Sparse moment-sum-of-squares relaxations of nonlinear dynamical systems with guaranteed convergence

Corbinian Schlosser[1], Milan Korda[1,2]
1: LAAS-CNRS Toulouse,
2: Faculty of Electrical Engineering, Czech Technical University in Prague

SMAI 2021
23.06.2021

Definitions

Il y a la systeme dynamique $\dot{x} = f(x)$ avec $X \subset \mathbb{R}^n$ compact. Le solutions s'appellent $\varphi_t(\cdot)$.

Definition (Maximum positively invariant (MPI) set)

Le set de tous les points $x_0 \in X$ tel que $\varphi_t(x_0) \in X$ pour tous les temps $t \in \mathbb{R}_+$, s'appelle MPI set.

Definitions

Given a dynamical system $\dot{x} = f(x)$ on $X \subset \mathbb{R}^n$ compact we denote the solution/flow map by $\varphi_t(\cdot)$.

Definitions

Given a dynamical system $\dot{x} = f(x)$ on $X \subset \mathbb{R}^n$ compact we denote the solution/flow map by $\varphi_t(\cdot)$.

Definition (Maximum positively invariant (MPI) set)

Set of initial conditions $x_0 \in X$, such that the solutions $\varphi_t(x_0)$ stay in X for all $t \in \mathbb{R}_+$, is called the MPI set.

A LP for the MPI set

In Korda et al. (2014) the following linear program for computing the MPI set is proposed

$$
\begin{aligned}
d^* := \quad & \inf \int_X w(x)\, d\lambda(x) \\
\text{s.t.} \quad & v \in \mathcal{C}^1(\mathbb{R}^n), w \in \mathcal{C}(X) \\
& \beta v(x) - \nabla v \cdot f(x) \geq 0 \quad \text{for } x \in X \\
& w \geq 0 \\
& w - v - 1 \geq 0
\end{aligned}
$$

A LP for the MPI set

In Korda et al. (2014) the following linear program for computing the MPI set is proposed

$$
\begin{aligned}
d^* := \quad & \inf \int_X w(x)\, d\lambda(x) \\
\text{s.t.} \quad & v \in \mathcal{C}^1(\mathbb{R}^n), w \in \mathcal{C}(X) \\
& \beta v(x) - \nabla v \cdot f(x) \geq 0 \quad \text{for } x \in X \\
& w \geq 0 \\
& w - v - 1 \geq 0
\end{aligned}
$$

Theorem

It holds $d^* = \lambda(MPI)$ and $w^{-1}([1,\infty)) \supset MPI$ for all feasible (v, w).

The corresponding hierarhy of SDPs

The corresponding semidefinite programs (SDPs) have the following form

$$
\begin{aligned}
d_k^* := \quad & \inf \int_X w(x)\, d\lambda(x) \\
\text{s.t.} \quad & v, w \in \mathbb{R}[x]_k \\
& \beta v - \nabla v \cdot f = \mathrm{SoS}_k \\
& w = \mathrm{SoS}_k \\
& w - v - 1 = \mathrm{SoS}_k
\end{aligned}
$$

The corresponding hierarhy of SDPs

The corresponding semidefinite programs (SDPs) have the following form

$$
\begin{aligned}
d_k^* := \qquad & \inf \int_X w(x)\, d\lambda(x) \\
\text{s.t.} \qquad & v, w \in \mathbb{R}[x]_k \\
& \beta v - \nabla v \cdot f = \mathrm{SoS}_k \\
& w = \mathrm{SoS}_k \\
& w - v - 1 = \mathrm{SoS}_k
\end{aligned}
$$

Theorem

It holds $d_k^* \to d^* = \lambda(MPI)$ and $w_k^{-1}([1,\infty)) \supset MPI$ for all feasible (v_k, w_k).

The corresponding hierarhy of SDPs

The corresponding semidefinite programs (SDPs) have the following form

$$
\begin{aligned}
d_k^* := \quad & \inf \int_X w(x)\, d\lambda(x) \\
\text{s.t.} \quad & v, w \in \mathbb{R}[x]_k \\
& \beta v - \nabla v \cdot f = \sum_i p_i^2 \\
& w = \mathrm{SoS}_k \\
& w - v - 1 = \mathrm{SoS}_k
\end{aligned}
$$

Theorem

It holds $d_k^* \to d^* = \lambda(MPI)$ and $w_k^{-1}([1,\infty)) \supset MPI$ for all feasible (v_k, w_k).

