

Optimising the Aero-suspension Interactions in a Formula One car

Mehdi Imani Masouleh and David J. N. Limebeer (*Fellow IEEE*)

Abstract—The interactions between the multi-link suspension system of an open-wheeled race car and the vehicle's aerodynamic performance are modelled and optimized. The closed kinematic loops associated with the suspension mechanisms are modelled as a set of nonlinear equations that are derived using a first principles kinematics analysis. For the purpose of optimal control calculations the solutions to these equations are approximated by multivariate polynomials in the normal tire forces. It is shown that the car's suspension motion can have a significant impact on the aerodynamic performance of the vehicle. Optimal control calculations are used to illustrate the impact of aero-suspension interactions and adjustments on the lap-time performance of the car.

Keywords: Vehicular and wireless technologies, vehicle dynamics, suspension systems, optimal control, lap-time simulation.

I. INTRODUCTION

SUSPENSION systems are sprung mechanisms that connect the chassis of a car to the wheel carriers. This flexible connection uses a parallel spring-damper combination to support the weight of the vehicle, improve road holding and isolate the driver from road disturbances. An introductory discussion of vehicular suspension systems can be found in [1]. A number of independent suspensions are surveyed in [2], where the synthesis of suspension linkages is studied, including their kinematic structure, their dimensions, and their compliance properties. The well-known MacPherson strut suspension is modelled in [3] as a three-dimensional kinematic mechanism; both linear and nonlinear analyses are provided. A MacPherson suspension mechanism model is also developed [4], where a parameter estimation technique based on the singular value decomposition is studied.

Multiple link suspension systems, of the type used in open-wheeled race cars, have been analysed in both industrial and academic papers. The design of a race car suspension is considered in [5]. This paper outlines a procedure for determining optimal suspension parameters, and contains an example that includes a heave spring, anti-roll bar, torsion bar and rocker assembly. The kinematic and dynamic behaviours of a five-link suspension unit is studied in [6], where the movement of the wheel carrier is constrained by five rods terminated at each end with a spherical joint. Rigid-body analyses of five-rod suspensions are extended to mechanisms with flexible joints in [7]. Flexible joints with experimentally determined linearized compliance matrices are considered, leading to an

algorithm for the elasto-kinematic analysis of the mechanism. The optimal synthesis of a five-link independent suspension system is considered in [8]. The synthesis objective is the minimisation of variations in the wheel track, and the toe and camber angles during the jounce and rebound of the wheel. The kinematic design of double-wishbone suspension systems is studied in [9], where multi-objective design optimisation is employed. The design parameters include the lengths of the links and the position of the ground joints. An in-plane 2-DOF kineto-dynamic quarter-car model for a double wishbone suspension was developed in [10] to obtain equivalent stiffness and damping rates for response analysis and suspension synthesis. The kinematics of a MacPherson strut mechanism and a five rod Mercedes 190 series rear wheel suspension are studied in [11] using symbolic computation and interval analysis.

Aerodynamics influences play a major role in the performance of Formula One cars. It is important for high-performance car designers to minimise drag, while simultaneously seeking to maximise down force effects in order to improve road holding and tire performance. Although Formula One regulations limit the use of ground effect aerodynamics, they are still an effective means of creating down force with only a small drag penalty. Since aerodynamic effects are strongly influenced by the orientation and proximity of the car to the road, the vehicle's suspension system plays an important part in determining the aerodynamic performance of the car. A typical Formula One car deploys a double wishbone suspension system with a push rod, or pull rod suspension system, and a rack-and-pinion steering arrangement. One thrust of this work is to model a general purpose double wishbone suspension system so that its effect on aerodynamic performance can be studied and optimised. Figure 1 shows the 24 elements of a Formula One car suspension; the wish bones are coloured black, the push rods are coloured green and the track rods are coloured red. The wheel carrier fixed points are shown as blue hexagrams. The car-body fixed points, the chassis ends of the push rods and the chassis ends of the track rods are marked with red hexagrams.

Following its development, the suspension model will be incorporated into a car model of the type described in [12]–[14]. This model will be used to study a number of optimal control problems including optimising the suspension set up and the car's aerodynamic configuration.

