On the maximum principle

Horia Cornean, d. 25/09/2008.

These notes are not completely rigorous, and we often use stronger conditions than necessary in order to simplify the presentation.

1 The one dimensional case

Let $f, f_0 \in C^2(\mathbb{R}^3)$ be two real functions, locally bounded in the C^2 norm. The control variables are generically denoted by $u : [0, \infty) \mapsto [-1, 1]$, where u is a piecewise continuous function.

For a given u and $T > 0$, let us consider the initial value problem:

$$\begin{aligned} &x'(t) = f(x(t), u(t), t), \quad 0 < t < T, \\ &x(0) = 0. \end{aligned} \tag{1.1}$$

There exists a unique solution x, continuous on $[0, T]$, and also continuously differentiable except at the discontinuities of u.

1.1 Fixed T

We want to maximize the following integral:

$$I(u) := \int_0^T f_0(x(t), u(t), t)dt, \tag{1.2}$$

by choosing a piecewise continuous "optimal" function $u^* : [0, T] \mapsto [-1, 1]$. We will provide sufficient conditions for the existence of such u^*'s. Clearly, u^* generates an optimal state function x^* which solves:

$$\begin{aligned} &{x^*}'(t) = f(x^*(t), u^*(t), t), \quad 0 < t < T, \\ &x^*(0) = 0. \end{aligned} \tag{1.3}$$

1.1.1 A regularization procedure

We allow here the image of u to be the whole real axis, and also modify the functional to be optimized.

If $M > 0$ is a large integer, consider the function $g_M : [-1, 1] \to [0, 1/(2M)]$, given by $g_M(v) := \frac{1}{2M}(1 - v^{2M^2})$. It is strictly concave, symmetric, has a maximum $1/(2M)$ at $v = 0$, $g_M(\mp 1) = 0$, $g_M'(\mp 1 \pm 0) = \pm M$, and $g_M''(\mp 1 \pm 0) = -M(2M^2 - 1) =: -C_M < 0$. Moreover, g_M' converges pointwise to 0 on $(-1, 1)$.

Define the function $G_M : \mathbb{R} \to \mathbb{R}$ such that:

$$\begin{aligned} &G_M(v) \\ &= [-C_M(v+1)^2/2 + M(v+1)]\theta(-v-1) - [C_M(v-1)^2/2 + M(v-1)]\theta(v-1) \\ &+ g_M(v)\theta(1 - v^2). \end{aligned} \tag{1.4}$$

G_M belongs to $C^2(\mathbb{R})$, is strictly concave and G_M' is a bijection. Inside the interval $[-1, 1]$ G_M is bounded by $1/M$, while outside the interval $[-1 - 1/\sqrt{M}, 1 + 1/\sqrt{M}]$ is as negative as $-\sqrt{M}$.

We consider the maximizing problem for:

$$I_M(u) := \int_0^T [f_0(x(t), u(t), t) + G_M(u(t))]dt, \tag{1.5}$$

where now $u : [0, T] \mapsto \mathbb{R}$, and u is piecewise continuous.

1.1.2 Approximation with a discrete model

Let $N > 1$ be an integer, and define:

$$\Delta t := T/N, \quad t_k = k\Delta t, \quad 0 \le k \le N. \tag{1.6}$$

Denote by u_M a piecewise continuous control function, and by x_M its corresponding state.

Consider the discrete dynamical system:

$$y(t_{k+1}) := y(t_k) + (\Delta t) f(y(t_k), u_M(t_k), t_k), \quad 0 \le k \le N-1, \tag{1.7}$$

$$y(t_0) = 0. \tag{1.8}$$

Lemma 1.1. *Denote by* $e_k := |x_M(t_k) - y(t_k)|$*. Then*

$$\lim_{N\to\infty} \max_{0\le k\le N} e_k = 0.$$

Proof. We can write

$$\begin{aligned} & x_M(t_{k+1}) - y(t_{k+1}) = x_M(t_k) - y(t_k) \\ & + \int_{t_k}^{t_{k+1}} \{f(x_M(t), u_M(t), t) - f(y(t_k), u_M(t_k), t_k)\}dt, \quad 0 \le k \le N-1. \end{aligned} \tag{1.9}$$

