Greedy Kernel Methods for Center Manifold Approximation

Bernard Haasdonk, Boumediene Hamzi, Gabriele Santin, and Dominik Wittwar

1 Introduction

Center manifold theory plays an important role in the study of the stability of dynamical systems when the equilibrium point is not hyperbolic. It isolates the complicated asymptotic behavior by locating the center manifold which is an invariant manifold tangent to the subspace spanned by the eigenspace of eigenvalues on the imaginary axis. Then, the dynamics of the original system will be essentially determined by the restriction of this dynamics on the center manifold since the local dynamic behavior "transverse" to this invariant manifold is relatively simple as it corresponds to the flows in the local stable (and unstable) manifolds. In practice, one does not compute the center manifold and its dynamics exactly since this requires the resolution of a quasilinear partial differential equation which is not easily solvable. In most cases of interest, an approximation of degree two or three of the solution is sufficient. Then, the reduced dynamics on the center manifold can be determined, its stability can be studied and then conclusions about the stability of the original system can be obtained [1, 3, 4, 6, 8].

In this article, we use greedy kernel methods to construct a data-based approximation of the center manifold. The present work is a preliminary study that is intended to introduce our concept and algorithm, and to test it on some examples.

B. Haasdonk · G. Santin · D. Wittwar
Inst. of Applied Analysis and Numerical Simulation, University of Stuttgart, Stuttgart, Germany
e-mail: bernard.haasdonk@mathematik.uni-stuttgart.de;
gabriele.santin@mathematik.uni-stuttgart.de; dominik.wittwar@mathematik.uni-stuttgart.de

B. Hamzi (✉)
Department of Mathematics, Imperial College London, London, UK

2 Background

We consider a large dimensional dynamical system

$$\dot{x} = f(x), \quad x \in D, \tag{1}$$

where $f : D \to \mathbb{R}^n$ is a continuously differentiable function over the domain $D \subset \mathbb{R}^n$ such that $0 \in D$. We are interested in the study of the behavior of the system around an equilibrium point $x \in D$, i.e., $f(x) = 0$, possibly analyzing a smaller dimensional system.

Without loss of generality, we may assume that the equilibrium is $x = 0$, and, letting $L = \frac{\partial f}{\partial x}(x)|_{x=0}$, we can rewrite (1) as

$$\dot{x} = f(x) = Lx + N(x),$$

with a suitable nonlinear component N, and denote as $\sigma_{\mathbb{R}}(L)$ the set of real parts of the eigenvalues of L. A classical result relates the stability of the equilibrium with the spectrum of L, and in particular it is known that if L has all its eigenvalues with negative real parts, i.e., $\sigma_{\mathbb{R}}(L) \subset \mathbb{R}_{<0}$, then the origin is asymptotically stable, and if L has some eigenvalues with positive real parts, then the origin is unstable. If instead $\sigma_{\mathbb{R}}(L) \subset \mathbb{R}_{\leq 0}$, the linearization fails to determine the stability properties of the origin, and thus the analysis of this situation requires to employ additional tools.

In this case, we can first use a linear change of coordinates to separate the zero and the negative eigenvalues, i.e., we can rewrite (1) as

$$\dot{x} = L_1 x + N_1(x, y)$$
$$\dot{y} = L_2 y + N_2(x, y) \tag{2}$$

where $L_1 \in \mathbb{R}^{d \times d}$ is such that $\sigma_{\mathbb{R}}(L_1) = \{0\}$ and $L_2 \in \mathbb{R}^{m \times m}$ with $m := n - d$ is such that $\sigma_{\mathbb{R}}(L_2) \subset \mathbb{R}_{<0}$. The nonlinear functions $N_1 : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ and $N_2 : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$ are continuously differentiable. Intuitively, we expect the stability of the equilibrium to only depend on the nonlinear term $N_1(x, y)$. This intuition turns out to be correct, and indeed it can be properly formalized by means of the center manifold theorem.

We start by recalling a sufficient condition for the existence of a center manifold.