The multi-link suspension system used in this study is analysed in Section II, with a multivariate polynomial approximation of the suspension kinematics given in Section II-A. The car and track models appear in Sections III and III-A respectively. Vehicle-related modelling features that have been used before are dealt with briefly with appropriate references

Mehdi Imani Masouleh (mehdi.imanimasouleh@eng.ox.ac.uk) and David J. N. Limebeer (david.limebeer@eng.ox.ac.uk) are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ.

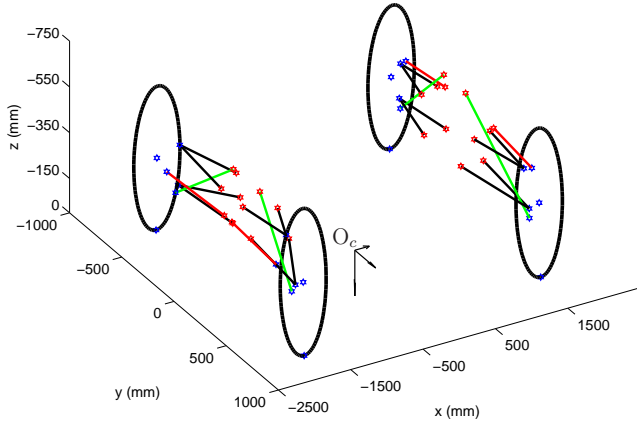


Fig. 1: The suspension system of a typical Formula One car. The wish bones are shown in black, the suspension push rods are shown in green and the track rods are shown in red. The origin of the SAE coordinate system is O_c .

provided. The optimal control processes employed here are summarised in Section IV. The results appear in Section V with the conclusions given in Section VI. The appendix details the least-squares parameter fitting procedure used in the development the suspension meta models.

II. SUSPENSION ANALYSIS

The kinematics of a suspension system are determined by its motion constraints. To fix ideas, we will focus on Figure 2(A) that shows the front-left suspension of a Formula One car. Rods one to four make up the double-wishbone suspension elements. Rod five is the steering track rod, which connects the wheel carrier to the steering rack. Rod six is the suspension push bar, which connects the wheel carrier to a suspension strut rocker mechanism mounted within the car's body; see Figure 3. In the following analysis all the joints in Figure 2 are treated as spherical.

In order to describe the suspension motion constraints in detail, we introduce car-fixed and wheel-carrier-fixed reference frames O_Cxyz and O_Wxyz respectively. As shown in Figure 2(A), points P_1 to P_4 are fixed in car chassis and are connected to Q_1 to Q_4 that are fixed in the wheel carrier using fixed-length wish bones of length l_i $i = 1, \dots, 4$. Points P_5 and Q_5 are at opposite ends of the steering track rod with Q_5 fixed in the wheel carrier. Point P_5 is connected to the steering rack and moves in the car body as the steering wheel position is varied. Points P_6 and Q_6 are at opposite ends of the suspension push rod with Q_6 fixed in the wheel carrier. Point P_6 is connected to a rocker mechanism and moves in the car body in sympathy with movements of the push rod; see Figure 3. The steering track rod and suspension push rod have lengths l_5 and l_6 respectively. This system has two degrees of freedom corresponding to changes in the steering angle and to changes in the angular position of the suspension rocker mechanism. The suspension rockers are also connected to a parallel heave spring and damper combination, a torsion bar and an anti-roll bar; the damper is not considered in the quasi-static analysis presented here. A tire moment M_{TL} is applied

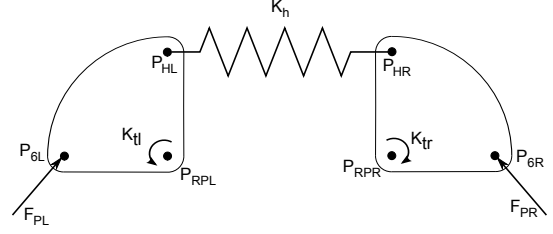


Fig. 3: Push-rod and rocker set-up. The heave spring has stiffness K_h . Right- and left-hand quantities are labelled with subscripts R and L respectively.

to the wheel carrier body, with a tire force F_{TL} applied at the wheel's ground contact point. Six motion constraints derive from the fixed lengths of the six suspension elements. If we suppose that P_i and Q_i are vector representations of the corresponding points in Figure 2(A) (expressed in a common reference frame), then the magnitudes of each of the six differences $P_i - Q_i$ (lengths) must be fixed too. That is

$$\|P_i - Q_i\|_2^2 = l_i^2 \quad i = 1, \dots, 6 \quad (1)$$

in which the lengths l_i are constant. A steering angle δ results in a displacement of the steering rod mount point in the y -axis direction, which is proportional to the radius r of the steering pinion

$$P_5 = [0 \quad -r\delta \quad 0]^T + P_5^0; \quad (2)$$

P_5^0 is the straight-running position of P_5 .

