EE C222/ME C237 - Spring'18 - Lecture 7 Notes[1]

Murat Arcak

February 7 2018

Existence and Uniqueness Theorems for ODEs

Khalil (Section 3.1), Sastry (Section 3.4)

$$\dot{x} = f(t,x) \quad x(0) = x_0 \tag{1}$$

<u>Theorem 1:</u> $f(t,x)$ locally Lipschitz in x and continuous in t

$\Rightarrow$ existence and uniqueness on some finite interval $[0,\delta]$.

<u>Sketch of the proof:</u> From the local Lipschitz assumption, we can find $r > 0$ and $L > 0$ such that

$$|f(t,x) - f(t,y)| \leq L|x-y| \;\; \forall x, y \in \{x \in \mathbb{R}^n : |x - x_0| \leq r\}.$$

If $x(t)$ is a solution, then:

$$x(t) = \underbrace{x_0 + \int_0^t f(\tau, x(\tau))d\tau.}_{=: T(x)(t)}$$

To apply the Contraction Mapping Theorem:

1. Choose δ small enough that T maps the following subset of $C^n[0,\delta]$ to itself :

   $$U = \{x \in C^n[0,\delta] : |x(t) - x_0| \leq r \; \forall t \in [0,\delta]\},$$

   i.e.

   $$|x(t) - x_0| \leq r \quad \forall t \in [0,\delta] \quad \Rightarrow \quad |T(x)(t) - x_0| \leq r \quad \forall t \in [0,\delta]. \tag{2}$$

   To find such a δ note that

   $$T(x)(t) - x_0 = \int_0^t f(\tau,x(\tau))d\tau = \int_0^t \Big(f(\tau,x(\tau)) - f(\tau,x_0) + f(\tau,x_0)\Big)d\tau$$

   $$\begin{aligned} |T(x)(t) - x_0| &\leq \int_0^\delta |f(\tau,x(\tau)) - f(\tau,x_0)|d\tau + \int_0^\delta |f(\tau,x_0)|d\tau \\ &\leq \int_0^\delta L|x(\tau) - x_0|d\tau + \int_0^\delta h d\tau \quad \text{where } h \text{ is a bound on } |f(\tau,x_0)| \\ &\leq (Lr+h)\delta. \end{aligned}$$

   Thus, by choosing $\delta \leq \frac{r}{Lr+h}$ we ensure that the implication (2) holds.

2. Show that T is a contraction in U, *i.e.*, there exists $\rho < 1$ s.t.

   $$x, y \in U \implies |T(x) - T(y)|_C \leq \rho |x-y|_C.$$

Note that, for all $t \in [0, \delta]$,

$$
\begin{aligned}
|T(x)(t) - T(y)(t)| &= \int_0^t |f(\tau, x(\tau)) - f(\tau, y(\tau))| d\tau \\
&\leq L \int_0^t |x(\tau) - y(\tau)| d\tau \\
&\leq \underbrace{L\delta}_{=:\rho} \max_{\tau \in [0,\delta]} |x(\tau) - y(\tau)| = \rho |x - y|_C.
\end{aligned}
$$

Therefore,

$$
|T(x) - T(y)|_C = \max_{t \in [0,\delta]} |T(x)(t) - T(y)(t)| \leq \rho |x - y|_C
$$

and $\rho < 1$ if $\delta \leq \frac{r}{Lr+h}$ as prescribed above.

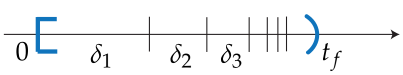
Theorem 2: $f(t, x)$ *globally* Lipschitz in x uniformly[2] in t, and continuous in $t \Longrightarrow$ existence and uniqueness on $[0, \infty)$.

[2] same L works for all t

Proof: Choose a δ that doesn't depend on x_0 and apply Theorem 1 repeatedly to cover $[0, \infty)$. This is possible because L works everywhere and we can pick r as large as we wish. Indeed, for any $\delta < \frac{1}{L}$, we can choose r large enough that $\delta \leq \frac{r}{Lr+h}$.

Q: Why can't we do this in Theorem 1?

A: δ depends on x_0 (no universal L) and x_0 changes at the next iteration. We can't use the same δ in every iteration:



- The theorems above are sufficient only, and can be conservative:

Example: $\dot{x} = -x^3$ is not globally Lipschitz but

$$
x(t) = \text{sgn}(x_0) \sqrt{\frac{x_0^2}{1 + 2tx_0^2}}
$$

is defined on $[0, \infty)$.

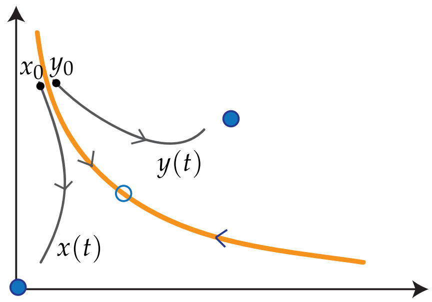
Continuous Dependence on Initial Conditions and Parameters

Theorem 3: (Continuous dependence on initial conditions) Let $x(t), y(t)$ be two solutions of $\dot{x} = f(t, x)$ starting from x_0 and y_0, and remaining in a set with Lipschitz constant L on $[0, \tau]$. Then, for any $\epsilon > 0$, there exists $\delta(\epsilon, \tau) > 0$ such that

$$
|x_0 - y_0| \leq \delta \Longrightarrow |x(t) - y(t)| \leq \epsilon \quad \forall t \in [0, \tau].
$$

- This conclusion does not hold on infinite time intervals (even if f is globally Lipschitz).

Example: bistable system



If ϵ is smaller than the distance between the two stable equilibria, no choice of δ guarantees $|x(t) - y(t)| \leq \epsilon \;\; \forall t \geq 0$.