Theorem 1 ([1]) *If N_1 and N_2 are twice continuously differentiable and are such that*

$$N_i(0, 0) = 0, \quad \frac{\partial N_i}{\partial x}(0, 0) = 0, \quad \frac{N_i}{\partial y}(0, 0) = 0, \quad i = 1, 2,$$

and if the eigenvalues of L_1 have zero real parts, and all the eigenvalues of L_2 have negative real parts, then there exists a neighbourhood $\Omega \subset \mathbb{R}^d$ of the origin $0 \in \mathbb{R}^d$

and a center manifold $h : \Omega \to \mathbb{R}^m$ for (2), i.e., $y = h(x)$ is an invariant manifold for (2),[1] h is smooth, and

$$h(0) = 0, \quad Dh(0) = 0. \tag{3}$$

Under the assumptions of this theorem, using (2) we deduce that h satisfies the PDE

$$L_2 h(x) + N_2(x, h(x)) = Dh(x) \left(L_1 x + N_1(x, h(x)) \right), \tag{4}$$

and the following center manifold theorem ensures that there are smooth solutions to this PDE. Moreover, it also allows to deduce the stability of the origin of the full order system (2) from the stability of the origin of a reduced order system called the *center dynamics*.

Theorem 2 (Center Manifold Theorem [1]) *The equilibria $x = 0$, $y = 0$ of the original dynamics is locally asymptotically stable (resp. unstable) if and only if the equilibrium $x = 0$ of the center dynamics (dynamics on the center manifold)*

$$\dot{x} = L_1 x + N_1(x, h(x)), \tag{5}$$

is locally asymptotically stable (resp. unstable).

In particular, this result guarantees that, after solving the PDE (4), the problem of analyzing the stability properties of the system (2) reduces to analyzing the nonlinear stability of the lower dimensional system (5). This second problem is of smaller dimension and thus, provided the knowledge of h, the approach is attractive to obtain information on the system (1) via a reduced model.

Moreover, we remark that an exact knowledge of h is not required for this purpose, i.e., it is sufficient to have an approximate solution of the PDE (4). Indeed, it is frequently sufficient to compute only the low degree terms of the Taylor series expansion of h around $x = 0$, i.e., if $(\cdot)^{[k]}$ is the degree k part of the Taylor series of h, the approximation

$$h(x) \approx h^{[1]} x + h^{[2]}(x) + h^{[3]}(x) + \ldots + h^{[d-1]}(x) \tag{6}$$

is sufficient to obtain an approximation of the dynamics of order ε^d as $\|x\| \leq \varepsilon$. The approximation (6) can be obtained by coefficient comparison, thus rewriting

---

[1] A differentiable manifold $\mathcal{M}$ is said to be invariant under the flow of a vector field X if for $x \in \mathcal{M}$, $F_t(x) \in \mathcal{M}$ for small $t > 0$, where $F_t(x)$ is the flow of X.

the PDE (4) as a set of algebraic equations as

$$L_2 h^{[1]} = h^{[1]} L_1$$

$$L_2 h^{[2]}(x) + N_2^{[2]}(x, h^{[1]}(x)) = \frac{\partial h^{[2]}}{\partial x}(x) \left(L_1 x_1 + N_1^{[2]}(x, h^{[1]}(x)) \right)$$

$$L_2 h^{[3]}(x) + \left(N_2(x, h^{[2]}(x)) \right)^{[2]} = \frac{\partial h^{[2]}}{\partial x}(x) \left(L_1 x + \left(N_1(x, h^{[2]}(x)) \right)^{[2]} \right)$$

. . . .

We remark that this methodology is valid for parameterized dynamical systems and is used to study the stability of dynamical systems with bifurcations.

Nevertheless, even this approximated knowledge of h can be difficult to obtain in practice for a general ODE. To overcome this limitation, and since an approximated knowledge of the manifold is sufficient, our goal in this paper is to find a data-based approximation of the center manifold. This approximation is based solely on the knowledge of the splitting (1) and on the numerical computation of a set of trajectories of the system, and it provides an approximation of h which can be used to study the system stability.

