Optimal control problems
with complementarity constraints

Mihai Anitescu

Argonne National Laboratory

Thanks: J. Peng, S. Akella (RPI), D. Stewart (Iowa)

Complementarity Constraints in Optimal Control

- Switching is an essential part of many dynamical systems.

- That can be modeled by complementarity constraints. The latter can appear from first-principles modeling as well (in Friction Problems).

- Therefore when doing optimal control on such dynamical systems, one will obtain a mathematical program with complementarity constraints.

- Prompted by the success of the elastic mode in solving mathematical programs with complementarity constraints, we have proceeded to a discretize and optimize, brute force approach in which to leverage these advances.

Organization

- Elastic Mode and Math Program with Complementarity Constraints

- A brute force application to control with contact collision and friction.

- **Maybe brute force is not the best idea and some problem-specific modeling may be necessary**... Theoretical analysis and results.

NLP Problem with some linear constraints

$$\begin{aligned}
\min_x \quad & f(x) \\
\text{subject to} \quad & g_i(x) \ \le \ 0, \ i = 1, 2, \ldots, n_i \\
& h_j(x) \ = \ 0, \ j = 1, 2, \ldots, n_e
\end{aligned}$$

We assume that

1. $g_i(x)$ is linear for $i = 1, 2, \ldots, l_i$,

2. $h_j(x)$ is linear for $j = 1, 2, \ldots, l_e$.

Elastic mode relaxation for NLP

$$(NLP(c)) \quad \begin{aligned} \min_{x,u,v,w} \quad & \tilde{f}(x) + c_1\left(e_{m-l_i}^T u + e_{r-l_e}^T (v+w)\right) \\ \text{subject to} \quad \tilde{g}_i(x) \quad &\leq \quad 0, \ i = 1, 2, \ldots, l_i, \\ \tilde{g}_i(x) \quad &\leq \quad u_i, \ i = l_i + 1, \ldots, m, \\ \tilde{h}_j(x) \quad &= \quad 0, \ j = 1, 2, \ldots, l_e \\ -v_j \leq \tilde{h}_j(x) \quad &\leq \quad w_j, \ j = l_e + 1, \ldots, r \\ u, v, w \quad &\geq \quad 0, \end{aligned}$$

An adaptive elastic mode approach

If *the QP is infeasible or its Lagrange multipliers are too large* then
      NLPC: *Find the solution* $(x^{c_1}, u^{c_1}, v^{c_1}, w^{c_1})$ *of the relaxed NLP*
      If $||(u^{c_1}, v^{c_1}, w^{c_1})|| = 0,$ then x^{c_1} *solves NLP.* Stop.
                              otherwise $c_1 = c_1 + K$, *return to* NLPC.

SNOPT uses that, with an active set algorithm for solving relaxed NLP.

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & g_i(x) && \leq && 0 && i = 1, 2, \ldots, n_i \\
& h_j(x) && = && 0 && j = 1, 2, \ldots, n_j \\
& x_{k1} && \leq && 0 && k = 1, 2, \ldots, n_c \\
& x_{k2} && \leq && 0 && k = 1, 2, \ldots, n_c \\
& \sum_{k=1}^{n_c} x_{1k} x_{2k} && \leq && 0
\end{aligned}
\tag{1}
$$

- The slack formulation is general and has certain algorithmic advantages over allowing the functions that enter the complementarity constraints to be nonlinear.

- The problem does not satisfy MFCQ at any feasible point. Linearization of the constraints may be infeasible arbitrarily close to the solution.

Applying the elastic mode to MPCC

- Under very weak assumptions (existence of a Lagrange multiplier), the relaxed problem has the same solution as the unrelaxed problem (2000).

- Superlinear convergence can be achieved if MPEC-LICQ holds, and certain strong second-order conditions. (2003).

- If the problem has certain structure, one can even obtain global convergence (2004).

