

Optimal control of Formula One car energy recovery systems

D. J. N. Limebeer, G. Perantoni and A. V. Rao *

March 7, 2016

Abstract

The utility of orthogonal collocation methods in the solution of optimal control problems relating to Formula One racing is demonstrated. These methods can be used to optimise driver controls such as the steering, braking and throttle usage, and the optimisation of vehicle parameters such as aerodynamic down force distributions. Of particular interest is the optimal usage of energy recovery systems (ERS). Contemporary kinetic energy recovery systems (KERS) are studied and compared with future hybrid kinetic-thermal/heat ERS known as (ERS-K) and (ERS-H). It is demonstrated that these systems, when properly controlled, can produce contemporary lap time using approximately two thirds of the fuel required by earlier generation (2013 and prior) vehicles.

Keywords: numerical optimal control; orthogonal collocation; pseudospectral methods; lap-time simulation; Formula One car modelling; energy recovery systems.

1 Introduction

MINIMUM-lap-time optimal control problems for Formula One cars are not easy to solve. A discussion of some of the technical difficulties

*David J. N. Limebeer (Corresponding Author) (david.limebeer@eng.ox.ac.uk) and Giacomo Perantoni (giacomo.perantoni@gmail.com) are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ. Anil V. Rao (anilvrao@gmail.com) is with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250. The work was supported by the UK Engineering and Physical Sciences Research Council.

related to the solution of these problems can be found in existing literature; see for example [1, 2]. A historical overview of some of the optimal control literature relating to vehicular applications can be found in [3].

A reader familiar with the classical theory of optimal control and the calculus of variations might think that the most natural approach to the solution of these problems is to use first-order necessary conditions to formulate and solve a two-point-boundary-value problem. This is the approach taken in [4], where two motorcycle manoeuvring problems are solved. In later developments of this line of research an indirect method is used to examine minimum time manoeuvres for a single-track car model through a U-turn [5].

In some recent work a direct transcription method is used to solve various minimum-lap-time optimal control problems for a complete lap [6]. This work employed a car model that reproduces the longitudinal, lateral and yaw dynamics of the vehicle, a realistic tyre model and a curvilinear (differential geometric) description of the track. This approach was successful in solving several minimum-lap-time optimal-control problems with a number of control and parameter optimisation variations. The Modelica modelling and optimization environment [7] has also been used to solve minimum lap time problems [8].

Much of the more recent work in this area is directed to improving the solubility of these problems and focusses on simplifying the vehicle and tyre models, and widening the use of linear optimal control techniques. These include the use of quasi-steady-state models [9] and the use of linear quadratic preview and model predictive control. One approach to minimising the burden associated with minimum-lap-time studies is the use of linear quadratic (LQ) preview to follow a prescribed driving line at a fixed speed [10]. The method proposed makes use of multiple linear models and a gain scheduling scheme, with its operation demonstrated on three track segments. Ideas based on linear approximations have been developed in [11, 12], where (linear time-varying) model predictive control (MP) rather than linear preview control is exploited. This approximation allows one to define the problem of finding a sub-optimal racing line and a speed profile as a convex optimisation problem.

In this paper we study the use of a direct orthogonal collocation method for solving racing-related optimal control problems. Apart from the use of a different optimal control methodology (as compared with [6]), the car model has been upgraded to use a more realistic aerodynamic model. A key focus of this paper is the optimal control of the ERS. Under the 2013 Formula One technical regulations a KERS is allowed to recover kinetic energy from the car during braking, store that energy (usually in a battery that we will refer to as the energy store (ES)), and then make it available to propel the car at

some later point in the race. The energy drawn from the ES to propel the car is limited to 400 KJ on a per-lap basis [13].

In the lead up to the 2008-2017 Formula One engine development freeze it was estimated that an average of 4 milliseconds per lap were gained for every million dollars spent on engine development [14]. This indicates that further engine development is bound not only to be prohibitively expensive, but also incremental in terms of future technical progress. In order to improve Formula One’s ‘green’ credentials, and accelerate road-car relevant power train developments, it was decided that from 2014 Formula One cars will have to recover thermal as well as kinetic energy that would otherwise be lost [15]. The hope is that the introduction of hybrid thermal-kinetic energy recovery systems will produce gains that are far more significant and cost-effective than engine development alone. The results of this paper show that contemporary lap times can be achieved using less powerful turbo-charged engines, which require only two-thirds of the fuel required by contemporary cars, when thermal and kinetic energy recovery systems are employed.

The vehicle and track models are covered in Section 2. Most of this material is standard, or has appeared elsewhere and so this treatment is correspondingly brief. The tyre modelling is covered in Appendix A with the vehicle parameters given in Appendix B. Section 3 briefly reviews the mathematical background underpinning orthogonal collocation transcription techniques. Key is the classical theory of orthogonal polynomials, which underpins most numerical integration schemes — particularly integration quadratures. The results appear in Section 4. The performance of the nominal car is studied first, with this preliminary investigation used as a vehicle for illustrating the use and features of adaptive pseudospectral optimization techniques. The main results of the paper relate to the optimal control of energy storage systems in Formula One racing. We begin with the optimal control of contemporary KERS, which is then followed by an investigation of the hybrid kinetic-thermal energy recovery system to be introduced in the 2014 racing season. The conclusions and acknowledgements appear in Sections 5 and 6.

2 Car and Track Model

The track and car kinematics are modelled using ideas from classical differential geometry. As was explained in [6], the track description is based on measured data with the curvature of the track found as the solution of a subsidiary optimal control problem. The car model is standard and is based on a rigid-body representation of a chassis with longitudinal, lateral and

yaw freedoms. We will use the tyre description given in [2] in combination with an upgraded aerodynamic model. The important geometric modelling quantities are shown in plan view in Figure 1.

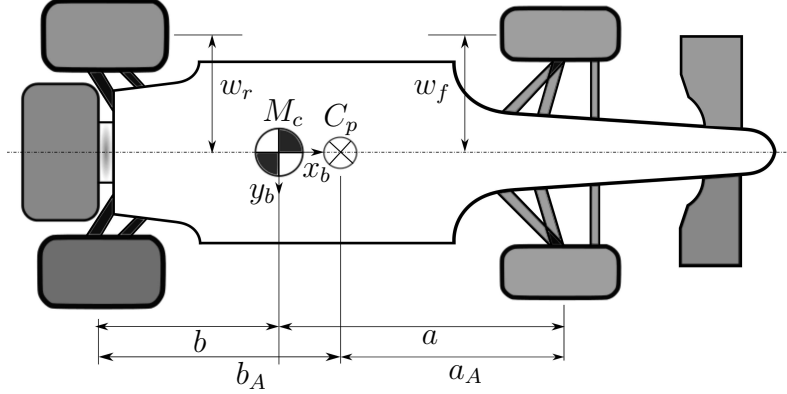


Figure 1: Plan view of a Formula One car with its basic geometric parameters. The car mass centre is at M_c , while the centre of pressure is located at C_p . The body-fixed axes x_b and y_b are in the ground plane.

2.1 Track Model

We will model the track using a curvilinear coordinate system that follows the vehicle using the track centre line position as the curvilinear abscissa [4].

Referring to Figure 2, we describe the location of the mass centre of the vehicle in terms of the curvilinear abscissa $s(t)$ and the vector $\mathbf{n}(s(t))$. The former quantity defines the distance travelled along the track centre line, while the latter gives the position of the vehicle's mass centre in a direction perpendicular to the track centre-line tangent vector $\mathbf{t}(s(t))$. It is assumed that the travelled distance $s(t)$ is an increasing function of time, and that 'time' and 'distance' can be thought of as alternative independent variables. The standard dot notation will be used to signify derivatives with respect to time. At any point s the track's curvature is given by \mathcal{C} (the s -dependence is implied) and its radius of curvature is given by \mathcal{R} . The track centre-line tangent vector \mathbf{t} will be described in terms of the track orientation angle θ , with the track's half-width given by \mathcal{N} . The yaw angle of the vehicle is given by ψ and the angle between the vehicle and the track by ξ ; $\psi = \theta + \xi$. In this coordinate system constraints on the track width are easily expressed in terms of constraints on the magnitude of \mathbf{n} .

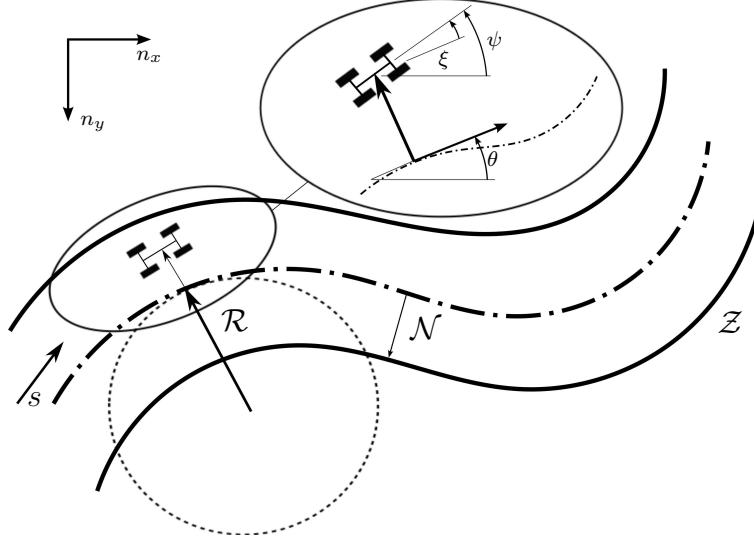


Figure 2: Curvilinear-coordinate-based description of a track segment \mathcal{Z} . The independent variable s represents the elapsed centre-line distance travelled, with $\mathcal{R}(s)$ the radius of curvature and $\mathcal{N}(s)$ the track half-width; n_x and n_y represent an inertial reference frame.

It follows by routine calculation [6] that

$$\dot{s} = \frac{u \cos \xi - v \sin \xi}{1 - n\mathcal{C}}, \quad (1)$$

in which u and v are the longitudinal and lateral components of the car's velocity. The rate of change of n is given by

$$\dot{n} = u \sin \xi + v \cos \xi. \quad (2)$$

Differentiating $\psi = \xi + \theta$ with respect to time results in

$$\dot{\xi} = \dot{\psi} - \mathcal{C}\dot{s}. \quad (3)$$

2.1.1 Change of Independent Variable

The 'distance travelled' will be used as the independent variable. This has the advantage of maintaining an explicit connection with the track position, as well as reducing (by one) the number of problem state variables. Suppose

$$dt = \frac{dt}{ds} ds = S_f(s) ds,$$

where S_f comes from (1) as follows

$$S_f = \left(\frac{ds}{dt} \right)^{-1} = \frac{1 - n\mathcal{C}}{u \cos \xi - v \sin \xi}. \quad (4)$$

The quantity S_f is the reciprocal of the component of the vehicle velocity in the track-tangent direction (on the centre line at s). There follows

$$\frac{dn}{ds} = S_f (u \sin \xi + v \cos \xi) \quad (5)$$

from (2), and

$$\frac{d\xi}{ds} = S_f \omega - \mathcal{C} \quad (6)$$

from (3); $\omega = \dot{\psi}$ is the vehicle yaw rate.