3   Kernel Approximation

We want to build a surrogate model $s_h : \Omega \to \mathbb{R}^m$ which approximates the center manifold h on a suitable set $\Omega \subset \mathbb{R}^d$, in the sense that $s_h(x) \approx h(x)$ for all $x \in \Omega$. This model is constructed in a data-based way, i.e., we assume some knowledge of the map h on a finite set of input parameters, or training data. In practice, such values are computed from high-fidelity numerical approximations, which will be discussed in detail in the following.

The surrogate is based on kernel approximation, which allows the use of scattered data, i.e., we do not require any grid structure on the set of training data. Moreover, since the unknown function h is vector-valued, we employ here matrix-valued kernels. Details on kernel-based approximation can be found e.g. in [9], and the extension to the vectorial case is detailed e.g. in [5, 10]. We recall here only that a positive definite matrix-valued kernel on Ω is a function $K : \Omega \times \Omega \to \mathbb{R}^{m \times m}$ such that $K(x, y) = K(y, x)^T$ for all $x, y \in \Omega$ and $[K(x_i, x_j)]_{i,j=1}^N \in \mathbb{R}^{mN \times mN}$ is positive semidefinite for any set $\{x_1, \dots, x_N\} \subset \Omega$ of pairwise distinct points, for all $N \in \mathbb{N}$. Associated to a positive definite kernel there is a unique Hilbert space $\mathscr{H}$ of functions $\Omega \to \mathbb{R}^m$, named native space, where the kernel is reproducing, meaning that $K(\cdot, x)\alpha$ is the Riesz representer of the directional point evaluation $\delta_x^\alpha(f) := \alpha^T f(x)$, for all $\alpha \in \mathbb{R}^m$, $x \in \Omega$.

We consider here a twice continuously differentiable matrix-valued kernel k on Ω, and we use a specific functional formulation for our approximation and a specific

cost function, in order to construct a surrogate that is well suited for the particular approximation task.

In detail, the approximant takes the form

$$s_h(x) = \sum_{i=1}^{n_1} K(x, x_i^{(1)})\alpha_i + \sum_{j=1}^{n_2}\sum_{i=1}^{m} \partial_i^{(2)} K(x, x_j^{(2)})\beta_{i,j},$$

with centers $x_i^{(1)} \in X^{(1)} = \left\{x_1^{(1)}, \ldots, x_{n_1}^{(1)}\right\}$, $x_j^{(2)} \in X^{(2)} = \left\{x_1^{(2)}, \ldots, x_{n_2}^{(2)}\right\}$ and coefficient vectors $\alpha_i, \beta_{i,j} \in \mathbb{R}^m$. Here the superscript $\partial^{(2)}$ denotes that the derivative with regards to the second kernel component is taken.

Subsequently, we assume to have a sufficient amount of data $X_{N^*} = \{x_1, \ldots, x_{N^*}\}$ and $Y_{N^*} = \{y_1, \ldots, y_{N^*}\}$ which, for example, is generated by running a numerical scheme to compute discrete trajectories for different initial values (x_0, y_0). For this step, we need to assume that the variable splitting (2) is known in advance. Note that this is not a severe restriction, as for a general ODE (1) the required state transformation can be determined by eigenvalue decomposition of L.

Observe that we do not know if a data pair (x_i, y_i) lies on the center manifold, i.e. if $y_i = h(x_i)$ holds. We only know that the data converges asymptotically to the center manifold as $x_i \to 0$. Thus, an interpolation-based surrogate which merely interpolates the data on a given subset $X \subset X_{N^*}$ seems ill-suited for our purposes. Instead we consider another set of conditions to define the approximant. First, we still require the conditions in (3) to be satisfied by our approximation. Moreover, for the given subsets $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$, we compute our approximant by minimizing the following functional $J : \mathscr{H} \to \mathbb{R}$ under the constraint $s(0) = 0$, $Ds(0) = 0$:

$$J(s) := \|s\|_{\mathscr{H}}^2 + \sum_{i=1}^{N} (s(x_i) - y_i)^T \omega_i (s(x_i) - y_i). \tag{7}$$

