A RIGOROUS NUMERICAL ALGORITHM FOR CONTROLLABILITY

FRITZ COLONIUS AND TOMASZ KAPELA

Abstract. The paper presents a rigorous numerical algorithm for controllability, based on the Krawczyk operator. Two examples are included.

1. Introduction

The purpose of this paper is to provide a rigorous numerical method that allows us to confirm controllability results. We consider the control system

\[ \dot{x}(t) = f(x(t), u(t)) \]

where \( f : \mathbb{R}^n \times U \rightarrow \mathbb{R}^n \) is a \( C^1 \)-function with respect to the first argument and \( U \subset \mathbb{R}^m \) is a set of admissible control values \( (u(t) \in U \text{ for all } t) \). For simplicity, we assume that the relevant initial value problems with \( x(0) = x_0 \) and control function \( u \) are uniquely solvable with controlled trajectories \( \varphi(t, x_0, u) \), \( t \in \mathbb{R} \).

A characteristic of controllability properties (in contrast to solution of differential equations) is that the control values can be chosen in dependence on time from a set of admissible control values. In order to establish rigorous controllability results we fix two subsets \( X \) and \( Y \) in the state space and ask, if we can connect each point in \( X \) to each point in \( Y \) by controlled trajectories, i.e., for all \( x_0 \in X \) and \( y_0 \in Y \) there are \( T > 0 \) and a control function \( u \) such that

\[ \varphi(T, x_0, u) = y_0. \]

In order to simplify the problem, we consider only piecewise constant controls and assume that the number of control values is finite. Thus we switch between a finite number of autonomous ODEs and our task is to compute the switching times. This can be reformulated as a problem to find a zero of a function. We use the Krawczyk operator [K] to solve this problem; the numerical implementation is based on the package CAPD, see [Capd]. The source code of the software performing proofs described in that article can be downloaded from Web page [TK].

Good initial guesses are of fundamental importance for rigorous numerical proofs. Here we need initial guesses for the desired trajectory or, more precisely, for the piecewise constant control and the switching times. For simplicity we only consider the two dimensional case \( (n = 2) \) and assume that \( u(t) \in \{u_-, u^+\} \subset \mathbb{R} \). We remark, that the restriction to two control values is not as restrictive as it may appear since in many applications it suffices to take only extremal values of the control range.

In order to check our method, we study two (two-dimensional) systems where delicate controllability problems occur. The problem is if two given
points in the state space can be connected by a trajectory corresponding to a control with one switching. The first system from H¨ackl/Schneider [HaSch] concerns the existence of control-homoclinic orbits. The second system from Gayer [Ga] is the escape equation and concerns bifurcation problems for control sets. In both examples we show rigorous results and compare them with the numerical results in the cited references.

We hope that in the future it will be possible to do computations for problems with more than one switching, in state spaces of dimensions higher than two, and also with higher dimensional controls.

We remark, that, in the control literature, the controllability problem for two points has also been considered as an optimal control problem: If the points can be connected by a trajectory, they can (under mild assumptions) also be connected by a time optimal trajectory. The resulting optimal control problem leads to a two-point boundary value problem, which then is solved numerically. While this yields a numerical approximation, it does not yield a rigorous proof confirming controllability, contrary to our approach. We are not aware of any other rigorous numerical results that allow to establish controllability.

The contents of this paper are as follows:
In the next section we recall the definition and properties of the Krawczyk operator. Then our algorithm is presented. Here we put particular emphasis on describing how to find good initial guesses. Finally, the announced examples are analyzed.

2. Interval Krawczyk Method

Our controllability result will be based on a transformation to a problem to find a zero of a function. A basic tool will be the Krawczyk operator. We cite the following result from Krawczyk [K].

Assume that:
- $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a $C^1$ function,
- $X \subset \mathbb{R}^n$ is an interval set,
- $\bar{x} \in X$
- $C \in \mathbb{R}^{n \times n}$ is a linear isomorphism.