Using the mean value theorem for f we can find three positive constants C_1, C_2, C_3 such that for every $t_k \le t \le t_{k+1}$ we have:

$$\begin{aligned} & |f(x_M(t), u_M(t), t) - f(y(t_k), u_M(t_k), t_k)| \\ & \le C_1 e_k + C_2 \sup_{t_k \le t \le t_{k+1}} |u_M(t) - u_M(t_k)| + C_3 \Delta t. \end{aligned} \tag{1.10}$$

Using this in (1.9) we can write:

$$\begin{aligned} & e_{k+1} \le \alpha e_k + \beta_k, \quad 0 \le k \le N-1, \\ & \alpha := 1 + C_1\Delta t, \quad \beta_k := C_2 \Delta t \sup_{t_k \le t \le t_{k+1}} |u_M(t) - u_M(t_k)| + C_3(\Delta t)^2. \end{aligned} \tag{1.11}$$

By induction we derive the inequality:

$$e_j \le \alpha^j e_0 + \sum_{k=0}^{j-1} \alpha^k \beta_{j-k-1}, \quad 1 \le j \le N. \tag{1.12}$$

Note that $e_0 = 0$. Then since

$$\alpha^j = e^{j \ln(1+C_1\Delta t)} \le e^{jC_1\Delta t} \le e^{C_1 T}, \quad j \le N,$$

there exists $C_4 > 0$ such that

$$\begin{aligned} e_j & \le C_4 \sum_{k=0}^{N-1} \beta_k \\ & \le C_4 C_3 T^2/N + C_4 C_2 T \frac{1}{N} \sum_{k=0}^{N-1} \sup_{t_k \le t \le t_{k+1}} |u_M(t) - u_M(t_k)|, \quad 1 \le j \le N. \end{aligned} \tag{1.13}$$

Now the lemma is concluded since u_M is piecewise uniformly continuous.

□

1.1.3 A finite dimensional optimization problem

Using the previous lemma, we have that for any piecewise continuous control function u_M:

$$\begin{aligned} I_M(u_M) &= \int_0^T \{f_0(x_M(t), u_M(t), t) + G_M(u_M(t))\}dt \\ &= \lim_{N\to\infty} \sum_{k=0}^{N-1} \{f_0(x_M(t_k), u_M(t_k), t_k) + G_M(u_M(t_k))\}\Delta t \\ &= \lim_{N\to\infty} \sum_{k=0}^{N-1} \{f_0(y(t_k), u_M(t_k), t_k) + G_M(u_M(t_k))\}\Delta t. \end{aligned} \tag{1.14}$$

Now for a fixed N, define $F_M : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ given by

$$F_M(a_1, \dots, a_N, b_0, \dots b_{N-1}) = \sum_{k=0}^{N-1} \{f_0(a_k, b_k, t_k) + G_M(b_k)\}\Delta t. \tag{1.15}$$

(Note that $a_0 = t_0 = 0$, while F_M does not depend on a_N). We will maximize F subject to the restrictions

$$\phi_k(\mathbf{a}, \mathbf{b}) := a_{k+1} - a_k - f(a_k, b_k, t_k)\Delta t, \quad 0 \le k \le N-1, \quad a_0 = 0. \tag{1.16}$$

We form the Lagrangian:

$$F_M(\mathbf{a}, \mathbf{b}) - \sum_{k=0}^{N-1} p_k \phi_k(\mathbf{a}, \mathbf{b}) \tag{1.17}$$

where the p_k's are Legendre multipliers. Differentiating with respect to a_N we obtain $p_{N-1} = 0$. Differentiating with respect to a_k, $1 \le k \le N-1$ we obtain:

$$(\partial_1 f_0)(a_k, b_k, t_k)\Delta t - p_{k-1} + p_k + p_k(\partial_1 f)(a_k, b_k, t_k)\Delta t = 0.$$

Introduce the Hamilton function:

$$H(a, b, p, t) := f_0(a, b, t) + pf(a, b, t). \tag{1.18}$$

Now we see that the above extremum conditions for p's can be written as:

$$\begin{aligned} p_k &= p_{k-1} - (\partial_1 H)(a_k, b_k, p_k, t_k)\Delta t, \quad 1 \le k \le N-1, \\ p_{N-1} &= 0. \end{aligned} \tag{1.19}$$

From now on we make the following assumption:

Assumption 1.2. *H is smooth and jointly concave in the first two variables a and b.*

We see from (1.17) that an optimal b_k must solve the equation $\partial_2 H(a_k, b, p_k, t_k) + G_M'(b) = 0$, which has a unique solution due to the fact that G_M is strictly concave. Moreover, this solution can be expressed as a smooth function of its variables $\tilde{b}_K(a_k, p_k, t_k)$.