These equations determine a hypersurface over which the twelve points in Figure 2(A) must move. In order to describe points that are fixed in the wheel carrier, in the car's reference frame, we need to recognise the relative motion of the two. There are six degrees of relative motion movement that are constrained by (1).

The translational freedoms of the wheel carrier (relative to the car) are

$$D = [\Delta x \quad \Delta y \quad \Delta z]^T. \quad (3)$$

The rotational freedoms of the wheel carrier relative to the vehicle's chassis are described in terms of the elementary rotations $R_x(\phi)$, $R_y(\theta)$ and $R_z(\psi)$, which are described by (4) in which s_ϕ and c_ϕ are the sine and cosine, respectively, of ϕ . The same notation is used for θ and ψ . As a result

$$x_{ci} = R_z(\psi)R_x(\phi)R_y(\theta)x_{wi} + D \quad (5)$$

describes a point x_{wi} that is fixed in the wheel carrier in the chassis frame. When working with (1) all points are referred to the chassis frame. Equations (1) and (5) provide six motion constraints that determine the kinematic behaviour of the suspension system.

As shown in Figure 3, the left- and right-hand sides of the suspension system are coupled through a heave spring (and a parallel damper which is not shown) and a rocker assembly. The right- and left-hand rockers are mounted on torsion bars with stiffnesses K_{tr} and K_{tl} respectively. The two rockers are also coupled by an anti-roll bar that resists chassis roll. This is represented by a moment $(\alpha_R + \alpha_L)K_{arb}$ applied to each rocker, in which α_R and α_L are the right- and left-hand

