OPTIMAL CONTROL OF ROAD VEHICLES:
DIRECT AND INDIRECT APPROACHES

Nicola Dal Bianco[1,3], **Matteo Massaro**[1], **Matteo Veneri**[1], **Roberto Lot**[2]

[1] Department of industrial engineering, University of Padova
[2] Faculty of Engineering and the Environment, University of Southampton
[3] e-mail address: nicola.dalbianco@phd.unipd.it

Keywords: Vehicle dynamics, optimal control, simulations, direct methods, indirect methods.

Abstract

Minimum lap time simulations are of great interest for the design and setup of race vehicles. These problems can be formulated and solved in various methods, like quasi-steady state models vs full dynamic models and pre-defined (fixed) trajectory problems vs free trajectory problems.

This work focuses on full dynamic models with free trajectory, that are the most appealing and challenging. This kind of problems can be solved using Optimal Control theory, that give rise to two numerical solving approaches: indirect and direct methods, the latter being the most widespread and used approach. This work aims to compare indirect and direct methods when applied to minimum lap time simulations, in order to understand the differences between the two approaches and whether one is more suitable than the other.

Introduction

Optimal control based minimum lap time simulations have been used for almost 30 years, with a growing interest in the last decade. This kind of dynamics simulation are used for different purposes, such as the set-up and optimization of race vehicles, development of energy usage or torque vectoring strategies for electric or hybrid vehicles, studying energy recovery systems benefits, and, more generally, investigation of control strategies that affect vehicle performance and dynamics.

Optimal control based simulations are basically a nonlinear constrained optimisation problem where the performance index can be the vehicle performance, the energy usage or a combination on both, and where the constraints include the equations of motion, power and energy limitation, tyre adherence, track boundaries, etc. Optimal control problems can be solved using different strategies including variational (indirect) approach [1,2], transcription to nonlinear programming NLP (direct approach) [1,3], dynamic programming [1] and evolutionary algorithm-based optimisation [4].

To the best of authors' knowledge, in the past decade the variational and the NLP approaches are the ones that demonstrated the best capabilities to solve minimum time problems when it comes to:

• dealing with non-trivial vehicle models with complex tyre and/or aerodynamics interaction;

• simultaneous optimisation of the racing line and the controls;

• computing a full lap simulation on a reasonable amount of time (less than few hours).

In particular evolutionary and dynamic programming algorithms usually require very long computation time. In [5,6] an evolutionary algorithm takes approximatively a day to simulate few turns, moreover in [7] dynamic programming shows to be significantly slower than the variational or NLP approach.

Indirect methods for optimal control problems are based on the Pontryagin Maximum Principle [2], which gives the optimality first order necessary conditions. Such conditions come into a set of ordinary differential equations with both initial and final boundary conditions, that is a so called two-points boundary value problem (TPBVP) plus a related minimisation problem to derive the optimal controls. Different numerical techniques can be used to solve this kind of problems. Early examples of application of indirect optimal control theory to minimum lap time simulations are reported in [8]. Since the late 90's various other works mainly have used this indirect methods for lap time simulations [9-16].

Differently, direct methods discretise the optimal control problem so as to give origin to a discrete constrained minimisation problem (i.e. direct transcription) also known as nonlinear programming problem (NLP). Controls only (i.e. sequential discretization) or both controls and states (i.e. full discretization) may be discretised, and various discretisation (i.e. integration) scheme may be used. Different works based on direct optimal control problem can be found since the beginning of the 00's [17-24].

In practice, almost all the minimum lap time simulations based on direct approach reported in literature have been performed using the software suite called PINS/XOptima [25]. On the other side, the most advanced lap time simulations based on indirect approach have been performed with the software GPOPS [26]. Due to the different methods adopted by the two solvers, the formulation of the minimum time problem should be changed accordingly (e.g. penalty functions), and thus differences in the calculated solutions may exist.