- The performance of SNOPT on these problems is remarkable (Fletcher and Leyffer, 2003).
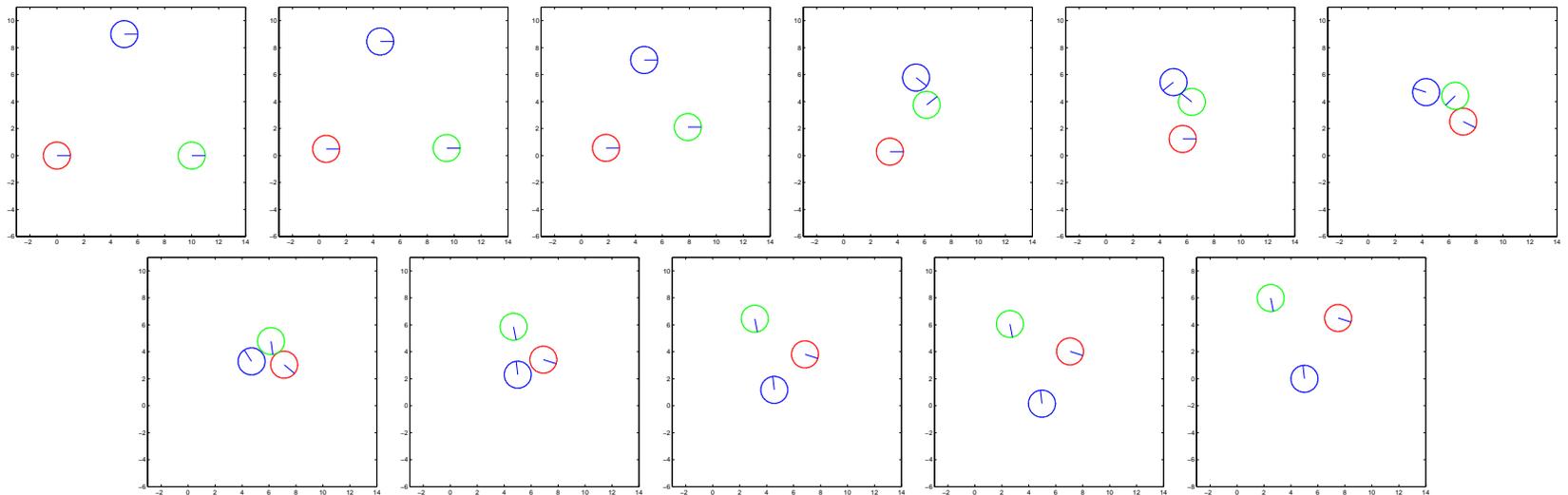
Discretize and Optimize Control with Switching

Figure 1: Three-robot coordination, with contact, collision and friction

Performance

Table 1: test examples in set B

Problem	N_r	D_{ef}	N_t	K	N_v	N_{tc}
B1	2	2	100	16	10665	7080
B2	3	2	50	16	9392	6564
B3	3	2	50	16	9392	6564

Table 2: computational results of examples in set B

Problem	G_s	G_t	LB
B1	9.78	3157.97	8.94
B2	9.48	4116.81	8.94
B3	9.39	4834.96	8.94

Intermediate conclusions

- SNOPT (with a few tricks) is used for this problem, and it does provide a reasonable solution. That is consistent with previous conclusions about elastic mode.

- The computational complexity of these problems grows very fast, and outcome tends to be highly protocol dependent.

- With David Stewart, we felt there was a need to further explore the nature of the difficulties.

$$\boxed{\textbf{Example}}$$

$$\min_{x_0} \quad (x(2) - 5/3)^2 + \int_0^2 x(t)^2 \, dt$$

$$\text{subject to} \quad \frac{dx}{dt} \in (1 + \alpha) - \text{Sgn}(x), \qquad x(0) = x_0$$

- The $\in$ symbol (differential inclusion) accommodates the potential situation where the system "stays" in the manifold (that is x=0 happens for an nonzero amount of time). That does not happen in this case.

- But if the ODE is solved, first, we get $x(t) = \alpha(t + x_0/(2 + \alpha))$ for $t \geq -x_0/(2 + \alpha)$. **The objective function is differentiable, (even analytical) with respect to x_0!!.**

Discretize and Optimize

For the objective function, it is straightforward. For the differential inclusions

$$x^{k+1} \in x^k + h(1 + \alpha) - h \operatorname{Sgn}(x^k + \theta(x^{k+1} - x^k)). \qquad (2)$$

- If $\theta = 1$, we have backward Euler.

- If $\theta = 0$ we have forward Euler.

