



HAL
open science

Data-driven Reachability using Christoffel Functions and Conformal Prediction

Abdelmouaiz Tebjou, Goran Frehse, Faïcel Chamroukhi

► **To cite this version:**

Abdelmouaiz Tebjou, Goran Frehse, Faïcel Chamroukhi. Data-driven Reachability using Christoffel Functions and Conformal Prediction. The 12th Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2023), Sep 2023, Limassol, Cyprus. hal-04095745

HAL Id: hal-04095745

<https://hal.science/hal-04095745>

Submitted on 16 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data-driven Reachability using Christoffel Functions and Conformal Prediction

Abdelmouaiz Tebjou

ABDELMOUAIZ.TEBJOU@IRT-SYSTEMX.FR

IRT SystemX, 2 boulevard Thomas Gobert, 91120 Palaiseau, France.

U2IS, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France.

Goran Frehse

GORAN.FREHSE@ENSTA-PARIS.FR

U2IS, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France.

Faïcel Chamroukhi

FAICEL.CHAMROUKHI@IRT-SYSTEMX.FR

IRT SystemX, 2 boulevard Thomas Gobert, 91120 Palaiseau, France.

Editor: Harris Papadopoulos, Khuong An Nguyen, Henrik Boström and Lars Carlsson

Abstract

An important mathematical tool in the analysis of dynamical systems is the approximation of the reach set, i.e., the set of states reachable after a given time from a given initial state. This set is difficult to compute for complex systems even if the system dynamics are known and given by a system of ordinary differential equations with known coefficients. In practice, parameters are often unknown and mathematical models difficult to obtain. Data-based approaches are promised to avoid these difficulties by estimating the reach set based on a sample of states. If a model is available, this training set can be obtained through numerical simulation. In the absence of a model, real-life observations can be used instead. A recently proposed approach for data-based reach set approximation uses Christoffel functions to approximate the reach set. Under certain assumptions, the approximation is guaranteed to converge to the true solution. In this paper, we improve upon these results by notably improving the sample efficiency and relaxing some of the assumptions by exploiting statistical guarantees from conformal prediction with training and calibration sets. In addition, we exploit an incremental way to compute the Christoffel function to avoid the calibration set while maintaining the statistical convergence guarantees. Furthermore, our approach is robust to outliers in the training and calibration set.

Keywords: data-driven reachability, Christoffel functions, conformal prediction, probably approximately correct analysis, statistical learning

1. Introduction

The problem of reach set approximation arises in different branches of applied mathematics and computer science, and in particular in control theory. In mathematics, the study of initial value problems and their guaranteed solution raises the question of which states can be reached under different configurations; see, for instance the work of [Berz and Makino \(1998\)](#). In computer science, the computation of reach sets is a fundamental operation in formal methods, which establish the correctness of a system with mathematical rigor. Initially, it was applied to program analysis, e.g., by [Halbwachs et al. \(1994\)](#). Later, the approach was extended to cyber-physical systems, which can involve interacting physical components, software, and communication channels, see [Alur \(2015\)](#). Reach set approximations may take different forms based on whether the focus is on scalability, tightness,

or efficient computability. Examples include polyhedra, ellipsoids, polynomial zonotopes, and others; see the overview by [Althoff et al. \(2021\)](#). In this paper, we establish reach set approximations that are sublevel sets of polynomials, more precisely, sum-of-squares (SOS) polynomials, which are computationally advantageous. Once established, these can readily be used to investigate properties of regions of attraction, stability, and safety or to solve optimization problems. To achieve this, polynomial reach set approximations have been used as barrier certificates, inductive invariants, or Lyapunov functions; see the survey by [Doyen et al. \(2018\)](#).

Traditionally, reach set approximations are established from first principles, starting from a mathematical model of the dynamics. This approach is limited to cases where sufficiently simple models are available and precise enough. More recently, data-based approaches have been used to deal with systems whose dynamics are too complex or where a model is not available and only observations are at hand. In the following, we provide a brief overview of such approaches.

Related Work The traditional approach to go from data to reach set approximations is to first identify a model of the system dynamics and then analyse the model. To give an example, a linear model can be identified efficiently by subspace identification as proposed by [Van Overschee and De Moor \(2012\)](#) and then one of the set-based techniques in the survey by [Althoff et al. \(2021\)](#) can be applied to approximate the reach set at a given time in the future. This can be extended to uncertain linear models and nonlinear systems based on linearization, as pursued by [Alanwar et al. \(2023\)](#). More recently, it has been proposed to derive reach set approximations more directly from data, e.g., the approach of [Djeumou et al. \(2021\)](#) uses Taylor series expansions and Lipschitz bounds to derive reach sets for nonlinear systems. These approaches can, in principle, bound the reach set over an arbitrary time horizon, but the approximation error may increase very rapidly with time. Furthermore, these approaches struggle with complex dynamics.

Our goal in this paper is different and more modest: We establish an SOS polynomial whose sublevel set contains the reachable set in the sense of a probably approximately correct (PAC) property. In particular, we consider the approximation of a single time step. This is sufficient for many of the applications considered above (as a first step in constructing barrier certificates, inductive invariants etc.), but in contrast to the approaches cited in the beginning of this section, it does not readily extend to extrapolating the reach set over longer time horizons (it would involve costly quantifier elimination).

One of the earliest data-driven approaches involving SOS polynomials was the construction of barrier certificates by [Prajna \(2006\)](#), e.g., to show that obstacles are avoided by a control system. The scalability was later improved by [Han et al. \(2015\)](#), but the optimisation problem remains somewhat challenging. Approximating the reach set is related to approximating the support of a probability measure, as observed by [Devonport et al. \(2021\)](#). Recent work by [Lasserre and Pauwels \(2019\)](#); [Lasserre \(2022\)](#) suggests that Christoffel functions are particularly useful for approximating the support. Our work is heavily inspired by [Devonport et al. \(2021\)](#), who proposed to approximate the one-step reach set with an SOS polynomial that is the superlevel set of the Christoffel function. The PAC guarantees provided by [Devonport et al. \(2021\)](#) are derived from measure theory and are, in practise, somewhat conservative. Based on conformal prediction, we propose significant improve-

ments that we outline below. Further work on conformal prediction will be cited in the text.

Contributions In this paper, we make the following contributions:

- We use conformal prediction to provide stronger and more sample-efficient guarantees on reach set approximation than those given by [Devonport et al. \(2021\)](#).
- We propose a version of reach set approximation that is robust to outliers, in contrast to the approach of [Devonport et al. \(2021\)](#).
- We exploit an incremental form of the Christoffel function for transductive conformal prediction, thanks to which we don't need to split the data set into training and calibration sets.
- To the best of our knowledge, this is the first use of the Christoffel function in conformal prediction. The particular properties of the Christoffel function in set and density approximation make it an excellent candidate for a nonconformity function.

Structure of the paper The paper is organized as follows. Section 2 presents the data-driven framework for reachability analysis using Christoffel functions. It describes the theoretical developments related to the reach set approximation and to Christoffel functions. In Section 3, we introduce our proposed approach to the reach set approximation with conformal prediction, whose statistical guarantees are presented in Section 3.1. Section 3.2 presents a technique to avoid the calibration set by using transductive conformal prediction and an incremental version of the Christoffel function. In Section 4, we discuss the robustness of our methodology to outliers. Section 5 provides numerical experiments on simulated data to support our theoretical results, and to highlight the effectiveness and potential of the proposed approach.