In conclusion, under the restrictions in (1.16), if the function F_M has a global maximum then it must be attained in some pair of vectors $\mathbf{a}$ and $\mathbf{b}$ which obey the following relations:

$$\begin{aligned} 0 &= \partial_2 H(a_k, \tilde{b}_k, p_k, t_k) + G_M'(\tilde{b}_k), \quad 0 \le k \le N-1, \\ p_k &= p_{k-1} - (\partial_1 H)(a_k, \tilde{b}_k, p_k, t_k)\Delta t, \quad 1 \le k \le N-1, \quad p_{N-1} = 0, \\ a_{k+1} &= a_k + f(a_k, \tilde{b}_k, t_k)\Delta t, \quad 0 \le k \le N-1, \quad a_0 = 0. \end{aligned} \tag{1.20}$$

It is not obvious how to solve such a system of equations. If we assume that p_0 is known, then we can uniquely solve the system by iteration. Then p_0 might be determined by imposing the condition $p_{N-1}(p_0) = 0$.

1.1.4 A Mangasarian type sufficiency condition

We will prove here that under Assumption 1.2, any solution to (1.20) provides a global maximum point for F_M. Let us assume that $\mathbf{a}^*, \mathbf{b}^*, \mathbf{p}$ is a solution of (1.20). Then we have:

$$
\begin{aligned}
& F_M(\mathbf{a},\mathbf{b}) - F_M(\mathbf{a}^*,\mathbf{b}^*) && (1.21)\\
&= \sum_{k=0}^{N-1} \{H(a_k,b_k,p_k,t_k) - H(a_k^*,b_k^*,p_k,t_k) + G_M(b_k) - G_M(b_k^*)\}\Delta t \\
&- \sum_{k=0}^{N-1} p_k\{f(a_k,b_k,t_k) - f(a_k^*,b_k^*,t_k)\}\Delta t.
\end{aligned}
$$

Using the concavity of H and G_M, and (1.20), we can write:

$$
\begin{aligned}
& H(a_k,b_k,p_k,t_k) - H(a_k^*,b_k^*,p_k,t_k) + G_M(b_k) - G_M(b_k^*) \\
&\leq (a_k - a_k^*)\partial_1 H(a_k^*,b_k^*,p_k,t_k) + (b_k - b_k^*)[\partial_2 H(a_k^*,b_k^*,p_k,t_k) + G_M'(b_k^*)] \\
&= -(a_k - a_k^*)(p_k - p_{k-1}), \quad 1 \leq k \leq N-1. && (1.22)
\end{aligned}
$$

Note that the above inequality is also formally true for $k = 0$, since $a_0 = a_0^* = 0$. Now using (1.16), (1.22), and $p_{N-1} = 0$ in (1.21) we obtain:

$$
\begin{aligned}
& F_M(\mathbf{a},\mathbf{b}) - F_M(\mathbf{a}^*,\mathbf{b}^*) && (1.23)\\
&\leq -\sum_{k=1}^{N-1} \{(a_k - a_k^*)(p_k - p_{k-1})\}\Delta t \\
&- \sum_{k=1}^{N-2} p_k\{[a_{k+1} - a_{k+1}^*] - [a_k - a_k^*]\}\Delta t = 0.
\end{aligned}
$$

Thus we have shown that F_M attains its global maximum at any solution of (1.20), provided that such solutions exist.

1.1.5 Back to the continuum

Now let us assume that the following system of equations:

$$
\begin{aligned}
0 &= \partial_2 H(x,u,p,t) + G_M'(u), && (1.24)\\
p' &= -(\partial_1 H)(x,u,p,t), \quad p(T) = 0, \\
x' &= f(x,u,t), \quad x(0) = 0,
\end{aligned}
$$

has a solution x_M^*, u_M^* and p_M^*. (The idea is to choose $p(0) = p_0$ as a parameter, solve the system, and then find a compatible p_0 from the condition $p(T;p_0) = 0$.)