Then the Krawczyk operator is given by

$$K(\bar{x}, X, F) := \bar{x} - CF(\bar{x}) + (Id - C[DF(X)])(X - \bar{x}).$$

Theorem 2.1. With the assumptions and notation introduced above, the following holds:

1. If $x^* \in X$ and $F(x^*) = 0$, then $x^* \in K(\bar{x}, X, F)$.
2. If $K(\bar{x}, X, F) \subset \text{int}X$, then there exists a unique $x^* \in X$ such that $F(x^*) = 0$.
3. If $K(\bar{x}, X, F) \cap X = \emptyset$, then $F(x) \neq 0$ for all $x \in X$.

Remark. To compute the Krawczyk operator we need the value of function $F$ in the point $\bar{x}$ and interval hull of its derivative $[DF]$ computed on the whole set $X$. Although $C$ can be any non-degenerate matrix, the best choice is to take an approximation of the inverse of $DF(\bar{x})$. 
3. Searching for the switching times

It is convenient to introduce the following notation. Consider the autonomous equations in \( \mathbb{R}^n \)
\[
\dot{x} = f_i(x) := f(x, u_i), \text{ where } i \in \{1, 2, \ldots, n\}.
\]
generating flows \( \varphi_i(t, x) \). Thus they correspond to the constant controls \( u_i \).
Since we assume that \( f \) is a \( C^1 \)-function with respect to \( x \), the flows are \( C^1 \).

Fix \( x_0, y_0 \in \mathbb{R}^n \). We want to find times \( t_1, t_2, \ldots, t_n \geq 0 \) such that we can go from the point \( x_0 \) to the point \( y_0 \) following the first flow for the time \( t_1 \) and then the second flow for the time \( t_2 \), etc. This means that we are searching for the solutions \((t_1, \ldots, t_n)\) of the equation
\[
\varphi_n(t_n, \varphi_{n-1}(t_{n-1}, \ldots, \varphi_1(t_1, x_0))) = y_0
\]
which is the same as searching for a zero of the function \( F_{x_0,y_0} : \mathbb{R}^n \to \mathbb{R}^n \) given by
\[
F_{x_0,y_0}(t_1, \ldots, t_n) := \varphi_n(t_n, \varphi_{n-1}(t_{n-1}, \ldots, \varphi_1(t_1, x_0))) - y_0.
\]
To this function we can apply the Interval Krawczyk method.

**Theorem 3.1.** If for some interval set \( T \subset \mathbb{R}^n_+ \) and a point \( t_0 \in T \) the rigorously computed Krawczyk operator \( K(t_0, T, F_{x_0,y_0}) \) is contained in the interior of \( T \), then there exists a trajectory joining the points \( x_0 \) and \( y_0 \). Explicitly, there exist times \((t_1, \ldots, t_n)\) such that \((t_1, \ldots, t_n) \in T \) and \( y_0 = \varphi_n(t_n, \varphi_{n-1}(t_{n-1}, \ldots, \varphi_1(t_1, x_0))) \).

**Proof:** For an application of Theorem 2.1 we have to show that \( F \) is a \( C^1 \)-map. This follows, since by assumption the flows are \( C^1 \) and, by the chain rule, the partial derivatives \( \frac{\partial F}{\partial t_i} \) exist and are continuous:
\[
\frac{\partial F}{\partial t_j} = (\varphi_n)_* \cdots (\varphi_{i+1})_* f(\varphi_i(t_1, \varphi_{i-1}(t_{i-1}, \ldots, \varphi_1(t_1, x_0)))�\]
here the linearized flows \((\varphi_j)_*\) are evaluated as
\[
(\varphi_j)_* = (\varphi_j)_* (t_{j+1}, f(\varphi_j(t_j, \varphi(t_{j-1}, \ldots, \varphi(t_1, x_0))))).
\]

We can easily generalize this result by replacing the initial and final points by interval sets.

Let \( X \) and \( Y \) be two interval sets in \( \mathbb{R}^n \) and define
\[
\tilde{F}_{X,Y}(t_1, \ldots, t_n) = \varphi_n(t_n, \varphi(t_{n-1}, \ldots, \varphi(t_1, X))) - Y.
\]
Then we have \( F_{x_0,y_0}(t_1, \ldots, t_n) \in \tilde{F}_{X,Y}(t_1, \ldots, t_n) \) for each \((x_0, y_0) \in X \times Y \).