2. Data-driven Reach Set Approximation with Christoffel Functions

Reachability analysis aims to determine the possible future states of a dynamical system starting from a given initial state. For our purposes, we consider the system to be defined (explicitly or implicitly) by a transition function

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^n,$$

which maps a state $\mathbf{x} \in \mathbb{R}^n$ to its successor state. We forego extending the notation to nondeterministic or stochastic systems, since our focus is on estimating the image of f applied to a set of initial states; in the case of a stochastic system we are interested in approximating the support of the image distribution. Beginning with a given initial set of states \mathcal{I} , we are interested in computing the reachable set

$$\mathcal{S} = \{f(\mathbf{x}) : \mathbf{x} \in \mathcal{I}\}.$$

When f is not precisely known or complex, obtaining the exact solution may not be possible or economical. Instead, we compute an approximation $\hat{\mathcal{S}}$ that covers most of \mathcal{S} . Every set S can be represented by a probability measure μ such as S is the support of μ . This motivated [Devonport et al. \(2021\)](#) to use the Christoffel function to approximate the set \mathcal{S} . In the following subsection, we introduce the Christoffel function, its empirical counterpart, and discuss how to compute it.

2.1. Preliminaries

We start by introducing some mathematical notation. Given a vector $\mathbf{x} \in \mathbb{R}^n$, we denote its elements as $\mathbf{x} = (x_1, \dots, x_n)$. An integer coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ defines the monomial $\mathbf{x}^\alpha = x_1^{\alpha_1} \times x_2^{\alpha_2} \dots \times x_n^{\alpha_n}$. For $d \in \mathbb{N}$, we consider $\mathbb{R}[\mathbf{X}]_d^n$ to be the vector space of n -variate polynomials whose degree is less or equal to d . With each coefficient vector $\boldsymbol{\alpha} \in \mathbb{N}^n$, we associate the monomial \mathbf{x}^α whose degree is equal to $\|\boldsymbol{\alpha}\| = \sum_{i=1}^n \alpha_i$. The monomials \mathbf{x}^α with $\|\boldsymbol{\alpha}\| \leq d$ form a canonical basis of $\mathbb{R}[\mathbf{X}]_d^n$. We denote the number of monomials of degree less or equal to d with

$$s(d) = \binom{n+d}{n}.$$

Let $\mathbf{v}_d(\mathbf{x}) \in \mathbb{R}^{s(d)}$ be the vector of monomials of degree less or equal to d evaluated at \mathbf{x} . For example, if $d = 2$ and $n = 2$, then $\mathbf{v}_d(\mathbf{x}) = [1 \ x_1 \ x_2 \ x_1x_2 \ x_1^2 \ x_2^2]$.

2.2. Christoffel Functions

Christoffel functions are a class of functions associated with a finite measure and a parameter degree $d \in \mathbb{N}$. They have a strong connection to approximation theory and in this section we briefly summarize some results by [Lasserre and Pauwels \(2019\)](#). For a finite measure μ on \mathbb{R}^n and an integer degree d , the Christoffel function $\Lambda_{\mu,d}(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ is defined in terms of the moment matrix of the measure μ :

$$\mathbf{M}_d = \int_{\mathbb{R}^n} \mathbf{v}_d(\mathbf{x}) \mathbf{v}_d(\mathbf{x})^\top d\mu(\mathbf{x}).$$

The moment matrix is semi-definite positive for all $d \in \mathbb{N}$. We furthermore assume that the matrix is positive definite, which ensures the invertibility of M_d .¹ With the help of the moment matrix, the Christoffel function is defined as :

$$\Lambda_{\mu,d}(\mathbf{x}) = \left(\mathbf{v}_d(\mathbf{x})^\top \mathbf{M}_d^{-1} \mathbf{v}_d(\mathbf{x}) \right)^{-1}. \quad (1)$$

The following alternative formulation of the Christoffel function can be useful when the moment matrix is large. It can be computed by solving a convex quadratic programming problem, which can be done efficiently using numerical techniques, even for high degrees d :

$$\Lambda_{\mu,d}(\mathbf{x}) = \inf_{P \in \mathbb{R}[\mathbf{X}]_d^n} \left\{ \int_{\mathbb{R}^n} P(\mathbf{z})^2 d\mu(\mathbf{z}), \quad \text{s.t.} \quad P(\mathbf{x}) = 1 \right\}$$

In a data-driven setting, the exact measure μ is unknown. One way to obtain information about μ is by sampling a set of points independently drawn from its distribution. For every $N \in \mathbb{N}$, when disposing of N i.i.d samples $\{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ from μ , we approximate μ with the empirical measure

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}^i},$$

1. In fact, the moment matrix of any finite measure μ is definite positive unless the support of μ is contained in the zeros of a polynomial; for a closer look at the moment matrix, we refer the reader to [Lasserre and Pauwels \(2019\)](#).

where $\delta_{\mathbf{x}}$ is the Dirac measure. The moment matrix $\widehat{\mathbf{M}}_d$ associated with the empirical measure $\hat{\mu}$ is

$$\widehat{\mathbf{M}}_d = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_d(\mathbf{x}^i) \mathbf{v}_d(\mathbf{x}^i)^T \quad (2)$$

Therefore, the empirical measure $\hat{\mu}$ defines an empirical Christoffel function. Since we are only interested in superlevel sets of the Christoffel function, we can forego the inversion and instead work with sublevel sets of what we call the empirical Christoffel polynomial:

$$\Lambda_{\hat{\mu},d}^{-1}(\mathbf{x}) = \mathbf{v}_d(\mathbf{x})^T \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}) \quad (3)$$

Note that the moment matrix $\widehat{\mathbf{M}}_d$ is almost surely invertible if the number of samples $N \geq s(d)$. The Christoffel polynomial is a sum-of-squares polynomial of degree $2d$. Consequently, it is nonnegative, and if $N > s(d)$, the empirical Christoffel polynomial is strictly positive. Note that, for increasing sample size N , the empirical Christoffel function converges uniformly to the Christoffel function of the exact measure.

2.3. Set Approximation with Christoffel Functions

Lasserre and Pauwels (2019) proposed various thresholding schemes for approximating the support of a probability measure using the Christoffel function or, more precisely, its empirical counterpart. This idea was applied by Devonport et al. (2021) to approximate the reachable set \mathcal{S} with the superlevel sets of the Christoffel function. In this section, we will briefly summarize the approach.

Let μ be the probability measure of the reachable set \mathcal{S} . For a given degree $d \in \mathbb{N}$, the reachable set can be approximated with the sublevel set

$$\hat{\mathcal{S}} = \{\mathbf{x} \in \mathbb{R}^n \mid \Lambda_{\mu,d}^{-1}(\mathbf{x}) \leq \alpha\} \quad (4)$$

for some $\alpha \in \mathbb{R}$. However, since the exactly reachable set \mathcal{S} is unknown, μ is unknown. Instead, the Christoffel function $\Lambda_{\mu,d}$ is approximated by an empirical Christoffel function using i.i.d generated samples \mathbf{x}^i from \mathcal{S} . We can obtain a conservative threshold α such that $\mathbf{x}^i \subseteq \hat{\mathcal{S}}$ by letting

$$\alpha = \max_i \Lambda_{\hat{\mu},d}^{-1}(\mathbf{x}^i). \quad (5)$$

Using methods from statistical learning theory, Devonport et al. (2021) proposed the following PAC guarantees:

Conjecture 1 (Thm. 1 in Devonport et al. (2021)) *Given a training set of i.i.d samples $\mathcal{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ from \mathcal{S} , let*

$$\hat{\mathcal{S}} = \{\mathbf{x} \in \mathbb{R}^n \mid \Lambda_{\hat{\mu},d}^{-1}(\mathbf{x}) \leq \max_i \Lambda_{\hat{\mu},d}^{-1}(\mathbf{x}^i)\}. \quad (6)$$

If $N \geq \frac{5}{\epsilon} \left(\log \frac{4}{\delta} + \binom{n+2d}{n} \log \frac{40}{\epsilon} \right)$, then $\mathbb{P}(\mu(\hat{\mathcal{S}}) \geq 1 - \epsilon) \geq 1 - \delta$.