A side-by-side comparison between the two methods (direct vs indirect) for optimal control problems has never been reported. Authors that are used to one solver typically never try the other and vice-versa. This gave the motivation for this contribution that aims to highlight similarities and differences between the two approaches. A minimum lap time problem of a GT series like race car is solved using both indirect and direct approaches. Numerical solvers Pins and Gpops have been used respectively for the indirect and direct approach. Results obtained with the two solvers are then analysed.

Car model and optimal control formulation

The car model used for the lap time simulation is described by 3 degrees of freedom (dof): the speed V, the sideslip angle λ and the yaw rate Ω. Three other variables track the car position on the circuit: the curvilinear abscissa s (that is the road centre line), the lateral displacement from the road centre line n and the heading angle α with respect to the road centre line. The first order differential equations that describes such variables are the following:

$$\begin{aligned}
\dot{s} &= \frac{V\cos(\alpha-\lambda)}{1-n\kappa} \\
\dot{n} &= V\sin(\alpha-\lambda) \\
\dot{\alpha} &= \Omega - \frac{\kappa V\cos(\alpha-\lambda)}{1-n\kappa} \\
M\Omega V\lambda + M\dot{V} &= S_{rr}+S_{rl}+S_{fr}+S_{fl}-\delta(F_{fr}+F_{fl})-D \\
M(\Omega V-\dot{V}\lambda-V\dot{\lambda}) &= \delta(S_{fr}+S_{fl})+F_{rr}+F_{rl}+F_{fr}+F_{fl} \\
I_z\dot{\Omega} &= a(F_{fr}+F_{fl})-b(F_{rr}+F_{rl})+t_w(-S_{rr}+S_{rl}-S_{fr}+S_{fl})
\end{aligned} \tag{1}$$

where M is the car mass, S_{ij} is the tyre longitudinal force where i = f , r indicates the front (f) or rear ® tyre, and j = r, l indicates the right (r) or left (l) side, F_{ij} is the tyre lateral force, δ is the steering angle, D is the drag force, Iz is the yaw inertia moment, a and b are respectively the distance of the front and rear axle form the centre of gravity and t_w is the car half width. In the previous equations the sideslip and steer angle have been assumed to be small, so as that $\cos\lambda \approx 1$, $\cos\delta \approx 1$, $\sin\lambda \approx \lambda$ and $\sin\delta \approx \delta$.

The drag force D is given by the classic law that depends on the square of the speed:

$$D = \frac{1}{2}\rho C_a V^2 \tag{2}$$

where Ca is the drag coefficient and ϱ is the air density. Tyre lateral forces are calculated with a linear tyre model, while longitudinal forces are given by the single control variable u_x that is equal to the normalized thrust:

$$\begin{aligned}
F_{ij} &= N_{ij}K_\lambda\lambda_{ij} \\
S_{fl} = S_{fr} &= \frac{Mg}{2}f^-(u_x)\beta \\
S_{rl} = S_{rr} &= \frac{Mg}{2}(f^+(u_x)+f^-(u_x)(1-\beta))
\end{aligned} \tag{3}$$

where β is the front braking bias (front over total barking force ratio), N_{ij} is the tyre load, λ_{ij} is the tyre sideslip angle, f^+ and f^- are functions that return respectively the positive and negative part of the argument and K_λ is the tyre sideslip stiffness. The tyre sideslip angles are given by the following kinematic expression:

$$
\begin{aligned}
\lambda_{rr} &= \lambda + \frac{\Omega(b+\lambda t_w)}{V} \\
\lambda_{rl} &= \lambda + \frac{\Omega(b-\lambda t_w)}{V} \\
\lambda_{fr} &= \lambda + \delta - \frac{\Omega(a-\lambda t_w)}{V} \\
\lambda_{fr} &= \lambda + \delta - \frac{\Omega(a+\lambda t_w)}{V}
\end{aligned} \tag{4}
$$