If the continuous solution exists and if $\partial_{p_0} p(T,p_0) \neq 0$, then if N is sufficiently large, the system in (1.20) will admit a solution which will "converge" to the continuous one in the same manner as described in Lemma 1.1. We do not prove this statement.

Now if we go back to (1.14), we see that we can write the inequality:

$$
I_M(u_M) \leq \lim_{N\to\infty} \sum_{k=0}^{N-1} \{f(a_k^*,b_k^*,t_k) + G_M(b_k^*)\}\Delta t = I_M(u_M^*). \tag{1.25}
$$

1.1.6 Lifting the regularization

Going back to (1.4), we see that the function G'_M is almost constant equal to zero on any interval of the form $[-1+\epsilon, 1-\epsilon]$ if M is large enough. On the other side, G'_M varies very sharply near ± 1.

We have the following result:

Lemma 1.3. *Fix x, p, t and assume that $H(x, \cdot, p, t)$ restricted to $[-1, 1]$ uniquely attains its maximum at $v \in [-1, 1]$. Then the unique solution $\tilde{b}_M$ of the equation*

$$\partial_2 H(x, \tilde{b}_M, p, t) + G'_M(\tilde{b}_M) = 0$$

fulfils:

$$\lim_{M\to\infty} \tilde{b}_M = v. \tag{1.26}$$

Proof. If $v \in (-1, 1)$, it means that $\partial_2 H(x, v, p, t) = 0$, $\partial_2^2 H(x, v, p, t) < 0$ and then we can use the implicit function theorem if M is large enough.

If $v = 1$, it means that $\partial_2 H(x, 1, p, t) \geq 0$. Because $-G'_M$ grows from almost zero to M on a very narrow interval near 1, then $\tilde{b}_M$ must be near 1. The same argument goes for $v = -1$. □

While $\tilde{b}_M$ is a smooth function of a, p, t, the maximum point v can vary sharply between the two extreme values -1 and $+1$. This is exactly what happens in the bang-bang controls.

Assumption 1.4. *Define the function $v(x, p, t)$ as the largest maximum point of $H(x, \cdot, p, t)$ on $[-1, 1]$. Assume that the system*

$$\begin{aligned} p' &= -(\partial_1 H)(x, v(x, p, t), p, t), \quad p(T) = 0, \\ x' &= f(x, v(x, p, t), t), \quad x(0) = 0, \end{aligned} \tag{1.27}$$

has a unique continuous, piecewise C^1 solution $x^*(t)$ and $p^*(t)$. Also assume that the function $u^*(t) := v(x^*(t), p^*(t), t)$ is piecewise continuous.

Then one can prove that the solution (x^*_M, p^*_M, u^*_M) of (1.24) converges uniformly to (x^*, p^*, u^*) on compacts avoiding discontinuities. Moreover, $u^*_M(t) \in (-1, 1)$ for all $0 \leq t \leq T$ if M is large enough.

Going back to (1.2), choose an arbitrary control function u. Using (1.5) we obtain:

$$I(u) \leq \limsup_{M\to\infty} I_M(u) \leq \limsup_{M\to\infty} I_M(u^*_M) = \limsup_{M\to\infty} I(u^*_M) = I(u^*).$$

1.2 Variable T, fixed end-point

Here we assume that f and f_0 do not depend on time. The state function starts again from $x_0 = 0$, and we want to get to a fixed $x_1 > 0$ such that a certain functional is maximized. For a given control function $u : [0, \infty) \to [-1, 1]$, denote by $T(u) > 0$ the first time when $x(t) = x_1$. Then we want to maximize:

$$I(u) = \int_0^{T(u)} f_0(x(t), u(t))dt. \tag{1.28}$$

The minimal arrival time problem is covered by the case $f_0(x, u) = -1$.

We introduce again the regularized functional I_M, and we discretize the problem. One important difference is that we must allow the different times t_k to be independent variables, although we still impose that they are a strictly increasing sequence and $t_0 = 0$.