The corresponding hierarhy of SDPs

The corresponding semidefinite programs (SDPs) have the following form

$$\begin{aligned} d_k^* := \quad & \inf \int_X w(x)\, d\lambda(x) \\ \text{s.t.} \quad & v, w \in \mathbb{R}[x]_k \\ & \beta v - \nabla v \cdot \mathbf{f} = \mathrm{SoS}_k \\ & w = \mathrm{SoS}_k \\ & w - v - 1 = \mathrm{SoS}_k \end{aligned}$$

Theorem

It holds $d_k^* \to d^* = \lambda(MPI)$ and $w_k^{-1}([1,\infty)) \supset MPI$ for all feasible (v_k, w_k).

Computational complexity

Solving the semidefinite program has

$$\text{complexity polynomial in } \binom{n+\frac{k}{2}}{k/2}$$

where k is the degree of the polynomials and n the space dimension.

Computational complexity

Solving the semidefinite program has

$$\text{complexity polynomial in } \binom{n+\frac{k}{2}}{k/2}$$

where k is the degree of the polynomials and n the space dimension.
Also other methods typically suffer from the curse of dimensionality.

$\rightsquigarrow$ reduction techniques are needed to solve larger problems

Sparsity graph

The sparsity graph of f represents the dependence of different states in the dynamics in the following way

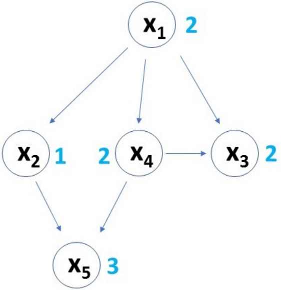


Figure: Example of a sparsity graph

where the nodes x_1, x_3, x_4 represent states in $\mathbb{R}^2$, x_2 represents a scalar state and x_3 a state in $\mathbb{R}^3$. The arrows indicate

$$\begin{aligned} \dot{x}_1 &= f_1(x_1),\ \dot{x}_2 = f_2(x_1, x_2),\ \dot{x}_3 = f_3(x_1, x_3, x_4) \\ \dot{x}_4 &= f_4(x_1, x_4),\ \dot{x}_5 = f_5(x_2, x_4). \end{aligned}$$

Sparsity graph

Let $x_1, \dots, x_k$ be the (grouped) states and $f = (f_1, \dots, f_k)$ then there is an edge between the nodes x_i and x_j if f_j depends on x_i.

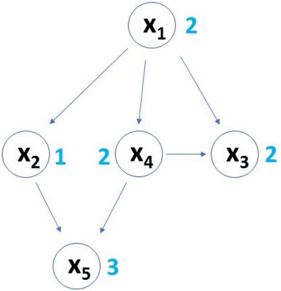


Figure: Example of a sparsity graph

where the nodes x_1, x_3, x_4 represent states in $\mathbb{R}^2$, x_2 represents a scalar state and x_3 a state in $\mathbb{R}^3$. The arrows indicate

$$
\begin{aligned}
\dot{x}_1 &= f_1(x_1),\ \dot{x}_2 = f_2(x_1, x_2),\ \dot{x}_3 = f_3(x_1, x_3, x_4) \\
\dot{x}_4 &= f_4(x_1, x_4),\ \dot{x}_5 = f_5(x_2, x_4).
\end{aligned}
$$

Complexity based on the graph structure of f

We will replace the large SDP by a collection of smaller SDPs. The size of the largest among these depends on the *largest weighted past* of the graph.

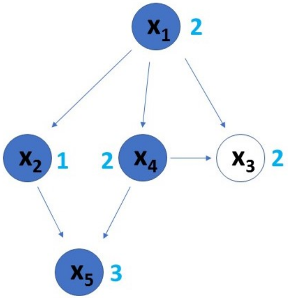


Figure: Largest weighted past of the example graph

The main theorem

Theorem

There exists a convergent hierarchy of sum-of-squares problems with the largest sum-of-squares multiplier containing ω variables, where ω is the longest weighted past in the graph.

The main theorem

Theorem

There exists a convergent hierarchy of sum-of-squares problems with the largest sum-of-squares multiplier containing ω variables, where ω is the longest weighted past in the graph.

1. Sparse description of the sets of interest

The main theorem

Theorem

There exists a convergent hierarchy of sum-of-squares problems with the largest sum-of-squares multiplier containing ω variables, where ω is the longest weighted past in the graph.

1. Sparse description of the sets of interest
2. Formulating and solving the corresponding linear optimization problems

The main theorem

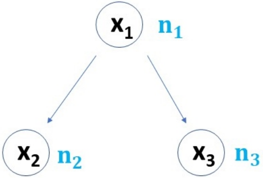
Theorem

There exists a convergent hierarchy of sum-of-squares problems with the largest sum-of-squares multiplier containing ω variables, where ω is the longest weighted past in the graph.

1. Sparse description of the sets of interest
2. Formulating and solving the corresponding linear optimization problems
3. "Glue" the results together to obtain (an approximation of) the MPI set.