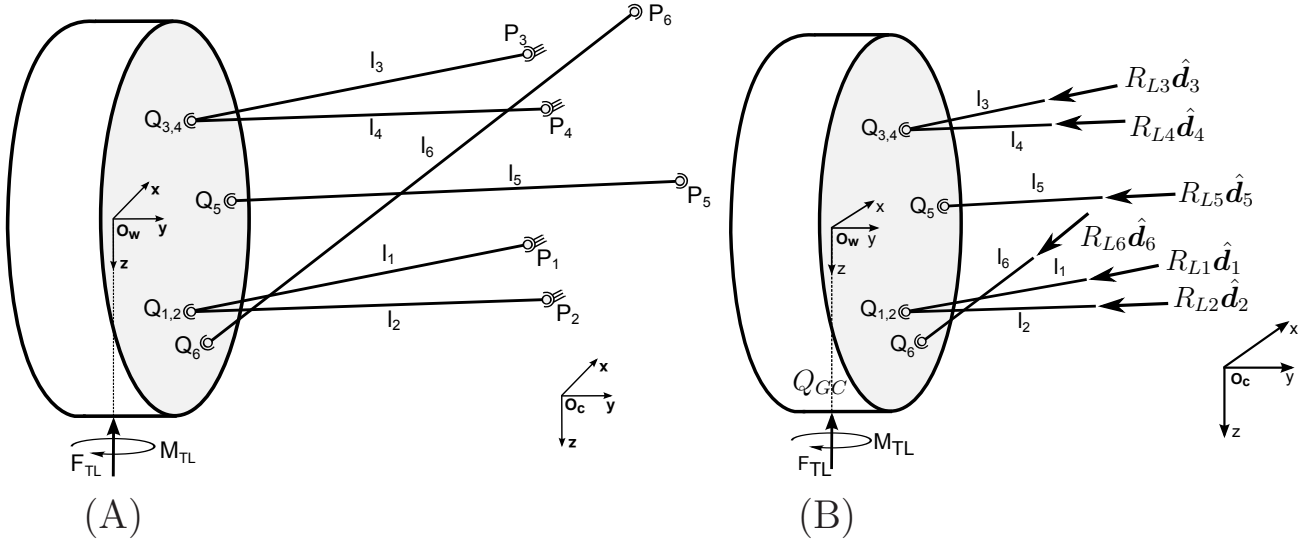


Fig. 2: Front-left suspension configuration of a typical Formula One car with a push rod suspension system. The axis system O_Cxyz is fixed in the car body, while O_Wxyz is fixed in the wheel carrier with its origin at the wheel's geometric centre. Figure (A) shows points P_i in the car chassis (with points $P_1 \dots, P_4$ fixed in the car chassis), the points Q_i are fixed in the wheel carrier, and the wishbones $l_1 \dots, l_4$, the push rod l_5 and the track rod l_6 . Figure (B) is a free body diagram of the front-left wheel carrier. The reaction force magnitudes are given by R_{Li} with their directions given by the vectors \hat{d}_i for $i = 1, \dots, 6$.

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\phi & -s_\phi \\ 0 & s_\phi & c_\phi \end{bmatrix} \quad R_y(\theta) = \begin{bmatrix} c_\theta & 0 & s_\theta \\ 0 & 1 & 0 \\ -s_\theta & 0 & c_\theta \end{bmatrix} \quad R_z(\psi) = \begin{bmatrix} c_\psi & -s_\psi & 0 \\ s_\psi & c_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

rocker angles. The anti-roll bar stiffness is denoted by K_{arb} . In Figure 3 F_{PL} and F_{PR} are the magnitudes of the right- and left-hand push rod forces, P_{6L} and P_{6R} are the push rod attachment points, P_{HL} and P_{HR} are the end points of the heave spring, R_{RPL} and R_{RPR} are the rocker pivot points and K_h is the heave spring stiffness.

In order to compute a moment balance for the rockers, we introduce the rotation matrix $R = R(\mathbf{n}, \alpha)$ that describes a rotation through α around the unit vector \mathbf{n} :

$$R(\mathbf{n}, \alpha) = \cos(\alpha)I + (1 - \cos(\alpha))\mathbf{n}\mathbf{n}^T + \sin(\alpha)S(\mathbf{n}) \quad (6)$$

in which the skew-symmetric matrix $S(\mathbf{n})$ is defined by

$$S(\mathbf{n}) = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix}. \quad (7)$$

Taking moments around the pin joint P_{RPL} gives

$$0 = (F_{PL}(R(\mathbf{n}_{RL}, \alpha_L)(P_{6L} - P_{RPL})) \times \hat{d}_6) \cdot \mathbf{n}_{RL} + (F_{heave}(R(\mathbf{n}_{RL}, \alpha_L)(P_{HL} - P_{RPL})) \times \hat{d}_{hs}) \cdot \mathbf{n}_{RL} - K_{tl}\alpha_L - K_{arb}(\alpha_L + \alpha_R) \quad (8)$$

in which \mathbf{n}_{RL} is the left rocker pivot rotation axis, \hat{d}_6 is a unit vector in the direction of l_6 , \hat{d}_{hs} is a unit vector in the heave spring direction pointing from P_{HL} to P_{HR} , F_{heave} is the heave spring force magnitude, \cdot represents the dot product and \times represents the cross product.

In the same way one can take moments around P_{RPR} to obtain

$$0 = (F_{PR}(R(\mathbf{n}_{RR}, \alpha_R)(P_{6R} - P_{RPR})) \times \hat{d}_6) \cdot \mathbf{n}_{RR} - (F_{heave}(R(\mathbf{n}_{RR}, \alpha_R)(P_{HR} - P_{RPR})) \times \hat{d}_{hs}) \cdot \mathbf{n}_{RR} - K_{tr}\alpha_R - K_{arb}(\alpha_L + \alpha_R) \quad (9)$$

in which the various term in (9) must be interpreted in the context of the right-hand suspension assembly. The heave spring force F_{heave} is a function of the change in its length, and is given by

$$F_{heave} = k_h(\Delta L)\Delta L \quad (10)$$

in which $\Delta L = |P_{HR} - P_{HL}| - L_0$ with L_0 the unloaded length. In general, the heave spring stiffness is non-linear and tends to 'harden' as it is compressed.