For a given u which generates a state x which will touch x_1 at $T(u)$, the discretization procedure from (1.6) and (1.7) will generate an approximate discrete solution which might not fulfil the endpoint condition: $y(t_N) = x_1$. The remedy is the following: we discretize exactly as before up to $N-1$, and then we choose t_N such that

$$x_1 = y(t_{N-1}) + (t_N - t_{N-1}) f_0(y(t_{N-1}), u_M(t_{N-1})).$$

Thus (1.14) can be written again, and we are left to a finite discretization problem, which we will describe next. For a fixed N, define $F_M : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \to \mathbb{R}$ given by

$$F_M(a_1, \dots, a_N, b_0, \dots b_{N-1}, t_1, \dots, t_N) = \sum_{k=0}^{N-1} \{f_0(a_k, b_k) + G_M(b_k)\}(t_{k+1} - t_k). \tag{1.29}$$

(Again $a_0 = t_0 = 0$, F_M does not depend on a_N). We will maximize F subject to the restrictions

$$\phi_k(\mathbf{a}, \mathbf{b}, \mathbf{t}) := a_{k+1} - a_k - f(a_k, b_k)(t_{k+1} - t_k), \quad 0 \le k \le N-1. \tag{1.30}$$

Note that we do not impose the t_k's to be increasing. We form the Lagrangian:

$$F_M(\mathbf{a}, \mathbf{b}, \mathbf{t}) - \sum_{k=0}^{N-1} p_k \phi_k(\mathbf{a}, \mathbf{b}, \mathbf{t}) - \lambda(a_N - x_1). \tag{1.31}$$

The necessary equations obeyed by a local extremum of this function are:

$$0 = \partial_2 H(a_k, \tilde{b}_k, p_k) + G'_M(\tilde{b}_k), \quad 0 \le k \le N-1, \tag{1.32}$$
$$p_k = p_{k-1} - (\partial_1 H)(a_k, \tilde{b}_k, p_k)(t_{k+1} - t_k), \quad 1 \le k \le N-1,$$
$$p_{N-1} = \lambda,$$
$$a_{k+1} = a_k + f(a_k, \tilde{b}_k)(t_{k+1} - t_k), \quad 0 \le k \le N-1,$$
$$H(a_k, \tilde{b}_k, p_k) + G_M(\tilde{b}_k) = H(a_{N-1}, \tilde{b}_{N-1}, p_{N-1}) + G_M(\tilde{b}_{N-1}) = 0, \tag{1.33}$$
$$a_N = x_1, \quad a_0 = 0.$$

One major difference compared to the fixed T case is the appearance of N conservation equations (1.33). Here we have N extra unknowns, the different times.

It is far from obvious how the general solution might look like, especially that we have not imposed monotonicity on t's. But if we are only searching for strictly increasing times, and if for instance we assume that $f > 0$ (which indicates that the a's can get closer to x_1 after each step), then we can get rid of the times and simplify the system:

$$0 = \partial_2 H(a_k, \tilde{b}_k, p_k) + G'_M(\tilde{b}_k), \quad 0 \le k \le N-1, \tag{1.34}$$
$$p_k = p_{k-1} - (\partial_1 H)(a_k, \tilde{b}_k, p_k)\frac{a_{k+1} - a_k}{f(a_k, \tilde{b}_k)}, \quad 1 \le k \le N-1,$$
$$p_{N-1} = \lambda,$$
$$H(a_k, \tilde{b}_k, p_k) + G_M(\tilde{b}_k) = H(a_{N-1}, \tilde{b}_{N-1}, p_{N-1}) + G_M(\tilde{b}_{N-1}) = 0, \tag{1.35}$$
$$a_N = x_1, \quad a_0 = 0.$$

This system can be solved in a recursive way. From (1.34) and (1.35) we obtain a_{N-1} and $\tilde{b}_{N-1}$ as functions of λ. Then we have p_{N-2} from the iteration, and we start again by finding a_{N-2} and $\tilde{b}_{N-2}$. Finally, we determine λ from the condition $a_0 = 0$. The time differences are then obtained as $\frac{a_{k+1}-a_k}{f(a_k,\tilde{b}_k)}$ and we are done.

1.2.1 The continuum case

Now let us assume that the following system of equations:

$$\begin{aligned} 0 &= \partial_2 H(x,u,p) + G'_M(u), \\ p' &= -(\partial_1 H)(x,u,p), \quad p(T) = \lambda, \\ x' &= f(x,u,t), \quad x(0) = 0, \quad x(T) = x_1, \\ 0 &= H(x,u,p) + G_M(u), \end{aligned} \tag{1.36}$$

has a solution with $x'(T) \neq 0$ where T is the first time when $x(T) = x_1$. Then the discrete problem will also have a solution which will "converge" to the continuous one when $N \to \infty$. We do not give details.