2.2 Car Model

Each tyre produces longitudinal and lateral forces that are responsive to the tyres' slip; see Appendix A. These forces together with the steer and yaw angle definitions are given in Figure 3.

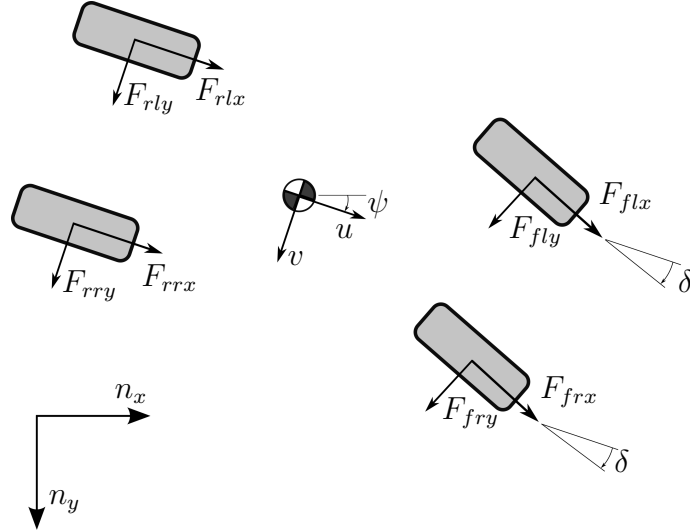


Figure 3: Tyre force system. The inertial reference frame is shown as n_x and n_y .

Balancing forces in the longitudinal and lateral directions, while also bal-

ancing the yaw moments, gives

$$\begin{aligned}
M \frac{d}{dt} u(t) &= M\omega v + F_x \\
M \frac{d}{dt} v(t) &= -M\omega u + F_y \\
I_z \frac{d}{dt} \omega(t) &= a (\cos \delta (F_{fry} + F_{fly}) + \sin \delta (F_{frx} + F_{flx})) + \\
&\quad w_f (\sin \delta F_{fry} - \cos \delta F_{frx}) - w_r F_{rrx} + \\
&\quad w_f (\cos \delta F_{flx} - \sin \delta F_{fly}) + w_r F_{rlx} - b (F_{rry} + F_{rly}), \quad (7)
\end{aligned}$$

in which F_x and F_y are the longitudinal and lateral forces, respectively, acting on the car. These forces are given by

$$F_x = \cos \delta (F_{frx} + F_{flx}) - \sin \delta (F_{fry} + F_{fly}) + (F_{rrx} + F_{rlx}) + F_{ax} \quad (8)$$

$$F_y = \cos \delta (F_{fry} + F_{fly}) + \sin \delta (F_{frx} + F_{flx}) + (F_{rry} + F_{rly}) \quad (9)$$

in which F_{ax} is the aerodynamic drag force. These equations can be expressed in terms of the independent variable s as follows:

$$\frac{du}{ds} = S_f(s) \dot{u} \quad (10)$$

$$\frac{dv}{ds} = S_f(s) \dot{v} \quad (11)$$

$$\frac{d\omega}{ds} = S_f(s) \dot{\omega}. \quad (12)$$

2.3 Tyre Forces

The tyre forces have normal, longitudinal and lateral components that act on the vehicle's chassis at the tyre-ground contact points and react on the inertial frame. The rear-wheel tyre forces are expressed in the vehicle's body-fixed reference frame, while the front tyre forces are expressed in a steered reference frame; refer again to Figure 3. In each case these forces are a function of the normal load and the tyre's longitudinal and lateral slip.

2.3.1 Load Transfer

In order to compute the time-varying tyre loads normal to the ground plane, we balance the forces acting on the car in the n_z direction and balance moments around the body-fixed x_b - and y_b -axes; see Figure 1. Balancing the vertical forces gives

$$0 = F_{rrz} + F_{rlz} + F_{frz} + F_{flz} + Mg + F_{az}, \quad (13)$$

in which the $F_{..z}$'s are the vertical tyre forces for each of its four wheels, g is the acceleration due to gravity and F_{az} is the aerodynamic down force acting on the car. Balancing moments around the car's body-fixed x_b -axis gives

$$0 = w_r(F_{rlz} - F_{rrz}) + w_f(F_{flz} - F_{frz}) + hF_y, \quad (14)$$

in which F_y is the lateral inertial force acting on the car's mass centre; see (9). Balancing moments around the car's body-fixed y_b -axis gives

$$0 = b(F_{rrz} + F_{rlz}) - a(F_{frz} + F_{flz}) + hF_x + (a_A - a)F_{az}, \quad (15)$$

where F_x is the longitudinal inertial force acting on the car's mass centre (see (8)), while F_{az} is the aerodynamic down force.

Equations (13), (14) and (15) are a set of linear equations in four unknowns. A unique solution for the tyre loads can be obtained by adding a suspension-related roll balance relationship, in which the lateral load difference across the front axle is some fraction of the whole

$$F_{frz} - F_{flz} = D_{roll}(F_{frz} + F_{rrz} - F_{flz} - F_{rlz}), \quad (16)$$

where $D_{roll} \in [0, 1]$.

2.3.2 Non-Negative Tyre Loads

The forces satisfying equations (13), (14), (15) and (16) are potentially both positive and negative. Negative forces are indicative of vertical reaction forces, while positive forces are fictitious 'forces of attraction'. Since the model being used here has no pitch, roll or heave freedoms, none of the wheels is free to leave the road, while also keeping faith with (13) to (16).

To cater for the possible 'positive force' situation within a nonlinear programming environment we introduce the vector

$$\bar{F}_z = \begin{bmatrix} \bar{F}_{frz} \\ \bar{F}_{flz} \\ \bar{F}_{rrz} \\ \bar{F}_{rlz} \end{bmatrix}, \quad (17)$$

and define a vector of non-positive loads

$$F_z = \min(\bar{F}_z, 0); \quad (18)$$

the minimum function $\min(\cdot, \cdot)$ is interpreted element-wise. It is clear that \bar{F}_z and F_z will be equal unless at least one entry of \bar{F}_z is positive (i.e. non-physical). We now argue that the model must respect the laws of mechanics

at all times and so equations (13), (14) and (15) must be enforced unconditionally. In contrast, we assume that the solution to (16), which is only an approximate representation of the suspension system, can be ‘relaxed’ in the event of a wheel load sign reversal.

Equations (13), (14) and (15) can be written in matrix form as

$$A_1 F_z = c, \quad (19)$$

while (16) is given by

$$A_2 F_z = 0. \quad (20)$$

In order to deal with the ‘light wheel’ situation, we combine (19) and (20)

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} F_z \\ \bar{F}_z \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}, \quad (21)$$

in which F_z in (20) has been replaced by \bar{F}_z . In the situation where all the wheels are normally loaded, $\bar{F}_z = F_z$ and (21) reduces to (13), (14), (15) and (16). On the other hand, if there is a ‘light wheel’, the mechanics equations (13), (14) and (15) will be satisfied by the non-positive forces F_z , while the roll balance equation is satisfied by the now fictitious forces \bar{F}_z that contain a force of attraction. The non-positive forces F_z are used to represent the normal tyre loads in the rest of the model. It is clear that the four components of \bar{F}_z have to satisfy the nonlinear circularly-dependent relationship (21), which will be solved by a NLP algorithm.

2.4 Aerodynamic Loads

The external forces acting on the car come from the tyres and from aerodynamic influences. The aerodynamic force is applied at the centre of pressure, which is located in the vehicle’s plane of symmetry. The drag and lift forces are given by

$$F_{ax} = -0.5 C_D(u) \rho A u^2, \quad (22)$$

and

$$F_{az} = 0.5 C_L(u) \rho A u^2, \quad (23)$$

respectively. The speed-dependent drag and down-force coefficients, and the speed-dependent location of the aerodynamic centre of pressure are given in Figure 4.

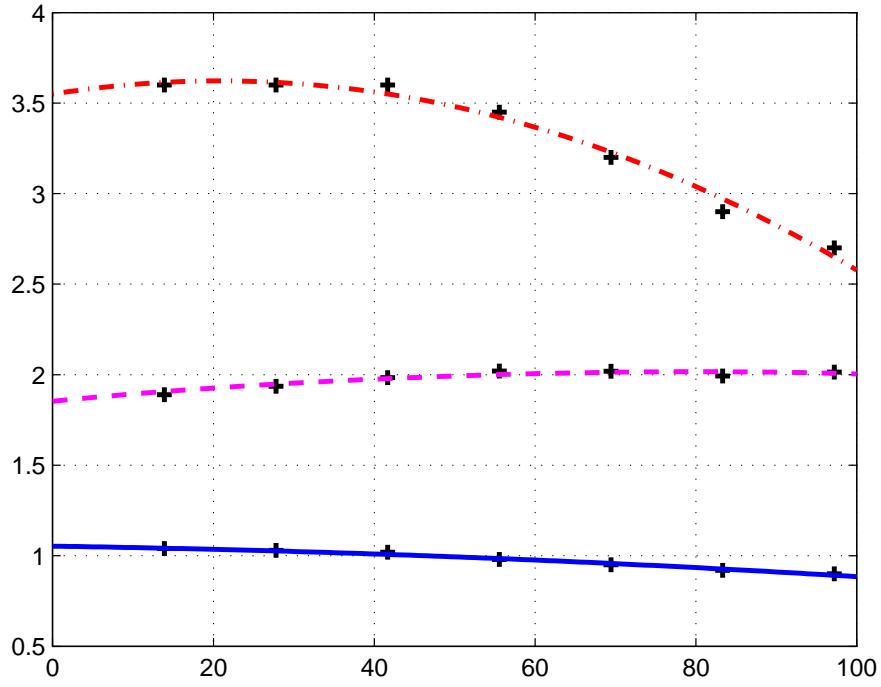


Figure 4: Car aerodynamic maps. The drag coefficient C_D is the solid (blue) curve, which is a function of speed (in m/s). The down-force coefficient C_L , also a function of speed, is given by the dot-dash (red) curve. The speed-depedent aerodynamic centre of pressure is given by the dashed (magenta) curve in metres from the front axle. The ‘+’ symbols represent measured data points.

2.5 Wheel Torque Distribution

In order to optimise the vehicle's performance, one needs to control the torques applied to the individual road wheels. The braking system applies equal pressure to the brake callipers on each axle, with the braking pressures between the front and rear axles satisfying a design ratio. The drive torques applied to the rear wheels are controlled by a differential mechanism.

2.5.1 Brakes

We approximate equal brake calliper pressures with equal braking torques when neither wheel on a particular axle is locked. If a wheel 'locks up', the braking torque applied to the locked wheel may be lower than that applied to the rolling wheel. For the front wheels this constraint is modelled as follows

$$0 = \max(\omega_{fr}, 0) \max(\omega_{fl}, 0)(F_{frx} - F_{flx}), \quad (24)$$

in which ω_{fr} and ω_{fl} are the angular velocities of the front right and front left wheel, respectively. If either road wheel 'locks up', the corresponding angular velocity will be non-positive and the braking torque constraint (24) will be inactive.