The equations of motion describing the suspension system will now be determined using force and moment balances with the aid of the free body diagram shown in Figure 2 (B). We will suppose that the reaction force acting at point Q_i has magnitude R_{Li} and acts in direction \hat{d}_i . It should be noted that although the reaction forces are vector-valued quantities, it is only the magnitudes of these vectors that are unknown, since their directions are determined by the kinematic constraints. The external forces acting on the wheel carrier are the tire force F_{TL} , and the gravitational force due to the wheel

carrier's mass m_L , which is given by

$$F_{Lg} = [0 \quad 0 \quad -m_L g]^T. \quad (11)$$

The tire force F_{TL} and the gravitational force F_{Lg} are expressed in the chassis reference frame, which is assumed for present purposes to be aligned with the inertial frame. Summing forces on the left-hand wheel carrier gives

$$0 = F_{TL} + F_{Lg} + R_{L1}\hat{\mathbf{d}}_1 + R_{L2}\hat{\mathbf{d}}_2 + R_{L3}\hat{\mathbf{d}}_3 + R_{L4}\hat{\mathbf{d}}_4 + R_{L5}\hat{\mathbf{d}}_5 + F_{PL}\hat{\mathbf{d}}_6 \quad (12)$$

in which $R_{L6} = F_{PL}$ is the force acting on the left-hand rocker. This equation provides three scalar equations in the six left-hand unknown reaction force magnitudes. A similar force balance on the right-hand front wheel carrier gives

$$0 = F_{TR} + F_{Rg} + R_{R1}\hat{\mathbf{d}}_1 + R_{R2}\hat{\mathbf{d}}_2 + R_{R3}\hat{\mathbf{d}}_3 + R_{R4}\hat{\mathbf{d}}_4 + R_{R5}\hat{\mathbf{d}}_5 + F_{PR}\hat{\mathbf{d}}_6, \quad (13)$$

where $R_{R6} = F_{PR}$ is the force acting on the right-hand suspension rocker. In this case

$$F_{Rg} = [0 \quad 0 \quad -m_R g]^T \quad (14)$$

with the $\hat{\mathbf{d}}_i$'s interpreted in a like manner in the context of the right-hand front wheel. This equation provides another three equations in the six right-hand reaction force magnitudes.

As we will now show the remaining equations come from moment balance calculations. To begin, we define a set of vectors that point from point Q_1 , which is fixed in the wheel carrier, to points Q_i $i = 2, \dots, 6$, O_W and the tire ground contact point Q_{GC} . These vectors are given by \mathbf{d}_{12} , \mathbf{d}_{13} , \mathbf{d}_{14} , \mathbf{d}_{15} , \mathbf{d}_{16} , \mathbf{d}_{1W} and \mathbf{d}_{1GC} respectively. With the exception of \mathbf{d}_{1GC} , these vectors are known and fixed in the wheel carrier.

In order to find the vector \mathbf{d}_{1GC} , which represents a moving point in the wheel frame, we need to study the geometry of the ground contact point. To that end we introduce the vector $\hat{\mathbf{e}}_w = [0 \ 1 \ 0]^T$, which is fixed in the wheel carrier, and points in the direction of the wheel spindle. The wheel spindle unit vector, expressed in the car frame, is thus given by

$$\hat{\mathbf{e}}_{cw} = R_z(\psi)R_x(\phi)R_y(\theta)\hat{\mathbf{e}}_w. \quad (15)$$

If the pitch and roll of the car relative to the road is 'small', the unit road-normal expressed in the chassis frame can be approximated by $\hat{\mathbf{e}}_{rn} = [0 \ 0 \ 1]^T$. This means that the vector pointing from the wheel centre to the ground contact point is given by

$$\mathbf{r}_w = -\frac{\hat{\mathbf{e}}_{cw} \times (\hat{\mathbf{e}}_{cw} \times \hat{\mathbf{e}}_{rn})}{\|\hat{\mathbf{e}}_{cw} \times (\hat{\mathbf{e}}_{cw} \times \hat{\mathbf{e}}_{rn})\|_2} R_{fw}(F_{TL}) \quad (16)$$

and so $\mathbf{d}_{1GC} = \mathbf{r}_w + \mathbf{d}_{1W}$. The loaded tire radius characteristic used here is given by

$$R_{fw} = 326.0 + 4.00 \times 10^{-3} F_{TL} + 4.075 \times 10^{-8} F_{TL}^2, \quad (17)$$

in which F_{TL} is the normal tire load expressed in Newtons; this empirical formula is based on tire-squash measurements.