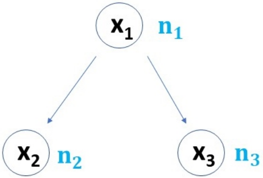
Prototype sparse structure

The most basic sparsity structure that allows a reduction is the following "cherry structure" from Chen et al. (2018)



Prototype sparse structure

The most basic sparsity structure that allows a reduction is the following "cherry structure" from Chen et al. (2018)

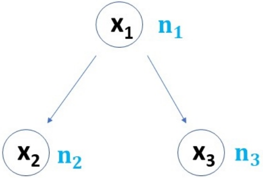


The corresponding dynamics are of the form

$$\begin{aligned}
\dot{x}_1 &= f_1(x_1) \quad \text{on } \mathbb{R}^{n_1} \\
\dot{x}_2 &= f_2(x_1, x_2) \quad \text{on } \mathbb{R}^{n_2} \\
\dot{x}_3 &= f_3(x_1, x_3) \quad \text{on } \mathbb{R}^{n_3}
\end{aligned}$$

Prototype sparse structure

The most basic sparsity structure that allows a reduction is the following "cherry structure" from Chen et al. (2018)



The corresponding dynamics are of the form

$$\begin{aligned}
\dot{x}_1 &= f_1(x_1) \quad \text{on } \mathbb{R}^{n_1} \\
\dot{x}_2 &= f_2(x_1, x_2) \quad \text{on } \mathbb{R}^{n_2} \\
\dot{x}_3 &= f_3(x_1, x_3) \quad \text{on } \mathbb{R}^{n_3}
\end{aligned}$$

Hence they induce the (almost) independent subsystems

$$\dot{(x_1, x_2)} = (f_1, f_2)(x_1, x_2)\ ,\ \dot{(x_1, x_3)} = (f_1, f_3)(x_1, x_3)\ ,\ \dot{x}_1 = f(x_1)$$

Sparse description of the MPI set

Assumption: Sparsity in the constraint set – $X = X_1 \times X_2 \times X_3$

Sparse description of the MPI set

Assumption: Sparsity in the constraint set – $X = X_1 \times X_2 \times X_3$

Lemma (Gluing)

Let $X = X_1 \times X_2 \times X_3$ and the dynamical system be sparse (in the sense as above) and let M_+^1 and M_+^2 denote the MPI sets for the subsystems on (x_1, x_2) and (x_1, x_3) then the MPI set M_+ of the whole system is given by

$$\{(x_1, x_2, x_3) \in X : (x_1, x_2) \in M_+^1, (x_1, x_3) \in M_+^2\}.$$

Proof.

Uses that the dynamical system decouples into subsystems (the state constraint decoupling is essential as well). □

Sparse description of the MPI set

Assumption: Sparsity in the constraint set – $X = X_1 \times X_2 \times X_3$

Lemma (Gluing)

Let $X = X_1 \times X_2 \times X_3$ and the dynamical system be sparse (in the sense as above) and let M_+^1 and M_+^2 denote the MPI sets for the subsystems on (x_1, x_2) and (x_1, x_3) then the MPI set M_+ of the whole system is given by

$$\{(x_1, x_2, x_3) \in X : (x_1, x_2) \in M_+^1, (x_1, x_3) \in M_+^2\}.$$

Proof.

Uses that the dynamical system decouples into subsystems (the state constraint decoupling is essential as well). □

This allows to solve separate SDPs for the subsystems instead of solving a SDP for the whole system.

Convergence

Proposition

For $i = 1, 2$ let M^i denote the MPI set for the two subsystems on (x_1, x_2) and (x_1, x_3) and M denotes the MPI set for the whole system. Let $M^i \subset M^i_k \subset X_1 \times X_{i+1} \subset \mathbb{R}^{n_1+n_{i+1}}$ for $i = 1, 2$ and

$$M_k^{1,2} := \{(x_1, x_2, x_3) \mid (x_1, x_2) \in M_k^1, (x_1, x_3) \in M_k^2\}.$$

Then

$$\lambda(M_k^{1,2} \setminus M) \leq \lambda(M_k^1 \setminus M^1)\lambda(X_3) + \lambda(M_k^2 \setminus M^2)\lambda(X_2). \qquad (1)$$

In particular if M^i_k converges to M^i with respect to λ for $i = 1, 2$ then $M_k^{1,2}$ converges to M with respect to λ.

General procedure

Given a dynamical system induced by f and a method for approximating/computing the region of attraction, MPI set or (global) attractors for an arbitrary dynamical system.

i. Find the minimal subsystems that *cover* the whole system.
ii. Compute (outer) approximations S_i for these subsystems.
iii. Glue the set S_i together as in the gluing lemma.