The tyre loads N_{ij} are related to the delayed longitudinal a_x and lateral a_y accelerations:

$$
\begin{aligned}
N_{rr} &= \frac{Mg}{4} + \frac{Mg}{4}\left(\frac{a-b+a_x h}{a+b} - a_y(1-\chi)\frac{h}{t_w}\right) \\
N_{rl} &= \frac{Mg}{4} + \frac{Mg}{4}\left(\frac{a-b+a_x h}{a+b} + a_y(1-\chi)\frac{h}{t_w}\right) \\
N_{fr} &= \frac{Mg}{4} + \frac{Mg}{4}\left(-\frac{a-b+a_x h}{a+b} - a_y\chi\frac{h}{t_w}\right) \\
N_{fl} &= \frac{Mg}{4} + \frac{Mg}{4}\left(-\frac{a-b+a_x h}{a+b} + a_y\chi\frac{h}{t_w}\right)
\end{aligned} \tag{5}
$$

where χ is the roll stiffness. The accelerations, both longitudinal a_x and lateral a_y, follow the actual vehicle accelerations through low band pass filter with time constant τ_{ax} and τ_{ay} in order to take into account for the suspension load transfer lag:

$$
\begin{aligned}
\tau_{a_x}\dot{a}_x + a_x &= \dot{V} + \Omega V \lambda \\
\tau_{a_y}\dot{a}_y + a_y &= \Omega V - (\dot{V}\lambda)
\end{aligned} \tag{6}
$$

Summarising, the state space model comprises eight dof (V, λ, Ω, s, n, α, a_x, a_y), with the corresponding eight first order equations (1), (6) and two controls (u_x, δ).

The equations so far described fully characterise the car model. In order to perform a minimum lap time problem some constraints must be included in order to ensure the simulated dynamics to be realistic. In particular it must be ensured that the power used is less than the motor maximum power, that the car never exceeds the track boundaries, and that the tyre forces are less than the tyre maximum adherence. These constraints can be expressed as follows:

$$
\begin{aligned}
c_p &= \frac{V(S_{rr}+S_{rl})}{P_{max}} \leq 1 \\
c_n &= \frac{n}{n_{max}} \leq 1 \\
c_n &= \frac{-n}{-n_{max}} \leq 1 \\
c_{t_{ij}} &= \left(\frac{S_{ij}}{N_{ij}\mu^x_{ij}}\right)^2 + \left(\frac{F_{ij}}{N_{ij}\mu^y_{ij}}\right)^2 \leq 1 \qquad i=f,r \;\; j=r,l
\end{aligned} \tag{7}
$$

where P_{max} is the motor maximum power, n_{max} is the car maximum lateral displacement from the road centre line (i.e. it is equal to half of the road width r_w minus half of the car width: $n_{max} = r_w/2 - t_w$), moreover μ_x and μ_y are respectively the tyre longitudinal and lateral adherence which depend on the tyre loads through the following relationship:

$$
\begin{aligned}
\mu^x_{ij} &= \mu^x_0 + K_\mu \frac{N_{ij}}{N_{0_{ij}}} \qquad i=f,r \;\; j=r,l \\
\mu^y_{ij} &= \mu^y_0 + K_\mu \frac{N_{ij}}{N_{0_{ij}}} \qquad i=f,r \;\; j=r,l
\end{aligned} \tag{8}
$$

whereN_{0ij} is the tyre load in static conditions and K_μ is a constant factor. The numerical data used to feed the car model is reported in table 6.

Results

The minimum lap time simulation has been performed on the Adria International Raceway. The lap time calculated by Pins (indirect) and Gpops (direct) are respectively 75.673s and 75.727s, the difference between the two solution is limited to approximately the 0.1%.

The car speed along the circuit is shown in figure 1 it can be noted that the speed difference between the two solvers is always less than 3km/h. Moreover, Pins simulated speed is generally higher than Gpops one in the middle of the turns but it is lower in the straights and in the first part of the braking manoeuvre.

