Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility

Optimal control for a Formula One car with variable parameters

Giacomo Perantoni[a] & David J.N. Limebeer[a]

[a] Department of Engineering Science, University of Oxford, Parks
Road, Oxford OX1 3PJ, UK
Published online: 27 Feb 2014.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Optimal control for a Formula One car with variable parameters

Giacomo Perantoni and David J.N. Limebeer*

Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

The minimum-lap-time optimal control problem for a Formula One race car is solved using direct transcription and nonlinear programming. Features of this work include significantly reduced full-lap solution times and the simultaneous optimisation of the driven line, the driver controls and multiple car set-up parameters. It is shown that significant reductions in the driven lap time can be obtained from track-specific set-up parameter optimisation. Reduced computing times are achieved using a combination of a track description based on curvilinear coordinates, analytical derivatives and model non-dimensionalisation. The curvature of the track centre line is found by solving an auxiliary optimal control problem that negates the difficulties associated with integration drift and trajectory closure.

Keywords: numerical optimal control; lap-time simulation; Formula One car modelling; parameter optimisation; nonlinear programming

1. Introduction

Optimal control calculations for road cars go back to the 1950s, when simple heuristic arguments were employed to estimate lap times and optimise set-up parameters.[1] By the early 1960s, shooting methods, multiple shooting methods and gradient descent methods were used routinely to solve a variety of aerospace flight path optimisation problems.[2] In the context of motor racing, the objective is to complete a given set of manoeuvres (laps) in the minimum possible time. This is commonly referred to as lap-time simulation, or minimum-time manoeuvring. In these studies, the driver is replaced by a control system that 'drives' a car model as hard as possible within the mechanical constraints of the vehicle.

The reader steeped in 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 [3], where two motorcycle manoeuvring problems are solved for the Luco–Poggiosecco curves on the Mugello raceway. This paper introduces an elegant curvilinear-coordinate-based description of the position and orientation of the vehicle on the track that is of general utility. An indirect method is also used in [4], which presents a numerical procedure for solving optimal control problems related to multibody systems. A motorcycle minimum-lap-time problem for the Adria Raceway in the Veneto region of

---

*Corresponding author. Email: david.limebeer@eng.ox.ac.uk

Northern Italy is solved, which compares favourably with data acquired from an on-board inertial measurement device. 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] Here, a series of nonlinear Pacejka-type tyre models, for varying road surfaces, and different transmission layouts are considered.

An early example of a direct numerical method for time-optimal control being applied to road vehicles is given in [6]. This paper considers a lane-change manoeuvre for a simple nonlinear single-mass car model with yaw, lateral, and longitudinal freedoms. The tyre lateral force model is based on classical magic formulae.[7] The optimal control problem was solved using a gradient projection algorithm.([8] p. 394–400) A key advantage of direct methods is that they do not require the adjoint equations, which become more and more difficult to derive as the complexity of the vehicle model and boundary constraints increase.

Recognising the disadvantages associated with conventional indirect methods, Casanova [9] invoked a direct multiple shooting technique that he deemed a sensible compromise between the sensitivity problems associated with conventional shooting methods [2] and the larger nonlinear programming (NLP) problems associated with transcription methods.[10] A disappointing feature of this approach is the long computation time required. Full-lap optimal control calculations for the Barcelona track using a simple four-wheeled car model required between 28 and 60 h to converge on a Sun Spark workstation.[9] Some related work is described in [11], which uses a vehicle model that is similar to the one used in this paper, although the track modelling is done in a different way. The method finds simultaneously the optimal racing line and the controls required to follow this line in minimum time. A full lap of the Jerez circuit with 10 m control spacing requires approximately 8 h of computing time on a single core 2.4 GHz processor.[11] The Modelica modelling and optimisation environment [12] was used to solve various minimum lap-time problems.[13] In order to solve the optimisation problem on closed tracks, the circuit is partitioned into a number of overlapping segments. The Modelica modelling environment was also used in this work to optimise the car set-up parameters.

The apparently unusable computation times associated with high-fidelity optimal control calculations motivated a search for faster optimisation methods. One hypothesis was that the computational time can be reduced by supplying the racing line rather than computing it as part of the lap-time optimisation. Another was the use of quasi-steady-state models, which are now widespread within the Formula One racing community. Both of these ideas were investigated in [14] and, not surprisingly, much shorter computation times were achieved. The reported difference between the predicted lap times using time-optimal and quasi-steady-state analysis for the Barcelona circuit was 2.19 s, which is a significant discrepancy. This discrepancy appears to be the result of different braking/acceleration crossover points used by the two methods.[14]

Another approach to minimising the burden associated with minimum-lap-time studies is the use of linear quadratic preview to follow a prescribed driving line at a fixed speed.[15] 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 further developed in [16], where (linear time-varying) model predictive control rather than linear preview control is exploited. This approximation allows one to define the problem of finding a sub-optimal racing line and speed profile as a sequence of convex optimisation problems. The computed trajectory is sub-optimal, because linear models are used (as opposed to nonlinear ones) in conjunction with control problems that are defined over a sequence of receding track segments (as opposed to the whole track).

The aim of this paper is to study the minimum-lap-time optimal control of a Formula One car with simultaneous parameter optimisation. Calculations of this type include the
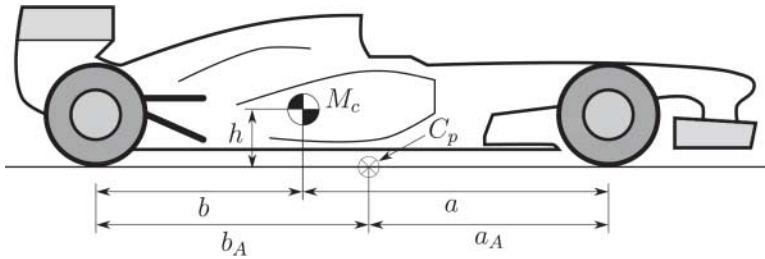
Figure 1. Side view of a Formula One car showing its mass centre M_c and its centre of pressure C_p.

simultaneous optimisation of the steering angle (as a function of the distance travelled from the start–finish line), the drive and braking of the vehicle, and one or more set-up parameters, such as the location of the car's centre of mass and/or its centre of pressure (Figure 1). By way of a theme example, we will seek to minimise the lap time of the Catalunya circuit in Barcelona. The optimal control calculations are carried out using a direct collocation scheme and a NLP algorithm.[10]

Section 2 describes the model of the track and car used in this study. The track is modelled using a curvilinear-coordinate-based description of the centre line. In order to control the noise level in the track curvature model, the curvature is computed by solving an auxiliary optimal control problem based on measured global positioning data (GPS) (see Section 2.1.1). The car model is described in Section 2.2 and has longitudinal, lateral, and yaw freedoms. The model has a combined-slip magic-formula-like tyre description, longitudinal and lateral load transfer effects, four-wheel braking, and a limited-slip differential. The optimal control calculation methodology is described in Section 3. Here, we summarise the optimal control problem formulation and the collocation procedure used. A brief account of our scaling procedures is also given. The results are given in Section 4. The optimisation of the track model is described in Section 4.1. A technique for including stability constraints in the optimal control problem formulation is considered in Section 4.2. Minimum-lap-time results are given in Section 4.3 with a parameter optimisation study given in Section 4.4. The possible treatment of neglected and slow dynamics is remarked on in Section 4.5 with the conclusions given in Section 5. The car parameters are given in Appendix 2.

2. Track and vehicle model

The optimal control calculations presented in this paper require mathematical models of both the track and the vehicle. Since the track modelling is purely kinematic, it will be dealt with first.

2.1. *Track model*

We will model the track using a curvilinear coordinate system that follows the vehicle using the track centre-line arc length as the abscissa. This approach provides a compact way of describing the vehicle's progress and of constraining it to remain within the track boundaries.[3] A curvilinear coordinate system also enables one to deal easily with a track that crosses over itself.

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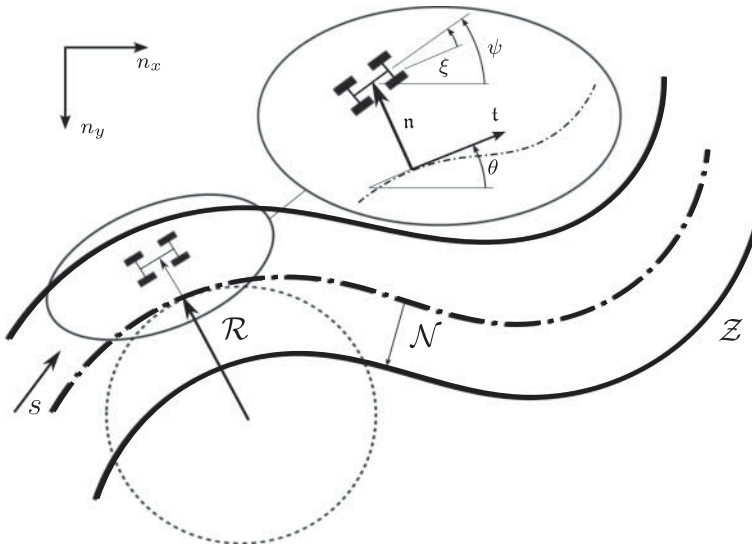
Figure 2.   Curvilinear-coordinate-based description of a track segment $\mathcal{Z}$. The independent variable s represents the elapsed centre-line distance, with $\mathcal{R}$ the radius of curvature and $\mathcal{N}$ the track half-width; n_x and n_y represent the inertial reference frame.

travelled along the track centre line from the start–finish line, while the latter gives the position of the vehicle's mass centre in a direction perpendicular to the track centre-line tangent vector $\mathfrak{t}(s(t))$. It is understood that the travelled distance $s(t)$ is a strictly increasing function of time, and that 'time' and 'distance' can be thought of as alternative independent variables. In order to simplify the notation, the track-related quantities will be described without explicit reference to an independent variable, although this could be time or arc length. The standard dot notation will be used to signify derivatives with respect to time. At any point s the track's radius of curvature is given by $\mathcal{R}$ and its curvature by $\mathcal{C} = 1/\mathcal{R}$. The track centre-line tangent vector $\mathfrak{t}$ will be described in terms of the track orientation angle θ, with the track's half-width given by $\mathcal{N}$.