In other words, if N, δ, ϵ satisfy the condition in Conjecture 1, then with probability bigger than $1 - \delta$ we are sure that $\hat{\mathcal{S}}$ contains more than $1 - \epsilon$ of the mass of \mathcal{S} . However, we believe this result neglects the dependencies between the empirical Christoffel polynomial and the

points used to construct the threshold α . As will be discussed in more detail in Section 3, different samples should be used for constructing the empirical Christoffel polynomial and for constructing the threshold α to ensure independence.

We informally note convergence results by Lasserre and Pauwels (2019), which hold for uniform probability measures (and some generalizations):

- As $d \rightarrow \infty$ and with an appropriately chosen threshold, the sublevel set of the (non-empirical) Christoffel polynomial converges to the support of the measure, i.e., to the exact reach set in these sense of a Hausdorff distance.
- For fixed d and $n \rightarrow \infty$, the empirical Christoffel function converges uniformly to the Christoffel function.
- For fixed d and $n \rightarrow \infty$, the border of the empirical Christoffel polynomial converges to the border of the Christoffel polynomial in these sense of a Hausdorff distance.

In consequence, we can informally expect that for a large enough degree d and large enough sample size N , the sublevel sets of the Christoffel polynomial are close enough to the reachable set.

We will use the following running example throughout the paper to illustrate the different concepts.

Example 1 (Four squares) *Let the transition function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be*

$$f(x, y) = (1 + \text{sign}(x) \cdot x^2, 1 + \text{sign}(y) \cdot y^2)$$

and let the initial set be $\mathcal{I} = [-1, 1]^2$. The reachable set consists of four squares, i.e.,

$$\mathcal{S} = [-3, -1]^2 \cup [-3, -1] \times [1, 3] \cup [1, 3] \times [-3, -1] \cup [1, 3]^2.$$

Figure 1 shows the reach set approximation given by (6), for a sample of size $N = 10\,000$ and different degrees d . The caption includes the corresponding uncertainty bound ε for confidence $1 - \delta = 0.99$ obtained by Conjecture 1.

We observe that, as intended by construction, all samples are included in \hat{S} . For increasing degrees, \hat{S} becomes more precise. However, the uncertainty in the covered probability mass ε , increases substantially. Indeed, the bound ε seems rather conservative since, in all instances, \hat{S} covers nearly 100% of \mathcal{S} .

3. Reach Set Approximation with Conformal Prediction

Following the reasoning of Section 2, we can expect a sublevel of the Christoffel polynomial to converge to the support of the distribution. Intuitively, the Christoffel polynomial takes high values where the density is low and low values where the density is high, which makes it a good candidate for a nonconformity function.

In this section, we briefly recall relevant results from conformal prediction and instantiate them to the special case of estimating the support of distribution, which in our setting is equivalent to approximating the reach set \mathcal{S} .

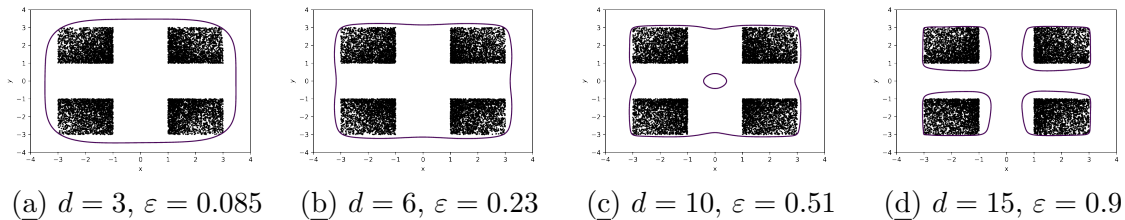


Figure 1: Reach set approximation \hat{S} for Example 1, using the sublevel set of the empirical Christoffel polynomial in (6) (purple outline) on a sample of size $N = 10000$ (black dots), for different degrees d and corresponding uncertainty bound ε , according to Conjecture 1.

Let $r : \mathbb{R}^n \rightarrow \mathbb{R}$ be a non-conformity function. Given a sample $\mathcal{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$, the p -value at \mathbf{x} is

$$pvalue(\mathbf{x}) = \frac{1}{N} \left| \left\{ i \mid r(\mathbf{x}^i) \geq r(\mathbf{x}) \right\} \right|$$

For $i \in \{0, \dots, N\}$, the conformal region is defined as

$$C_{\mathcal{D}}^{\frac{i}{N}} = \left\{ \mathbf{x} \in \mathbb{R}^n \mid pvalue(\mathbf{x}) \geq \frac{i}{N} \right\}$$

According to conformal prediction theory, see Shafer and Vovk (2008); Angelopoulos and Bates (2021), a new i.i.d sample \mathbf{x}^{N+1} satisfies

$$\mathbb{P} \left(\mathbf{x}^{N+1} \in C_{\mathcal{D}}^{\frac{i}{N}} \right) \geq 1 - \frac{i+1}{N+1}. \quad (7)$$

Note that in (7), the set \mathcal{D} is also subject to randomness. In other words, (7) stands on average only if the set \mathcal{D} is re-sampled for each \mathbf{x}^{N+1} . However, in reachability analysis and data-driven applications more generally, we may be restricted to a single, fixed data set \mathcal{D} . Therefore, we need to take into account the probability on the left hand side of (7), conditioned on the sample \mathcal{D} .

3.1. Statistical Guarantees

In this section, we ensure statistical independence between the nonconformity function r and the set \mathcal{D} by splitting it into a training set $\mathcal{D}_{\text{train}}$ and a calibration set \mathcal{D}_{cal} . The use of distinct sets of samples from the same measurement (i.e., a training set and a calibration set) is essential to ensure the independence of the samples used for computing the p -values and conformal regions from the nonconformity function, which is computed based on the training set, see Angelopoulos and Bates (2021) and Bates et al. (2023). This is a special case of conformal prediction called split conformal prediction or inductive conformal prediction. The computational advantage of this method lies in its requirement to fit the model only once. However, this comes at the cost of statistical efficiency as the method necessitates the division of the data into separate, and therefore smaller, training and calibration data sets.

A smaller calibration set increases the coverage error, while a smaller training set reduces the tightness of the approximation. An alternative to this trade-off will be examined in section 3.2. Here, we use the training set for computing the empirical Christoffel polynomial, while the calibration set is used to compute the conformal region. This will lead to bounds on the conditional probability

$$\mathbb{P}\left(\mathbf{x}^{N+1} \in C_{\mathcal{D}}^{\frac{i}{N}} \mid \mathcal{D}_{\text{cal}}\right).$$

Note that the theorems in this section apply to any choice of nonconformity function. The following theorem provides PAC guarantees for conformal regions that are defined with suitably chosen probability thresholds b_1, \dots, b_n . We will afterwards propose values for b_1, \dots, b_n that correspond to the special case of set approximation.