2.5.2 Differential

The drive torque is delivered to the rear wheels through a limited-slip differential, which is modelled by

$$R(F_{lrx} - F_{rrx}) = -k_d(\omega_{lr} - \omega_{rr}), \quad (25)$$

in which ω_{lr} and ω_{rr} are the rear-wheel angular velocities, R is the wheel radius and k_d is a torsional damping coefficient. The special cases of an open- and a locked-differential correspond to $k_d = 0$ and k_d arbitrarily large respectively. Limited slipping occurs between these extremes.

3 Orthogonal Collation Methods

The use of orthogonal collocation (pseudospectral) methods in trajectory optimisation problems has been gathering pace since the 1980's. An early example was the solution of the classical brachistochrone problem using a Chebyshev polynomial expansion description of the state and control [16]. In these techniques a series of node points, called *collocation* points, are defined at which the state and control vectors are represented in discrete form. In

collocation methods implicit integration schemes such as Gauss-Legendre quadrature rules [17, 18] are used to solve the system equations and integrate the stage cost. In [19] the pseudospectral method is generalized to consider collocation based on the roots of the derivatives of Jacobi polynomials. The Legendre nodes can be obtained as a particular case of this more general formulation. We do not propose to review this literature, but the interested reader will find much of this work described in [20, 21]. The work presented in this paper makes use of a Radau pseudospectral method [22, 23] that has been implemented in the software package GPOPS-II [24]. The reader may also like to consult [25, 26].

3.1 Optimal control

The purpose of an optimal control computation is to determine the state and control associated with a system in order to minimise a performance index. When expressed in Bolza form, the performance index is given by

$$J = \Phi(t_0, x(t_0), t_f, x(t_f), p) + \int_{t_0}^{t_f} l(t, x(t), u(t), p) dt, \quad (26)$$

while the system and operating constraints are given by

$$\begin{cases} \frac{dx}{dt} - f(t, x(t), u(t), p) = 0 \\ g(t, x(t), u(t), p) = 0 \\ h(t, x(t), u(t), p) \leq 0 \\ g_b(x(t_0), x(t_f), u(t_0), u(t_f), p) = 0 \end{cases} \quad (27)$$

where $t_0 \leq t \leq t_f$ is the optimisation interval with t_f either fixed, or a free optimisation parameter. The vector $p \in \mathbb{R}^{n_p}$ contains fixed parameters to be optimised¹, and $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are the state and control vectors respectively. The vector-valued function $f(\cdot) \in \mathbb{R}^n$ describes the system dynamics. The vector functions $g(\cdot) \in \mathbb{R}^{n_g}$ and $h(\cdot) \in \mathbb{R}^{n_h}$ define the equality and inequality constraints for the system. The subscript b refers to the boundary constraints with $g_b(\cdot) \in \mathbb{R}^{n_{g_b}}$. The scalar function $l(\cdot)$ is the stage cost that is a function of the state, the controls and the parameters.

Certain numerical techniques (like orthogonal collocation methods) were developed with a fixed optimisation interval in mind. In order to normalise the optimisation interval in the general case, one replaces $t \in [t_0, t_f]$ with $\tau \in [-1, 1]$ by using the change of variable

$$t = \frac{t_f - t_0}{2} \tau + \frac{t_f + t_0}{2}. \quad (28)$$

¹ \mathbb{R}^n denotes the set of n -dimensional real vectors.

This mapping can be used with any free finite initial and terminal times.

Direct methods of the type employed here transcribe, or convert, infinite-dimensional optimal control problems into finite-dimensional optimization problems with algebraic constraints.

3.2 Grid Refinement

The numerical solution of an optimal control problem involves three steps: (A) The transcription of the optimal control problem into a nonlinear programming (NLP) problem; (B) the solution of the (sparse) NLP; (C) a review of the accuracy of the solution and, if necessary, a refinement of the mesh followed by a repeat of steps (A) to (C).

The accuracy and efficiency of this process can be influenced by many things including the transcription process itself. Factors favouring the use of simple low-order integration schemes such as the trapezoidal rule include sparse Jacobian matrices, sparse Hessians and fewer right-hand-side evaluations (when finite differencing is used to compute derivatives) as compared with more complex integration schemes. Set against these advantages are the fact that these simple integration schemes have relatively poor accuracy and so a finer grid is needed, thereby inflating the number of NLP decision variables required.

Global orthogonal collocation methods lie at the other end of the spectrum, since a single integration segment is used. In this case increased accuracy comes with increasing the degree of the interpolating polynomials used. For problems whose solutions are smooth pseudospectral methods converge rapidly. The disadvantages associated with (global) pseudospectral methods derive from the fact that even smooth problems may require high-order polynomials. In the case of problems with non-smooth solutions, the convergence rate of a pseudospectral methods may be slow (as the degree of the approximating polynomial increases) and a poor approximation results even if high degree polynomials are used.

A second limitation of global orthogonal collocation methods is that the use of a high degree polynomial results in a dense nonlinear programming problem. An alternative is to segment the optimal control problem (as with conventional integration algorithms), and then employ orthogonal collocation techniques within each segment. GPOPS-II uses a two-tiered grid refinement strategy that refines both the problem segmentation and the orthogonal polynomial orders [27, 28]. If the integration error across a particular segment is uniform the number of collocation points may be increased. If the error at an isolated point within the segment is significantly larger than those at other points within the segment it may be subdivided (at these large-error

Corner	Distance (m)	Corner	Distance
1	725	9	2750
2	800	10	3350
3	1050	11	3500
4	1650	12	3650
5	2000	13	3900
6	2250	14	4000
7	2420	15	4050
8	2500	16	4200

Table 1: Approximate distances to mid corner on the ‘Circuit de Catalunya’ track (in metres from the start-finish line).

points).

4 Results

We begin with a brief study of the nominal vehicle performance with a view to introducing some of the technical features of orthogonal collocation methods with adaptive grid refinement [27,28]. These preliminary results are also used to reconcile the results obtained with orthogonal collocation methods with the simpler techniques employed in [6]. The main purpose of this section is to study racing-related kinetic and thermal energy recovery systems.

4.1 Nominal Car Performance

Our study of the use of orthogonal collocation methods in the solution of minimum-lap-time problems will make use of the Barcelona Formula One Circuit. We begin this section by briefly reviewing the key features of that circuit. Figure 5 shows a plan view of the circuit together with its numbered corners. In preparation for the solution of the optimal control problem, one can also see an initial mesh that comprises twenty segments with four mesh points per segment — the mesh points lie along the track’s centre line. Approximate distances from the start-finish line to each of the corners are given in Table 1. The car and tyre data used in this study is given in Appendix B.

A magnified view of turns ten to sixteen is shown in Figure 6. This figure shows the centre line and a refined grid that follows fifteen grid refinement

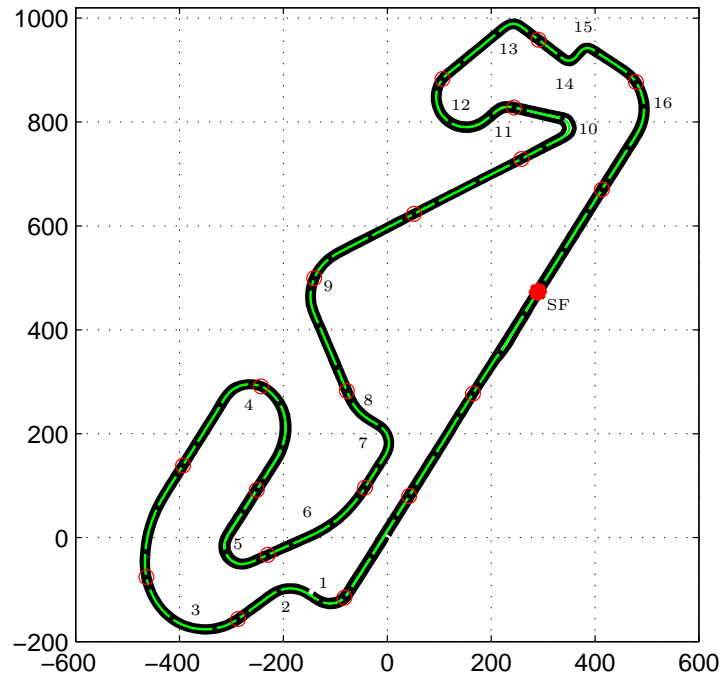


Figure 5: Plan view of the ‘Circuit de Catalunya’ (Barcelona) with the start-finish line shown as $\textcircled{\text{SF}}$. Shown also are the track boundaries, the track centre line (green), the initial mesh segment boundaries (red circles) and the collocation points (black dots); a magnified version of part of the track is given in Figure 6. All distances are in metres.

steps. The final segmentation is uneven as are the orders of the approxim-

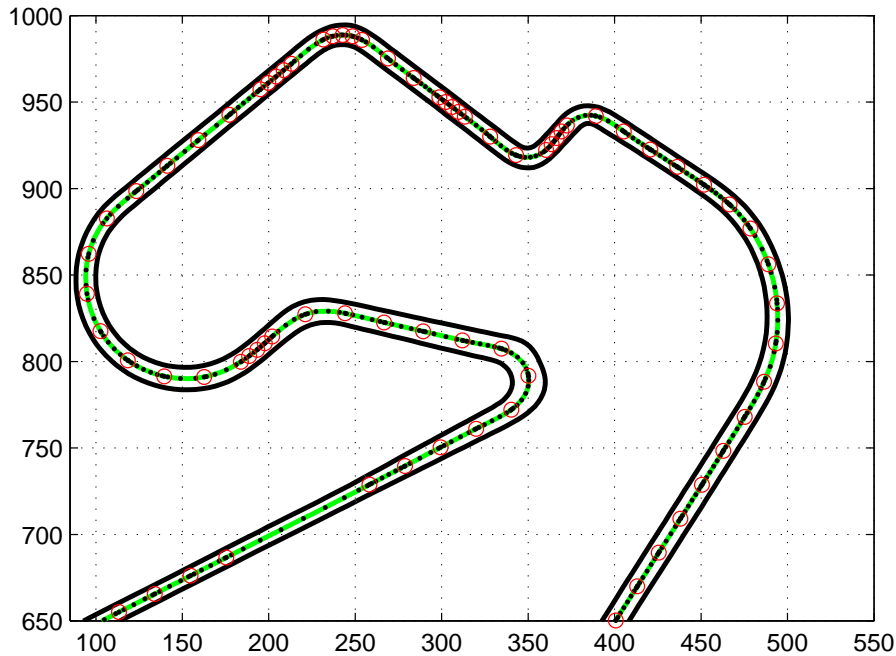


Figure 6: The final optimisation mesh showing the segment boundaries and collocation points. The track centre line (green) with the segment boundaries shown as (red) circles. The black dots are the collocation points.

The solution of the minimum-lap-time problem for the nominal car results in the vehicular speed profile given in Figure 7; the locations of each of the sixteen corners are also shown. The predicted lap time in this case is 80.50 s, with the racing line for part of the circuit shown in Figure 8. These results are in broad agreement with those given in [6], with exact agreement achieved when the simple constant-coefficient aerodynamic model given in [6] is used.