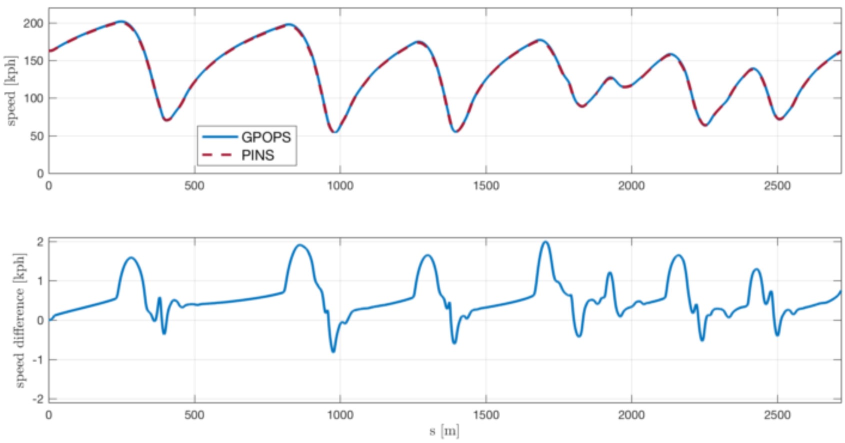

Figure 1 Speed profile on the Adria International Raceway. Differences between Gpops and Pins solution are lower than 2kph

Figure 2 shows the resulting optimal trajectory; again there is no noticeable difference between the two solvers. The major discrepancies are noticed in the two (straights) track sections comprised between the black and the green diamonds. This difference is more noticeable when looking at the car lateral displacement n from the road centre line, that is reported in the top plot of figure 3. In the two straight sections it is clear that Pins trajectory moves towards the road center line. This difference can be explained by the penalty terms of the indirect approach: while Gpops remain close to the track border in the two straights, Pins tends to reduce the penalty by moving to the road center line. This effect is probably evident in these two sections because here this manoeuvre affects only marginally the lap time.

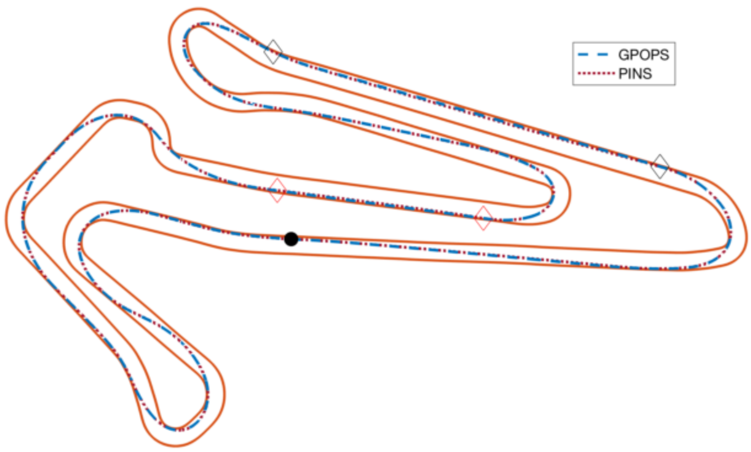

Figure 2 Optimal trajectory along the circuit. Small differences between the trajectories obtained with the two software arise in the straights between the black and red rhombuses.

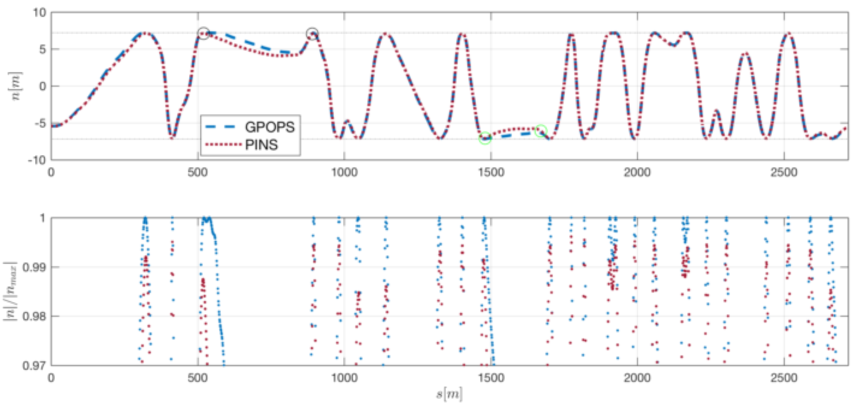

Figure 3 In the top plot the car lateral displacement is shown; the trajectory differences in the two straights between the two rhombuses are noticeable. The road-borders-constraint enforcement c_n (7) is shown in the bottom plot (c_n = 1 is the constraint limit): Pins solution doesn't touch the constraint limit.