Theorem 2 (Thm. 4 from Bates et al. (2023)) *Consider N uniform random samples $U_1, \dots, U_N \stackrel{i.i.d.}{\sim} \text{Unif}([0, 1])$, with order statistics $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N)}$, and fix any $\delta \in (0, 1)$. Suppose $0 \leq b_1 \leq b_2 \leq \dots \leq b_N \leq 1$ are reals such that*

$$\mathbb{P}\left[U_{(1)} \leq b_1, \dots, U_{(N)} \leq b_n\right] \geq 1 - \delta.$$

Let also $b_0 = 0$. Then for any i.i.d vector \mathbf{x} sampled from μ :

$$\mathbb{P}\left[\mathbb{P}\left(\mathbf{x} \in C_{\mathcal{D}_{\text{cal}}}^{\frac{i}{N}} \mid \mathcal{D}_{\text{cal}}\right) \geq 1 - b_i\right] \geq 1 - \delta \quad (8)$$

We propose an analogous, theorem to bound the conditional probability from above.

Theorem 3 *Under the assumptions of Thm. 2, suppose further that the nonconformity function $r(\mathbf{x})$ is continuous, the measure μ is continuous, and that α is a real such that*

$$\mathbb{P}(U_{(N)} \leq \alpha) \geq 1 - \delta.$$

Then for any i.i.d vector \mathbf{x} sampled from μ :

$$\mathbb{P}\left[\mathbb{P}\left(\mathbf{x} \in C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \mid \mathcal{D}_{\text{cal}}\right) \leq \alpha\right] \geq 1 - \delta \quad (9)$$

Proof Under the assumptions, $r(\mathbf{x})$ has a continuous distribution. Let F_μ be the cumulative distribution function of $r(\mathbf{x})$. Since $r(\mathbf{x})$ has a continuous distribution, $F_\mu(r(\mathbf{x}))$ follows $\text{Unif}([0, 1])$ and $F_\mu(r(\mathbf{x}^1)), F_\mu(r(\mathbf{x}^2)), \dots, F_\mu(r(\mathbf{x}^N))$ all follow $\text{Unif}([0, 1])$. Without loss of generality, we assume $r(\mathbf{x}^1) \leq r(\mathbf{x}^2) \leq \dots \leq r(\mathbf{x}^N)$. Letting $U_N = F_\mu(r(\mathbf{x}^N))$, we obtain

$$\mathbb{P}\left[F_\mu(r(\mathbf{x}^N)) \leq \alpha\right] \geq 1 - \delta.$$

Considering \mathbf{x} sampled i.i.d from μ , we get

$$\mathbb{P}\left(\mathbf{x} \in C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \mid \mathcal{D}_{\text{cal}}\right) = \mathbb{P}\left(r(\mathbf{x}) \leq r(\mathbf{x}^N) \mid \mathcal{D}_{\text{cal}}\right) = F_\mu\left(r(\mathbf{x}^N)\right)$$

Combining the latter two results, we obtain (9). ■

We now use the results of Thm. 2 and Thm. 3 to provide a guarantee on the accuracy of approximated reachable set $\hat{\mathcal{S}}$ in Algorithm 1. Note that Thm. 3 requires the nonconformity function to be continuous, which is the case for the empirical Christoffel polynomial.

Theorem 4 Suppose that the nonconformity function $r(\mathbf{x})$ is continuous. $\forall \delta \in (0, 1)$,

$$\mathbb{P} \left[\mu \left(C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \right) \geq \exp \left(\frac{\log(\delta)}{N} \right) \right] \geq 1 - \delta, \quad (10)$$

If the measure μ is continuous, then

$$\mathbb{P} \left[\exp \left(\frac{\log(1-\delta)}{N} \right) \geq \mu \left(C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \right) \right] \geq 1 - \delta, \quad (11)$$

Combining these results, we obtain $\forall \delta \in (0, 1/2)$:

$$\mathbb{P} \left[\exp \left(\frac{\log(1-\delta)}{N} \right) \geq \mu \left(C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \right) \geq \exp \left(\frac{\log(\delta)}{N} \right) \right] \geq 1 - 2\delta. \quad (12)$$

Proof We instantiate Theorem 2 for a particular choice of $b_1 \dots, b_N$. Since we are interested in the support of the measure, we take b_1 as the smallest possible value and set the other values $b_2 \dots, b_N = 1$. To satisfy the conditions of Theorem 2, we first show the following intermediate result: Let $U_1, \dots, U_N \stackrel{\text{i.i.d.}}{\sim} \text{Unif}([0, 1])$, with order statistics $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N)}$. Fixing $b_1 = 1 - \delta^{\frac{1}{N}}$ and $b_2 = \dots = b_N = 1$, it is straightforward that

$$\mathbb{P}(U_{(1)} \leq b_1, \dots, U_N \leq b_N) = \mathbb{P}(U_{(1)} \leq b_1) = 1 - \mathbb{P}(U_{(1)} \geq b_1).$$

Since $U_{(1)}$ is the smallest of the random variables, U_1, \dots, U_N , $\mathbb{P}(U_{(1)} \geq b_1)$ is equivalent to all of the U_i being greater or equal to b_1 :

$$1 - \mathbb{P}(U_{(1)} \geq b_1) = 1 - \prod_{i=1}^N \mathbb{P}(U_i \geq b_1) = 1 - (1 - b_1)^N = 1 - \delta.$$

Applying the above in Theorem 2, we obtain

$$\mathbb{P} \left[\mathbb{P} \left(\mathbf{x} \in C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \mid \mathcal{D}_{\text{cal}} \right) \geq \exp \left(\frac{\log(\delta)}{N} \right) \right] \geq 1 - \delta.$$

As $\mu \left(C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \right) = \mathbb{P} \left[\mathbf{x} \in C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} \mid \mathcal{D}_{\text{cal}} \right]$ we obtain the result in (10).

Fixing $\alpha = \exp \left(\frac{\log(1-\delta)}{N} \right)$, we have $\mathbb{P}[U_N \leq \alpha] = \alpha^N = 1 - \delta$, since $U_{(N)} \leq \alpha$ means all U_i have to be lower than α . Substituting the above value of α in Theorem 3, we obtain the result in (11). Combing (10) and (11), we obtain the result in (12). ■

Example 2 We illustrate Algorithm 1 on the running Example 1. We take $M = 10000$ i.i.d samples from the reachable set \mathcal{S} by sampling uniformly M i.i.d samples in \mathcal{I} , which we then split into a calibration set of size $N = 2000$ and a training set of size $M - N$. Figure 2 shows the approximated reachable set produced by Algorithm 1 for various degrees d . Theorem 4 guarantees that with confidence $1 - \delta = 99\%$, the coverage error ϵ is lower than $\epsilon \leq 0.002$. Notably, in contrast to the algorithm presented in Devonport et al. (2021), this guarantee is independent of the dimension of the samples n and the degree of the empirical

Algorithm 1: Reach set approximation (without outliers)

Input: An i.i.d data sample $\mathcal{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^M\}$, drawn from the reach set $\mathcal{S} = f(\mathcal{I})$, the degree d , the size N of the calibration set with $N < M$

Output: ϵ -accurate approximation $\hat{\mathcal{S}}$ of \mathcal{S} with confidence $1 - \delta$ and coverage error $\epsilon = 1 - \delta^{1/N}$

Construct the training set of $M - N$ samples and the calibration set of N samples:

$$\mathcal{D}_{\text{train}} = \{\mathbf{x}^{N+1}, \dots, \mathbf{x}^M\} \text{ and } \mathcal{D}_{\text{cal}} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$$

1. Compute the empirical moment matrix $\widehat{\mathbf{M}}_d$ and its inverse

(a) $\widehat{\mathbf{M}}_d = \frac{1}{M-N} \sum_{i=N+1}^M \mathbf{v}_d(\mathbf{x}^i) \mathbf{v}_d(\mathbf{x}^i)^\top$, with $\mathbf{x}^i \in \mathcal{D}_{\text{train}}$

(b) Compute $\widehat{\mathbf{M}}_d^{-1}$.