Here $\omega_i \in \mathbb{R}^{m \times m}$ is a positive definite weight matrix. It can be shown that (7) has a unique minimizer s_h (see [11]). In particular s_h and its derivative Ds_h have the form

$$s_h(x) = \sum_{i=1}^{N+1} K(x, x_i)\alpha_i + \sum_{i=1}^{m} \partial_i^{(2)} K(x, 0)\beta_i, \tag{8}$$

$$Ds_h(x) = \sum_{i=1}^{N+1} D^{(1)} K(x, x_i)\alpha_i + \sum_{i=1}^{m} D^{(1)} \partial_i^{(2)} K(x, 0)\beta_i,$$

where we set $x_{N+1} := 0$. The coefficient vectors α_i, β_i can be computed by solving the system

$$\begin{pmatrix} A + W & B \\ B^T & C \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} Y \\ Z \end{pmatrix}, \tag{9}$$

with

$$A := \left(K(x_i, x_j) \right)_{i,j} \in \mathbb{R}^{m(N+1) \times m(N+1)},$$

$$W := \operatorname{diag} \left(\omega_1^{-1}, \dots, \omega_N^{-1}, 0 \right) \in \mathbb{R}^{m(N+1) \times m(N+1)},$$

$$B := \left(\partial_j^{(2)} K(x_i, 0) \right)_{i,j} \in \mathbb{R}^{m(N+1) \times m^2},$$

$$C := \left(\partial_i^{(1)} \partial_j^{(2)} k(0, 0) \right)_{i,j} \in \mathbb{R}^{m^2 \times m^2},$$

$$Y := (y_1^T, \dots, y_n^T, 0)^T \in \mathbb{R}^{m(N+1)},$$

$$Z := 0 \in \mathbb{R}^{m^2 \times m}.$$

The weight matrices ω_i can either be chosen manually, or a regularizing function $r : \Omega \to \mathbb{R}^{m \times m}$ can be prescribed such that $\omega_i = r(x_i)$ is symmetric and positive definite. In our numerical examples in Sect. 4 we chose a constant regularization function, i.e.

$$\omega_i = r(x_i) = \lambda I_m$$

for some $\lambda > 0$. However, one might consider a more general approach, where the weight increases as the data tends to the origin, i.e. $\omega_i \succeq \omega_j$ if $\|x_i\| \leq \|x_j\|$.

3.1   Greedy Approximation

If the technique of the previous section is used as it is, the surrogate (8) is given by an expansion with N^* terms, where N^* is the number of points in the training set. Therefore, the model evaluation might not be efficient enough if the model is built using a too large dataset. Furthermore, the computation of the coefficients in (8) requires the solution of the linear system (9), whose size again scales with the size of the training set, and which can be severely ill-conditioned for non well-placed points.

   To mitigate both problems, we employ an algorithm that aims at selecting small subsets X_N, Y_N of points such that the surrogate computed with these sets is a sufficiently good approximation of the one which uses the full sets. The algorithm

selects the points in a greedy way, i.e., one point at a time is selected and added to the current training set. In this way, it is possible to identify a good set without the need to solve a nearly infeasible combinatorial problem.

The selection is performed using the P-greedy method of [2] applied to the kernel K, such that the set of points is selected before the computation of the surrogate. The number of points, and therefore the expansion size and evaluation speed, is depending on a prescribed target accuracy $\varepsilon_{tol} > 0$. For details on the method implementation and its convergence properties we refer to [7].

4 Numerical Examples

We test now our method on three different examples. In each of them, we specify the setting and the parameters used to build the surrogate and visualize our approximation to the center manifold. Additionally, we compute the pointwise residual

$$r(x) = Ds_h(x)\left(L_1 x + N_1(x, s_h(x))\right) - \left(L_2 s_h(x) + N_2(x, s_h(x))\right),$$

which measures how well the surrogate s_h satisfies the ODE (4).