- The backward Euler scheme mimics the hard constraint approach developped by several authors recently for multibody dynamics with contact and friction.

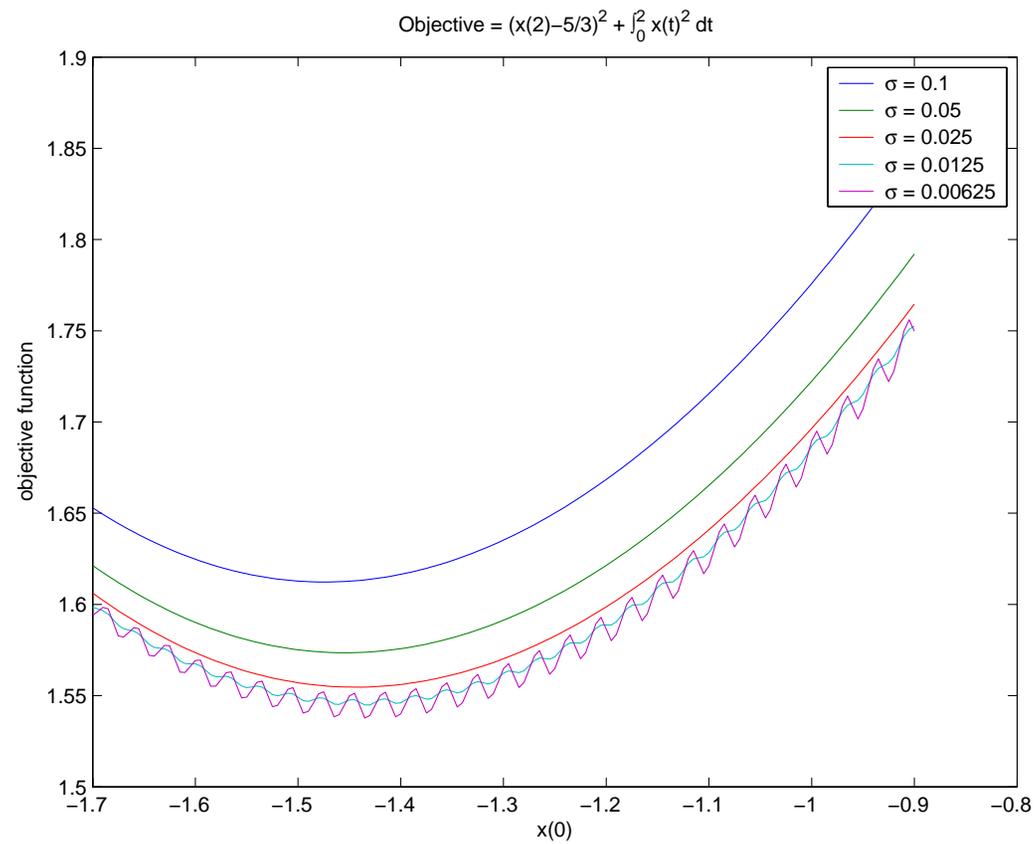- To make the problem differentiable optimization, **we model the signum function with complementarity constraints**.