The absolute yaw angle of the vehicle is given by ψ and the angle between the vehicle and the track by ξ, thus $\psi = \theta + \xi$. In this coordinate system constraints on the track width are easily expressed in terms of constraints on the magnitude of $\mathfrak{n}$. If the half-track width is described by $\mathcal{N}$, then we can constrain the vehicle to drive on the track by simply enforcing $|\mathfrak{n}| \leq \mathcal{N}$.[1] It is clear that in this coordinate system each increment of progress $\mathrm{d}s$ along the track centre line produces increments $\mathrm{d}x$ and $\mathrm{d}y$ in the conventional rectangular inertial system as follows:

$$\mathrm{d}x = \mathrm{d}s \cos\theta, \tag{1}$$

$$\mathrm{d}y = \mathrm{d}s \sin\theta. \tag{2}$$

This gives

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \tan\theta. \tag{3}$$

It follows from elementary geometry that the track curvature is given by

$$\mathcal{C} = \frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}\left(\arctan\frac{\mathrm{d}y}{\mathrm{d}x}\right), \tag{4}$$

which provides a conceptual way of finding the track curvature from GPS data expressed in rectangular coordinates.

Resolving the vehicle's velocity in the t-direction (Figure 2) gives

$$\dot{s} - n\dot{\theta} = u\cos\xi - v\sin\xi, \tag{5}$$

where *u* and *v* are, respectively, the longitudinal and lateral components of the vehicle's mass centre velocity. The lateral position of the vehicle's mass-centre relative to the track centre line is given by *n*, which is in the direction of n; note that $d(n\theta)/dt = n\dot{\theta} + \dot{n}\theta$ with $\dot{n}\theta$ orthogonal to t. Equation (5) gives

$$\dot{s} = \frac{u\cos\xi - v\sin\xi}{1 - n\mathcal{C}} \tag{6}$$

using Equation (4). The rate of change of *n* is given by

$$\dot{n} = u\sin\xi + v\cos\xi. \tag{7}$$

Finally, differentiating $\psi = \xi + \theta$ with respect to time results in

$$\dot{\xi} = \dot{\psi} - \mathcal{C}\dot{s} \tag{8}$$

using Equation (4).

2.1.1. *Track curvature from optimal control*

The track centre line can be reconstructed using a centre-line curvature estimate $\mathcal{C}$ by integrating Equation (4) to find the track orientation angle, and then integrating Equations (1) and (2) to find the track centre line in terms of the rectangular coordinates *x* and *y*. In order to implement this process for closed tracks, the track curvature estimate must satisfy the closure condition $\int_\Omega \mathcal{C}ds = 2n\pi$.[2] This track reconstruction process may fail as a result of integration drift, or if the curvature estimate does not satisfy the closure condition. As we will now show, a track curvature estimate can be found by solving an optimal control problem that explicitly enforces the closure condition while solving Equations (1), (2) and (4) using a single numerical integration process in order to minimise integration drift.

Suppose that the control input $(\tilde{u})$[3] to an optimal control problem is the rate of change of the centre-line curvature

$$\frac{d\mathcal{C}}{ds} = \tilde{u}. \tag{9}$$

It then follows from Equation (4) that

$$\frac{d\theta}{ds} = \mathcal{C}, \tag{10}$$

and from Equations (1) and (2) one obtains

$$\frac{dx}{ds} = \cos\theta, \tag{11}$$

$$\frac{dy}{ds} = \sin\theta. \tag{12}$$

Finally, minimising the performance index

$$J = \int_{s_0}^{s_f} (r\tilde{u}^2 + (x_c - x)^2 + (y_c - y)^2) \, ds, \tag{13}$$

in which x_c and y_c are given GPS data for the track centre line, yields an optimal estimate of the track curvature $\mathcal{C}$ by minimising the error between the measured and estimated centre-line data sets. The constant r in Equation (13) is a weighting factor that determines the importance of the rate of change of the track curvature relative to the reconstruction accuracy. In a sense, the constant r can be used to low-pass filter the estimate of $\mathcal{C}$, but this will be at the expense of increasing the centre-line estimation error.

2.1.2. *Change of independent variable*

Since 'time' and the 'distance travelled' are closely related independent variables, one would naturally assume that either could be dispensed with in the formulation of optimal control problems. Using the 'distance travelled' as the independent variable has the advantage of maintaining an explicit connection with the track position, as well as reducing (by one) the number of problem variables. In order to make this change of variables, one has to ensure that there is a one-to-one correspondence between s and t. Since $ds = \dot{s} \, dt$, one need only to ensure that the vehicle velocity in the track-tangent direction is positive; this will be assumed from now on. Next, we observe that

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

where S_f comes from Equation (6) as follows:

$$S_f = \left(\frac{ds}{dt}\right)^{-1} = \frac{1 - n\mathcal{C}}{u \cos \xi - v \sin \xi}. \tag{14}$$

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), which defines a one-to-one map between s and t when the numerator and denominator are both positive. Selecting the 'distance travelled' as the independent variable, one obtains

$$\frac{dn}{ds} = S_f(u \sin \xi + v \cos \xi) \tag{15}$$

from Equation (7), and

$$\frac{d\xi}{ds} = S_f \omega - \mathcal{C} \tag{16}$$

from Equation (8); $\omega = \dot{\psi}$ is the vehicle yaw rate. Expressing the controls and problem states in terms of the distance travelled allows one to assess the vehicle's performance anywhere on the track.

2.2. *Dynamic model*

The car model we use here is standard and is based on a rigid-body representation of a chassis with longitudinal, lateral and yaw freedoms. The important geometric modelling quantities are shown in plan view in Figure 3.
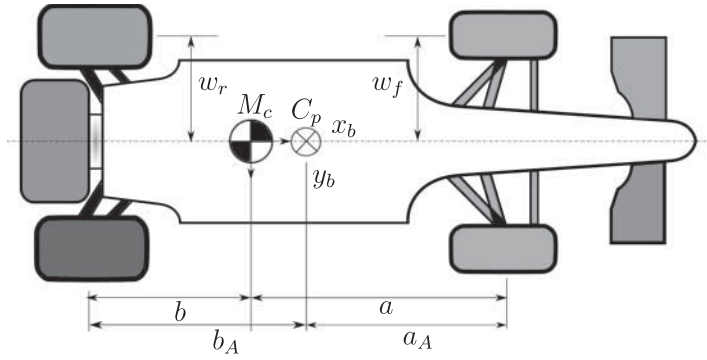
Figure 3. Plan view of a Formula One car with its basic geometric parameters. The body-fixed axes x_b and y_b are in the ground plane.
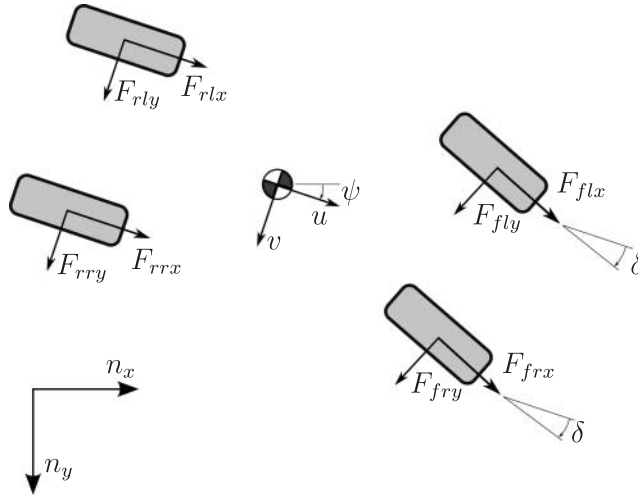
Figure 4. Tyre force system.

Each tyre produces longitudinal and lateral forces in response to slip quantities (see Appendix 1). These forces together with the steer (δ) and yaw (ψ) angle definitions are given in Figure 4; the inertial reference frame is shown as n_x and n_y.