**Corollary 3.2.** If for some interval set \( T \subset \mathbb{R}^n_+ \) and a point \( t_0 \in T \) the rigorously computed Krawczyk operator \( K(t_0, T, \tilde{F}_{X,Y}) \) is contained in the interior of \( T \), then for each \((x_0, y_0) \in X \times Y \) there exists trajectory from the point \( x_0 \) to \( y_0 \).

**Proof.** Let \((x_0, y_0) \in X \times Y \). Then \( K(t_0, T, F_{x_0,y_0}) \subset K(t_0, T, \tilde{F}_{X,Y}) \subset intT \). Theorem 2.1 yields the existence of a point \( t^* = (t_1^*, \ldots, t_n^*) \in T \) such that \( F_{x_0,y_0}(t^*) = 0 \). Hence from the point \( x_0 \) we can reach the point \( y_0 \) by
first going for time $t_1^*$ with the first flow, then with second one for time $t_2^*$, etc.

**Remark.** In order to obtain a function $F$ on $\mathbb{R}^n$, we need that the number of switching times is equal to the dimension of the state space. If more switching times are needed, we have to split up the problem into subproblems. This, however, will require good estimates for the intermediate end points.

### 4. Algorithm isReachable

From now on we restrict our attention to two-dimensional case ($n = 2$). For two given sets $X$ and $Y$ if algorithm `isReachable` returns `true` then each point in the set $Y$ can be reached from each point in the set $X$. On the other hand if this algorithm returns `false` we cannot conclude that there are points in $Y$ that are not reachable from some points in $X$.

**INPUT**
- $\varphi_1(t, x)$, $\varphi_2(t, x)$ - two dynamical flows,
- $X$ - an interval set of "starting" points,
- $Y$ - an interval set, whose reachability we want to check,

**OUTPUT**
- `true` if we succeeded to show that every point in $Y$ can be reached from each point in $X$,
- `false` otherwise.

We return `true` if and only if for some interval set $T$ and $t_0 \in T$ the rigorously computed Krawczyk operator $K(t_0, T, \bar{F}_{X,Y})$ is subset of $\text{int}T$. But for this we need good approximations of times $T$ and $t_0 \in T$. On one hand the set $T$ should be as small as possible to enable rigorous computation and to provide good approximation of the Krawczyk operator. On the other hand it should be also big enough to guarantee that all "switching" times are included in $T$ and that the values of the Krawczyk operator, even if it is overestimated, form a subset of the interior of $T$.

In the first step of algorithm we want to approximate the "switching" times and define a "good" set $T$. There are several possibilities:
- we may already have very good approximations from analytic knowledge or from numerical simulations and we want to define set $T$ manually.
- we may have only rough estimates and we want to improve them.
- we do not have any initial guesses and we need an algorithm to find estimates and to define the set $T$.

In the first case we just use those approximations. A possible method for the third case is presented in subsection 4.2. This method can be also used to improve the existing estimates. If the used approximations method fails and we do not have any initial guesses for the set $T$ then we return `false`.

The main part of this algorithm is to compute rigorously the Krawczyk operator and to check a suitable inclusion. We use the $C^1$-Lohner algorithm (see [ZLo]) implemented in the CAPD package [Capd], which provides rigorous bounds for trajectories and derivatives of dynamical flows. We can perform computations using various values of parameters such as the time.
step and the order of the Taylor method. We can also try to increase or decrease the set \( T \).

If one of these attempts succeeds (which means that values of the Krawczyk operator are contained in the interior of the set \( T \)) then we return \texttt{true}. If all of them fail we return \texttt{false}.

4.1. Method of intersection times approximation. We have two dynamical flows \( \varphi_1(t, x), \varphi_2(t, x) \) and two points \( x_0 \) and \( y_0 \). We want to find approximations \( \tilde{t}_1 \) and \( \tilde{t}_2 \) of times \( t_1 \) and \( t_2 \), respectively, such that \( \varphi_2(t_2, \varphi_1(t_1, x)) = y_0 \). For this we perform a simulation of the trajectory starting in the point \( x_0 \) and the backwards trajectory starting in the point \( y_0 \). Then we search for an estimated intersection point and corresponding times \( \tilde{t}_1 \) (from point \( x_0 \) to the intersection point) and \( \tilde{t}_2 \) (from the intersection point to \( y_0 \)).