4.2 Energy Recovery Systems

The prohibitively high cost of engine development, in combination with a drive towards a ‘greener’ sport, resulted in the introduction of kinetic and thermal energy recover devices into Formula One. The underlying concept is that energy recovered during braking, and/or from the engine exhaust gasses, should be redeployed to improve both the fuel efficiency and the car’s lap times. Hopefully the continued development of energy recovery systems

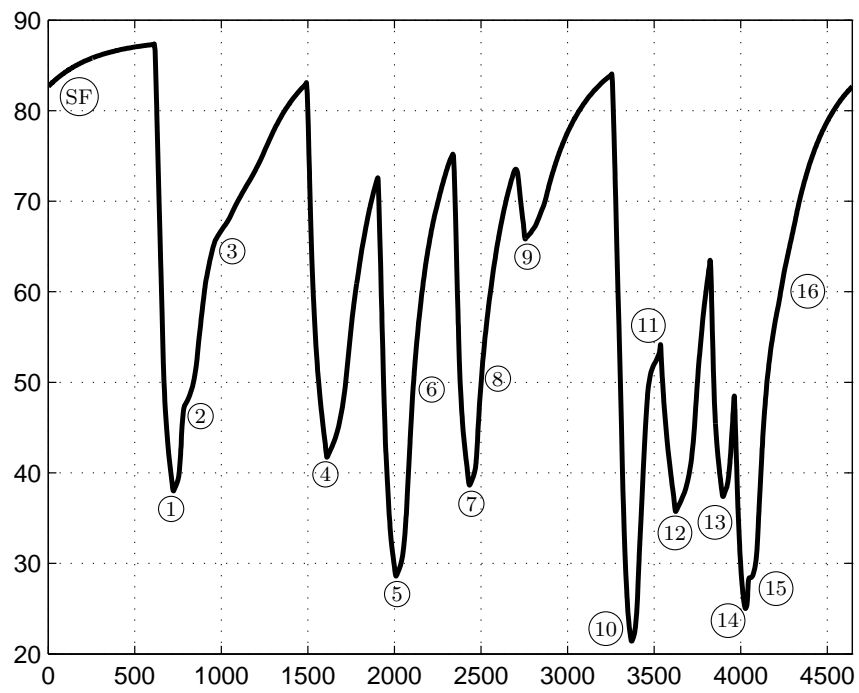


Figure 7: Car speed (in m/s) as a function of position on the Barcelona Formula One track (measured in metres from the start-finish line). The corner numbers correspond to those given in Figure 5.

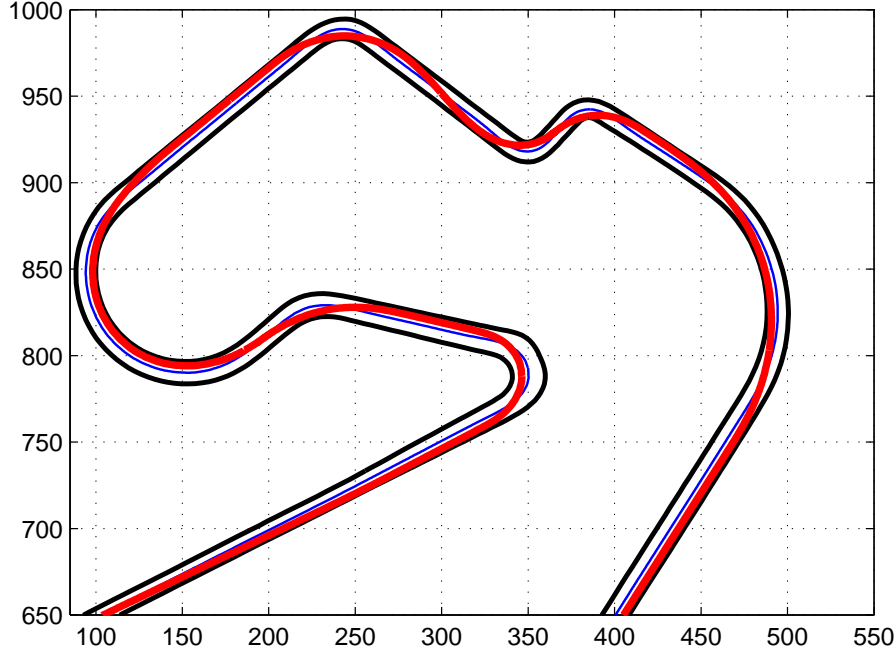


Figure 8: Racing line (red) through tuns ⑩ to ⑯; the track boundaries are shown in (black) with the centre line (blue).

of this type will produce cost-effective gains that are also relevant to road cars. We will study the optimal control of these systems by recognising the power transfer and energy storage restrictions imposed by the technical regulations relating to Formula One [13, 15].

4.2.1 Kinetic Energy Recovery

Contemporary (2013 and prior) KERS facilitate the capture of kinetic energy that derives from braking the moving vehicle. The recovered energy can be stored for later use in propelling the car. Referring to Figure 9 the reader will appreciate that a vehicle without KERS can operate only along the horizontal axis as far to the right as P_{IC}^{max} ; the engine's peak power limit. The left-hand limit will be dictated by the braking capacity of the tyres. With the introduction of KERS the vehicle can operate within the cross-hatched region; the KERS power limits are $P_{kers}^{max} = \pm 60$ kW, while the energy released from the ES may not exceed 400 kJ in any one lap. In order to enforce these restrictions the following path constraints are introduced:

$$\left. \begin{aligned} P_{IC}^{max} + P_{kers} - P_m &\geq 0 \\ P_{kers} - P_m H(-P_m) &\geq 0 \end{aligned} \right\}, \quad (29)$$

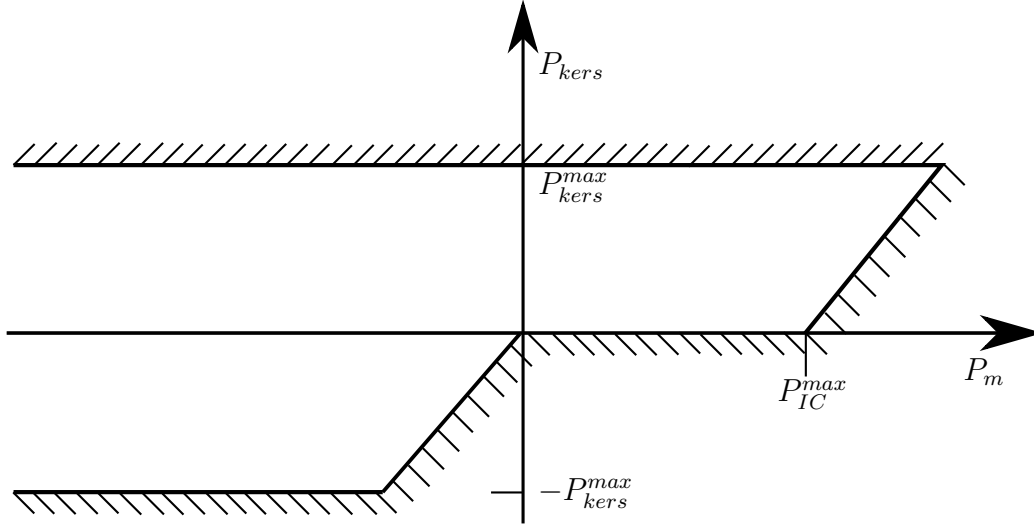


Figure 9: Operating regime of the 2013 KERS. The ordinate represents the KERS power flow P_{kers} , while the abscissa represents the power P_m delivered to the rear wheels.

in which $H(\cdot)$ is the Heaviside step function². It is also necessary to enforce the KERS power limit $-P_{kers}^{max} \leq P_{kers} \leq P_{kers}^{max}$, the ES capacity limit $0 \leq \int_0^s P_{kers} ds \leq 400 \text{ kJ}$ for all values of s to race distance, and the per-lap energy-store discharge limit $0 \leq \int_{lap} H(P_{kers}) P_{kers} ds \leq 400 \text{ kJ}$ required in the regulations [13]; $\int_{lap}(\cdot) ds$ are line integrals carried out around a lap of the circuit.

The minimum-lap-time optimal control problem was solved using GPOPS-II with an optimal lap time of 80.23 s achieved. This is 0.27 s faster than the lap time achieved without the KERS. The KERS power usage is shown in Figure 10, while the energy usage is illustrated in Figure 11. It is clear from Figure 10 that the KERS operates against the allowable power limits using a bang-bang (switching) strategy³, with maximum power drawn from the KERS on fast portions of the track. Figure 10 shows that the KERS is being

² $H(\cdot) = 0$ when the argument is non-positive, with $H(\cdot) = 1$ otherwise. The Heaviside step function is not differentiable at the origin and so is approximated by a smooth function such as $H(x) \approx (1 + x/\sqrt{x^2 + \epsilon})/2$ in which ϵ is ‘small’.

³Optimal control problems with costs and dynamics that are linear in the control usually involve switching strategies. In these problems the control takes on limit values dictated by a switching function that is derivable from the Minimum Principle; this is the well-known *bang-bang principle*. In the case that the switching function is zero for a finite time, the problem has a singular arc and a bang-singular-bang strategy results — see page 246 in [29].

deployed between turns ⑩ and ⑪, through turn ⑥ and between turns ⑨ and ⑩. During periods of rapid deceleration the KERS is recharged, which is evident on the entry to most corners. Bang-bang behaviour is not always possible during recharging, because the KERS regeneration power cannot exceed the braking power available.

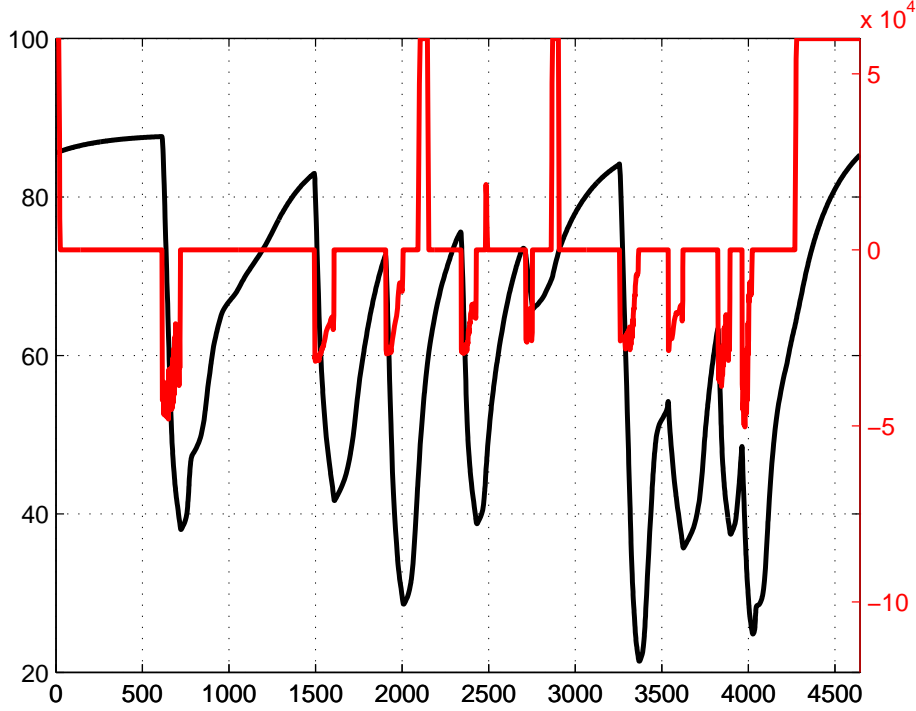


Figure 10: Power usage with conventional KERS. The solid (black) curve shows the vehicle's speed (left-hand axis in m/s), while the solid (red) curve is the power delivered to/from the KERS energy store (right-hand axis in J).

