1} w_i H_i^1\,(1) = C \\ \sum_{(i=1)}^{N+1} w_i \mathrm{H}_i^1(11) = D \end{cases}$$

Solution of ordinary differential equations of arbitrary order

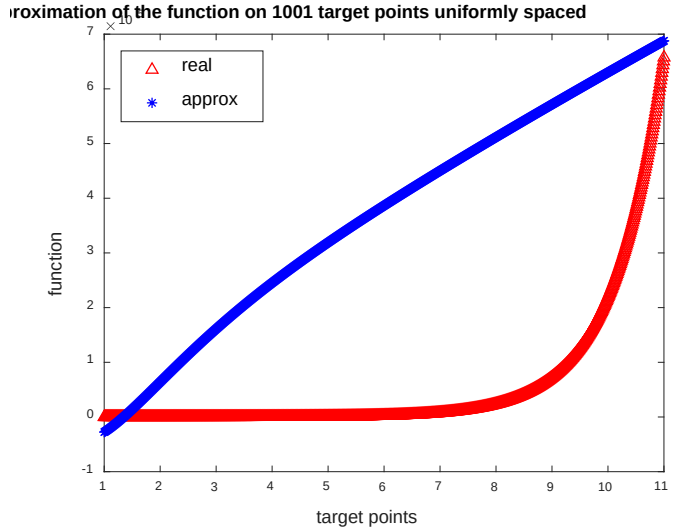
By “collocating” in the point of the discretization of the domain we obtain a linear system of the form:

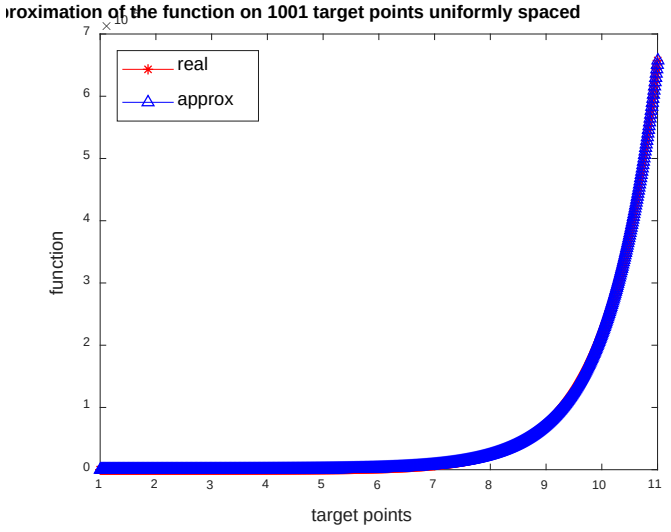
$$Gw = \begin{pmatrix} G1 \\ cond1 \\ cond2 \\ cond3 \\ cond4 \end{pmatrix} w = F$$

$$G1 = x^4\mathcal{F} + 4x^3\mathcal{H}_1 + x^2(12 - x^2)\mathcal{H}_2 + 2x(x^2 - 12)\mathcal{H}_3 + 2(12 - x^2)\mathcal{H}_4$$

Numerical experiments

DRBFN: $cond \sim 10^8$
IRBFN: $cond \sim 10^{13}$





Numerical solution DRBFN (left) and IRBFN (right) with multiquadrics

Conditioning number

SCALING THE DOMAIN: $D = [1, 11] \rightarrow D = \left[\frac{1}{11}, 1\right]$ ⟹ $IRBFNcond \sim 10^8$ but little lost of accuracy

RILEY METHOD: want to solve $Ax = b$, A ill-conditioned

Perturbed matrix: $C = A + \mu I$

Solve: $Cy = b$

$$A = C - \mu I \rightarrow \mathrm{A}^{-1} = \frac{1}{\mu}\sum_{k=1}^{\infty}(\mu C^{-1})^k$$

$$x = A^{-1}b = \frac{1}{\mu}\sum_{k=1}^{\infty}(\mu C^{-1})^k\, b = \frac{1}{\mu}\sum_{k=1}^{\infty}(\mu C^{-1})^{k-1}y = y + (\mu C^{-1})y + (\mu C^{-1})^2 y + \cdots$$

$$x_{k+1} = x_k + (\mu C^{-1})^k y$$

Conditioning number

By a trial and error approach varying μ we see that the value of μ which guarantees the best error does not improve the conditioning number