In all the three examples, the greedy algorithm is used to select a suitable subset of the points, and in all cases the procedure is stopped with a prescribed ε_{tol}. In the first two examples we set $\varepsilon_{tol} := 10^{-15}$, while $\varepsilon_{tol} := 10^{-10}$ is used in the last one.

4.1 Example 1

We consider the 2-dimensional system

$$\begin{aligned} \dot{x} &= L_1 x + N_1(x, y) = 0 + xy \\ \dot{y} &= L_2 y + N_2(x, y) = -y + x^2. \end{aligned} \tag{10}$$

We generate the training data by solving (10) with an implicit Euler scheme for initial time $t_0 = 0$, final time $T = 1000$ and with the time step $\Delta t = 0.1$. We initiate the numerical procedure with initial values $(x_0, y_0) \in \{\pm 0.8\} \times \{\pm 0.8\}$ and store the resulting data pairs in X and Y after discarding all data whose x-values are not contained in the neighborhood $[-0.1, 0.1]$ which results in $N^* = 38{,}248$ data pairs.

We run the greedy algorithm for the kernels $k_1(x, y) := \left(1 + xy/2\right)^4$ and $k_2(x, y) = e^{-(x-y)^2/2}$. This results in the sets X_1 and X_2 which contain 14 and 6 points, respectively. The corresponding approximations s_1 and s_2 for the constant

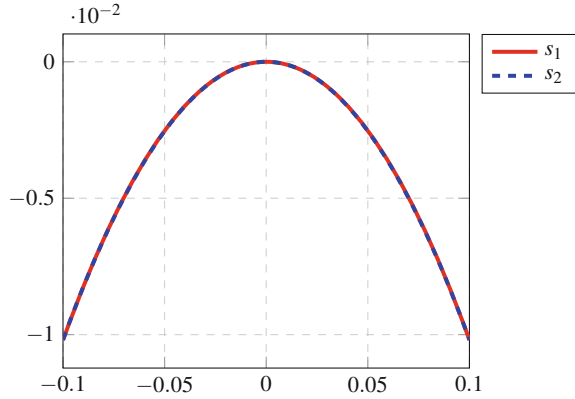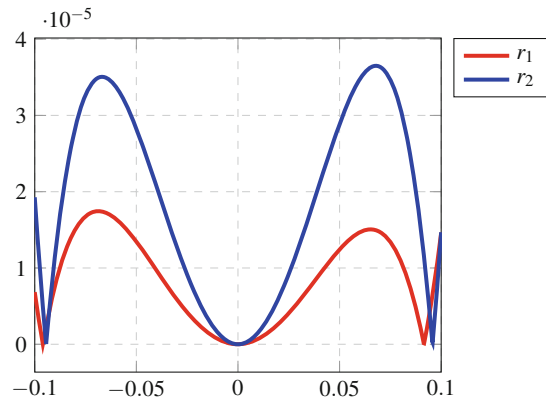**Fig. 1** Approximations s_1 and s_2 of the center manifold

Fig. 2 Residuals r_1 and r_2 of the center manifold

regularization function $r \equiv 10^{-10}$ are plotted in Fig. 1 over the domain $[-0.1, 0.1]$. The pointwise residual is depicted in Fig. 2.

4.2   Example 2

We consider the 2-dimensional system

$$
\begin{aligned}
\dot{x} &= L_1 x + N_1(x, y) = 0 - xy \\
\dot{y} &= L_2 y + N_2(x, y) = -y + x^2 - 2y^2.
\end{aligned}
\tag{11}
$$

The training data is generated the same way as in Example 1. We again use the kernels k_1 and k_2. The greedy algorithm gives sets X_1 and X_2 of size 12 and 6, respectively. The evaluation of the approximations s_1 and s_2 over the neighborhood