Figure 11 shows the vehicle speed, the state of charge of the KERS⁴ and the energy transferred from the ES to the KERS drive motor as a function of track position. The optimal strategy is to maximize the stored energy on the entry to turn ⑭ in preparation for a period of prolonged KERS usage out of turn ⑯. The KERS is heavily utilised, although its full energy storage capacity is not required. The KERS per-lap discharge limit is an active constraint which is met at the start-finish line; this constraint restricts the KERS power usage to the three full-power pulses shown in Figure 10.

⁴This is defined as a number between 0 and 1; 0 represents a completely discharged ES, while 1 corresponds to a fully charged ES.

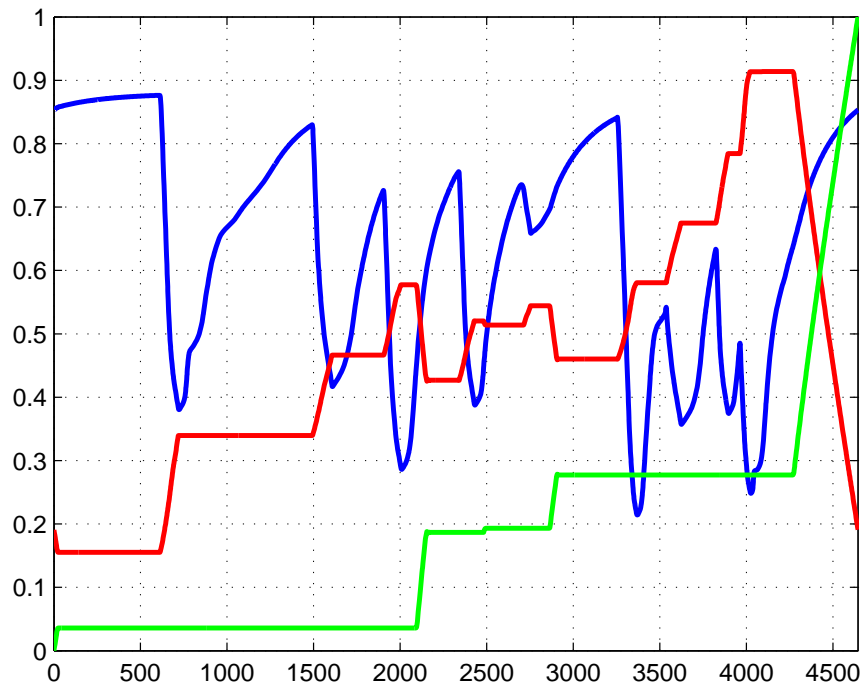


Figure 11: Energy usage with conventional KERS. The (blue) curve is the vehicle speed divided by 100 m/s, the (red) curve is the ES's state of charge, while the (green) curve is the energy transferred from the ES to the KERS divided by 400 KJ for a racing lap.

4.2.2 Kinetic and Thermal Energy Recovery

The energy recovery rules for 2014 are more complex than those used previously, with the optimal control problem correspondingly more intricate. The key differences are these:

1. The car is restricted to a maximum 100 kg of fuel per race;
2. the fuel mass flow rate must not exceed 0.028 kg/s (or 100 kg/hour);
3. turbocompounded engines [30] have been introduced that include a turbine, a compressor and a motor-generator on a common drive shaft; this motor-generator is referred to as the motor-generator-unit-heat (MGU-H). The turbine is bypassed by a controllable waste gate; closing the waste gate will increase the power generated by the MGU-H, but decrease the power generated by the engine. There is no limit on the amount of energy than can be recovered from the exhaust gases. The operation of the thermal energy recovery system is summarised in Figure 12 that shows a plot of power from the internal-combustion engine P_{IC} versus power supplied by the MGU-H; denoted P_h . When the waste gate is open the MGU-H requires 60 kW to operate the compressor and the IC engine output power is boosted by 20 kW. When the waste gate is closed 40 kW is recovered from the MGU-H and the IC engine output power drops to approximately 440 kW. All the figures given here are representative values only.
4. a motor-generator-unit-kinetic (MGU-K) is coupled directly to the engine crankshaft. As before, the MGU-K can be used to recover vehicular kinetic energy during braking, or boost the power to the rear wheels during periods of firm acceleration;
5. the car's ES can store up to 4 MJ;
6. the ES can accept up to 2 MJ per lap from the MGU-K;
7. the MGU-K can draw up to 4 MJ per lap from the ES;
8. power flows to and from the MGU-K is restricted to ± 120 kW;
9. the power and energy flow to and from the MGU-H is unrestricted; see Appendix 3 in [15].

It is clear that operating the engine with an open waste gate is very inefficient from an energy conservation point of view. The engine controls

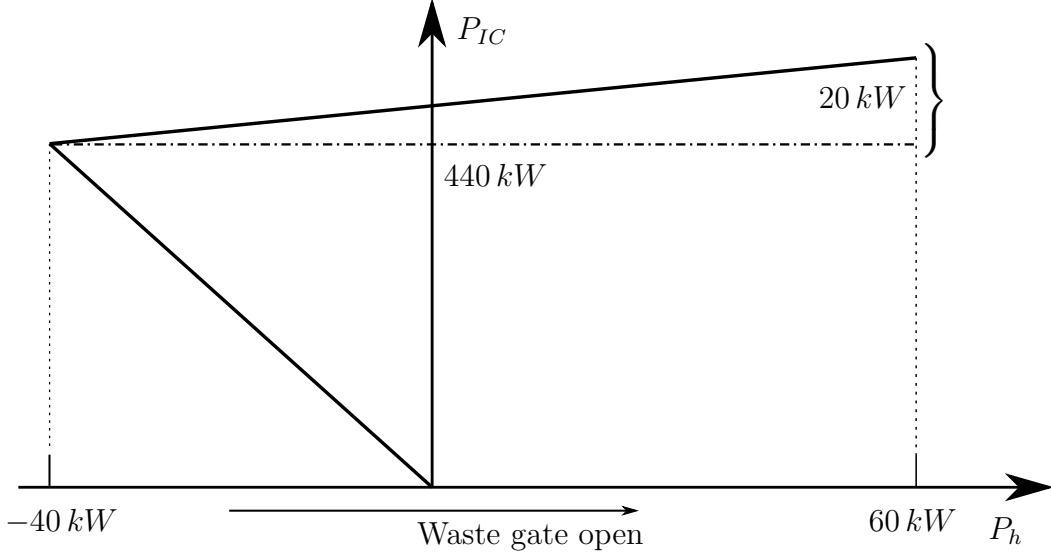


Figure 12: Operating regime of the 2014 engine and thermal energy recovery system (ERS) at full power; power absorbed by the MGU-H is deemed positive.

are the fuel-flow rate and the waste gate opening. It is assumed that the engine full-speed rotational losses are 40 kW.

We will now explain how these various constraints were set up in the optimal control calculation. The power available to the rear wheels, P_m , is constrained by the following inequality:

$$(P_{max} + P_{W_g} W_g) \dot{F} - P_{loss} + P_k - P_m \geq 0. \quad (30)$$

The P_{max} term in equation (30) represents the power generated by the IC engine under full fuelling when the waste gate is closed — $\dot{F} \in [0, 1]$ is the normalized fuel mass flow rate. The second term in (30) represents the power boost resulting from opening the waste gate; $W_g \in [0, 1]$ is the waste gate control with $P_{W_g} = 20$ kW. The third term represents the engine's rotational losses and is set at a constant 40 kW for illustrative purposes, P_k is the power delivered to the MGU-K (P_k is positive when the MGU-K is motoring) and P_m represents the mechanical power delivered to the rear wheels (positive values accelerate the car). The power generated/absorbed by the MGU-H is given by

$$P_h = ((P_h^{max} - P_h^{min}) W_g + P_h^{min}) \dot{F} \quad (31)$$

and is a function of the waste gate opening and the fuel flow rate; $P_h^{min} = -40$ kW and $P_h^{max} = 60$ kW.

In order to monitor the resource constraints associated with fuel usage, ES usage and MGU-K usage, four auxiliary state variables are introduced as follows:

$$F = \int_{lap} \dot{F} ds, \quad (32)$$

This state is used to monitor the fuel consumption and is constrained by $0 \leq F \leq 100/n_l \text{ kg}$, in which n_l is the number of laps in the race. As pointed out in point 2 above, \dot{F} , when normalised, is constrained to the interval $[0, 1]$. The second auxiliary state is given by

$$\dot{E}_s = -(P_h + P_k). \quad (33)$$

This state is used to monitor the stored energy and is constrained by $4 \text{ MJ} \geq E_s \geq 0$ over the whole circuit. The third auxiliary state is described by

$$\dot{E}_{ES2K} = \begin{cases} P_k & P_k > 0, P_h > 0 \\ P_k + P_h & P_k > -P_h > 0 \\ 0 & \text{Otherwise.} \end{cases} \quad (34)$$

This state is used to monitor the energy supplied to the MGU-K from the ES and is constrained by $4 \text{ MJ} \geq E_{ES2K} \geq 0$. The first alternative corresponds to the case when both electrical machines are running as motors. In this case it is only the MGU-K energy usage that is ‘taxed’. The second alternative corresponds to the case in which the MGU-K is motoring, but the MGU-H is generating. In this case the power generated by the MGU-H is off set against the MGU-K power requirements, with the difference attracting a being monitored.

The fourth auxiliary state is used to tax energy flows from the MGU-K into the ES and is described by

$$\dot{E}_{K2ES} = \begin{cases} P_k & P_k < 0, P_h < 0 \\ P_k + P_h & 0 > -P_h > P_k \\ 0 & \text{Otherwise;} \end{cases} \quad (35)$$

this state is constrained by $0 \geq E_{ES2K} \geq -2 \text{ MJ}$. The first alternative corresponds to the case when both electrical machines are operating as generators. In this case it is only the MGU-K energy generation that is penalised. The second alternative corresponds to the case in which the MGU-K is generating, while the MGU-H is motoring. In this case the power required by the MGU-H is off set against the power generated by the MGU-K, with the difference monitored. As before, differentiable approximations to the Heaviside step function are used to enforce constraints (34) and (35). Inequality

constraint (30), in combination with the four auxiliary states (32) to (35) and their associated box constraints are used to model the power and energy constraints required by the 2014 Formula One rules [15].