Better to use SCALING OF THE DOMAIN technique

RKHS for vector valued functions

Definition: let $\mathcal{Y}$ be a real Hilbert space with inner product $(\cdot,\cdot)$, $\mathcal{X}$ a set and $\mathcal{H}$ an Hilbert space of functions defined on $\mathcal{X}$ with inner product $\langle\cdot,\cdot\rangle$. We say that $\mathcal{H}$ is a reproducing kernel Hilbert space if $\forall y \in \mathcal{Y}$ and $x \in \mathcal{X}$ we have that

$$f \to (y, f(x))$$

Is continuous.

by Riesz theorem $\exists\, K_x y \in \mathcal{H}$ s.t $\boxed{\big(y, f(x)\big) = \langle K_x y, f\rangle}$ (REPRODUCTION PROPERTY)

RKHS for vector valued functions

We introduce the linear operator $K(x,t): \mathcal{Y} \to \mathcal{Y}$ for every $x, t \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $K(x,t)y \coloneqq K_t y(x)$. The following properties are satisfied:

- $\forall y, z \in \mathcal{Y}$ we have that $(y, K(x,t)z) = \langle K_t z, K_x y \rangle$
- $K(x,t) = K(t,x)^*$
- $K(x,x) \in \mathcal{L}_+(\mathcal{Y})$
- $\forall m \in \mathbb{N}, \{x_j : j \in \mathbb{N}_m\} \subset \mathcal{X}, \{y_j : \in \mathbb{N}_m \subset \mathcal{Y}\} \rightarrow \sum_{j,l \in \mathbb{N}_m} (y_j, K(x_j, x_l) y_l) \geq 0$
- $\|f(x)\|_{\mathcal{Y}} \leq \|f\|_{\mathcal{H}} \|K(x,x)\|_{\mathcal{Y}}^{\frac{1}{2}}$

If $\mathcal{Y} = \mathbb{R}^n \rightarrow \big(K(x,t)\big)_{kl} = \langle K_x e_k, K_t e_l \rangle$, $n \times n$ matrix of scalar valued functions

Interpolation problem for vector valued functions-general framework

Let $f_1 \dots f_n : \mathbb{R}^s \to \mathbb{R}$ mapping observed at sampling points $X_i = \{x_{ir} : r = 1 \dots N_i\}$ such that

$$f_i(x_{ir}) = d_{ir} \quad r = 1 \dots N_i, \quad i = 1 \dots n$$

We search for:

$$I_i(x) = \sum_{j=1}^{n} \sum_{r=1}^{N_i} u_{jr} \phi_{ij}(x, x_{jr}) + \sum_{l=1}^{Q} c_{il} p_l(x) \quad i = 1 \dots n$$

Subjected to side conditions:

$$\sum_{r=1}^{N_j} u_{jr} q(x_{jr}) = 0 \qquad \forall q \in \Pi_{k-1}^{s} \quad j = 1 \dots n$$

Interpolation problem for vector valued functions-general framework

$$(*)\underbrace{\begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,n} & P_1 & 0 & \cdots & 0 \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,n} & 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{n,1} & \psi_{n,2} & \cdots & \psi_{n,n} & 0 & 0 & \cdots & P_n \\ P_1^T & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & P_2^T & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_n^T & 0 & 0 & \cdots & 0 \end{pmatrix}}_{\mathcal{F}} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\psi_{ij} \in \mathbb{R}^{N_i \times N_j};\ \ P_i = \mathbb{R}_i^{N_i \times Q};\ \ \ u_i, d_i \in \mathbb{R}^{N_i};\ \ c_i \in \mathbb{R}^Q$$

Interpolation problem for vector valued functions-solvability of the system

Definition: an $n \times n$ matrix of kernels $\Phi = \left[\phi_{i,j}\right]_{i,j=1}^{n}$ with $\phi_{i,j}: \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}$ is said to be matrix conditionally positive definite of order k on$\mathbb{R}^s$ if for any n sets of distinct points X_i the matrix $\Psi = \left[\psi_{i,j}\right]_{i,j=1}^{n}$ where:

$$\psi_{i,j} = \left[\phi_{i,j}(x_{i,r}, x_{j,s})\right]_{r,s=1}^{N_i,N_j} \in \mathbb{R}^{N_i \times N_j} \quad 1 \leq i,j \leq n$$

Is conditionally positive semidefinite with respect to the subspace $V_{X_1,k} \times \cdots \times V_{X_n,k}$ where:

$$V_{X_i,k} \coloneqq \{\left[P\left(x_{r,i}\right)\right]_{r=1}^{n} : p \in \Pi_{k-1}^{s}, x_{ri} \in X_i\}$$

Proposition: if Φ is conditionally positive definite of order k and the X_i's are unisolvent for Π_{k-1}^{s} then $\mathcal{F}$ is invertible.