Balancing forces in the longitudinal and lateral directions, while also balancing the yaw moments, gives

$$M \frac{\mathrm{d}}{\mathrm{d}t} u(t) = M\omega v + F_x,$$

$$M \frac{\mathrm{d}}{\mathrm{d}t} v(t) = -M\omega u + F_y,$$

$$I_z \frac{\mathrm{d}}{\mathrm{d}t} \omega(t) = a(\cos\delta(F_{\mathrm{fr}y} + F_{\mathrm{fl}y}) + \sin\delta(F_{\mathrm{fr}x} + F_{\mathrm{fl}x}))$$

$$+ w_\mathrm{f}(\sin\delta F_{\mathrm{fr}y} - \cos\delta F_{\mathrm{fr}x}) - w_\mathrm{r} F_{\mathrm{rr}x}$$

$$+ w_\mathrm{f}(\cos\delta F_{\mathrm{fl}x} - \sin\delta F_{\mathrm{fl}y}) + w_\mathrm{r} F_{\mathrm{rl}x} - b(F_{\mathrm{rr}y} + F_{\mathrm{rl}y}) \qquad (17)$$

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

$$F_x = \cos\delta(F_{\text{fr}x} + F_{\text{fl}x}) - \sin\delta(F_{\text{fr}y} + F_{\text{fl}y}) + (F_{\text{rr}x} + F_{\text{rl}x}) + F_{ax}, \qquad (18)$$

$$F_y = \cos\delta(F_{\text{fr}y} + F_{\text{fl}y}) + \sin\delta(F_{\text{fr}x} + F_{\text{fl}x}) + (F_{\text{rr}y} + F_{\text{rl}y}) \qquad (19)$$

in which F_{ax} is the aerodynamic drag force, defined in Section 2.4. These equations can be expressed in terms of the independent variable *s* as follows:

$$\frac{du}{ds} = S_{\text{f}}(s)\dot{u}, \qquad (20)$$

$$\frac{dv}{ds} = S_{\text{f}}(s)\dot{v}, \qquad (21)$$

$$\frac{d\omega}{ds} = S_{\text{f}}(s)\dot{\omega}. \qquad (22)$$

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 4). In each case, these forces are functions 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 (Figure 3). These calculation must recognise the gravitational, inertial, and aerodynamic forces acting on the car. The vertical force balance gives

$$0 = F_{\text{rr}z} + F_{\text{rl}z} + F_{\text{fr}z} + F_{\text{fl}z} + Mg + F_{az}, \qquad (23)$$

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, defined in Section 2.4. Balancing moments around the car's body-fixed x_b-axis gives

$$0 = w_{\text{r}}(F_{\text{rl}z} - F_{\text{rr}z}) + w_{\text{f}}(F_{\text{fl}z} - F_{\text{fr}z}) + hF_y, \qquad (24)$$

in which F_y is the lateral inertial force acting on the car's mass centre (see Equation (19)). Balancing moments around the car's body-fixed y_{b} -axis gives

$$0 = b(F_{\text{rr}z} + F_{\text{rl}z}) - a(F_{\text{fr}z} + F_{\text{fl}z}) + hF_x + (a_{\text{A}} - a)F_{az}, \qquad (25)$$

where F_x is the longitudinal inertial force acting on the car's mass centre (see Equation (18)).

Equations (23)–(25) 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 total lateral load difference

$$F_{\text{fr}z} - F_{\text{fl}z} = D(F_{\text{fr}z} + F_{\text{rr}z} - F_{\text{fl}z} - F_{\text{rl}z}), \qquad (26)$$

where $D \in [0, 1]$. If this fourth equation is recognised, and Equations (23)–(26) are linearly independent, they have a unique solution for the tyre normal load distribution.

2.3.2. *Non-negative tyre loads*

The forces satisfying Equations (23), (24), (25), and (26) are potentially both positive and negative. If these forces are negative, the corresponding wheels are subject to vertical reaction forces. On the other hand, positive forces represent non-physical 'forces of attraction' that will not be experienced by a real vehicle and should therefore not be allowed in the car model. Since the model being used here has no pitch, roll, or heave freedoms, none of the wheels is free to leave the road, while simultaneously keeping faith with Equations (23)–(26). In order to cater for the possible 'positive force' situation within a nonlinear programming environment, we begin by introducing the vector

$$\bar{F}_z = \begin{bmatrix} \bar{F}_{\mathrm{fr}z} \\ \bar{F}_{\mathrm{fl}z} \\ \bar{F}_{\mathrm{rr}z} \\ \bar{F}_{\mathrm{rl}z} \end{bmatrix}, \tag{27}$$

and by defining the vector of non-positive loads

$$F_z = \min(\bar{F}_z, 0); \tag{28}$$

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 (23)–(25) must be enforced. In contrast, we assume that the solution to Equation (26), which is only an approximate representation of the suspension system, can be 'relaxed' in the event of a wheel load sign reversal.

Equations (23)–(25) can be assembled in matrix form as

$$A_1 F_z = c, \tag{29}$$

while Equation (26) can be written as

$$A_2 F_z = 0. \tag{30}$$

The entries in A_1 and A_2, and the vector c are assembled from Equations (23) to (26).

In order to deal with the 'light wheel' situation, we combine Equations (29) and (30) as

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} F_z \\ \bar{F}_z \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}; \tag{31}$$

with F_z replaced with $\bar{F}_z$ in the second block row. In the situation where all the wheels are normally loaded, $\bar{F}_z = F_z$ and Equation (31) reduces to Equations (23)–(26). On the other hand, if there is a 'light wheel', the mechanics equations (23)–(25) 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. It is clear that the four components of $\bar{F}_z$ have to satisfy the nonlinear relationship (31), which will be solved by a NLP algorithm.

There is a circular dependency between tyre loads and tyre forces inherent in the tyre description. This circularity is treated as a nonlinear set of algebraic equations that is solved as part of the optimising NLP; the tyre friction model is given in Appendix 1.

2.4.   *Aerodynamic loads*

The external forces acting on the car come from the tyres and from aerodynamic influences. The aerodynamic model used here is limited to speed-dependent down force and drag influences acting in the car's plane of symmetry; aerodynamic side forces, yawing moments, and pitching effects have been neglected. As is standard in models of this type, the aerodynamic force is assumed to be quadratic in the vehicle's speed and is applied at the centre of pressure, which is located in the vehicle's plane of symmetry (Figure 1). The down force and drag are modelled as

$$F_{az} = 0.5 C_1 \rho A u^2 \tag{32}$$

and

$$F_{ax} = -0.5 C_d \rho A u^2, \tag{33}$$

respectively, which act in the car's body-fixed z- and x-axis directions, respectively. The constants C_1 and C_d are the aerodynamic down force and drag coefficients, ρ is the air density, and A is the frontal area of the vehicle.

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. In some applications, including current F1 racing, the braking system has to be designed so that equal pressure is applied to the brake callipers of each axle, with the braking pressures between the front and rear axles satisfying some design ratio. The drive torques applied to the rear wheels are controlled by a differential mechanism.

2.5.1.   *Brakes*

We equate equal brake calliper pressures with equal braking torques when neither wheel on a particular axle is locked. If one wheel is locked, the braking torque applied to the locked wheel may be lower than that applied to the rolling wheel. In the case of the front wheels, this constraint is modelled as

$$0 = \max(\omega_{fr}, 0) \max(\omega_{fl}, 0)(F_{frx} - F_{flx}), \tag{34}$$

in which ω_{fr} and ω_{fl} are the angular velocities of the front right and front left wheels, respectively. If either road wheel 'locks up', the corresponding angular velocity will be non-positive and the braking torque constraint (34) becomes inactive.

2.5.2.   *Differential*

We assume that the drive torque is delivered to the rear wheels through a limited-slip differential that is modelled as

$$R(F_{lrx} - F_{rrx}) = -k_d(\omega_{lr} - \omega_{rr}), \tag{35}$$

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

3.   Numerical optimal control

The minimum-lap-time optimal control problem includes nonlinear and non-smooth dynamics, as well as state and control constraints. Care is required when formulating these problems in order that a reliable solution is obtained in a reasonable time. Key steps include selecting the optimisation strategy to be used, selecting a simple vehicle model that captures all the important problem features, finding an efficient and numerically stable track centre-line description, scaling the problem correctly, and developing techniques for 'smoothing out' non-smooth features of the problem so that the gradients required by the optimisation algorithm are properly defined.

3.1.   *Optimal control problem formulation*

The optimal control problem formulation used here involves minimising a Lagrange cost over a finite horizon, subject to position-dependent differential-algebraic equations with equality and inequality constraints on the problem parameters, states, and controls. An optimal control problem general enough for our purposes is

$$\min_{\{p,\tilde{u}(\cdot),\tilde{x}(s_0),\tilde{x}(s_f)\}} \int_{s_0}^{s_f} l(s,p,\tilde{x}(s),\tilde{u}(s))\, \mathrm{d}s, \tag{36}$$

subject to the constraints

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}s} - f(s,p,\tilde{x}(s),\tilde{u}(s)) &= 0, \\
g(s,p,\tilde{x}(s),\tilde{u}(s)) &= 0, \\
h(s,p,\tilde{x}(s),\tilde{u}(s)) &\leq 0, \\
g_{\mathrm{b}}(p,\tilde{x}(s_0),\tilde{x}(s_f)) &= 0,
\end{aligned} \tag{37}$$

where s is the distance travelled along the track centre line; in this case s_0 and s_f are on the start–finish line. The vector $p \in \mathbb{R}^{n_p}$ contains the vehicle parameters to be optimised,[4] and $\tilde{x}(s) \in \mathbb{R}^n$ and $\tilde{u}(s) \in \mathbb{R}^m$ are the vehicle state and control vectors, respectively. The vector function $f(\cdot) \in \mathbb{R}^n$ describes the vehicle 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 (e.g. the track topology and box constraints on the states, controls, and parameters); the subscript b refers to the boundary constraints with $g_{\mathrm{b}}(\cdot) \in \mathbb{R}^{n_{g_{\mathrm{b}}}}$. The scalar function $l(\cdot)$ is the stage cost that is a function of the state, the controls, and the parameters.

In the case of the work presented here, $f(\cdot)$ contains Equations (15), (16), (20), (21), and (22), with the associated states given by n, ξ, u, v, and ω. The normal tyre loads appear in $g(\cdot)$ and are constrained by Equations (28) and (31), with the tyre longitudinal slips constrained by conditions of the form (34). The rear tyre slips are constrained by the differential according to Equation (35). The track width constraint $-\mathcal{N} \leq n \leq \mathcal{N}$ appears in $h(\cdot)$, with the track closure condition $\tilde{x}(s_f) - \tilde{x}(s_0)$ given in $g_{\mathrm{b}}(\cdot)$. The Lagrange cost function is $l(\cdot) = S_{\mathrm{f}}^{-1}$.