The bottom graph in figure 3 shows how much the two solutions get close to the road boundaries. It is expected, indeed, that the minimum time is achieved touching the road borders at some points. This is what Gpops solution does (but never exceeds the track boundaries), while Pins ones goes up to ≈ 0.995 times the maximum lateral displacement. This difference is consequence of the penalty approach that makes the constraints to be "soft". A similar behaviour can be observed also in the other constraints, that are the maximum power limit and the tyre adherence limit.

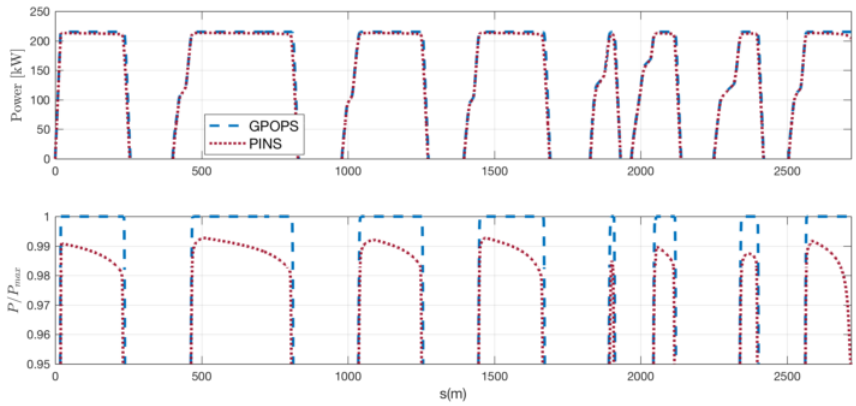

Figure 4 Power used (top plot) and power limit constraint enforcement (bottom plot). Pins solution uses only up to 99% of the maximum power, while Gpops ones arrives up to approximately 100%.

The power limit usage constraint is shown in figure 4. Again, Gpops solution uses up to 100% (up to machine precision) of the available power, while Pins never uses more than ≈ 99%. Moreover, similar conclusions can be stated for the tyre engagement constraints c_{tij} 5, which are reported in figure 5: Pins solution uses up to ≈ 99.7% of the maximum tyre adherence. The fact that Pins solution does not reach the exact constraints bounds may seem a consequence of the penalty approach used by indirect methods. However even Ipopt, that is based on an Interior Point algorithm, uses a similar penalty approach [27]. The main difference between Pins and Ipopt is that the latter implements an algorithm that automatically sharpens as much as possible each penalty separately on each mesh point. On the contrary, Pins sharpens each penalty globally on the entire mesh. In different words, Ipopt can tune a constraint penalty differently point per point, while Pins can tune it only in the same way on all mesh points.

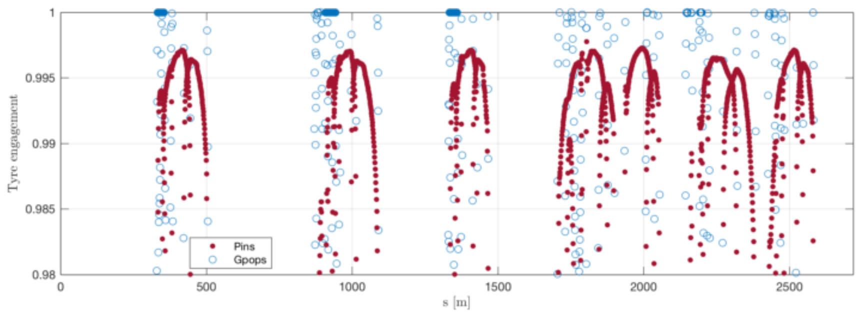

Figure 5 Maximum tyre engagement versus distance travelled. Again, Gpops solutions exploit all tyre adherence ellipses, while Pins stops at 99.7%.