Let \( \Psi(h,x) \) be a non-rigorous method for integrating ordinary differential equations (for a given time step \( h \) and starting point \( x \) it returns an estimation of \( \varphi(h, x) \)). We follow the trajectory of the flow \( \varphi_1 \) starting in the point \( x_0 \) up to time \( t_1^{\text{max}} \) using a time step \( h_1 > 0 \) and the backwards trajectory of the point \( y_0 \) under the flow \( \varphi_2 \) with a time step \( h_2 > 0 \) up to time \( t_2^{\text{max}} \). We have \( x_n = \Psi_1(h_1, x_{n-1}) \) for \( n = 1, 2, \ldots, N_1 \) and \( y_k = \Psi_2(-h_2, y_{k-1}) \) for \( k = 1, \ldots, N_2 \) where \( N_1 = [t_1^{\text{max}} / h_1] \) and \( N_2 = [t_2^{\text{max}} / h_2] \). We assume now that during a single time step the trajectories are straight lines. Then for each \( k \) and \( n \) we check if the segments \( \overline{x_{n-1}x_n} \) and \( \overline{y_{k-1}y_k} \) intersect. Thus we search for a solution \((s_1, s_2)\) of the equation

\[
x_{n-1} + s_1 \cdot x_{n-1}x_n = y_{k-1} + s_2 \cdot y_{k-1}y_k.
\]

If for some \( n \) and \( k \) we have \( 0 \leq s_1 \leq 1 \) and \( 0 \leq s_2 \leq 1 \) then we set \( \tilde{t}_1 = h_1 \cdot [(n-1) + s_1] \) and \( \tilde{t}_2 = h_2 \cdot [(k-1) + s_2] \).

This method needs a lot of computation to get good approximations, because small time steps lead to many segments to be checked. A method to circumvent this is to use first a large time step to get rough estimates; once we have intersecting segments for some \( n \) and \( k \), we can approximate the trajectory between \( x_{n-1} \) and \( x_n \) and between \( y_{k-1} \) and \( y_k \) with a smaller time step and again search for intersecting segments. By iterating this division procedure we can come very close to the real intersection point. The same can be done if we already have some rough estimates. We can search for the intersection only locally, in order to decrease the number of segments. In the implementation we take \( M \) segments before the approximate intersection point and \( M \) segments after (usually \( M \) was set to 100).

It is possible that there exist more than one intersection point. In this case we can apply several strategies: just take first one found; take the one with the smallest possible \( \tilde{t}_1 \) or \( \tilde{t}_2 \); or each of them or just pick one manually. The Krawczyk method needs that this intersection point is isolated and that, at least in the set \( T \), there is exactly one ”switching” time for each pair \((x_0, y_0)\) in \( X \times Y \).

4.2. Times estimates for the Krawczyk method. We have two interval sets \( X \) and \( Y \) and we want to find a set \( T \) and a point \( t_0 \), which are good initial approximations needed by Krawczyk method.
If the sets $X$ and $Y$ consist of one point $x_0$ and $y_0$ respectively, we can use the algorithm described in previous section to find approximations $\tilde{t}_1$ and $\tilde{t}_2$. Then we set $t_0 = (\tilde{t}_1,\tilde{t}_2)$ and $T = [\tilde{t}_1 - \delta_1, \tilde{t}_1 + \delta_1] \times [\tilde{t}_2 - \delta_2, \tilde{t}_2 + \delta_2]$. The real positive numbers $\delta_1$ and $\delta_2$ are chosen manually by a process of trial and error.

If the set $X$ or $Y$ consists of more than one point, then we can take one point from each of them and proceed like for one point sets choosing appropriate $\delta_1$ and $\delta_2$. We can also perform more simulations to obtain better approximations. Usually we have infinite sets, so we have to choose some finite subsets.

Let $X_0$ and $Y_0$ be finite subsets of $X$ and $Y$, respectively (if $X$ or $Y$ is an interval set then we can take subset of all vertices). For each $x_0 \in X_0$ and $y_0 \in Y_0$ we search for times $\tilde{t}_1$ and $\tilde{t}_2$ using the algorithm from previous section. Then we define $T$ to be an interval set containing all pairs $(\tilde{t}_1,\tilde{t}_2)$. It should not be the smallest possible interval set, because we want to show that Krawczyk operator is contained in the interior of this set (if our estimates are accurate, then they will also be contained in the values of the Krawczyk operator). In our implementation we calculate the interval hull of all pair $(\tilde{t}_1,\tilde{t}_2)$ and then multiply the diameter of that interval hull by some factor (usually equal to 2). As $t_0$ we take the middle point of $T$.