Interpolation problem for vector valued functions-construction of the kernel

Theorem: let $C \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix, ϕ a conditionally positive definite kernel of order k and at least one of C or ϕ be symmetric. Then:

$$\Psi := C\phi$$

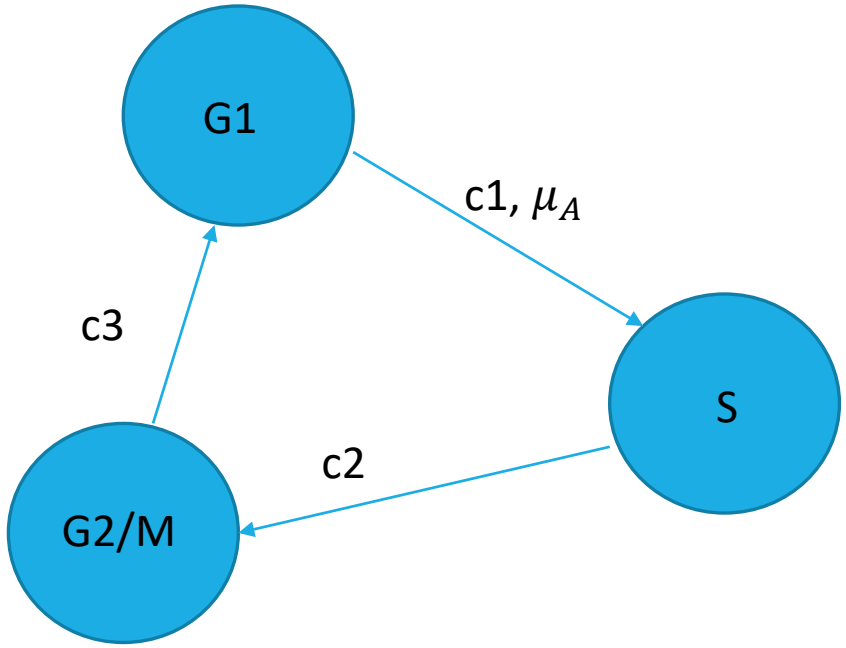
Is a matrix conditionally positive definite kernel of order k.

Solution of ODEs system

- Choose the matrix C (example: n=2→$C = \begin{pmatrix} 1 & \alpha \\ \beta & 1 \end{pmatrix}$)→PARAMETER ANALISYS
- Substitute the RBF interpolant in to the system and construct the matrix like $(*)$
- Multiply such a matrix by C
- Solve the linear system

Biological model

- β-cells are responsible for the production and the store of insulin → we want to model the β-cells cycle
- Three phases:
  - G1: the cells grow
  - S: DNA reduplicates
  - G2/M: mitosis