3.2.   *Transcription*

Betts [10] gives a detailed overview and classification of the various numerical methods that could be used for solving optimal control problems. There are a number of issues in favour of using a direct solution method that we will mention briefly. Firstly, direct methods do not

require one to find the necessary conditions for optimality. Secondly, direct methods are relatively easy to implement and a wide variety of 'ready-to-use' software is available. Thirdly, parameter optimisation is easily accommodated by simply adding new decision variables to the transcribed optimisation problem. In the work presented here, the MATLAB-based optimal control transcription toolbox ICLOCS [17] was used with trapezoidal integration to convert the optimal control problem into a NLP problem. The ICLOCS software also constructs data structures containing the derivative information required by the NLP algorithm (IPOPT [18]). The required derivatives include first- and second-order partial derivatives of the stage cost, the path constraints, the system dynamics, and the boundary constraints with respect to the decision variables that include the inputs, the states, and the parameters. The interested reader can consult the ICLOCS manual [17] (downloadable), or [10] for a full account of the derivatives required in transcription methods.

In order to negate the influences of the truncation errors associated with finite-differencing, analytical derivatives were used which enhance the convergence of the NLP solution. Since analytical expressions are available for the stage cost, the path constraints, the system dynamics, and the boundary constraints, they can be differentiated symbolically using a symbolic mathematics package such as Maple. These symbolic derivatives were coded in C and compiled into executable functions that are used to generate numerical values of the derivatives for any choice of the decision variables.

The transcription of the optimal control problem given in Equations (36) and (37) into a NLP problem was achieved using the methodology described in [10]. To begin, the dynamic system is discretised in terms of arc length increments, while the continuous state and control functions are quantised into a finite set of decision variables. First, let us introduce a strictly increasing finite sequence of arc length samples $S' = [s_1, s_2, \ldots, s_N] \in \mathbb{R}^N$, with $s_1 = s_0$, $s_N = s_f$, and every other entry $\in (s_0, s_f)$. The sequence S defines the collocation points for the trapezoidal approximation, where N is the number of sampling points. The arc length difference sequence is then defined as $S'_d = [s_2 - s_1, s_3 - s_2, \ldots, s_N - s_{N-1}] \in \mathbb{R}^{N-1}$. If

$$
T_1 = \begin{bmatrix} -1 & +1 & & \\ & \ddots & \ddots & \\ & & -1 & +1 \end{bmatrix}_{(N-1) \times N}, \tag{38}
$$

then

$$
S_d = T_1 S. \tag{39}
$$

Next, we define two parameter vectors that represent the values of the states and the controls at the collocation points, $U' = [\tilde{u}'_1, \ldots, \tilde{u}'_N] \in \mathbb{R}^{Nm}$ and $X' = [\tilde{x}'_1, \ldots, \tilde{x}'_N] \in \mathbb{R}^{Nn}$. We will also require sampled versions of $l(\cdot), f(\cdot), g(\cdot), h(\cdot)$, and $g_b(\cdot)$, which will replace their continuous counterparts in Equation (37). These are given in vector form as

$$
L(S, p, U, X) = \begin{bmatrix} l(s_1, p, \tilde{u}_1, \tilde{x}_1) \\ \vdots \\ l(s_N, p, \tilde{u}_N, \tilde{x}_N) \end{bmatrix} \in \mathbb{R}^N, \tag{40}
$$

$$
F(S, p, U, X) = \begin{bmatrix} f(s_1, p, \tilde{u}_1, \tilde{x}_1) \\ \vdots \\ f(s_N, p, \tilde{u}_N, \tilde{x}_N) \end{bmatrix} \in \mathbb{R}^{Nn}, \tag{41}
$$

$$G(S, p, U, X) = \begin{bmatrix} g(s_1, p, \tilde{u}_1, \tilde{x}_1) \\ \vdots \\ g(s_N, p, \tilde{u}_N, \tilde{x}_N) \end{bmatrix} \in \mathbb{R}^{Nn_g}, \tag{42}$$

$$H(S, p, U, X) = \begin{bmatrix} h(s_1, p, \tilde{u}_1, \tilde{x}_1) \\ \vdots \\ h(s_N, p, \tilde{u}_N, \tilde{x}_N) \end{bmatrix} \in \mathbb{R}^{Nn_h}. \tag{43}$$

The integral cost in Equation (36) is approximated using the trapezoidal rule as follows:

$$\int_{s_0}^{s_f} l(s, p, \tilde{u}(s), \tilde{x}(s)) \, ds \simeq \sum_{i=1}^{N-1} (s_{i+1} - s_i) \frac{l(s_i, p, \tilde{u}_i, \tilde{x}_i) + l(s_{i+1}, p, \tilde{u}_{i+1}, \tilde{x}_{i+1})}{2}$$

$$= S_d' T_2 L(S, p, U, X), \tag{44}$$

with

$$T_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & \\ & \ddots & \ddots & \\ & & \frac{1}{2} & \frac{1}{2} \end{bmatrix}_{(N-1) \times N} . \tag{45}$$

The constraints associated with the system dynamics are approximated using a trapezoidal rule approximation of the differential operator $(d/ds)(\cdot)$:

$$(T_1 \otimes I_n)X = (\text{diag}(S_d)T_2 \otimes I_n)F(S, p, U, X), \tag{46}$$

in which $\otimes$ is the Kronecker product.[19] After transcription, the optimal control problem given in Equations (36) and (37) is re-cast as the NLP problem

$$\min_{[p,U,X]} S_d' T_2 L(S, p, U, X)$$

$$\text{such that} \begin{cases} (T_1 \otimes I_n)X = (\text{diag}(S_d)T_2 \otimes I_n)F(S, p, U, X), \\ G(S, p, U, X) = 0, \\ H(S, p, U, X) \leq 0, \\ g_b(p, \tilde{x}_1, \tilde{x}_N) = 0 \end{cases} . \tag{47}$$

in which the decision variables appear in p, U, and X.

3.3. *State and control scaling*

Scaling can have a significant influence on the performance of optimisation algorithms. Since convergence tolerances and other criteria are necessarily based on some notion of 'small' and 'large' quantities, problems with unusual or unbalanced scaling may cause difficulty. One notion of scaling is to transform the variables from their original representation, which may reflect the physical nature of the problem, to variables that have desirable properties in terms of optimisation.

3.3.1. *Non-dimensionalisation*

In problems that involve physical quantities, poor scaling can be a direct consequence of the choice of the problem units. In our case, using SI units, the vehicle mainframe has a mass

of approximately 600 kg and the acceleration due to gravity is approximately $g = 9.8\,\text{ms}^{-2}$. When the effect of aerodynamic down force is included, the normal tyre loads and the tyre friction forces can be of the order of $10^4\,\text{N}$ in some cases. The forces acting on the vehicle are thus 4 orders of magnitude larger than length-related decision variables, angles (e.g. steering), and/or tyre slip coefficients. For these reasons, we non-dimensionalise the model by normalising the basic physical units of length, mass, and time. All the mass- and length-related quantities are scaled so that the vehicle has unit mass and unit length. Time is scaled so that the acceleration due to gravity is unity too; after non-dimensionalisation, angles are still expressed in radians.

3.4. *Smoothing functions*

Interior point NLP algorithms such as IPOPT need first- and second-order derivative information for the functions defining the cost and constraints. For this reason, non-smooth problem features have to be approximated in a way that does not change significantly the problem's solution. Functions such as $\min(x, 0)$ and $\max(x, 0)$ have undefined derivatives at $x = 0$ and are therefore approximated using

$$\max(x, 0) \approx \frac{x + \sqrt{x^2 + \epsilon}}{2}, \quad \min(x, 0) \approx -\frac{-x + \sqrt{x^2 + \epsilon}}{2}, \tag{48}$$

in which ϵ is a 'small' constant. Since

$$\frac{\partial}{\partial x}\left(\frac{1}{2}(x + \sqrt{x^2 + \epsilon})\right) = \frac{1}{2}\left(\frac{x}{\sqrt{x^2 + \epsilon}} + 1\right), \tag{49}$$

one sees that the derivative approximation for $\max(x, 0)$ at $x = 0$ is well defined as $\frac{1}{2}$. These approximations are more accurate for smaller values of ϵ, with values in the range $10^{-4} \leq \epsilon \leq 10^{-2}$ used for the computation of the results given here.

4. Results

In this section, we present results relating to several minimum-lap-time and parameter optimisation studies. Before presenting the main results, we will deal with some preliminaries relating to the track centre-line modelling and possible dynamic instabilities. The vehicle's geometric, mechanical, tyre, and aerodynamic data are given in Appendix 2.

4.1. *Optimal track description*

We will now find an optimised curvilinear coordinate description of the centre line of the Circuit de Catalunya using measured GPS data and the optimal control problem described in Section 2.1.1. Figure 5 shows the measured Barcelona raceway and its centre line. In order to solve minimum-time optimal control problems, the track curvature is required (as a function of the elapsed centre-line distance). This has been computed for two values of the curvature-rate weighting factor r given in Equation (13). As expected, the fit between the 'true' centre line and the curvilinear approximation degrades when r is increased; this discrepancy is best seen in turns ⑩, ⑭, and ⑮ as shown on the right-hand side of Figure 5.