The behavior of the objective function

Objective = $(x(2)-5/3)^2 + \int_0^2 x(t)^2 \, dt$

Legend:
- h = 4e−2
- h = 2e−2
- h = 1e−2
- h = 5e−3

objective function vs x(0)
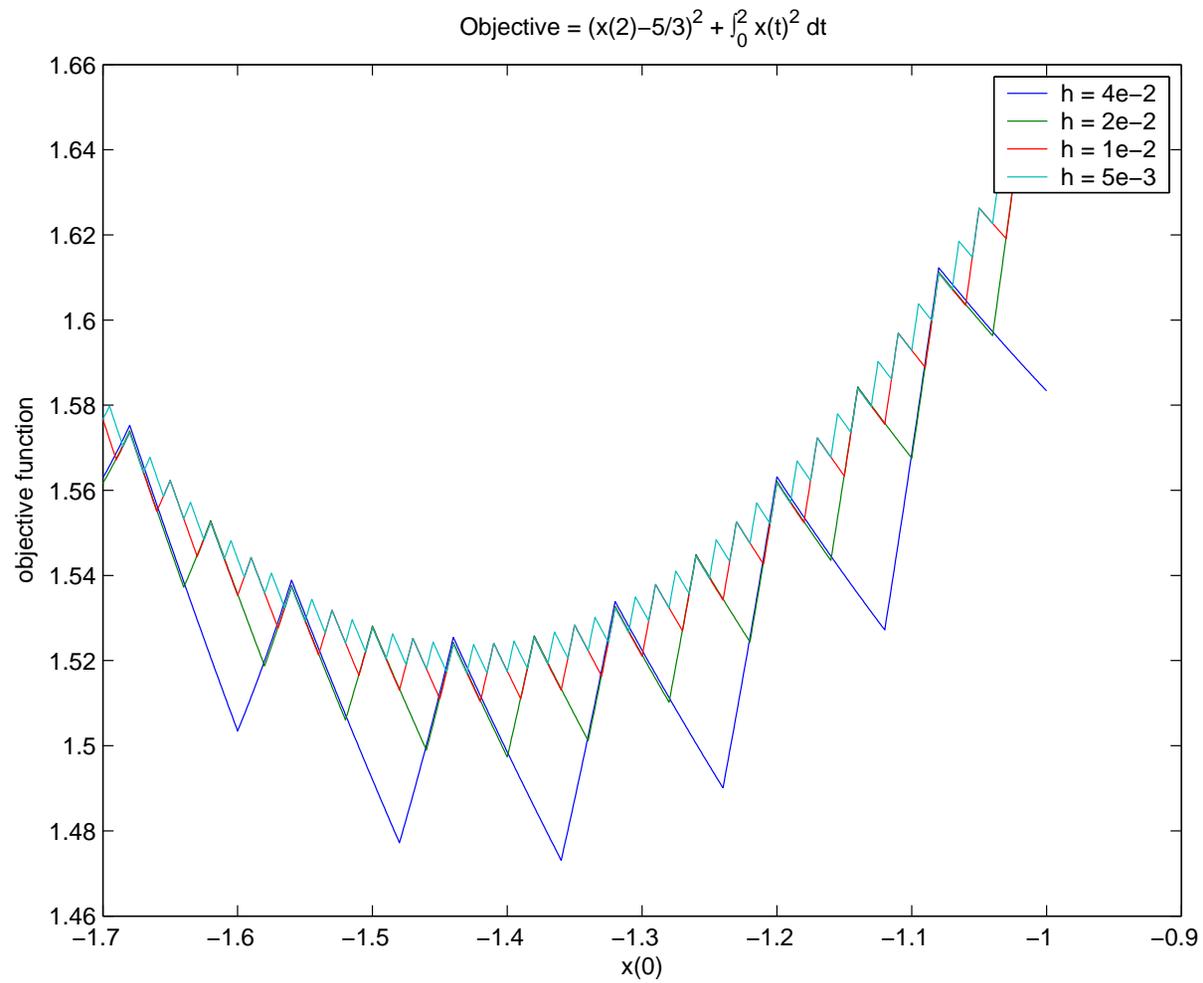
Observations

- **There is no hope of convergence, arbitrarily close to the actual solution !!**

- The problem is the incorect sensitivity jump due to the transition.

- So we either detect and track all transitions, or **we do some modeling**.

Smoothing example, for $h = 0.01$

Objective = $(x(2)-5/3)^2 + \int_0^2 x(t)^2 \, dt$

Legend:
- $\sigma = 0.1$
- $\sigma = 0.05$
- $\sigma = 0.025$
- $\sigma = 0.0125$
- $\sigma = 0.00625$

x-axis: x(0)
y-axis: objective function

For smoothing the example

- We use $\tanh(x/\sigma)$ to approximate the $\text{sgn}(x)$.

- Note that the size of the smoothing parameter does matter.

- **Even stiff systems integrated implicitly are not immune to this behavior**.