- Theorem 3 also shows continuous dependence on parameter μ in $f(t, x, \mu)$ if we rewrite the system equations as:

$$\begin{aligned} \dot{x} &= f(t,x,\mu) \\ \dot{\mu} &= 0 \end{aligned} \qquad X = \begin{bmatrix} x \\ \mu \end{bmatrix} \qquad \dot{X} = F(t,X) \triangleq \begin{bmatrix} f(t,x,\mu) \\ 0 \end{bmatrix},$$

  where μ appears as a state variable with initial condition $\mu(0) = \mu$.

Q: How do you reconcile bifurcations with continuous dependence on parameters? We could pick two values of the bifurcation parameter arbitrarily close, but one below and one above the critical value, thereby expecting a drastic difference in the solutions.

A: The two solutions are close in the short term (Theorem 3 holds on finite time intervals); the drastic difference builds up over time.

Sensitivity to Parameters

Consider the system

$$\dot{x} = f(t,x,\mu) \quad x \in \mathbb{R}^n, \mu \in \mathbb{R}^p \tag{3}$$

where μ is a vector of p parameters, and let $\phi(t, x_0, \mu)$ denote the trajectories starting at the initial condition x_0.

To determine to what extent this trajectory depends on the parameters we define the $n \times p$ *sensitivity matrix*:

$$S(t, x_0, \mu) := \frac{\partial \phi(t, x_0, \mu)}{\partial \mu} = \left[\frac{\partial \phi(t, x_0, \mu)}{\partial \mu_1} \cdots \frac{\partial \phi(t, x_0, \mu)}{\partial \mu_p} \right], \tag{4}$$

where each column is the sensitivity with respect to a particular parameter.

To see how $S(t, x_0, \mu)$ can be computed numerically, first note that $\phi(t, x_0, \mu)$ satisfies the equation (3), that is,

$$\frac{\partial \phi(t, x_0, \mu)}{\partial t} = f(t, \phi(t, x_0, \mu), \mu).$$

Next, differentiate both sides with respect to μ:

$$\frac{\partial^2 \phi(t, x_0, \mu)}{\partial t \partial \mu} = \frac{\partial f}{\partial x}(t, \phi(t, x_0, \mu), \mu)\frac{\partial \phi(t, x_0, \mu)}{\partial \mu} + \frac{\partial f}{\partial \mu}(t, \phi(t, x_0, \mu), \mu)$$

and use the definition of the sensitivity matrix to rewrite this as

$$\frac{\partial S(t, x_0, \mu)}{\partial t} = \frac{\partial f}{\partial x}(t, \phi(t, x_0, \mu), \mu)S(t, x_0, \mu) + \frac{\partial f}{\partial \mu}(t, \phi(t, x_0, \mu), \mu).$$

Thus, S can be computed by numerical integration of (3) simultaneously with

$$\dot{S} = \frac{\partial f}{\partial x}(t, x, \mu)S + \frac{\partial f}{\partial \mu}(t, x, \mu).$$

The initial condition for S is $\frac{\partial x_0}{\partial \mu} = 0$, assuming that x_0 is independent of the parameters.

Example: For the harmonic oscillator

$$\begin{aligned} \dot{x}_1 &= -\mu x_2 \\ \dot{x}_2 &= \mu x_1 \end{aligned}$$

we have

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 0 & -\mu \\ \mu & 0 \end{bmatrix} \qquad \frac{\partial f}{\partial \mu} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

Thus the sensitivity equation is

$$\dot{S} = \begin{bmatrix} 0 & -\mu \\ \mu & 0 \end{bmatrix} S + \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

Logarithmic Sensitivity

To compare the sensitivity with respect to multiple parameters $\mu_1, \ldots, \mu_p$ it is preferable to use the *logarithmic sensitivity*

$$\frac{\partial \phi(t, x_0, \mu)}{\partial \mu_i / \mu_i} = \frac{\partial \phi(t, x_0, \mu)}{\partial \ln \mu_i}$$

so that the denominator is dimensionless and represents the change in the parameter μ_i relative to its nominal value. This means that the ith column of the sensitivity matrix S in (4) must be multiplied with the nominal parameter μ_i, $i = 1, \ldots, p$, before these columns are compared for the relative significance of the parameters.

Application to Parameter Tuning and Identification

Sensitivity equations are useful for solving a class of optimization problems of the form

$$\begin{array}{c}\min_\mu J(\mu) = \int_{t_0}^{t_1} q(t, x(t), \mu) dt \\ \text{subject to } \dot{x} = f(t, x, \mu), \ x(t_0) = x_0.\end{array}$$

For example, one may take $q(t, x) = |h(x(t)) - r(t)|^2$ to penalize the error between the output $y(t) = h(x(t))$ of a control system and a reference trajectory $r(t)$ to be followed. In this example $\dot{x} = f(t, x, \mu)$ represents the closed loop model with tunable control parameters μ.

In other applications $\dot{x} = f(t, x, \mu)$ may represent the model of a physical process with unknown parameters, and $q(t, x) = |h(x(t)) - r(t)|^2$ penalizes the error between the model prediction for a variable, $y(t) = h(x(t))$, and the experimental observation $r(t)$. Then the optimization problem above aims to find parameters that best fit the experimental data.

A typical optimization algorithm requires the gradient $\frac{\partial J(\mu)}{\partial \mu}$, which can be obtained with the help of the chain rule and the sensitivity equations:

$$\frac{\partial J(\mu)}{\partial \mu} = \int_{t_0}^{t_1} \left(\frac{\partial q}{\partial x}(t, x(t), \mu) S(t, x_0, \mu) + \frac{\partial q}{\partial \mu}(t, x(t), \mu) \right) dt.$$