When solving the curvature optimisation problem there is a trade-off between the accuracy of the centre-line estimate and noise in the curvature estimate. Figure 6 illustrates this
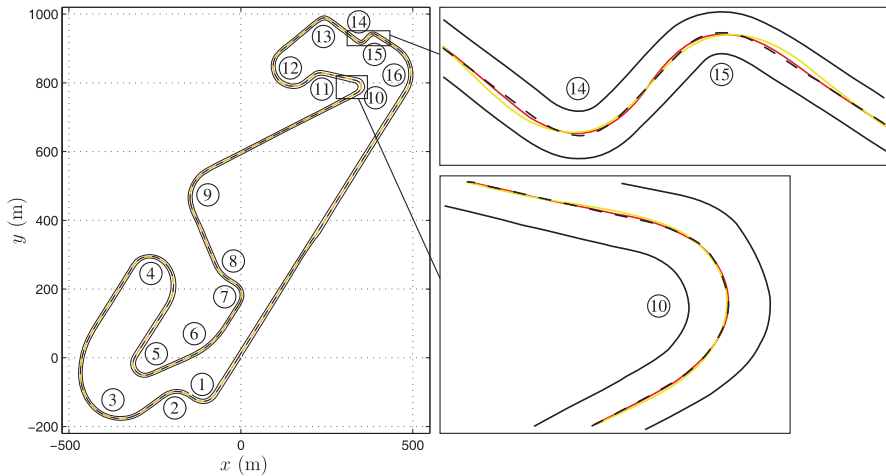
Figure 5. Plan view of the 'Circuit de Catalunya' (Barcelona). The solid black lines are the track boundaries derived from GPS data, with the black dashed line the track centre line. The light (yellow) and dark (red) solid curves are the track centre lines computed for weights of $r = 10^6$ and $r = 10^5$, respectively, in Equation (13). The insets on the right-hand side show magnified versions of two track sectors.
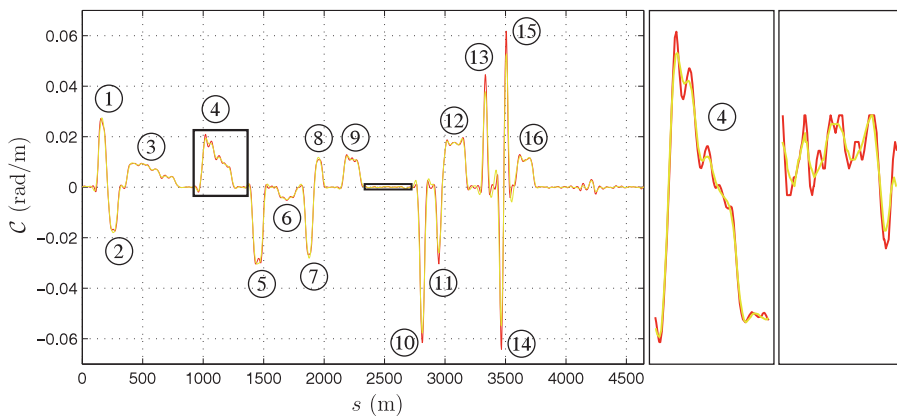

Figure 6. Optimised curvature estimates as functions of the elapsed distance for the centre-line descriptions given in Figure 5. The colour coding is consistent with Figure 5 for different values of r. The insets on the right-hand side show two track sections.

compromise, with magnified views shown for turn ④, and the straight between turns ⑨ and ⑩; the higher accuracy centre-line fit is accompanied by increased noise levels in the track curvature estimate.

4.2. *Vehicle stability*

Under over-steering conditions four-wheeled vehicles can be unstable in yaw.[7] While stability-related constraints have not been included explicitly in the optimal control problem formulation described in Section 3.1, it is possible to include them implicitly by restricting the range over which parameters can be optimised. To show how this might be accomplished we consider briefly the impact of the location of the aerodynamic centre of pressure on the yaw-mode stability of the car under high-speed cornering. The conditions we will examine
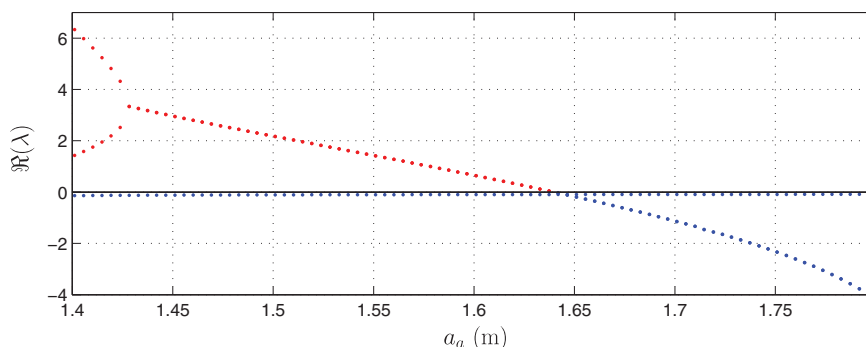
Figure 7. Yaw-mode stability as a function of the location of the aerodynamic centre of pressure. The plot shows the real part of the yaw-mode eigenvalue as a function of the distance between the front axle and the centre of pressure (Figure 1); negative real parts are displayed in blue, while positive ones are red. The model is linearised at a constant speed of 65 m/s in a corner with radius of curvature 150 m, with the aerodynamic centre of pressure moved towards the front axle at 0.0033 m/s.

are representative of the racing conditions in turn ③ of the Barcelona track, which we will assume is a 65 m/s corner with a constant radius of curvature of 150 m.

For this analysis we use a vehicle model that includes a parallel spring–damper double-wishbone suspension system, and the gyroscopic effects related to the road wheels. The car's parameters are based on those given in Appendix 2, with the suspension parameters used in the sprung model consistent with the nominal roll balance used for the optimal control calculations. In order to establish the vehicle's stability properties, the sprung vehicle was driven at constant speed (65 m/s) around a circular trajectory of constant radius (150 m). The equilibrium states thus obtained were used to find a small-perturbation linear model that is continuously updated as the centre of pressure is moved slowly towards the front axle. The vehicle's yaw-mode stability is analysed as a function of the location of the centre of pressure by plotting the centre-of-pressure-dependent eigenvalues of the linearised model in Figure 7. For the cornering conditions considered, when the aerodynamic centre of pressure is behind $a_A = 1.64$ m the vehicle is stable, but the vehicle becomes unstable in yaw when the centre of pressure is forward of this point. Stability information of this type can be used to restrict the search range over which parameters might be optimised. The nominal location of the aerodynamic centre of pressure is at $a_A = 1.9$ m and one could constrain any parameter optimisation problem so that $a_A \geq 1.7$ m, for example. Once potential instabilities of this type have been identified, it is a straightforward matter to constrain parameter optimisation studies to avoid dynamic instabilities in parameter space.

4.3. *Optimal lap*

In this section, we study the behaviour of the optimally controlled car using the nominal parameter set given in Appendix 2. The centre-line curvature estimate used here corresponds to the case $r = 10^6$ as shown in Figure 5. The speed profile for a computed minimum-time lap of the Barcelona track is given in Figure 8. The computed lap time is 82.43 s, which was obtained using a fixed uniform grid that will be used for all the calculations in this section.

The optimal racing line for four track sections is shown in Figure 9. In each case, the cornering manoeuvre starts with a ◯ and ends with a ★. As one would expect, the optimal racing line seeks to 'smooth out' the corners by maximising the radius of curvature of the trajectory throughout the manoeuvre.
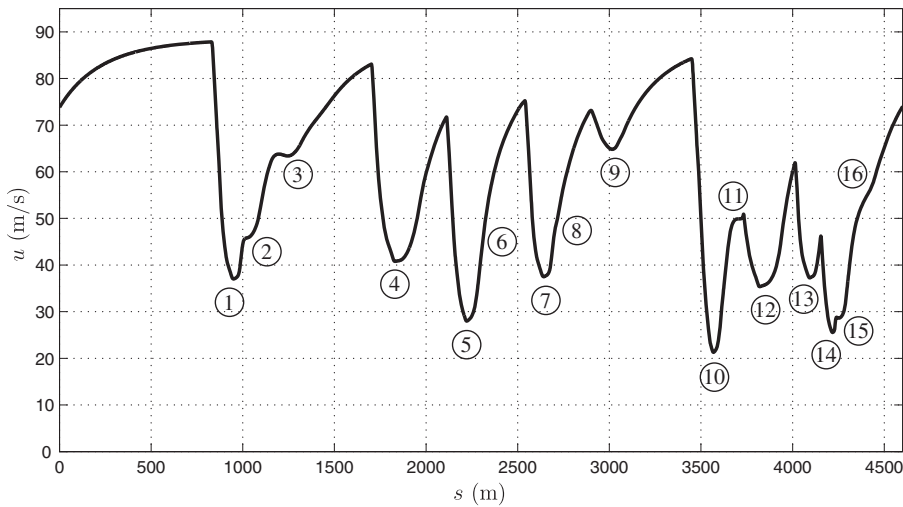
Figure 8. Car speed as a function of the distance travelled from the start–finish line on the Barcelona Formula One track. The centre line is estimated using a curvature-rate weighting of $r = 10^6$ in the performance index (13) as shown in Figure 5. The corner numbers correspond to those given in the same figure.

Figure 10 focuses on the tyre force system behaviour in turn ① at the six points highlighted in Figure 9. At point ◯ the front tyres are essentially free-rolling while they are both under equal normal loads of just over 6000 N. At this point the rear tyres are supporting equal longitudinal drive forces of just under 3000 N and normal loads of ≈7500 N. During the braking phase between △ and □ large longitudinal braking forces are produced by all four tyres with a simultaneous reduction in the normal loads resulting from reduced aerodynamic down force. There is also a load transfer from the rear to the front axle during braking. During the mid-corner at ◇ the tyres are experiencing almost no longitudinal force and large lateral forces on the left-hand (outer) tyres that are also experiencing high normal loads due to a lateral load transfer. Once the car starts to exit the corner between ▽ and ★ the lateral tyre forces reduce, the drive forces on the rear tyres increase, and the lateral load transfer is reversed in preparation for the following left-hand turn ②.