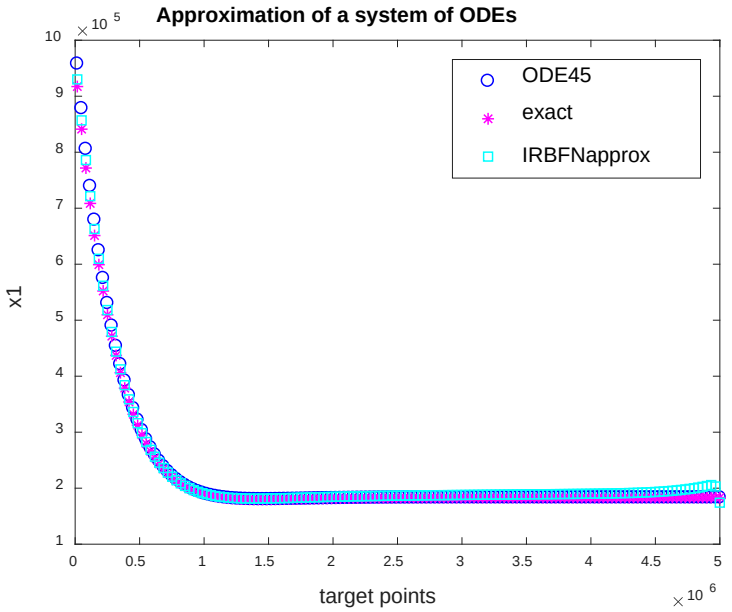
Biological model

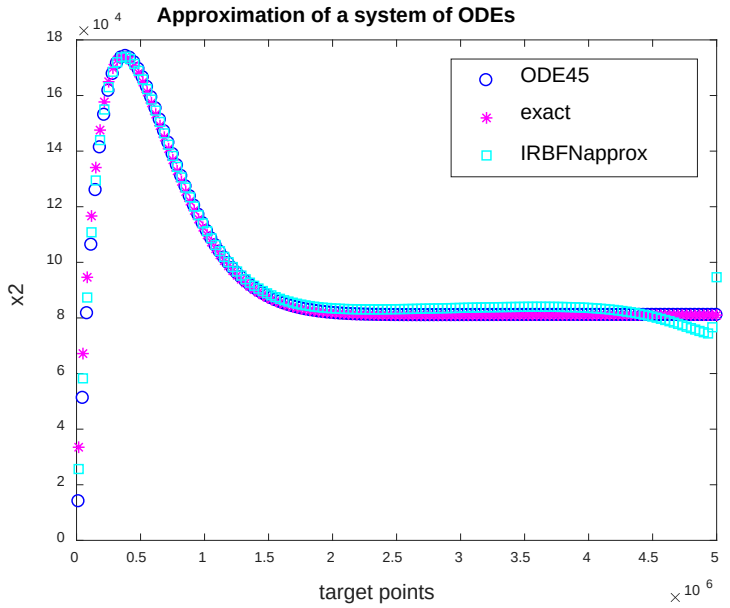
$$\begin{cases} \frac{dG_1(t)}{dt} = 2c_3 G_2 M(t) - (c1 + \mu_A) G_1(t) \\ \frac{dS(t)}{dt} = c_1^*(t) G_1(t) - c_2 S(t) \\ \frac{dG_2 M(t)}{dt} = c_2 S(t) - c_3 G_2 M(t) \end{cases}$$

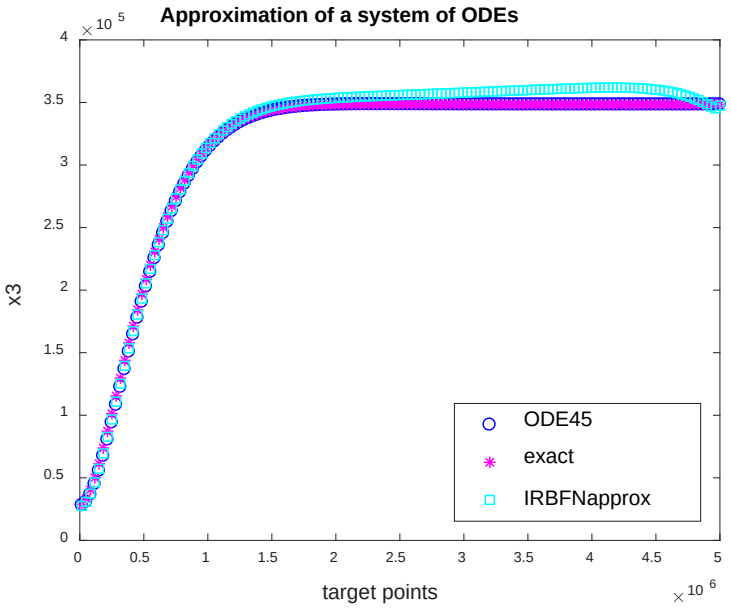
$$c_1^*(t) = c_1\big(1 + rI(t)\big)$$

- G_1, G_2M, S are densities of cells in the different phases
- $I(t)$ is the insulin concentration in the blood

Numerical solution







Numerical solution of the ODEs system using multiquadrics and 5×10^6 equally spaces interpolation points

Numerical solution

error	$x_1 - component$	$x_2 - component$	$x_3 - component$
IRBFN	2.8×10^{-2}	2.9×10^{-2}	2.5×10^{-2}
ODE45	3.5×10^{-2}	2.8×10^{-2}	7.1×10^{-3}

conclusions

- IRBFN approach is definitely better than DRBFN one
- Some kernels behave better because of
  - Better localization properties
  - Derivatives that belongs to the native space
- SCALING OF THE DOMAIN vs RILEY METHOD
- There exist a very natural extension of the concept of RKHS for vector valued functions

Solid theory for vector valued functions interpolations

Future directions

- Trying grids different from equally spaced points
- Changing method for parameter analysis for the determination of matrix C
- Trying different kernels from multiquadrics
- Solve the whole system of equations of the biological model

Thank you for your attention!