If for some point $y_0$ and $x_0$ we do not find an intersection point, then we say that our estimation procedure failed.

5. Examples

The first example is the Takens-Bogdanov oscillator, where Häckl/Schneider [HaSch] (compare also Colonius/Kliemann [CK]) could analytically confirm the existence of a control range such that two given points can be connected by a trajectory with one switching (while constant controls do not allow this). Furthermore, they established a nonrigorous numerical example of such a control range. The second one is the escape equation. Here, for controllability problems, Gayer [Ga] could establish numerical estimates for the required size of the control range.

5.1. The Takens-Bogdanov oscillator. The controlled Takens-Bogdanov oscillator is given by

$$\begin{align*}
\dot{x} &= y \\
\dot{y} &= \lambda_1 + \lambda_2 x + x^2 + xy + u(t), \\
u(t) &\in [-\rho, \rho]
\end{align*}$$

with real parameters $\lambda_1, \lambda_2, \rho$. It is known (cp. [CK, Section 9.4] for the following assertions) that for parameters $(\lambda_1, \lambda_2)$ in a certain subset $S_2 \subset \mathbb{R}^2$ the uncontrolled system has a stable fixed point and a hyperbolic fixed point without homoclinic orbit. For small $\rho > 0$ there are two simply connected control sets (i.e., maximal subsets of complete controllability), one around each of the equilibria. These control sets contain all stable and all hyperbolic equilibria, respectively, corresponding to constant controls in $[-\rho, \rho]$. For the parameter values in $S_2$ given by

$$\lambda_1 = -0.3, \lambda_2 = -1$$
one can prove that there is $\rho > 0$ such that the control set containing the hyperbolic equilibrium is not simply connected and surrounds the other control set. A numerical computation for the control range given by $\rho = 0.05$ shows this behavior; it is reproduced in Figure 1: There are control homoclinic orbits, while it is known that for constant controls $u \in [-\rho, \rho]$ there is no homoclinic orbit.

Our goal was to prove that for these parameter values there is a control heteroclinic orbit.

We will show that there exists a trajectory joining the point $x_0 = (1.26, 0.0)$, which is the hyperbolic fixed point for $u(t) = -0.0276$ (it is near the hyperbolic fixed point for $u(t) = -\rho$, and the fixed point $y_0 = (1.207106781, 0.0)$ for $u(t) = \rho$. This trajectory goes around the invariant control set.

To show this we will use the following two control values: first $u_1(t) = \rho = -0.05$ and then $u_2(t) = 0.04$. So we have two vector fields

$$f_1(x, y) = (y, x^2 + xy - x - 0.35),$$
$$f_2(x, y) = (y, x^2 + xy - x - 0.26).$$

Our initial guesses for the times are $t_0 = (t_1, t_2) = (2.9039196889, 10.559934)$. We set $T = [t_1 - \delta_1, t_1 + \delta_1] \times [t_2 - \delta_2, t_2 + \delta_2]$ where $\delta_1 = 2. e - 9$ and $\delta_2 = 2. e - 3$. In the first approximation we obtain two intersection points of trajectory of $x_0$ and backwards trajectory of $y_0$. We choose the point with the smaller...
time $t_1$ and then we start several times our approximation procedure with decreasing time steps.

To compute the Krawczyk operator we use the $C^1$ Lohner algorithm with time step equal to 0.01 and a 6th order Taylor method. We obtain this values by performing computations with different time steps and orders and then comparing diameters of computed Krawczyk operators.

Finally, we obtain that

$$K(t_0, T, F_{x_0,y_0}) \subset [2.9039196882987688, 2.9039196895134038] \times [10.558743378108383, 10.56125701344494] \subset int(T).$$

From Theorem 3.1 it follows that there exists a trajectory joining the points $x_0$ and $y_0$.