Figure 11 shows the effect of the mesh size variations on the predicted lap time and the solution time. The important message from this figure is that the predicted optimal lap time approaches an asymptotic value of 82.57 s. A subsidiary point is that the solution time increases significantly for finer meshes, but without any apparent benefit in terms of predic- tion accuracy. Once the mesh length falls below 0.5 m, solution times of the order of hours can be expected. This figure also shows that for the computation meshes considered the solu- tion time grows approximately according to $t_{NLP} \approx 5 \times 10^{-3}N^{1.5}$ in which N is the number of mesh points. This provides evidence that the NLP solver (IPOPT) is exploiting the sparsity of the gradient matrices, since otherwise the computation time would increase according to $t_{NLP} \propto N^3$, the time taken to solve a dense set of linear equations. If for the Barcelona circuit Δs is a (fixed) mesh length, the number of samples is approximately $N = 4655/\Delta s$.

4.4. *Parameter optimisation*

One of the important attributes of direct methods for optimal control is the ease with which they optimise simultaneously parameter values, and optimal control and state trajectories. We will exploit this facility by optimising a number of set-up parameter values. The goal
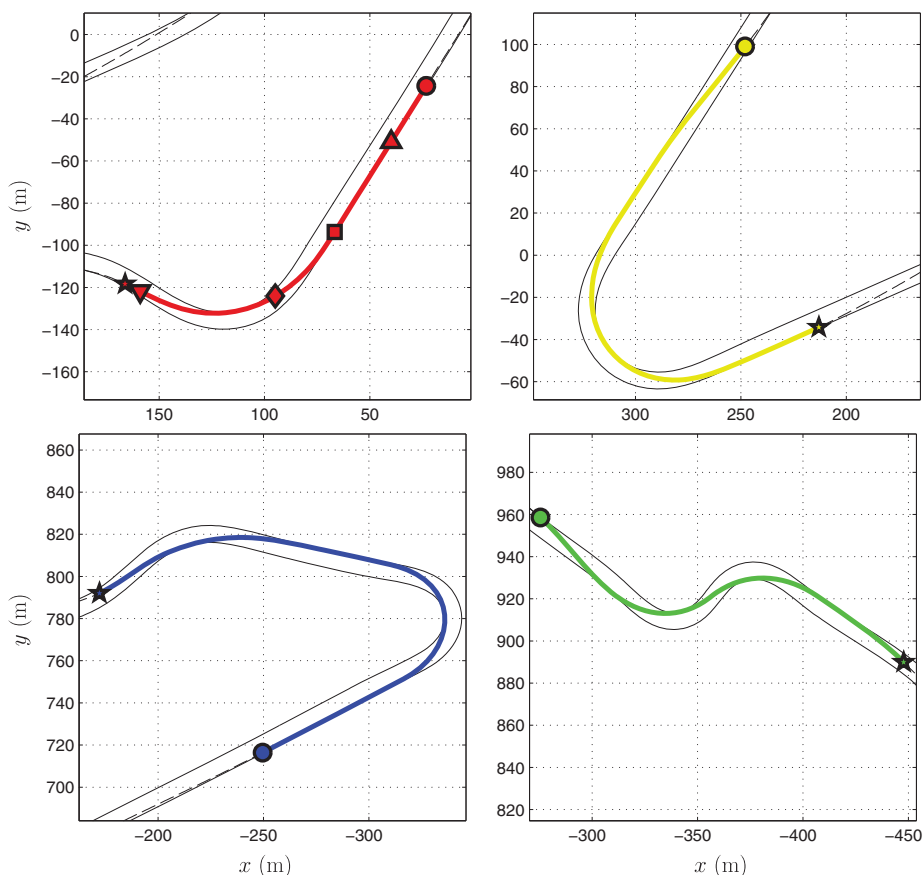
Figure 9. Optimal trajectory of the car's centre of mass through turns ①, ⑤, ⑩–⑪, and ⑭–⑮. The markers associated with different points in the corners are as follows: the start of the manoeuvre is marked with a ◯; the start of the braking phase is marked with a △; the start of the cornering phase is marked with a □; the ◇ marks the beginning of the corner exit acceleration phase; the ▽ symbol marks a reversal in the direction of the tyre side forces, while the ★ is the end of the manoeuvre. The marker points defined for turn ① correspond to specific transition points in the tyre forces shown in Figure 10.

is to assess the sensitivity and importance of these parameters in minimising optimal lap times. Figure 12 considers the optimisation of the location of the vehicle's mass centre, its aerodynamic centre of pressure, its roll balance, and finally, the differential viscosity constant. Each of these parameters is first adjusted on a one-parameter-at-a-time basis with multi-parameter optimisation considered after that.

The first plot in Figure 12 shows that the nominal location of the vehicle's mass centre is almost optimal, with little to be gained from a small 1.6 cm rearward adjustment to its optimal location. The optimisation of the aerodynamic centre of pressure location requires a rearward shift of 12.5 cm, which has the effect of reducing the car's optimum lap time by approximately 0.22 s. The third plot shows that the roll stiffness (see Equation (26)) should be increased from its nominal value of 0.5 to 0.629 thereby reducing the optimal lap time by 0.49 s in this case. The single parameter in the group studied that appears to be most influential in reducing the optimum lap time is the viscosity constant of the differential (see Equation (35)).[5] By running the car with an almost locked differential, rather than with a viscosity constant of $k_d = 10.47$ Nms/rad, a reduction of 1.08 s in the minimum lap time can be achieved. In another single-parameter study, the change in lap time was found to vary almost linearly with fuel load at a
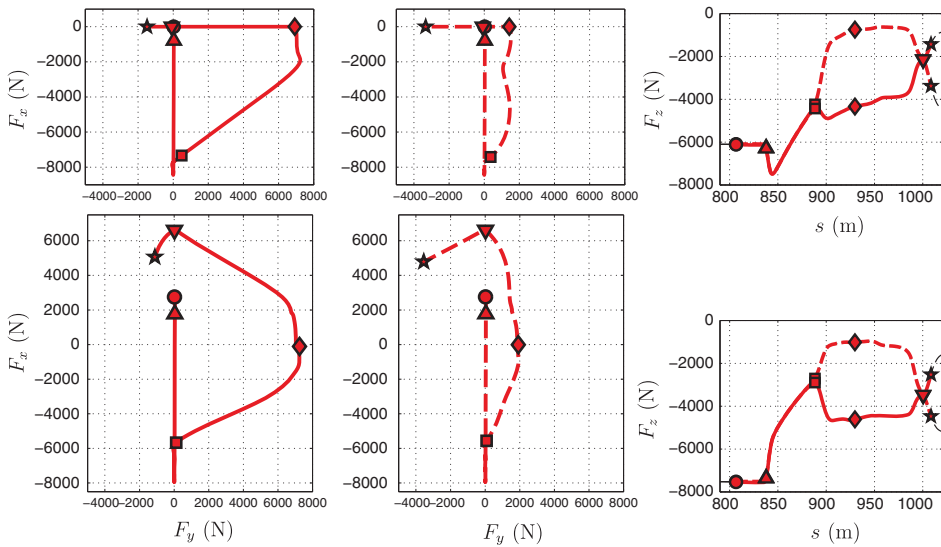
Figure 10. Tyre forces in turn ①. The first column gives the longitudinal and lateral tyre forces for the left-hand wheels, with the left-front wheel at the top. The longitudinal forces are on the vertical axes, while the side forces are on the horizontal axes. The second column gives the longitudinal and lateral tyre forces for the right-hand wheels, with the right-front wheel at the top. The third column gives the normal loads as functions of the distance travelled from the start–finish line, with the front wheels at the top; the left-hand wheels are the solid lines, while the right-hand wheels are the dashed lines.
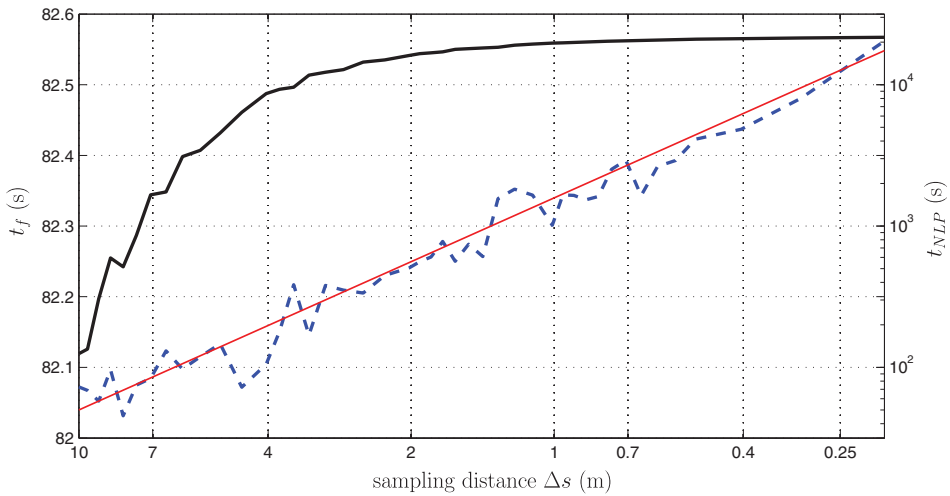

Figure 11. Sensitivity of the minimum lap time to arc length mesh size. The solid black line shows the optimal lap time as a function of the arc length sampling distance (on a semi-log scale). The dashed blue line shows the NLP solution time (on a log–log scale) as a function of the arc length sampling distance. The red line is the approximate computation time given by $t_{NLP} \approx 5 \times 10^{-3}(4655/\Delta s)^{1.5}$.

rate of 0.035 s/kg, which is in close agreement with the commonly employed 'rule of thumb (0.03 s/kg)' that is based on years of track-side measurements. The results of this study are summarised in Table 1.