Our smoothing approach

.

$$\frac{dx}{dt} \in \begin{cases} \{f_1(x)\}, & \psi(x) < 0, \\ \{f_2(x)\}, & \psi(x) > 0, \\ \mathrm{co}\{f_1(x),\ f_2(x)\}, & \psi(x) = 0. \end{cases} \qquad (3)$$

We can approximate this by the smoothed system

$$\frac{dx_\sigma}{dt} = \varphi_\sigma(\psi(x_\sigma))\, f_2(x_\sigma) + (1 - \varphi_\sigma(\psi(x_\sigma)))\, f_1(x_\sigma). \qquad (4)$$

Here

$$\varphi_\sigma(w) = \int_{-\infty}^{w} \theta_\sigma(r')\, dr' \qquad (5)$$

where $\theta_\sigma(r) = (1/\sigma)\theta(r/\sigma)$ and $\theta \geq 0$,

$$\boxed{\textbf{Our theoretical setup}}$$

Consider.

$$\frac{dx}{dt} = f(x), \qquad x(t_0) = x_0, \tag{6}$$

where f may have switching branches but satisfies a one-sided Lipschitz constant L as in

$$y_i \in f(x_i) \text{ for } i = 1, 2 \quad \text{implies} \quad (y_2 - y_1)^T (x_2 - x_1) \leq L \, \|x_2 - x_1\|^2.$$

Smoothing and discretization

We consider the smoothed differential equation.

$$\frac{dx_\sigma}{dt} = f_\sigma(x_\sigma), \qquad x_\sigma(t_0) = x_0.$$

and its fixed step Euler discretization.

$$x_\sigma^{k+1} = x_\sigma^k + h\, f_\sigma(x_\sigma^k)$$

We denote by $x^h(t)$ the interpolated numerical scheme.

Main result

Assume that $h = o(\sigma)$. Then, away from the transition points we have that

$$\frac{\partial x_\sigma^h(t)}{\partial x_0} \xrightarrow{\sigma \to 0} \frac{\partial x(t)}{\partial x_0}$$

. **Observations**

- The proof is horrendous, since when the solution stays in the discontinuity manifold, the sensitivities of the constraint systems, which include projections, need to pop out of the proof.

- A similar result holds for adjoints, and their jump is computable.

- Under regularity assumptions of the objective function, the minima themselves will converge.

- We are now investigating infinite dimensional optimal control.

The Michael Schumacher problem (simplified)

The simplified dynamics of a sports car.

$$\dot{\mathbf{x}} = \mathbf{v} \tag{7}$$

$$\dot{\mathbf{v}} = a(t)\,\mathbf{t}(\theta) + F\,\mathbf{n}(\theta), \tag{8}$$

$$\dot{\theta} = s(t)\,(\mathbf{t}(\theta)^T\mathbf{v}), \tag{9}$$

$$F \in -\mu N \mathrm{Sgn}(\mathbf{n}(\theta)^T\mathbf{v}). \tag{10}$$

μ is the coefficient of friction and N is the normal contact force (assumed constant). **Track constraints**.

$$\mathbf{x} \in C = \left\{\,(x,y) \mid |y - y_{cl}(x)| \leq w/2\,\right\} \tag{11}$$

Track equation.

$$y_{cl}(x) = \begin{cases} \sin(x), & x \leq \pi, \\ \pi - x, & \pi \leq x \leq 2\pi, \\ -\pi - \sin(x), & 2\pi \leq x. \end{cases} \tag{12}$$

Control constraints

$$|a(t)| \;\leq\; a_{\max} \qquad \text{for all } t, \tag{13}$$

$$|s(t)| \;\leq\; s_{\max} \qquad \text{for all } t. \tag{14}$$

Objective function

$$g(T, x(T)) = \alpha \|\mathbf{x}(T) - \mathbf{x}_{tgt}\|^2 + T.$$

σ	N	objective	T	# iter'ns	CPU time
0.1	250	5.544654	5.53393	303	14.1
	500	5.549708	5.53877	531	53.3
	1000	5.552177	5.54111	349	53.3
	2000	5.553451	5.54239	420	227.0
0.05	500	5.590849	5.58285	1501	153.5
	1000	5.409497	5.39842	977	242.2
	2000	5.409183	5.39813	643	300.8
0.025	1000	5.353886	5.34289	1240	184.5
	2000	5.354256	5.34321	1759	913.2
	4000	5.354451	5.34341	1368	1552.8

Trajectory of "race-car"