2. Calculate the threshold α : $\alpha = \max_{i=1, \dots, N} \mathbf{v}_d(\mathbf{x}^i)^\top \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}^i)$, with $\mathbf{x}^i \in \mathcal{D}_{\text{cal}}$

3. Given the returned $\widehat{\mathbf{M}}_d^{-1}$ and α , record the conformal region:

$$C_{\mathcal{D}_{\text{cal}}}^{\frac{1}{N}} = \hat{\mathcal{S}} = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{v}_d(\mathbf{x})^\top \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}) \leq \alpha \right\}$$

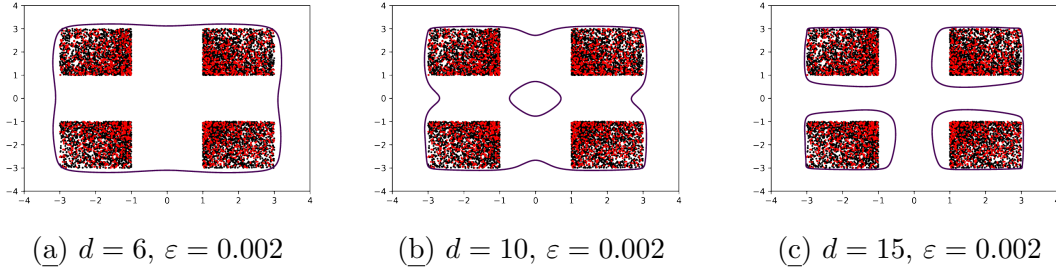


Figure 2: Reach set approximations (outlined in purple) from Example 2, obtained with Algorithm 1, which uses the Christoffel polynomial as a nonconformity function, for $M = 10000$ samples, of which $N = 2000$ are the calibration set (red dots) and the remainder the training set (black dots). Higher degrees d lead to tighter approximation.

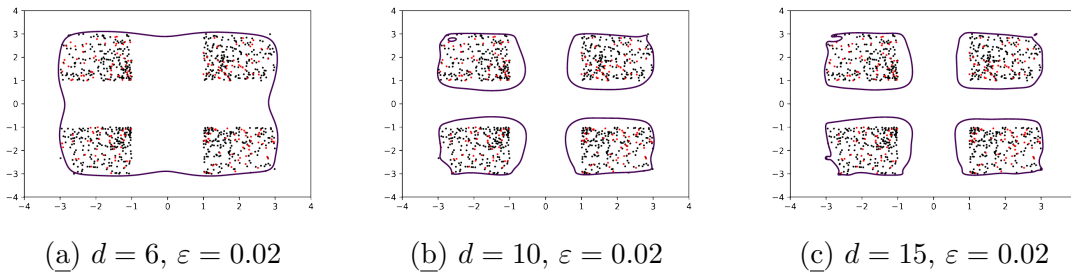


Figure 3: Reach set approximations (outlined in purple) from Example 2, with a reduced sample size of $M = 1000$, of which $N = 200$ are used as a calibration set.

Christoffel polynomial d . It only depends on the confidence parameter δ and the size N of the calibration set. Figure 3 shows the same result for $M = 1000$ samples, of which $N = 200$ samples were utilized as a calibration set. Here, the coverage error ϵ will be lower than $\epsilon \leq 0.02$. To empirically verify the theoretical guarantees obtained in Theorem 4, we repeated this experiment 1000 times. The empirical error was computed by checking how many of these 10000 samples were not contained in the approximated reachable set. In only 6 experiments, the coverage error exceeded $\epsilon = 0.02$, confirming that the confidence $1 - \delta$ is greater than 99%.

3.2. Avoiding the Calibration Set

In this section, we circumvent split between training and calibration sets by using transductive conformal prediction [Vovk \(2013\)](#). Transductive conformal prediction is a method used to construct prediction regions for a new data point without relying on a separate training set or calibration set. The calibration set is taken to be the entire training set plus the point at which the function is evaluated, in other words a new non conformity is modulated by the data point. The statistical guarantees of the previous section, and in particular of Theorem 4, hold also for this choice of nonconformity function, with $\mathcal{D}_{\text{cal}} := \mathcal{D}$. This approach allows us to use all the available sample points from the measure μ to train the Christoffel function and compute the conformal region, but at the price of higher computational cost, as will be discussed below.

Let the training set be $\mathcal{D} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$ be N i.i.d samples from the probability distribution μ . To compute the p-value at any point $\mathbf{x} \in \mathbb{R}^n$, we add \mathbf{x} to the set \mathcal{D} before computing the empirical Christoffel polynomial. Let $\mathcal{D}_x = \mathcal{D} \cup \{\mathbf{x}\}$, let the empirical measure for \mathcal{D}_x be $\hat{\mu}_x$, and let $\widehat{\mathbf{M}}_x$ be its moment matrix. Using \mathcal{D}_x in the empirical Christoffel polynomial, we get the nonconformity function

$$r(\mathbf{x}) = \Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}) = \mathbf{v}_d(\mathbf{x})^T \widehat{\mathbf{M}}_x^{-1} \mathbf{v}_d(\mathbf{x}).$$

We now have to evaluate a different empirical Christoffel polynomial each time we evaluate the p-value

$$pvalue(\mathbf{x}) = \frac{1}{N} \left| \left\{ i \mid \Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}^i) \geq \Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}) \right\} \right|$$

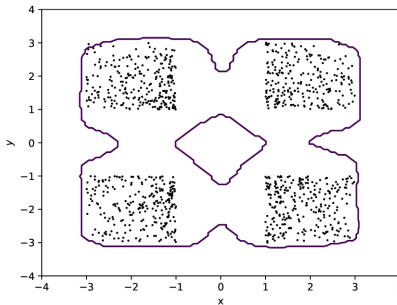


Figure 4: Reach set approximation of example 1 using the transductive conformal prediction and a Christoffel polynomial of degree $d = 15$, which avoids the split into training and calibration sets.

In particular, we need to compute a new moment matrix and invert it for each evaluation. This is computationally expensive, on the order of $\mathcal{O}(s(d)^3)$. To avoid this, we compute the inverse moment matrix of the set \mathcal{D}_x incrementally using the Sherman-Morrison formula, as proposed by Ducharlet et al. (2022). This allows us to replace the evaluation of $\Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x})$, which depends on \mathbf{x} , with evaluations of the original Christoffel polynomial $\Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x})$, plus one additional product:

$$\Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}) = \frac{\Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x})}{1 + \Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x})}, \quad \Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}^i) = \Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x}^i) - \frac{(\mathbf{v}_d(\mathbf{x})^\top \mathbf{y}^i)^2}{1 + \Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x})}, \quad (13)$$

where $\mathbf{y}^i = \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}^i)$ are vectors that can be precomputed. The cost of precomputing $\Lambda_{\hat{\mu}, d}^{-1}(\mathbf{x}^i)$ and the vectors \mathbf{y}^i is $\mathcal{O}(Ns(d)^2)$, with storage requirements $\mathcal{O}(Ns(d))$. This reduces the cost of evaluating $\Lambda_{\hat{\mu}_x, d}^{-1}(\mathbf{x}^i)$ for a given \mathbf{x} to $\mathcal{O}(s(d))$. The resulting cost of evaluating $p_{value}(\mathbf{x})$ is $\mathcal{O}(Ns(d) + s(d)^2)$.

Example 3 Building on example 1, Figure 4 shows the reachable set approximation obtained using transductive conformal prediction with a Christoffel function of degree 15. In this case, we use the same $M = N = 1000$ sample points to train the Christoffel function and compute the set approximation. The guarantees provided by Theorem 4 assert that, using a training set of 1000 samples, the coverage error is below 0.45% with confidence $1 - \delta = 0.99$.