Figure 13 shows that if the four parameters considered in Figure 12 are optimised in concert, the nominal minimum lap time can be reduced by approximately 1.5 s, which is significantly larger than the improvement achieved in any of the one-parameter-at-a-time
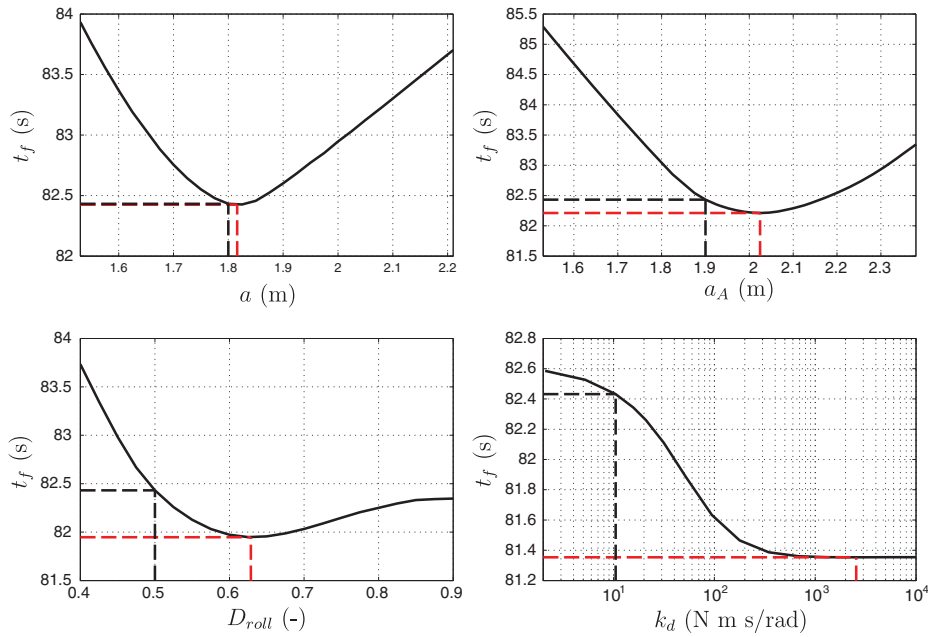

Figure 12. Minimum lap time as a function of the centre of mass position a, the position of the centre of pressure a_A, the roll balance D_{roll}, and the differential viscosity k_d. These curves were computed by varying each parameter in isolation. The black dashed lines correspond to the nominal configuration and its corresponding lap time, while the dashed light (red) lines correspond to optimised parameter values.

Table 1. Independently optimised parameter values with corresponding lap times (Figure 12).

Parameter	Optimal value	Lap time
a	1.816 m	82.425 s
a_A	2.025 m	82.212 s
D_{roll}	0.629	81.947 s
k_d	2552.3 N m s/rad[a]	81.353 s
M	—	$\sim$0.035 s/kg

[a]Dierential essentially locked.

studies. In the context of competitive motor sport, a lap-time reduction of this magnitude is substantial. The results of the multiple parameter optimisation study are given in Table 2, which are in broad agreement with Table 1 in trend-predicting sense, although the specific parameter values are different. The solid black curve in Figure 13 shows the time difference at each point on the track between the nominal parameter study and the times achieved with optimised parameters. Although the parameter-optimised car is substantially faster than the nominal vehicle in terms of lap time, there are isolated sections on the track, for instance, turns ①, ④, ⑤, and ⑫, where the nominal car is quicker. The most substantial performance gain appears to result from a higher exit speed from turn ⑯, which is then carried down the long high-speed straight towards turn ①; there is a 0.3 s gain in this section alone. Further gains are achieved between turns ⑩ and ⑮ where a lot of braking, accelerating, and steering is required.

It is common lore that the racing line is relatively insensitive to the dynamic properties of individual cars. While this belief is generally borne out by this study, it is not always true. Figure 14 shows that the racing line for the nominal car in turn ④ can vary by over
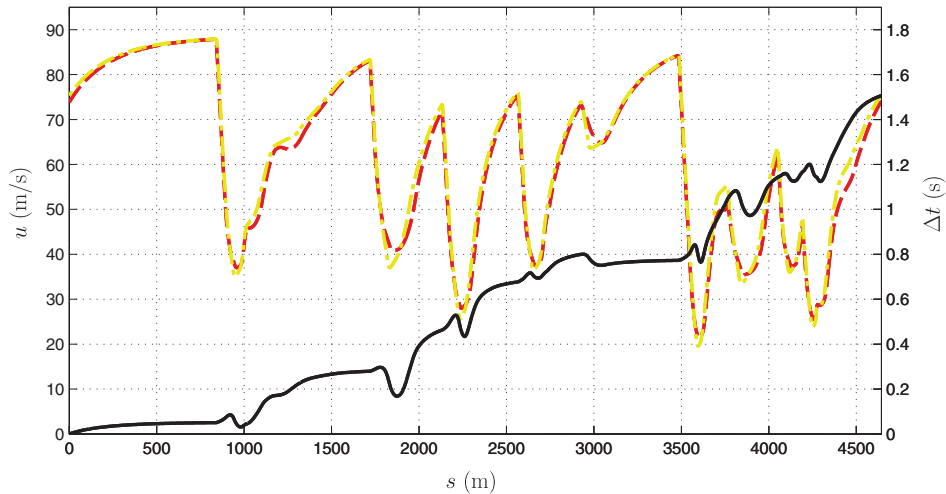
Figure 13. Speed profiles for the nominal (dark red line) and optimised (light yellow line) parameter sets. The solid black line shows the time gain for the optimised vehicle with distance travelled from the start–finish line; the lap time for the vehicle with optimised parameters is 80.94 s. The left-hand *y*-axis gives the speed, while the right-hand *y*-axis is the time gained by the parameter optimised car.

Table 2. Parameter values from a multi-parameter opti-misation study.

Parameter	Optimal value
a	1.9^a m
a_A	2.1789 m
D_{roll}	0.4537
k_d	2218.1 N m s/rad^b

Note: The resulting lap time is 80.94 s (Figure 13).
[a]Restricted by a box constraint on the allowed parameter range.
[b]Differential essentially locked.

half a track width when compared with the vehicle using optimised parameter settings. If one studies Figures 13 and 14 in combination, it is evident that the braking/driving strategy is adjusted to take full advantage of the vehicle's enhanced performance. In the case of the parameter-optimised car, the braking phase in turn ④ starts slightly later and lasts longer than with the nominal car, which results in a significantly lower minimum speed in the corner; the optimised car loses time during the braking phase. In compensation, the optimised vehicle undergoes a more rapid transition from braking to driving that provides a higher exit speed from the corner, which is carried on to the straight to turn ⑤. The higher exit speed and shorter trajectory (Figure 14) gain the optimised vehicle almost 0.2 s between the braking points on turns ④ and ⑤. Similar behaviours appear elsewhere, with even more pronounced gain in time for the slow-speed corners.

4.5. *Neglected and slow dynamics*

The vehicle model used here captures the essential behaviour of a racing car, but it also neglects a number of dynamic features that may be of interest in some studies. These include, for example, the suspension dynamics, limits on fuel usage, vehicle mass variations due to reducing fuel loads, and the effects of tyre wear and tyre temperature changes. The inclusion
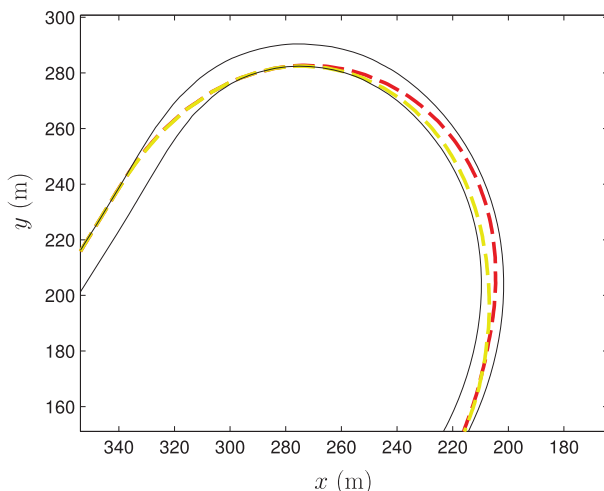
Figure 14.   Change in the racing line in turn ④ due to parameter variations. The centre of mass trajectory for the nominal car is shown in red, while the trajectory for the optimised car is shown in yellow.

of the slow dynamics associated with system states such as the amount of fuel on board, or the temperature of the tyres would require only routine changes to the model and the optimal control problem set-up. There is no apparent need to change the way in which the model is transcribed, although any new dynamics may affect the convergence of the NLP solution, or its sensitivity to initial conditions. The inclusion of faster (and therefore 'stiffer') dynamics, such as those associated with suspension systems, or tyre carcass deformation, may require a change in the transcription method used. At this point in time the difficulties that these changes may produce are largely unknown (at least to us). Another challenge relates to the use of tables that may be associated with such things as engine and fuel consumption maps, the thermal properties of the tyres, and the heave- and pitch-dependent aerodynamic drag and down-force properties. Future studies will investigate the use of alternative transcription methods, such as orthogonal collocation, and of adaptive grid refinement schemes.

5.   Conclusions

We have shown that full-lap vehicular optimal control problems can be solved quickly using direct collocation methods without recourse to expedients such as linearisation, preview, or track segmentation. While it is difficult to make quantitative statements about performance gains of this type, prior-art calculations that took many hours can now be completed in less than 15 min on standard desktop computing facilities using a sampling distance of approximately 2 m for both the control and the dynamics. These gains come from a combination of several refinements including the use of curvilinear coordinates for the track model; a car model that has no 'stiff' dynamics, which require a very fine solution mesh; model non-dimensionalisation and scaling; the approximation of non-smooth modelling features with differentiable functions; and the use of computer-generated analytic derivatives in combination with an open-source sparse large-scale NLP solver. Our focus has been on minimum-lap-time optimal control problems for Formula One cars. We have shown that the simultaneous optimisation of the driver controls (and therefore the racing line) and several set-up parameters can be easily achieved using standard optimal control techniques.