Fig. 3 Approximations s_1 and s_2 of the center manifold
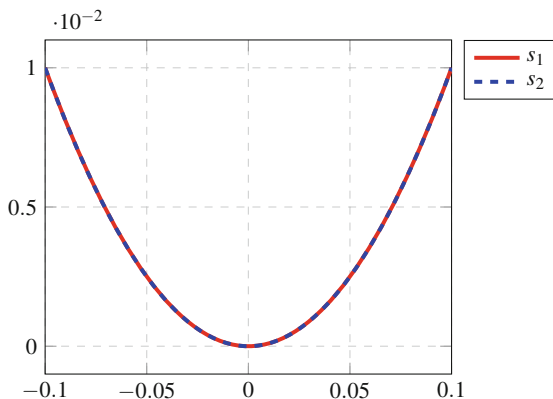
Fig. 4 Residuals r_1 and r_2 of the center manifold
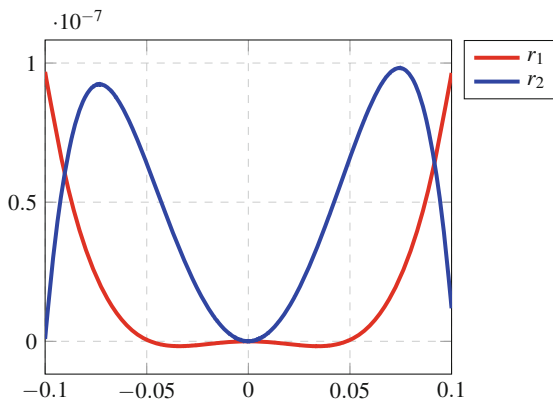
$[-0.1, 0.1]$ can be seen in Fig. 3, while the respective pointwise residuals are plotted in Fig. 4.

4.3 Example 3

We consider the $(2 + 1)$-dimensional system

$$\dot{x} = L_1 x + N_1(x, y) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} x_1 y \\ x_2 y \end{pmatrix}$$

$$\dot{y} = L_2 y + N_2(x, y) = -y - x_1^2 - x_2^2 + y^2. \tag{12}$$

We generate the training data in a similar fashion as before. We again use the implicit Euler scheme with start time $t = 0$, final time $T = 1000$ and with time step
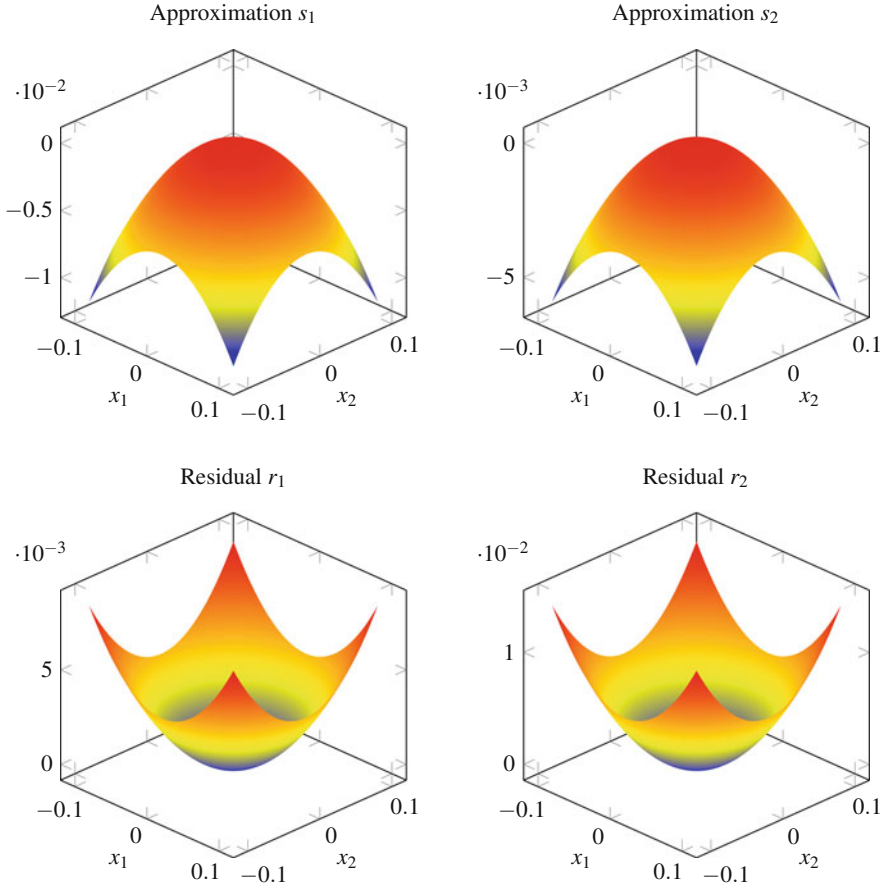
Fig. 5 Approximations s_1 and s_2 of the center manifold and corresponding residuals r_1 and r_2