4. Robustness to Outliers

In this section, we address the presence of outliers in the data set. As data may not be very abundant in real-life applications, one may have to work with a calibration set containing outliers without knowing which data point is an outlier and which one isn't. The presence of outliers in the training set does not affect the theoretical guarantees obtained

using conformal prediction theory, though it will affect the tightness of the approximated reachable set. On the other hand, the presence of outliers in the calibration set will impact those guarantees.

The following theorem provides PAC guarantees on the reach set approximation even with outliers in the calibration set. Under the assumption that no more than p outliers are in the calibration set \mathcal{D} , the confidence in the result depends on ϵ , p , and the size of the calibration set N .

Theorem 5 *Consider a set of points $\mathcal{D} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$ containing no more than p outliers, with $2p+1 < N$, and where the rest of samples are i.i.d from a probability measure μ . Then for any i.i.d vector \mathbf{x} sampled from μ and $\epsilon \in (0, 1)$,*

$$\mathbb{P}\left(\mu\left(C_{\mathcal{D}}^{\frac{p+1}{N}}\right) \geq 1 - \epsilon\right) \geq \sum_{i=p+1}^{N-p} \binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i} \quad (14)$$

This bound is tight in the sense that for $p = 0$, (14) is identical to the case without outliers, i.e., we obtain (10).

Proof Let $\mathcal{D} = \mathcal{D}_{inlier} \cup \mathcal{D}_{outlier}$, with $m \leq p$ being the unknown real size of $\mathcal{D}_{outlier}$. Let $U_1, \dots, U_{N-m} \stackrel{\text{i.i.d.}}{\sim} \text{Unif}([0, 1])$, with order statistics $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(N-m)}$. For $\epsilon \in (0, 1)$ let $b_1 = \dots = b_{p+1} = \epsilon$ and $b_{p+2} = \dots = b_{N-p} = \dots = b_{N-m} = 1$. Then $\forall m \leq p$:

$$\begin{aligned} \mathbb{P}\left[U_{(1)} \leq b_1, \dots, U_{(N-m)} \leq b_{N-m}\right] &\geq \mathbb{P}\left[U_{(1)} \leq b_1, \dots, U_{(N-p)} \leq b_{N-p}\right] \\ &\geq \sum_{i=p+1}^{N-p} \binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i}. \end{aligned}$$

The above result is obtained by the following reasoning: let $0 < i \leq N-p$, if we have $N-p$ random variable $V_1, \dots, V_{N-p} \stackrel{\text{i.i.d.}}{\sim} \text{Unif}([0, 1])$ the probability to have exactly i of them below ϵ is equal to $\binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i}$, therefore, the probability of having at least $p+1$ of them below ϵ is equal to $\sum_{i=p+1}^{N-p} \binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i}$.

Let \mathbf{x} be an i.i.d vector sampled from μ . By definition,

$$\mu\left(C_{\mathcal{D}_{inliers}}^{\frac{p+1}{N-m}}\right) = \mathbb{P}\left(\mathbf{x} \in C_{\mathcal{D}_{inliers}}^{\frac{p+1}{N-m}} \mid \mathcal{D}_{inliers}\right).$$

Using Theorem 2, we get :

$$\mathbb{P}\left(\mu\left(C_{\mathcal{D}_{inliers}}^{\frac{p+1}{N-m}}\right) \geq 1 - \epsilon\right) \geq \sum_{i=p+1}^{N-p} \binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i}$$

Since $C_{\mathcal{D}_{inliers}}^{\frac{p+1}{N-m}} \subseteq C_{\mathcal{D}}^{\frac{p+1}{N}}$, we have $\mu\left(C_{\mathcal{D}}^{\frac{p+1}{N}}\right) \geq \mu\left(C_{\mathcal{D}_{inliers}}^{\frac{p+1}{N-m}}\right)$, which leads us to (14). \blacksquare

Note that the bound in Theorem 5 (14) is tight in the sense that for $p = 0$ we obtain the same lower bound as in Theorem 4 (10). Table 4 shows the confidence bound of (14) for different values of the calibration set size and the approximation uncertainty ϵ under the assumption that no more than 5% of the calibration set are outliers. We observe that the confidence rapidly approaches 100% when the admissible coverage error is above the ratio of outliers; it rapidly drops to 0% when it is below.

Table 1: The confidence bound of (14) for different sizes N of the calibration set and the desired coverage error ϵ for a calibration set with 5% outliers or less

size N	confidence in %			
	$\epsilon = 4\%$	$\epsilon = 5\%$	$\epsilon = 6\%$	$\epsilon = 10\%$
100	33	51	68	96
500	10	42	77	99.99
1000	3	37	84	99.99
2000	0.4	31	92	99.99

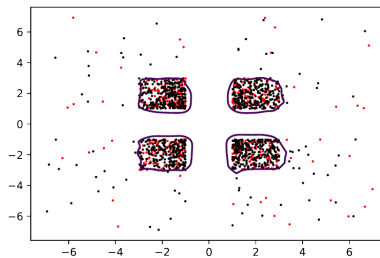


Figure 5: An approximation of the reach set of example 1 (purple outline) obtained with Algorithm 2 using a Christoffel polynomial of degree 15, on a data set with 10% outliers. The training set is shown in black, the calibration set in red.

Example 4 *To evaluate the performance of Algorithm 2 on example 1, we construct a data set from $M = 1500$ samples of the reach set and substitute 10% with outliers, i.e., i.i.d. samples outside the reachable set. We use a calibration set of size $N = 500$, and the rest of the samples are used as a training set to compute the empirical Christoffel polynomial. Figure 5 shows the resulting approximation. With Theorem 5, the coverage error $\epsilon = 0.15$ with a confidence = 98.9%. To empirically confirm these bounds, as in Example 2, we repeat the experiment 1000 times with different samples. For each experiment, we take 10000 samples of the reach set in order to compute the empirical coverage error. None of the experiments resulted in an empirical coverage error above 15%, which is consistent with the theoretical guarantee of 98.9% confidence.*

5. Experiments

We now turn our focus to the suitability of the empirical Christoffel polynomial as a non-conformity function.

Algorithm 2: Reachability analysis with outliers

Input: Transition function f ; initial set $\mathcal{I} \subset \mathbb{R}^n$; Christoffel function order d , N the size of the calibration set and p the upper bound number of outliers in the calibration; M the total number of simulations with $M > N$ and an i.i.d data sample $\mathcal{D} = \{\mathbf{x}^{(i)}\}$ for $i \in \{1, \dots, M\}$. The sample $\mathbf{x}^{(i)}$ is an inlier if $\mathbf{x}^{(i)} \in f(\mathcal{I})$ and an outlier otherwise. .

Output: Set $\hat{\mathcal{S}}$ representing an ϵ -accurate approximation of the true reachable set \mathcal{S} with confidence $\sum_{i=p+1}^{N-p} \binom{N-p}{i} \epsilon^i (1-\epsilon)^{N-p-i}$

1. Compute the empirical moment matrix with the associated Christoffel degree d using $M - N$ samples: $\widehat{\mathbf{M}}_d = \frac{1}{M-N} \sum_{i=N+1}^M \mathbf{v}_d(\mathbf{x}^{(i)}) \mathbf{v}_d(\mathbf{x}^{(i)})^\top$
 2. Use a calibration set of N samples: $\mathcal{D}_{\text{cal}} = \{\mathbf{x}^{(i)} \mid i \in \{1, \dots, N\}\}$ and:
 - (a) Compute the scores: $\text{score}_i = \mathbf{v}_d(\mathbf{x}^{(i)})^\top \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}^{(i)})$ for $i = 1, \dots, N$
 - (b) Sort the scores in descending order such that: $\text{score}_1 \geq \text{score}_2 \geq \dots \geq \text{score}_N$
 - (c) Set the conformal region $C_{\mathcal{D}}^{\frac{p+1}{N}}$ as $\hat{\mathcal{S}} = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{v}_d(\mathbf{x})^\top \widehat{\mathbf{M}}_d^{-1} \mathbf{v}_d(\mathbf{x}) \leq \text{score}_{p+1} \right\}$
-

5.1. Empirical False Positive Rate

We start by examining the tightness of the reachable set approximation in example 2 through the empirical measurement of false positives.