The set-up parameters we have studied here include the location of the vehicle's mass centre, the location of the aerodynamic centre of pressure, the roll balance, and the differential viscosity constant. The value of this work lies in the fact that it represents an optimal control and parameter optimisation basis that can be developed further. The solution of more difficult problems, such as three-dimensional minimum-lap-time problems, and the optimal control of combined kinetic/thermal energy recovery systems are now in prospect.

Funding

Notes

1. The real track boundary constraints apply to the positions of the vehicle's tyres with respect to the track edges. In order to simplify the problem the boundary constraints are enforced on the vehicle's centre of mass, with the track half-width reduced by the vehicle half-width.
2. The left-hand side of this equation is the line integral of the track curvature around the track centre line Ω, in which n is an integer (usually $-1 \leq n \leq 1$).
3. We will use $\tilde{u}$ for the input to control systems and $\tilde{x}$ will be used to denote the state vector in control-system-related contexts.
4. $\mathbb{R}^n$ denote the set of n-dimensional real vectors.
5. Since tyre wear is not modelled, the differential serves only to maximise the drive torque. In practice, a fully locked differential will result in different longitudinal rear-tyre slips in corners that will give rise to unproductive tyre wear and energy dissipation. If tyre wear were recognised in the optimal control problem set-up, a lower value for the differential viscosity constant would inevitably result.

References

[1] Scherenberg H. Mercedes-Benz racing design and cars experience. SAE Transactions, no. 580042; 1958. p. 414–420.
[2] Bryson AE. Optimal control – 1950 to 1985. IEEE Control Syst Mag. 1996;13(3):26–33.
[3] Cossalter V, Lio MD, Lot R, Fabbri L. A general method for the evaluation of vehicle manoeuvrability with special emphasis on motorcycles. Veh Syst Dyn. 1999;31:113–135.
[4] Bertolazzi E, Biral F, Lio MD. Symbolic-numeric indirect method for solving optimal control problems for large multibody systems. Multibody Syst Dyn. 2005;13:233–252.
[5] Tavernini D, Massaro M, Velenis E, Katzourakis DI, Lot R. Minimum time cornering: the effect of road surface and car transmission layout. Veh Syst Dyn. 2013. Available from: http://dx.doi.org/10.1080/00423114.2013.813557
[6] Hendrikx JM, Meijlink T, Kriens RFC. Application of optimal control theory to inverse simulation of car handling. Veh Syst Dyn. 1996;26:449–461.
[7] Pacejka HB. Tyre and vehicle dynamics. 2nd ed. Oxford: Butterworth-Heinemann; 2008.
[8] Kirk DE. Optimal control theory: an introduction. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1970.
[9] Casanova D. On minimum time vehicle manoeuvring: the theoretical optimal lap [PhD thesis]. Bedford: Cranfield University, School of Engineering; 2000.
[10] Betts JT. Practical methods for optimal control and estimation using nonlinear programming. 2nd ed. Philadelphia, PA: SIAM; 2001. Available from: http://dx.doi.org/10.1137/1.9780898718577
[11] Kelly DP. Lap time simulation with transient vehicle and tyre dynamics [PhD thesis]. Bedford: Cranfield University, School of Engineering; 2008.
[12] Åkesson J, Årzén KE, Gäfvert M, Bergdahl T, Tummescheit H. Modeling and optimization with optimica and jmodelica.org-languages and tools for solving large-scale dynamic optimization problems. Comput Chem Eng. 2010;34:1737–1749.
[13] Gustaffsson T. Computing the ideal racing line using optimal control [Master's thesis]. Linköping: Linköping University; 2008.
[14] Brayshaw DL, Harrison MF. 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 D: J Automob Eng. 2005;19: 725–739.
[15] Thommyppillai M, Evangelou SA, Sharp RS. Car driving at the limit by adaptive linear optimal preview control. Veh Syst Dyn. 2009;479:1535–1550.

[16] Timings JP, Cole DJ. Minimum maneuver time calculation using convex optimization. J Dyn Syst Meas Control. 2013;135(3):031015–031015-9.

[17] Falugi P, Kerrigan E, van Wyk E. Imperial college london optimal control software user guide (ICLOCS). London: Imperial College London; May 2010. Available from: http://www.ee.ic.ac.uk/ICLOCS

[18] Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program. 2006;106:25–57. doi:10.1007/s10107-004-0559-y. Available from: http://dx.doi.org/10.1007/s10107-004-0559-y

[19] Horn RA, Johnson CR. Topics in matrix analysis. New York: Cambridge University Press; 1991.

Appendix 1. Tyre friction

The tyre frictional forces are modelled using empirical formulae that are responsive to the tyre's normal loads and 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.[7] Following standard conventions, we use

$$\kappa = -\left(1 + \frac{R\omega_w}{u_w}\right) \tag{A1}$$

and

$$\tan\alpha = -\frac{v_w}{u_w}, \tag{A2}$$

where R is the wheel radius and ω_w the wheel's spin angular velocity. The terms u_w and v_w are the longitudinal and lateral components of the velocity of the wheel centre in body-fixed coordinates.

The following formulae give the peak values and locations of the lateral and longitudinal friction coefficients using linear interpolation[11]:

$$\mu_{x\max} = (F_z - F_{z1})\frac{\mu_{x\max2} - \mu_{x\max1}}{F_{z2} - F_{z1}} + \mu_{x\max1}, \tag{A3}$$

$$\mu_{y\max} = (F_z - F_{z1})\frac{\mu_{y\max2} - \mu_{y\max1}}{F_{z2} - F_{z1}} + \mu_{y\max1}, \tag{A4}$$

$$\kappa_{\max} = (F_z - F_{z1})\frac{\kappa_{\max2} - \kappa_{\max1}}{F_{z2} - F_{z1}} + \kappa_{\max1}, \tag{A5}$$

$$\alpha_{\max} = (F_z - F_{z1})\frac{\alpha_{\max2} - \alpha_{\max1}}{F_{z2} - F_{z1}} + \alpha_{\max1}, \tag{A6}$$

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:

$$\kappa_n = \frac{\kappa}{\kappa_{\max}}, \tag{A7}$$

$$\alpha_n = \frac{\alpha}{\alpha_{\max}}. \tag{A8}$$

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

$$\rho = \sqrt{\alpha_n^2 + \kappa_n^2}. \tag{A9}$$

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

$$\mu_x = \mu_{x\max}\sin\left(Q_x\arctan(S_x\rho)\right), \tag{A10}$$

$$\mu_y = \mu_{y\max}\sin\left(Q_y\arctan(S_y\rho)\right), \tag{A11}$$

with

$$S_x = \frac{\pi}{2\arctan(Q_x)}, \tag{A12}$$

$$S_y = \frac{\pi}{2\arctan(Q_y)}. \tag{A13}$$

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

$$F_x = \mu_x F_z \frac{\kappa_n}{\rho}, \tag{A14}$$

$$F_y = \mu_y F_z \frac{\alpha_n}{\rho}. \tag{A15}$$

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

$$\alpha_{rr} = \arctan\left(\frac{v - \dot{\psi}b}{u - \dot{\psi}w_r}\right),$$

$$\alpha_{rl} = \arctan\left(\frac{v - \dot{\psi}b}{u + \dot{\psi}w_r}\right),$$

$$\alpha_{fr} = \arctan\left(\frac{\sin\delta(\dot{\psi}w_f - u) + \cos\delta(\dot{\psi}a + v)}{\cos\delta(u - \dot{\psi}w_f) + \sin\delta(\dot{\psi}a + v)}\right),$$ (A16)

$$\alpha_{fl} = \arctan\left(\frac{\cos\delta(\dot{\psi}a + v) - \sin\delta(\dot{\psi}w_f + u)}{\cos\delta(\dot{\psi}w_f + u) + \sin\delta(\dot{\psi}a + v)}\right),$$

with the longitudinal slip coefficients given by

$$\kappa_{rr} = -\left(1 + \frac{R\omega_{rr}}{u - \dot{\psi}w_r}\right),$$

$$\kappa_{rl} = -\left(1 + \frac{R\omega_{rl}}{u + \dot{\psi}w_r}\right),$$

$$\kappa_{fr} = -\left(1 + \frac{R\omega_{fr}}{\cos\delta(u - \dot{\psi}w_f) + \sin\delta(\dot{\psi}a + v)}\right),$$ (A17)

$$\kappa_{fl} = -\left(1 + \frac{R\omega_{fl}}{\cos\delta(u + \dot{\psi}w_f) + \sin\delta(\dot{\psi}a + v)}\right).$$

As is evident from Equations (A14) and (A15) the tyre model is responsive to normal load variations and combined slip effects. It also allows one to specify locations and peak values for the coefficients of friction in the longitudinal and lateral directions, which are treated as linear functions of the normal load. Longitudinal and lateral 'shape factors' are used to determine the cornering stiffness and the asymptotic 'roll-off' characteristics of the tyre. Due to a paucity of publicly available Formula One tyre data, we have used the same tyre model on the front and rear wheels.

Appendix 2. Vehicle and tyre data

This appendix contains the nominal values of the tyre, the vehicle, and the aerodynamic parameters used in this study.

Table A1. Tyre friction parameters.

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 A2.   Vehicle parameters.

Symbol	Description	Value
M	Vehicle mass	660 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	10.47 N m s/rad
C_d	Drag coefficient	0.9
C_l	Downforce coefficient	3.0
A	Frontal area	1.5 m^2
ρ	Air density	1.2 kg/m^3
a_A	Centre of pressure to front axle distance	1.9 m
b_A	Centre of pressure to rear axle distance	$w - a_A$