Lap time simulations are most used to optimise the design and setup of vehicles, therefore it comes natural to wonder if the different simulated dynamics of Pins and Gpops give different optimisation outcomes. This has been tested on the parametric analysis of the braking bias, i.e the ratio between front and rear braking torque. The minimum lap time has been calculated for different values of the braking bias, so as to find out the optimal value. The obtained results are shown in figure 6: it can be noticed that the best performance is achieved almost for the same value of the braking bias, 0.64. In particular, Pins optimum braking bias is 0.639, while Gpops one is 0.641.

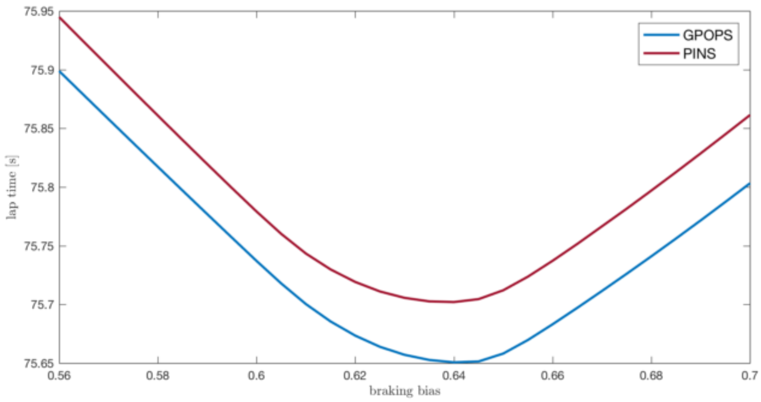

Figure 6 Lap time versus braking bias. Despite the absolute time provided by the two solvers is slightly different, the optimum braking bias is almost the same.

Conclusions

In this work indirect and direct optimal control methods have been compared on a minimum lap time simulation of a race car on the Adria International Raceway. Pins and Gpops software have been chosen respectively for the indirect and direct methods for being the most used software in literature for this purpose.

The two solvers provided very similar results: speed profile, trajectory, power and tyre usage does not show appreciable differences at first sight.

Constraints enforcement analysis revealed that Pins solutions almost never uses up to 100% of the available power, tyre engagement or track width. This is not a direct consequence of the penalty approach of the indirect methods, rather of the fact that Pins does not adjust the penalty independently on each mesh point.

The parametric analysis performed on the braking bias showed that, despite Pins uses slightly less power, tyres, and track width, it gives the same optimum value for the braking bias. This suggests that the different usage of the constraints does not affect the car design optimisation.

Finally, Pins has been significantly faster than Gpops in computing the solution (Gpops takes 8 times Pins to find the solution). This is not an intrinsic advantage of indirect method over direct but more likely due to the different technologies used.

References

[1] Bryson, A. E. (1999). Dynamic optimization, volume 1. Prentice Hall.

[2] Pontryagin, L. S. (1987). Mathematical theory of optimal processes. CRC Press.

[3] Betts, J. T. (2010). Practical methods for optimal control and estimation using nonlinear programming. Society for Industrial and Applied Mathematics.

[4] Branke, J. (2012). Evolutionary optimization in dynamic environments (Vol. 3). Springer Science & Business Media.

[5] Maniowski, M. (2016). Optimization of driver and chassis of fwd racing car for faster cornering. In The Dynamics of Vehicles on Roads and Tracks: Proceedings of the 24th Symposium of the International Association for Vehicle System Dynamics (IAVSD 2015), Graz, Austria, 17-21 August 2015, page 77. CRC Press.

[6] Maniowski, M. (2016). Optimisation of driver actions in rwd race car including tyre thermodynamics. Vehicle System Dynamics, 54(4):526–544.

[7] Biral, F., Bertolazzi, E., and Bosetti, P. (2016). Notes on numerical methods for solving optimal control problems. IEEJ Journal of Industry Applications, 5(2):154–166.