We compare the empirical Christoffel polynomial with other prevalent nonconformity functions: one-class SVM, Isolation Forest (Liu et al., 2008), and Local Outlier Factor (LOF), as shown in Figure 6. Only the approximation using LOF seems comparable to that of the Christoffel polynomial, while Isolation Forest exhibits significant variability depending on the random seed.

To gauge the number of false positives and assess the accuracy of the reachable set approximation, we generated 10,000 uniformly distributed samples within the domain $[-4, 4]^2$. The false-positive rate was empirically determined for various degrees d , as shown in Table 2. As observed in earlier plots, a higher degree results in a more accurate fit of the reachable set. The false-positive rates for the other algorithms can also be observed in Table 2 for varying sizes of the training and calibration sets. Consistent with the findings from the Figure 6, only the LOF provides results that are comparable in quality to those obtained using the Christoffel polynomial.

To further demonstrate the effectiveness of the empirical Christoffel polynomial as a non-conformity function, we examine its robustness in the presence of outliers within the training set. Although the theoretical guarantees discussed in this article and in general conformal prediction hold for any choice of non-conformity function, even with outliers in the training set, the presence of these outliers can impact the accuracy of the model. To compare the empirical Christoffel polynomial with LOF, we conducted two experiments. In the first experiment, we considered the region $[-1, 1]^2$ as the reachable set to approximate. We focused on comparing the performance of the algorithms under the presence of outliers

Table 2: Experimentally estimated false-positive rates for different algorithms applied to the reach set approximation of Example 1, with confidence $1 - \delta = 99\%$

Nonconformity function	$ \mathcal{D} $	$ \mathcal{D}_{\text{train}} $	$ \mathcal{D}_{\text{cal}} $	ϵ in %	FP%
Christoffel with $d = 6$	10000	8000	2000	0.2	49.5
Christoffel with $d = 10$				0.2	39.5
Christoffel with $d = 15$				0.2	11.7
Christoffel with $d = 18$	\vdots	\vdots	\vdots	0.2	7.2
LOF score	\vdots	\vdots	\vdots	0.2	3.4
IsolationForest score				0.2	92.9
Oneclass SVM score				0.2	65.7
Christoffel with $d = 6$	1000	800	200	2.2	44.6
Christoffel with $d = 10$				2.2	20
Christoffel with $d = 15$				2.2	12.7
Christoffel with $d = 18$	\vdots	\vdots	\vdots	2.2	12.4
LOF score	\vdots	\vdots	\vdots	2.2	10.6
IsolationForest score				2.2	86.8
Oneclass SVM score				2.2	60.7
Transduct. Christ. with $d = 15$	1000	1000	1000	0.5	46.6

ϵ = Coverage error, at least $1 - \epsilon$ of the measure is covered; FP% = False positives in %, measured by uniform sampling of a sufficiently large bounding box and counting samples in $\hat{S} \setminus S$

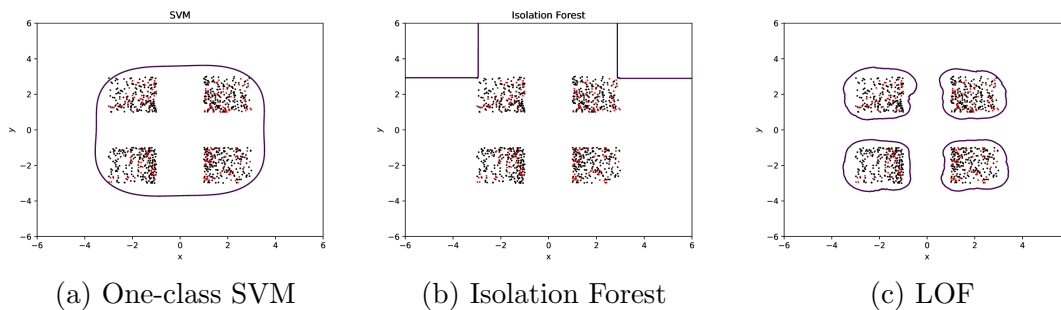


Figure 6: Reach set approximations (purple outline) of Example 1 using one-class SVM, Isolation Forest, and Local Outlier Factor (LOF) as nonconformity functions, for a common training set of size 800 (black dots) and calibration set of size 200 (red dots).

in the training set. We generated a training set of size 1,200 containing 200 outliers and a calibration set of size 200, all belonging to the reachable set. The second experiment was similar to the first one, with a star-shaped region as the reachable set. We generated a training set of size 900 containing 100 outliers and a calibration set of size 200. Figure 9 il-

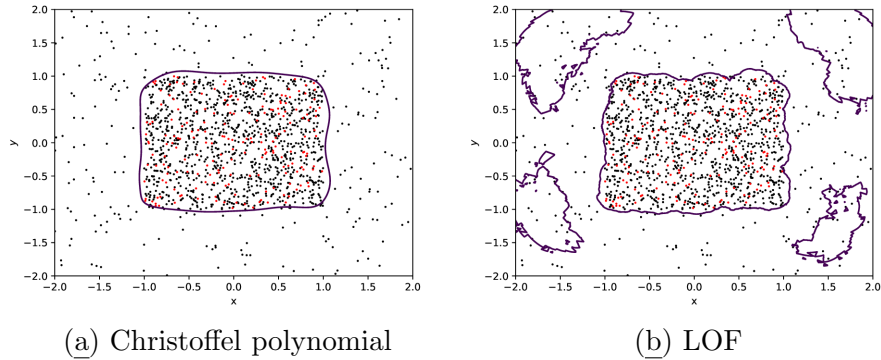


Figure 7: Comparison of reachable set approximations for the empirical Christoffel polynomial (degree 10) and LOF in the first experiment, with the region $[-1, 1]^2$ as the target. The training set, containing outliers, is represented by black dots, while the calibration set is shown in red. The plot highlights the performance differences and robustness of both methods in the presence of outliers, demonstrating how the empirical Christoffel polynomial is far more robust.

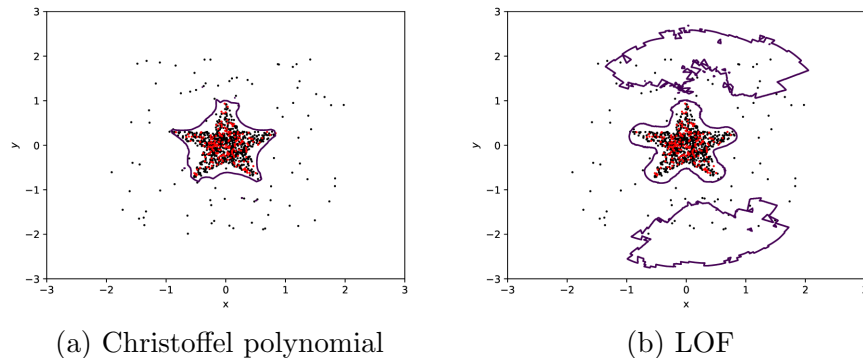


Figure 8: A comparison of reach set approximation (purple outline) using the Christoffel polynomial with degree 15 and LOF for the second experiment, which targets a star-shaped region. Training set samples are in black and calibration set in red. This plot highlights the performance and robustness of both methods when encountering outliers in a complex geometric scenario, illustrating the effectiveness of the empirical Christoffel polynomial under the presence of outliers.

illustrates how the empirical Christoffel polynomial and LOF approximate the true reachable set in the presence of outliers.