Optimal time trajectory
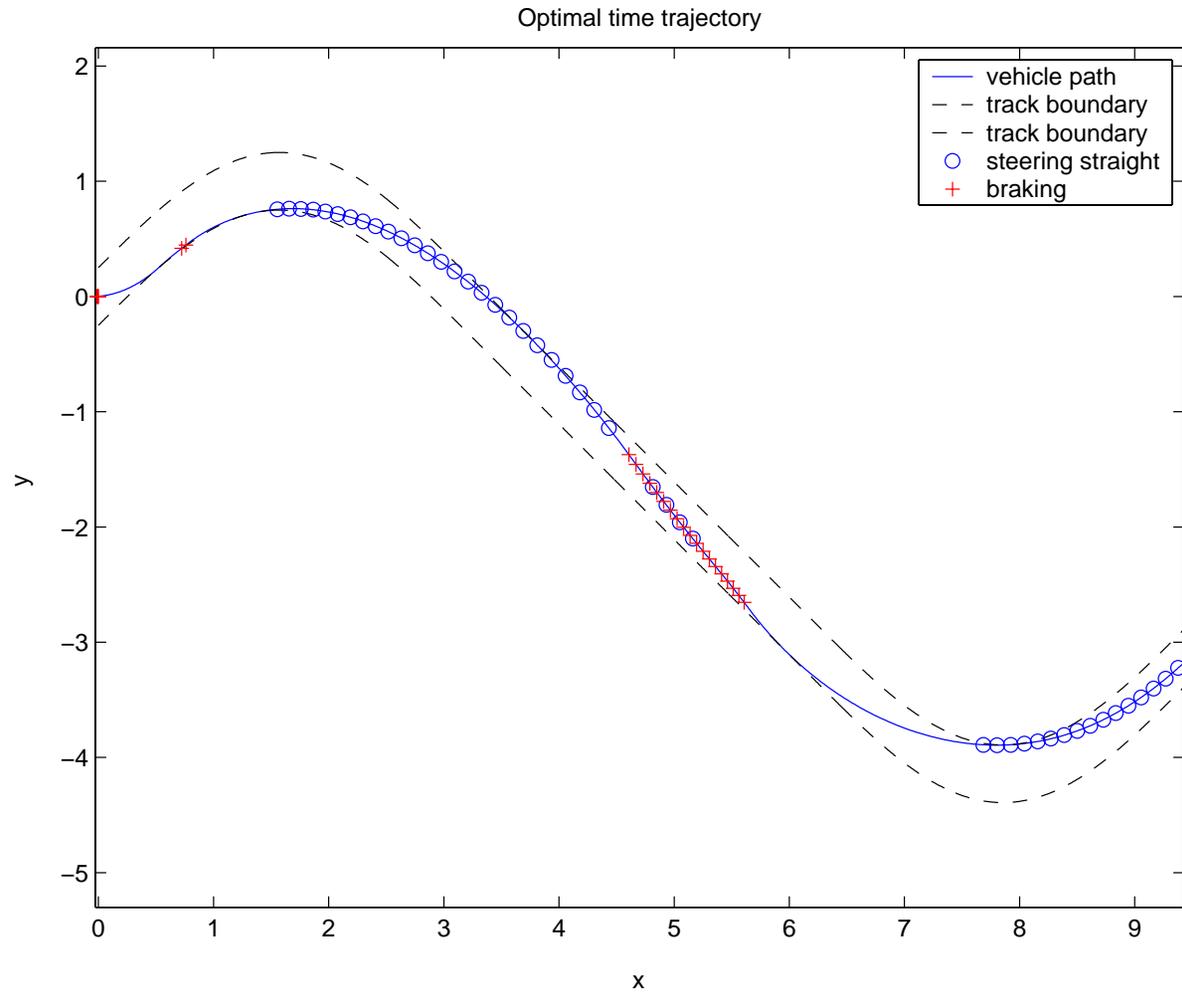
Computed optimal control functions and velocities