The minimum-lap-time optimal control problem for the 2014 car was solved using GPOPS-II. The resulting optimal lap time was 82.43 s, which is 1.93 s slower than the lap time achieved with the nominal car studied in Section 4.1. Contributing to this downgraded performance is the fact that the 2014 car is 50 kg heavier than its predecessor and its internal combustion engine is 120 kW less powerful. The commonly-used 0.03 s/kg ‘rule of thumb’ would cut the car’s lap time to 80.94 s if it had a mass of 660 kg rather than 710 kg.

In order to analyse the performance of the 2014 car we will begin with the full lap energy usage as illustrated in Figure 13. The first thing to observe

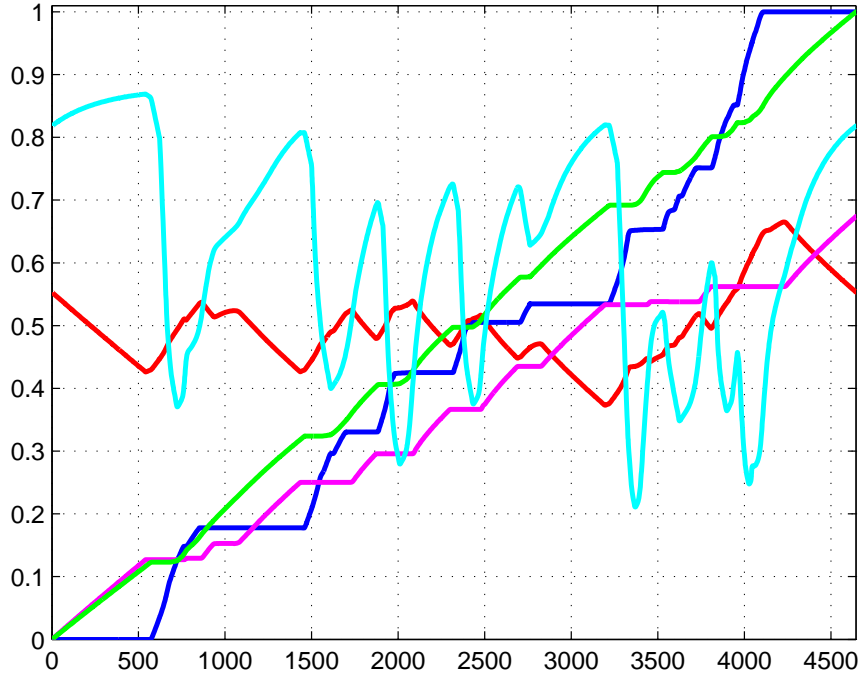


Figure 13: Energy usage in the 2014 ERS. The (cyan) plot is the vehicle speed divided by 100 m/s; the (red) line is the ES state of charge; the (green) plot is the fuel used divided by the race fuel limit/66 - Barcelona is a 66 lap race; the (blue) line is the energy transferred from MGUK to ES divided by 2 MJ; the (magenta) line is the energy transferred from ES to MGUK divided by 4 MJ.

from the red, blue and magenta traces is that an optimal lap is utilising fully the fuel and the MGU-K generation allowances, but not the MGU-K

motoring allowance. Not surprisingly there are braking periods when no fuel is being used — this occurs for example on the entries to turns ①, ④, ⑤ and ⑩; see Table 5. Also as expected, the MGU-K is being used to increase the vehicle speed on the fast straights between for example turns ⑩ and ①, turns ③ and ④, and between turns ⑨ and ⑩. The MGU-K is being used to recharge the ES during braking sections that correspond approximately to the periods of zero fuel usage; see for example the sections entering turns ①, ④, ⑤. There is also intermittent energy recover occurring on the slower section between turns ⑩ and ⑮. The gradients of the magenta and blue curves, which represent the MGU-K usage, differ by a factor of two due to the way in which they have been normalised. Notable is the fact that the ES is under utilised on an optimal lap of the Barcelona circuit; indeed only 1.12 MJ of its 4 MJ capacity is being used. This is a result of restrictions on the total amount of MGU-K generation and motoring allowed, and by the fact that the state of charge at the beginning and end of the lap are constrained to be equal. It is likely that greater use of the ES will occur on sub-optimal laps involving the over taking of other vehicles.

The vehicle’s fuel usage, MGU-K and waste gate controls, and the power delivered to the rear-wheels are illustrated in Figure 14, where one observes the liberal use of the MGU-K to boost the total drive power available at the rear wheels. Again, a bang-bang-type strategy is again being employed by the MGU-K during motoring. The waste gate is essentially unused due to its poor energy efficiency. The MGU-K is again being used to recharge the ES on the entry to corners; several particular instances of this are discussed above. In sum, these results are encouraging and suggest that contemporary car performances can be achieved, on two-thirds of the current fuel usage, especially if the vehicle’s aerodynamic drag can be improved.

Figure 15 considers the case of a qualifying lap for which there is a fuel mass flow rate limit, but no total fuel usage restriction. Another fundamental difference between this case and the one considered in Figure 13 is the freedom to start the lap with a fully charged ES, and to then end it with the ES fully discharged. The qualifying lap is completed in 82.06 s, which is 0.37 s faster than an optimal racing lap with 100 kg of fuel. In this case the MGU-K motoring energy quota is fully utilised, with the MGU-K regeneration quota only partially used. In its qualifying configuration the engine is run with the waste gate open for sustained periods of time when maximum engine power is needed. During these periods of time the ES will be supplying both the MGU-K and the MGU-H, with the latter used to drive the engine boost compressor. The main benefit of using the MGU-K, in combination with an open waste gate during qualifying, is the higher top speeds achieved between turns ⑩ and turn ①, between turns ③ and ④, and between turns ⑨ and

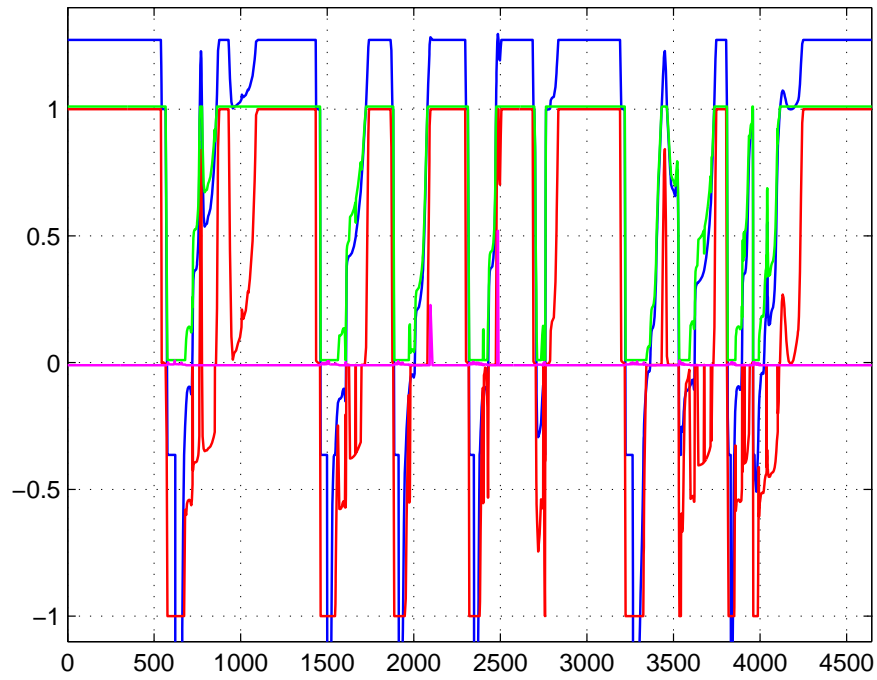


Figure 14: Normalised power usage for the 2014 ERS. The (blue) plot is the mechanical power delivered to the rear wheels divided by 440 kW; the (green) plot is normalised fuel usage rate ('1' corresponds to the maximum allowed rate); the (red) plot is KERS power divided by 120 kW; the (magenta) plot is the waste gate position (0-closed 1-open).

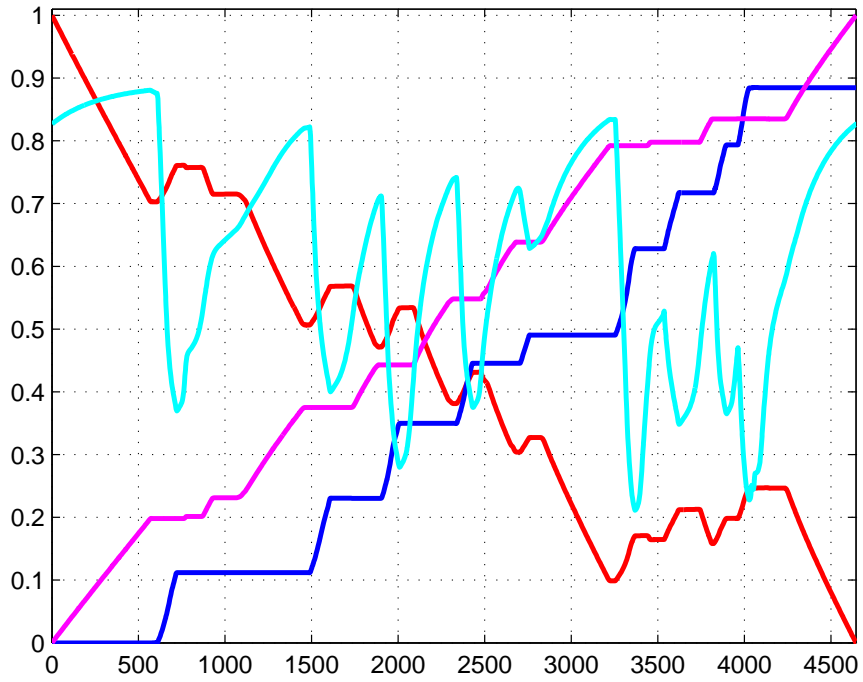


Figure 15: Energy usage for a 2014 ERS qualifying lap. The (cyan) plot is the vehicle speed divided by 100 m/s; the (red) line is the ES state of charge; the (blue) line is the energy transferred from MGUK to ES divided by 2 MJ; the (magenta) line is the energy transferred from ES to MGUK divided by 4 MJ.

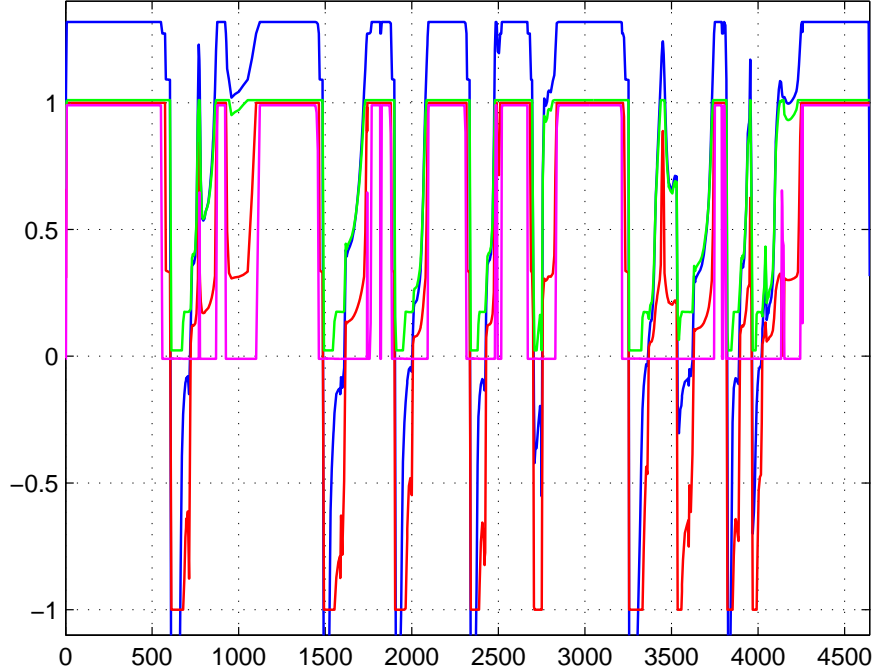


Figure 16: Normalised power usage for the 2014 ERS in qualifying. The (blue) plot is the mechanical power delivered to the rear wheels divided by 440 kW; the (green) plot is normalised fuel usage rate ('1' corresponds to the maximum allowed rate); the (red) plot is KERS power divided by 120 kW; the (magenta) plot is the waste gate position (0-closed 1-open).