Figures 7 and 8 display the performance of both the empirical Christoffel polynomial and LOF in handling outliers within the training set across distinct and complex geometric situ-

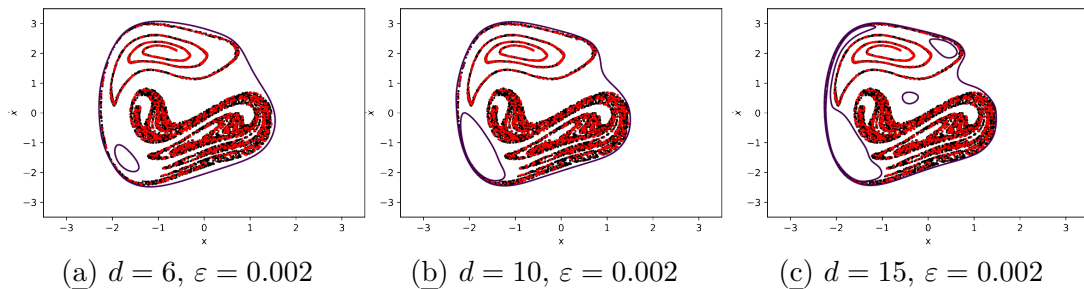


Figure 9: Reach set approximation (purple outline) of the duffing oscillator using the Christoffel polynomial with the data set split into training (black) and calibration set (red), for different degrees d of the Christoffel function, with corresponding coverage error ε for confidence $1 - \delta = 0.99$.

ations. When employed as a non-conformity function, the empirical Christoffel polynomial demonstrated greater robustness in the presence of outliers across both experiments.

5.2. Duffing oscillator

The Duffing oscillator is a nonlinear mathematical model that captures the behavior of a system that oscillates when subject to an external force. It has been used in a variety of physical systems, from mechanical vibrations to biological dynamics. The Duffing oscillator is described by the following nonlinear second-order differential equation:

$$\ddot{x} = -\delta\dot{x} + \alpha x - \beta x^3 + \gamma \cos(\omega t)$$

Similar to [Devonport et al. \(2021\)](#), we take $\alpha = 1, \beta = 1, \delta = 0.05, \gamma = 0.4$ and $\omega = 1.3$. We choose the initial set to be $\mathcal{I} = [-0.95, 1.05] \times [-0.05, 0.05]$. Figure 9 shows an approximation of the reach set, computed with the Christoffel function as nonconformity function for different degrees. We observe that for increasing degrees, the approximation is more precise and is able to recover holes. The results are comparable to those reported by [Devonport et al. \(2021\)](#), where no split into training and calibration sets was carried out.

6. Conclusion

In this paper, we studied the mathematical reach set approximation in the analysis of dynamical systems based on conformal prediction. We consider for the first time the use of the Christoffel function as a nonconformity function, thanks to its attractive properties in set and density approximation. Our conformal prediction approach provides stronger and more sample-efficient guarantees on reach set approximation and proposed a version of reach set approximation that is robust to outliers, compared that the most relevant approaches in the literature. We exploited an incremental form of the Christoffel function for transductive conformal prediction that avoids splitting the data into training and calibration sets. Extensive illustrative numerical experiments show the effectiveness and the performance of our proposed approach and its associated algorithms.

The theoretical results that we presented here in the context of reach set approximation are equally valid to approximate compact sets, or the support of probability distributions, in other application domains. Naturally, the computation of the Christoffel function is subject to numerical errors. The impact of such numerical issues will be studied in future work.

Acknowledgments

This work has been supported by the French government under the “France 2030” program as part of the SystemX Technological Research Institute. This work was conducted as part of the Con fiance.AI program, which aims to develop innovative solutions for enhancing the reliability and trustworthiness of AI-based systems.

References

- Amr Alanwar, Anne Koch, Frank Allgower, and Karl Henrik Johansson. Data-driven reachability analysis from noisy data. *IEEE Transactions on Automatic Control*, pages 1–16, 2023. doi: 10.1109/tac.2023.3257167.
- Matthias Althoff, Goran Frehse, and Antoine Girard. Set propagation techniques for reachability analysis. *Annual Review of Control, Robotics, and Autonomous Systems*, 4: 369–395, 2021.
- Rajeev Alur. *Principles of cyber-physical systems*. MIT press, 2015.
- Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *CoRR*, abs/2107.07511, 2021. URL <https://arxiv.org/abs/2107.07511>.
- Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178, 2023.
- Martin Berz and Kyoko Makino. Verified integration of odes and flows using differential algebraic methods on high-order taylor models. *Reliable Computing*, 4(4):361–369, 1998. doi: 10.1023/A:1024467732637. URL <https://doi.org/10.1023/A:1024467732637>.
- Alex Devonport, Forest Yang, Laurent El Ghaoui, and Murat Arcak. Data-driven reachability analysis with christoffel functions. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5067–5072, 2021. doi: 10.1109/CDC45484.2021.9682860.
- Franck Djeumou, Aditya Zutshi, and Ufuk Topcu. On-the-fly, data-driven reachability analysis and control of unknown systems: An F-16 aircraft case study. In *Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control, HSCC ’21*, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3447928.3457355.
- L Doyen, G Frehse, GJ Pappas, and A Platzer. *Verification of Hybrid Systems*, chapter 28. Springer, 2018.

- Kévin Ducharlet, Louise Travé-Massuyès, Jean-Bernard Lasserre, Marie-Véronique Le Lann, and Youssef Miloudi. Leveraging the Christoffel-Darboux kernel for online outlier detection, 2022.
- Nicolas Halbwachs, Yann-Eric Proy, and Pascal Raymond. Verification of linear hybrid systems by means of convex approximations. In International Static Analysis Symposium, SAS'94, Namur (Belgium), September 1994.
- Shuo Han, Ufuk Topcu, and George J. Pappas. A sublinear algorithm for barrier-certificate-based data-driven model validation of dynamical systems. In 2015 54th IEEE Conference on Decision and Control (CDC), pages 2049–2054, 2015. doi: 10.1109/CDC.2015.7402508.
- Jean-Bernard Lasserre. On the Christoffel function and classification in data analysis. Comptes Rendus. Mathématique, 360:919–928, 2022. doi: 10.5802/crmath.358.
- Jean-Bernard Lasserre and Edouard Pauwels. The empirical christoffel function with applications in data analysis. Advances in Computational Mathematics, 45(3):1439–1468, 2019.
- Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In 2008 eighth IEEE international conference on data mining, pages 413–422. IEEE, 2008.
- Stephen Prajna. Barrier certificates for nonlinear model validation. Automatica, 42(1): 117–126, 2006.
- Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. Journal of Machine Learning Research, 9(3), 2008.
- Peter Van Overschee and Bart De Moor. Subspace identification for linear systems: Theory—Implementation—Applications. Springer Science & Business Media, 2012.
- Vladimir Vovk. Transductive conformal predictors. In Artificial Intelligence Applications and Innovations: 9th IFIP WG 12.5 International Conference, AIAI 2013, Paphos, Cyprus, September 30–October 2, 2013, Proceedings 9, pages 348–360. Springer, 2013.