Optimal controls and velocities
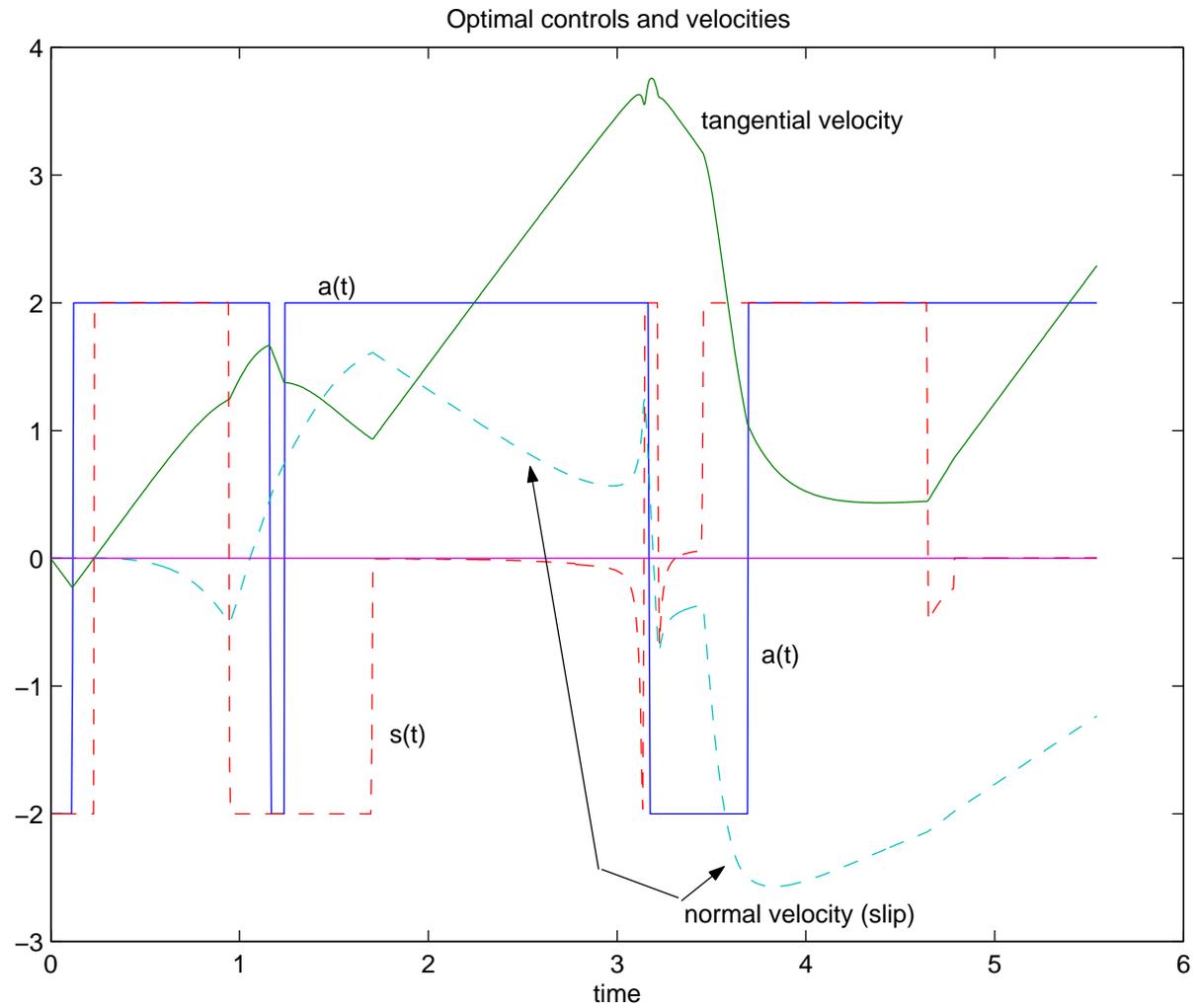
Convergence..

We do not have a criterion that we do not encounter the case we were warning against, **but the plots stay the same with varying σ and N**. So we seem to converge.

Conclusions

- In spite of their modeling appeal simplicity, optimize and discretize approaches are not well suited for optimal control with switching dynamics, such as contact and friction **even if they can be integrated stably**.

- We presented an example of such a problem for which the derivatives do not converge as time step goes to 0. The example has sequences of local minimizers that converge to points arbitrarily close to the local minimum sought.

- We proved that in the case of parametric control this difficulty is eliminated if the problem is smoothed and the time step is taken to 0 faster than the smoothing parameter.

- We presented a simplified example of the Michael Schumacher problem.