We also performed computations with the same parameters for small interval sets $X$ and $Y$ around the points $x_0$ and $y_0$ (we took radius $10^{-14}$) and we obtained

$$K(t_0, T, F_{X,Y}) \subset [2.903919687831086, 2.9039196899790425] \times [10.557940607196224, 10.5619284723529] \subset int(T).$$

Theorem 3.2 implies the existence of trajectories from each point in the set $X$ to each point in $Y$. In this example, the sets $X$ and $Y$ are very small because we are very close to the hyperbolic point.

In both cases the computation times was approximately 1 second.

5.2. **Escape equation.** The controlled escape equation is given by
\[ \dot{x} = y \]
\[ \dot{y} = \lambda y + x^2 - x + u(t), \]
\[ u(t) \in [-\rho, \rho] \]

where \( \lambda \) and \( \rho \) are real parameters.

Figures 3 and 5 show simulations of control sets for \( \lambda = -0.1 \) and two values of \( \rho \) due to T. Gayer (for more details see [Ga]). For \( \rho = 0.041 \) there are two control sets, variant control set surrounds invariant one. For \( \rho = 0.04 \) we also have two control sets, but the variant control set is simple connected. For \( \lambda = -0.1 \) and \( \rho = -0.041 \) we will prove the existence of a periodic orbit (see Figure 4), going around the invariant control set. For \( \rho = 0.04 \) we proved that the simulation is not exact and that for this parameter there still exists the controlled trajectory starting and ending in the variant control set and going around the invariant control set (see Figure 6).

We set \( x_0 = y_0 = (0.95,0) \), this is a point near the leftmost hyperbolic equilibrium. We have two vector fields

\[ f_1(x, y) = (y, -0.1y + x^2 - x - 0.041), \]
\[ f_2(x, y) = (y, -0.1y + x^2 - x + 0.041) \]
and approximations of switching times $t_1 = 5.18658693445$ and $t_2 = 7.18515699298$. Again we take $T = [t_1 - \delta_1, t_1 + \delta_1] \times [t_2 - \delta_2, t_2 + \delta_2]$ to be interval ball around our time approximations ($\delta_1 = 2.\varepsilon - 9$, $\delta_2 = 2.\varepsilon - 6$). The first approximation brings us two intersection points between the trajectory of $x_0$ and the backwards trajectory of $y_0$, but this time we take the point with larger $t_1$ and start our approximation method near this point (the same can be done with the other point).

We check that for parameters: time step equal to 0.0425 and the 6th order Taylor method, the values of the Krawczyk operator are contained in interior of the set $T$. Hence Theorem 3.1 implies the existence of a trajectory joining the points $x_0$ and $y_0 = x_0$. Thus a periodic orbit is obtained (cp. Figure 4).

In the case $\rho = 0.04$, to prove that the variant control set is not simple connected we take two hyperbolic stable points $x_0 = (1.0385, 0.0)$ and $y_0 = (0.9582575695, 0.0)$ which correspond to controls $u(t) = -0.03998225$ and $u(t) = 0.04$ respectively and we show that there is a controlled trajectory joining this points going around the invariant control set. Our simulations shows that for that range of the control it does not exist the controlled periodic orbit with only one switching. To reach the point $y_0$ in a finite time we cannot use the maximal value of the control. We use controls $u_1 = -0.04$ and $u_2 = 0.0399$. Our approximation procedure brings us $t_1 = 13.769247218$ and $t_2 = 11.400673878$. Using time step 0.0825 and a 10th order of the Taylor method we showed that the interval Krawczyk operator is contained in the interior of the set $T = [t_1 - \delta_1, t_1 + \delta_1] \times [t_2 - \delta_2, t_2 + \delta_2]$ where
\( \delta_1 = 1 \cdot e^{-8} \) and \( \delta_2 = 2 \cdot e^{-3} \). Hence the existence of the trajectory joining points \( x_0 \) and \( y_0 \) follows from Theorem 3.1.

**Acknowledgement.** Tomasz Kapela was supported by CTS (European, Control Training Site) in the form of a fellowship. We thank Tobias Gayer for the permission to use Figure 3 from [Ga].

**References**

Figure 6. Controlled heteroclinic orbit for the controlled escape equation with $\rho = 0.040$

Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany, fritz.colonius@math.uni-augsburg.de

Instytut Matematyki, Akademia Pedagogiczna, Kraków, Polska, tkapela@ap.krakow.pl