General procedure

Given a dynamical system induced by f and a method for approximating/computing the region of attraction, MPI set or (global) attractors for an arbitrary dynamical system.

i. Find the minimal subsystems that *cover* the whole system.
ii. Compute (outer) approximations S_i for these subsystems.
iii. Glue the set S_i together as in the gluing lemma.

Theorem (Main theorem)

This procedure produces convergent approximations of the desired sets if the sets S_i are converging approximations of the desired sets for the subsystems. In case of the sum-of-squares approach we have convergence w.r.t Lebesgue measure discrepancy and the size of the largest occuring SDP depends only on the largest weighted pasts of the sparsity graph of f.

Finding a optimal decomposition

Remark

We call a set of states $(x_i)_{i \in I}$ *for some index set* $I \subset \{1, \dots, n\}$ *a subsystem of* $\dot{x} = f(x)$ *if we have*

$$f_I \circ P_I = P_I \circ f$$

where $f_I := (f_i)_{i \in I}$ *denotes the components of* f *according to the index set* I *and* $P_I : \mathbb{R}^n \to \mathbb{R}^{|I|}$ *denotes the canonical projection onto the states* x_I*, i.e.* $P_I(x) := (x_i)_{i \in I}$.

Finding a optimal decomposition

Remark

We call a set of states $(x_i)_{i \in I}$ *for some index set* $I \subset \{1, \dots, n\}$ *a subsystem of* $\dot{x} = f(x)$ *if we have*

$$f_I \circ P_I = P_I \circ f$$

where $f_I := (f_i)_{i \in I}$ *denotes the components of* f *according to the index set* I *and* $P_I : \mathbb{R}^n \to \mathbb{R}^{|I|}$ *denotes the canonical projection onto the states* x_I*, i.e.* $P_I(x) := (x_i)_{i \in I}$.

Remark

Finding an optimal decomposition is only based on the sparsity graph of f *and the product structure of* X*. First decompose* $X = X_1 \times \ldots \times X_k$ *(up to permutation of coordinates) and construct the sparsity graph of* f *with respect to the states* $x_i \in X_i$*. The minimal subsystems are characterized by the resulting sparsity graph and can be found fast using well known algorithms.*

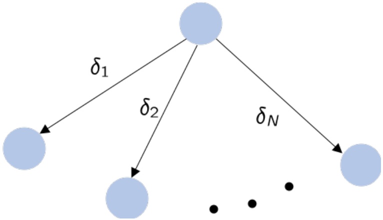
A hard example

The graph of the dynamics considered in Tacchi et al. (2019) is given by

Our approach does not lead to a reduction for this example.

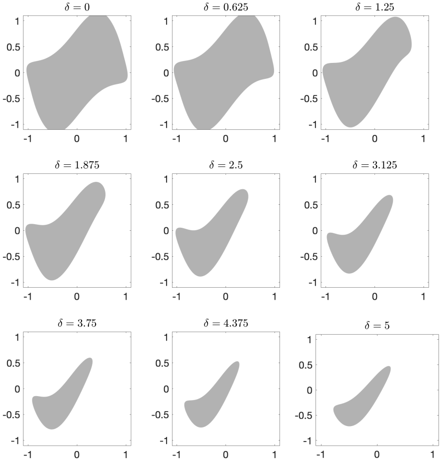
Numerical example

For a Van der Pol oscillator with cherry structure of the form

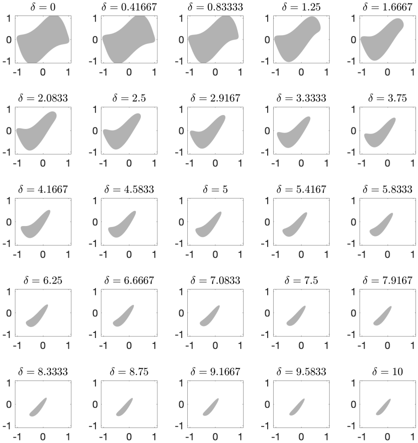


Numerical example

We get



for degree $k = 8$ and $N = 9$ and total dimension 20 and



for degree $k = 8$ and $N = 25$ and total dimension 52.

Conclusion

Contribution of this work

1. Can be applied in a similar way for the region of attraction and global attractors
2. First sparse method to approximate the MPI set, global attractors and the region of attraction with guaranteed convergence

Conclusion

Contribution of this work

1. Can be applied in a similar way for the region of attraction and global attractors
2. First sparse method to approximate the MPI set, global attractors and the region of attraction with guaranteed convergence

Outlook and perspectives

1. Extending to sparse control systems.
2. Improve/increase exploiting sparse structures.
3. Exploit polynomial structure of f and not only focusing sparse coupling; as for example term sparsity.
4. Coordinate free formulation.