The vehicle's fuel usage, MGU-K and waste gate controls, and the power delivered to the rear-wheels during a qualifying lap are illustrated in Figure 16. Unlike the racing lap, the car is repeatedly brought to its top speed by the simultaneous use of the MGU-K and a fully open waste gate; this can be observed on the straight between turns ⑩ and ①, and intermittently, most of the way between turns ③ and ⑩. This extravagant use of the waste gate derives from the ability to simply 'run down' the ES on a qualifying lap. In contrast to the racing lap, the waste gate is typically open when the engine is fully fuelled. On the entry to turns ①, ④, ⑦ and ⑩ the waste gate is being closed a little before simultaneously cutting the fuel and the MGU-K. We believe this strategy is necessary in order to save power as the ES is fully drained at the end of the lap. Also evident is the fact that the MGU-K is used regeneratively on the entry to corners when possible, although its generation quota is not fully utilised.

Figure 17 shows a speed comparison between the qualifying lap with 100 kg of fuel, the racing lap with 100 kg of fuel and a racing lap with only 75 kg of fuel (with a lap time of 85.23 s). The energy management strategy used in the 75 kg case is broadly the same as that employed in the 100 kg case. All the fuel is used, the 2 MJ MGU-K generation quota is fully utilized, while the MGU-K motoring quota is not fully employed; only 0.8 MJ of the energy storage capacity is required. The higher straight line speed for the qualifying lap over the racing lap is clearly evident. The need to ‘lift off’ the throttle into turns ①, ④ and ⑩, thus conserving fuel, is evident from the speed profile in the 75 kg of fuel case

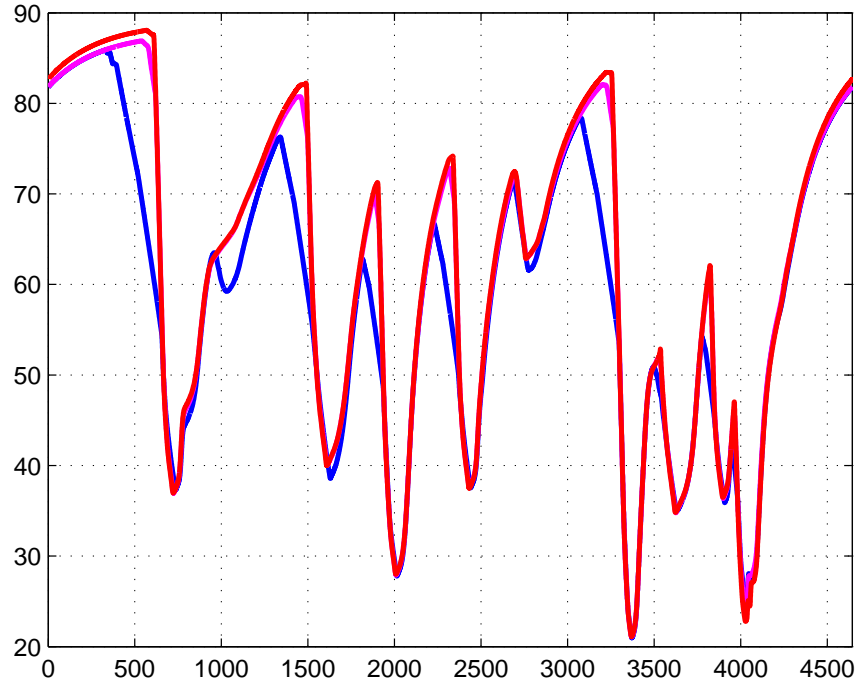


Figure 17: Speed comparison for the three cases studied. The (red) curve is a qualifying lap, the (cyan) curve is the racing lap with 100 kg of fuel, and the (blue) curve is the racing lap with 75 kg of fuel.

5 Conclusion

The use of pseudospectral methods for vehicular optimal control problems have been increasing since the early 1980s. While the majority of these applications relate to aerospace problems, we have demonstrated that they are also well suited to ground-vehicle-related optimisation problems. The

results in this paper focus on the control of energy recovery systems in Formula One racing cars, where the emphasis is on minimising the lap time. This minimum-time type of performance index leads to the use of a bang-bang strategy during KERS-assisted and ERS-assisted driving. A strategy reminiscent of bang-bang control is also employed during braking, although regenerative braking is limited by the braking power available making the limits for the KERS bang-bang control variable. Following the analysis of contemporary KERS, we analysed the control issues relating to the 2014 ERS that has both kinetic and thermal energy recovery features. Unlike the conventional KERS, the upcoming ERS has an explicit fuel-saving role. It is demonstrated that the ERS in combination with a lower-powered turbo-boosted engines can provide contemporary levels of lap-time performance with approximately 100 kg (rather than the current 150 kg) of fuel per race. It is also shown that the internal combustion engine and the ERS have to be operated in a very different manner during qualifying, when the short-term, but profligate use of stored electrical energy, is allowed. The potential for redeploying these technologies in every-day road cars is self-evident. Future work is expected to include a three-dimensional track, more complicated (and realistic) aerodynamic maps and a suspension system.

6 Acknowledgement

This work was supported by the UK Engineering and Physical Sciences Research Council.

References

- [1] D. Casanova, *On Minimum Time Vehicle Manoeuvring: The Theoretical Optimal Lap*. Cranfield University School of Engineering, 2000, PhD Thesis.
- [2] D. P. Kelly, *Lap Time Simulation with Transient Vehicle and Tyre Dynamics*. Cranfield University School of Engineering, 2008, PhD Thesis.
- [3] R. S. Sharp and H. Peng, “Vehicle dynamics applications of optimal control theory,” *Vehicle System Dynamics*, vol. 49, no. 7, pp. 1073 – 1111, 2011.
- [4] V. Cossalter, M. D. Lio, R. Lot, and L. Fabbri, “A general method for the evaluation of vehicle manoeuvrability with special emphasis on motorcycles,” *Vehicle System Dynamics*, vol. 31, pp. 113–135, 1999.
- [5] D. Tavernini, M. Massaro, E. Velenis, D. I. Katzourakis, and R. Lot, “Minimum time cornering: the effect of road surface and car transmission layout,” *Vehicle System Dynamics*, 2013.
- [6] G. Perantoni and D. J. N. Limebeer, “Minimum-lap-time optimal control for a formula one car with variable parameters,” *Vehicle System Dynamics*, 2013, In Press.
- [7] J. Åkesson, K. E. Årzén, M. Gäfvert, T. Bergdahl, and H. Tummescheit, “Modeling and optimization with optimica and jmodelica.org-languages and tools for solving large-scale dynamic optimization problems,” *Computers and Chemical Engineering*, vol. 34, pp. 1737–1749, 2010.
- [8] T. Gustaffsson, “Computing the ideal racing line using optimal control,” Master’s thesis, Linköping University, 2008.
- [9] D. L. Brayshaw and M. F. Harrison, “A quasi steady state approach to race car lap simulation in order to understand the effects of racing line and centre of gravity location,” *Proc. IMechE Part D: J. Automobile Engineering*, vol. 219, pp. 725–739, 2005.
- [10] M. Thommyppillai, S. A. Evangelou, and R. S. Sharp, “Car driving at the limit by adaptive linear optimal preview control,” *Vehicle System Dynamics*, vol. 47, no. 12, pp. 1535 — 1550, 2009.
- [11] J. P. Timings and D. J. Cole, “Vehicle trajectory linearisation to enable efficient optimisation of the constant speed racing line,” *Vehicle System Dynamics*, vol. 50, no. 6, pp. 883 – 901, 2012.

- [12] ———, “Minimum maneuver time calculation using convex optimization,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 135, pp. 0310151–1 to 0310151–9, 2013.
- [13] Fédération Internationale de l’Automobile, “2013 formula one technical regulations,” Tech. Rep., 2012.
- [14] A. Trabesinger, “Power games,” *NATURE*, vol. 447, pp. 900–903, June 2007.
- [15] Fédération Internationale de l’Automobile, “2014 formula one technical regulations,” Tech. Rep., 2013.
- [16] R. V. Dooren and J. Vlassenbroeck, “A new look at the brachistochrone problem,” *Journal of Applied Mathematics and Physics*, vol. 31, pp. 785–790, 1980.
- [17] C. Lanczos, *Applied Analysis*. Englewood Cliffs, N. J.: Prentice Hall, 1956.
- [18] F. B. Hildebrand, *Introduction to Numerical Analysis*, 2nd ed. New York: Dover Publications, 1974.
- [19] P. Williams, “Jacobi pseudospectral method for solving optimal control problems,” *Journal of Guidance, Control and Dynamics*, vol. 27, no. 2, p. 293 to 297, 2004.
- [20] M. A. Patterson and A. V. Rao, “Exploiting sparsity in direct collocation pseudospectral methods for solving optimal control problems,” *Journal of Spacecraft and Rockets*, vol. 39, no. 2, pp. 364–377, March-April 2012.
- [21] D. Garg, W. W. Hager, and A. V. Rao, “Pseudospectral methods for solving infinite-horizon optimal control problems,” *Automatica*, vol. 47, no. 4, pp. 829–837, April 2011.
- [22] D. Garg, M. A. Patterson, C. L. Darby, C. Françolin, G. T. Huntington, W. W. Hager, and A. V. Rao, “Direct trajectory optimization and costate estimation of finite-horizon and infinite-horizon optimal control problems using a radau pseudospectral method,” *Computational Optimization and Applications*, vol. 49, no. 2, pp. 335–358, 2011.
- [23] D. Garg, M. A. Patterson, W. W. Hager, A. V. Rao, A. V. Benson, and G. T. Huntington, “A unified framework for the numerical solution of

- optimal control problems using pseudospectral methods,” *Automatica*, vol. 46, no. 11, pp. 1843–1851, 2010.
- [24] M. A. Patterson and A. V. Rao, “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*, December 2013, Accepted for Publication.
 - [25] J. T. Betts, *Practical methods for Optimal Control and estimation Using Nonlinear Programming*, 2nd ed. Philadelphia, PA: SIAM, 2001.
 - [26] —, “Survey of numerical methods for trajectory optimization,” *Journal of Guidance, Control and Dynamics*, vol. 21, no. 2, pp. 193 – 207, 1998.
 - [27] C. L. Darby, W. W. Hager, and A. V. Rao, “An hp-adaptive pseudospectral method for solving optimal control problems,” *Optim. Control Appl. Meth.*, vol. 32, pp. 476–502, 2011.
 - [28] M. A. Patterson, W. W. Hager, and R. A. V., “A *ph* mesh refinement method for optimal control,” *Optimal Control Applications and Methods*, January 2013, Accepted for Publication.
 - [29] D. E. Kirk, *Optimal Control Theory: An Introduction*. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1970.
 - [30] A. M. Mamat, A. Romagnoli, and R. F. Martinez-Botas, “Design and development of a low-pressure turbine for turbocompounding applications,” *International Journal of Gas Turbine, Propulsion and Power Systems*, vol. 4, no. 3, pp. 1–8, 2012.
 - [31] H. B. Pacejka, *Tyre and Vehicle Dynamics*, second edition ed. Butterworth-Heinemann, 2008.