$\Delta t = 0.1$. The Euler method is performed for initial data $(x_0, y_0) \in \{\pm 0.8\}^3$ and the resulting trajectories are stored in X and Y, where only data with $x \in [-0.1, 0.1]^2$ was considered; this leads to $N^* = 78,796$ data pairs. We use the kernels $k_1(x, y) = (1 + x^T y/2)^4$ and $k_2(x, y) = e^{-\|x-y\|_2^2/2}$, and the greedy-selected sets have the size 21 (for k_1) and 25 (for k_2), respectively. The approximations s_1, s_2 and their corresponding residuals r_1 and r_2 computed over the domain $[-0.1, 0.1]^2$. The results can be seen in Fig. 5.

We remark that in all the three experiments both kernels give comparable results in terms of error magnitude, and they both provide a good approximation of the manifold.

5 Conclusions

In this paper we introduced a novel algorithm to approximate the center manifold of a given ODE using a data-based surrogate.

This algorithm computes an approximation of the manifold from a set of numerical trajectories with different initial data. It is based on kernel methods, which allow the use of the scattered data generated by these simulations as training points. Moreover, an application-specific ansatz and cost function have been employed in order to enforce suitable properties on the surrogate.

Several numerical experiments suggested that the present method can reach a significant accuracy, and that it has the potential to be used as an effective model reduction technique. It seems promising to apply this approach to high dimensional systems as the approximation technique straightforwardly can be extended and is less prone to the curse of dimensionality than grid-based approximation techniques. An interesting extension would consist of determining the decomposition (2) in a data-based fashion by suitable processing of the trajectory data.

References

 1. Carr, J.: Applications of Centre Manifold Theory. Applied Mathematical Sciences, vol. 35. Springer, Berlin (1981)
 2. De Marchi, S., Schaback, R., Wendland, H.: Near-optimal data-independent point locations for radial basis function interpolation. Adv. Comput. Math. **23**(3), 317–330 (2005)
 3. Henry, D.: Geometric Theory of Semilinear Parabolic Equations. Lecture Notes in Mathematics, vol. 840. Springer, Berlin (1981)
 4. Kelley, A.: The stable, center-stable, center, center-unstable, unstable manifolds. J. Differ. Equ. **3**, 546–570 (1967)
 5. Micchelli, C.A., Pontil, M.: On learning vector-valued functions. Neural Comput. **17**(1), 177–204 (2005)
 6. Pliss, V.A.: A reduction principle in the theory of stability of motion. Izv. Akad. Nauk SSSR Ser. Mat. **28**, 1297–1324 (1964)
 7. Santin, G., Haasdonk, B.: Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. Dolomites Res. Notes Approx. **10**, 68–78 (2017)
 8. Shoshitaishvili, A.N.: Bifurcations of topological type of singular points of vector fields that depend on parameters. Funkcional. Anal. i Priložen. **6**(2), 97–98 (1972)
 9. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge (2005)
10. Wittwar, D., Santin, G., Haasdonk, B.: Interpolation with uncoupled separable matrix-valued kernels. Dolomites Res. Notes Approx. **11**, 23–29 (2018)
11. Wittwar, D., Santin, G., Haasdonk, B.: Weighted regularized interpolation with matrix valued kernels. Technical Report, University of Stuttgart, (2019, in preparation)