[8] Metz, D. and Williams, D. (1989). Near time-optimal control of racing vehicles. Automatica, 25(6):841–857.

[9] Hendrikx, J., Meijlink, T., and Kriens, R. (1996). Application of optimal control theory to inverse simulation of car handling. Vehicle System Dynamics, 26(6):449–461.

[10] Da Lio, M. (1997). Analisi della manovrabilita dei veicoli un approccio basato sul controllo ottimo. ATA-TORINO-, 50:35–42.

[11] Cossalter, V., Da Lio, M., Lot, R., and Fabbri, L. (1999). A general method for the evaluation of vehicle manoeuvrability with special emphasis on motorcycles. Vehicle system dynamics, 31(2):113–135.

[12] Cossalter, V., Peretto, M., and Bobbo, S. (2010). Investigation of the influences of tyre—road friction and engine power on motorcycle racing performance by means of the optimal manoeuvre method. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 224(4):503–519.

[13] Cossalter, V., Lot, R., and Tavernini, D. (2013). Optimization of the centre of mass position of a racing motorcycle in dry and wet track by means of the "optimal maneuver method". In Mechatronics (ICM), 2013 IEEE International Conference on, pages 412–417. IEEE.

[14] Tavernini, D., Massaro, M., Velenis, E., Katzourakis, D. I., and Lot, R. (2013). Minimum time cornering: the effect of road surface and car transmission layout. Vehicle System Dynamics, 51(10):1533–1547.

[15] Tavernini, D., Velenis, E., Lot, R., and Massaro, M. (2014). The optimality of the handbrake cornering technique. Journal of Dynamic 24Systems, Measurement, and Control, 136(4):041019.

[16] Lot, R. and Dal Bianco, N. (2015). The significance of high-order dynamics in lap time simulations. In The Dynamics of Vehicles on Roads and Tracks: Proceedings of the 24th Symposium of the International Association for Vehicle System Dynamics (IAVSD 2015), Graz, Austria, 17-21 August 2015. CRC Press.

[17] Casanova, D. (2000). On minimum time vehicle manoeuvring: The theoretical optimal lap. PhD thesis, School of Engineering, Cranfield University.

[18] Casanova, D., Sharp, R. S., and Symonds, P. (2000). Minimum time manoeuvring: The significance of yaw inertia. Vehicle system dynamics, 34(2):77–115.

[19] Brayshaw, D. (2004). Use of numerical optimisation to determine on-limit handling behaviour of race cars. PhD thesis, School of Engineering, Cranfield University.

[20] Brayshaw, D. and Harrison, M. (2005a). A quasi steady state approach to race car lap simulation in order to understand the effects of racing line and centre of gravity location. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 219(6):725–739.

[21] Brayshaw, D. and Harrison, M. (2005b). Use of numerical optimization to determine the effect of the roll stiffness distribution on race car performance. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 219(10):1141–1151.

[22] Limebeer, D. (2014). Optimising the aero-suspension interactions in a forumla one car. IEEE Transactions on Control Systems Technology.

[23] Limebeer, D. J., Perantoni, G., and Rao, A. (2014). Optimal control of formula one car energy recovery systems. International Journal of Control, 87(10):2065–2080.

[24] Limebeer, D. and Perantoni, G. (2015). Optimal control of a formula one car on a three-dimensional track—part 2: Optimal control. Journal of Dynamic Systems, Measurement, and Control, 137(5):051019.

[25] Bertolazzi, E., Biral, F., and Da Lio, M. (2006). Symbolic-numeric efficient solution of optimal control problems for multibody systems. Journal of computational and applied mathematics, 185(2):404–421.

[26] Patterson, M. A. and Rao, A. V. (2014). Gpops-ii: A matlab software for solving multiple-phase optimal control problems using hp-adaptive gaussian quadrature collocation methods and sparse nonlinear programming. ACM Transactions on Mathematical Software (TOMS), 41(1):1.

[27] Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical programming, 106(1):25–57.