A Tyre Friction

The tyre frictional forces are modelled using empirical formulae that are responsive to the tyre normal loads and the combined slip. The tyre's longitudinal slip is described by a longitudinal slip coefficient κ , while the lateral slip is described by a slip angle α [31]. Following standard conventions we use

$$\kappa = - \left(1 + \frac{R\omega_w}{u_w} \right), \quad (36)$$

$$\tan \alpha = - \frac{v_w}{u_w}, \quad (37)$$

where R is the wheel radius and ω_w the wheel's spin angular velocity. The quantities u_w and v_w are the absolute speed components of the wheel centre in a wheel-fixed coordinate system. The following formulae give the peak values and locations of the lateral and longitudinal friction coefficients using linear interpolation [2]

$$\mu_{x\max} = (F_z - F_{z1}) \frac{\mu_{x\max2} - \mu_{x\max1}}{F_{z2} - F_{z1}} + \mu_{x\max1}, \quad (38)$$

$$\mu_{y\max} = (F_z - F_{z1}) \frac{\mu_{y\max2} - \mu_{y\max1}}{F_{z2} - F_{z1}} + \mu_{y\max1}, \quad (39)$$

$$\kappa_{\max} = (F_z - F_{z1}) \frac{\kappa_{\max2} - \kappa_{\max1}}{F_{z2} - F_{z1}} + \kappa_{\max1}, \quad (40)$$

$$\alpha_{\max} = (F_z - F_{z1}) \frac{\alpha_{\max2} - \alpha_{\max1}}{F_{z2} - F_{z1}} + \alpha_{\max1}, \quad (41)$$

where the quantities containing a '1' or a '2' in the subscript are measured tyre parameters. Next, the tyre slip is normalised with respect to the peak slip:

$$\kappa_n = \kappa / \kappa_{\max}, \quad (42)$$

$$\alpha_n = \alpha / \alpha_{\max}. \quad (43)$$

$$(44)$$

Following normalisation, the slip is characterised by a combined-slip coefficient

$$\rho = \sqrt{\alpha_n^2 + \kappa_n^2}. \quad (45)$$

The friction coefficients in the longitudinal and lateral directions are described by

$$\mu_x = \mu_{x\max} \sin(Q_x \arctan(S_x \rho)), \quad (46)$$

$$\mu_y = \mu_{y\max} \sin(Q_y \arctan(S_y \rho)). \quad (47)$$

with

$$S_x = \frac{\pi}{2 \arctan(Q_x)}, \quad (48)$$

$$S_y = \frac{\pi}{2 \arctan(Q_y)}. \quad (49)$$

Finally, the longitudinal and lateral components of the tyre forces are given by

$$F_x = \mu_x F_z \frac{\kappa_n}{\rho}, \quad (50)$$

$$F_y = \mu_y F_z \frac{\alpha_n}{\rho}. \quad (51)$$

In the car model used here the normal loads are determined by solving the load balance equations given in Sections 2.3.1 and 2.3.2. The four wheels slip angles are given by

$$\begin{aligned} \alpha_{rr} &= \arctan \left(\frac{v - \dot{\psi} b}{u - \dot{\psi} w_r} \right), \\ \alpha_{rl} &= \arctan \left(\frac{v - \dot{\psi} b}{u + \dot{\psi} w_r} \right), \\ \alpha_{fr} &= \arctan \left(\frac{\sin \delta (\dot{\psi} w_f - u) + \cos \delta (\dot{\psi} a + v)}{\cos \delta (u - \dot{\psi} w_f) + \sin \delta (\dot{\psi} a + v)} \right), \\ \alpha_{fl} &= \arctan \left(\frac{\cos \delta (\dot{\psi} a + v) - \sin \delta (\dot{\psi} w_f + u)}{\cos \delta (\dot{\psi} w_f + u) + \sin \delta (\dot{\psi} a + v)} \right), \end{aligned} \quad (52)$$

with the longitudinal slip coefficients given by

$$\begin{aligned} \kappa_{rr} &= - \left(1 + \frac{R \omega_{rr}}{u - \dot{\psi} w_r} \right), \\ \kappa_{rl} &= - \left(1 + \frac{R \omega_{rl}}{u + \dot{\psi} w_r} \right), \\ \kappa_{fr} &= - \left(1 + \frac{R \omega_{fr}}{\cos \delta (u - \dot{\psi} w_f) + \sin \delta (\dot{\psi} a + v)} \right), \\ \kappa_{fl} &= - \left(1 + \frac{R \omega_{fl}}{\cos \delta (u + \dot{\psi} w_f) + \sin \delta (\dot{\psi} a + v)} \right). \end{aligned} \quad (53)$$

The left-hand part of Figure 18 shows the load dependence of the longitudinal tyre force under pure longitudinal slip conditions; larger longitudinal forces are available when the normal load is increased. The main reason for introducing aerodynamic down-force generating systems on Formula One cars is to exploit this effect. The right-hand part of Figure 18 shows how the longitudinal force is compromised when the tyre is side slipping. As the side-slip angle increases, the longitudinal peak force reduces and moves towards higher slip values.

The normal-load-dependent inverted witch's hat in Figure 19 gives a three-dimensional representation of the combined slip characteristics of the tyre model at a fixed normal load of 2000 N. The tyre's pure slip characteristics are obtained by taking vertical cuts through the lines $\alpha = 0$ and $\kappa = 0$.

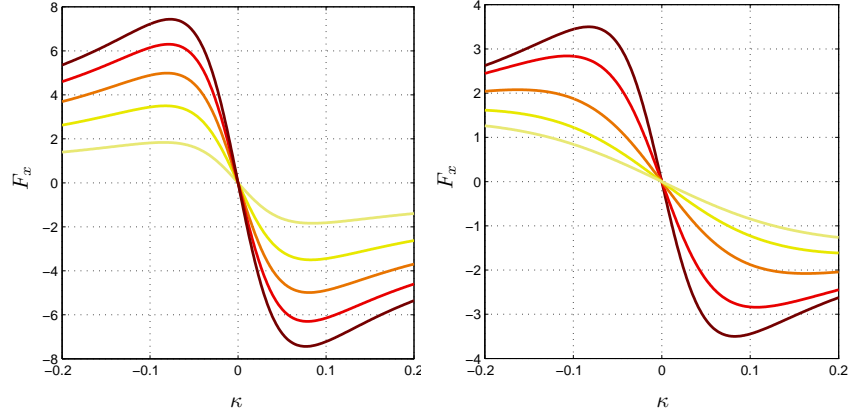


Figure 18: Longitudinal tyre force versus longitudinal slip; F_x is given in kN. The left-hand diagram considers five normal loads (at $\alpha = 0^\circ$); the lightest curve corresponds to 1000 N with the normal load then increased in steps of 1000 N. The right-hand diagram considers combined slip for five values of side-slip angle at a normal load of 2000 N; the darkest curve corresponds to a slip angle of 0° with the slip angle increased in steps of 5° .

B Vehicle and Tyre Data

This appendix contains nominal values for the tyre and vehicle parameters used in this study. The air density is assumed to be $\rho = 1.2 \text{ kg/m}^3$, while the car's frontal area is $A = 1.5 \text{ m}^2$.

Symbol	Description	Value
P_{max}	Peak Engine Power	560/440* kW
P_{kers}^{max}	Maximum KERS Power	60/120* kW
M	vehicle mass	660/710* kg
I_z	moment of inertia about the z -axis	450 $kg\,m^2$
w	wheelbase	3.4 m
a	distance of the mass centre from the front axle	1.8 m
b	distance of the mass centre from the rear axle	$w - a$
h	centre of mass height	0.3 m
D_{roll}	roll moment distribution (fraction at the front axle)	0.5
w_f	front wheel to car centre line distance	0.73 m
w_r	rear wheel to car centre line distance	0.73 m
R	wheel radius	0.33 m
k_d	differential friction coefficient	100 N/rpm

Table 2: Vehicle parameters; values marked with an asterisk relate to the 2014 car.

Symbol	Description	Value
F_{z1}	reference load 1	2000 N
F_{z2}	reference load 2	6000 N
μ_{x1}	peak longitudinal friction coefficient at load 1	1.75
μ_{x2}	peak longitudinal friction coefficient at load 2	1.40
κ_1	slip coefficient for the friction peak at load 1	0.11
κ_2	slip coefficient for the friction peak at load 2	0.10
μ_{y1}	peak lateral friction coefficient at load 1	1.80
μ_{y2}	peak lateral friction coefficient at load 2	1.45
α_1	slip angle for the friction peak at load 1	9°
α_2	slip angle for the friction peak at load 2	8°
Q_x	longitudinal shape factor	1.9
Q_y	lateral shape factor	1.9

Table 3: Tyre friction parameters.

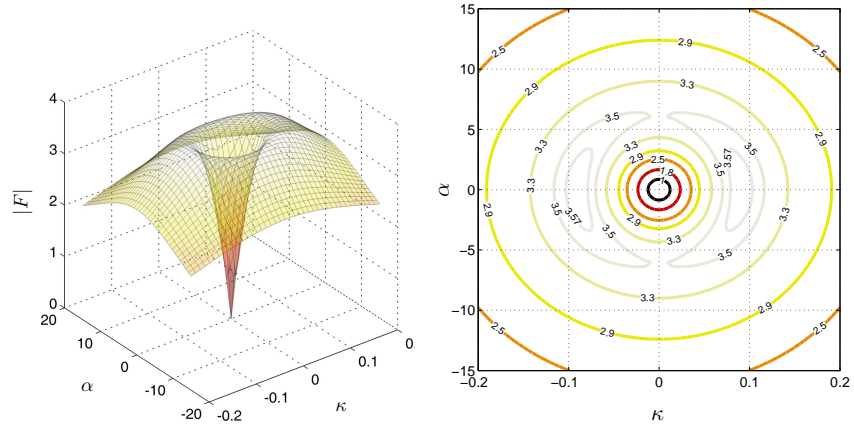


Figure 19: Modulus of the tyre force as function of the longitudinal slip κ and the slip angle α (in degrees) for a normal load of 2000 N. The right-hand diagram shows contours of equal tyre-force.