Taking moments around Q_1 in Figure 2 (B) gives

$$0 = M_{TL} + \mathbf{d}_{1GC} \times F_{TL} + \mathbf{d}_{1W} \times F_{Lg} + R_{L2}(\hat{\mathbf{d}}_2 \times \mathbf{d}_{12}) + R_{L3}(\hat{\mathbf{d}}_3 \times \mathbf{d}_{13}) + R_{L4}(\hat{\mathbf{d}}_4 \times \mathbf{d}_{14}) + R_{L5}(\hat{\mathbf{d}}_5 \times \mathbf{d}_{15}) + R_{L6}(\hat{\mathbf{d}}_6 \times \mathbf{d}_{16}). \quad (18)$$

A similar calculation for the front right wheel gives

$$0 = M_{TR} + \mathbf{d}_{1GC} \times F_{TR} + \mathbf{d}_{1W} \times F_{Rg} + R_{R2}(\hat{\mathbf{d}}_2 \times \mathbf{d}_{12}) + R_{R3}(\hat{\mathbf{d}}_3 \times \mathbf{d}_{13}) + R_{R4}(\hat{\mathbf{d}}_4 \times \mathbf{d}_{14}) + R_{R5}(\hat{\mathbf{d}}_5 \times \mathbf{d}_{15}) + R_{R6}(\hat{\mathbf{d}}_6 \times \mathbf{d}_{16}) \quad (19)$$

in which the vectors $\hat{\mathbf{d}}_2, \dots, \hat{\mathbf{d}}_6$, \mathbf{d}_{1GC} and $\mathbf{d}_{12}, \dots, \mathbf{d}_{16}$ must be interpreted in an analogous manner in the context of the right-hand suspension.

In sum, equations (1) are used to determine the right- and left-hand Euler angles and the right- and left-hand wheel carrier displacements in (3); twelve variables in total. Equations (8) and (9) determine the two rocker angles. Equations (12), (13), (18) and (19) determine the twelve right- and left-hand push rod force magnitudes. Equations (1), (8), (9), (12), (13), (18) and (19) can be solved for the twenty six unknowns associated with each axle using the nonlinear equation solver *fsolve* in MATLABTM. Figure 4 illustrates the force-displacement behaviour of the front-left suspension system as a function of the normal tire load; the front-right hand tire is assumed un-loaded. It is evident from Figure 4 that the wheel's camber and toe angles vary very little as a result of normal load variations. This figure also shows that the suspension movement resulting from normal load variations is almost entirely in the vertical direction. The conclusions relating to Figure 4 are not necessarily general ones, and relate only to the particular race car suspension geometry being studied here.

A. Aerodynamic and suspension metamodels

Given its complexity, the suspension model described by Equations (1), (8), (9), (12), (13), (18) and (19), and the tire squash, are replaced by a metamodel, which acts as a "surrogate" for these equations. The intention is to preserve the accuracy of the physical kinematic model, while seeking simultaneously to significantly increase its solution speed in the context of optimal control calculations. The interested reader is referred to [15] for a review of *metamodelling* in the context of general engineering design optimisation. Models based on approximating polynomials, splines, radial basis functions, neural networks and Kriging are all discussed.

Since the suspension deflection is dominated by the normal tire load, the tire moments M_{TL} and M_{TR} , the x- and y-axis components of F_{TL} and F_{TR} , and the steering angle δ are all set to zero in the aerodynamic force studies presented in this paper. In these studies we make use of a SAE coordinate system that is fixed in the road.

Figure 5 shows the rear axle of the car with its associated roll angle and ride height. In its nominal configuration¹ the front ride height is $h_{f0} = -14$ mm, while the rear ride height is $h_{r0} = -74.1$ mm. If the tire normal loads are swept

¹The car is stationary on the garage floor.

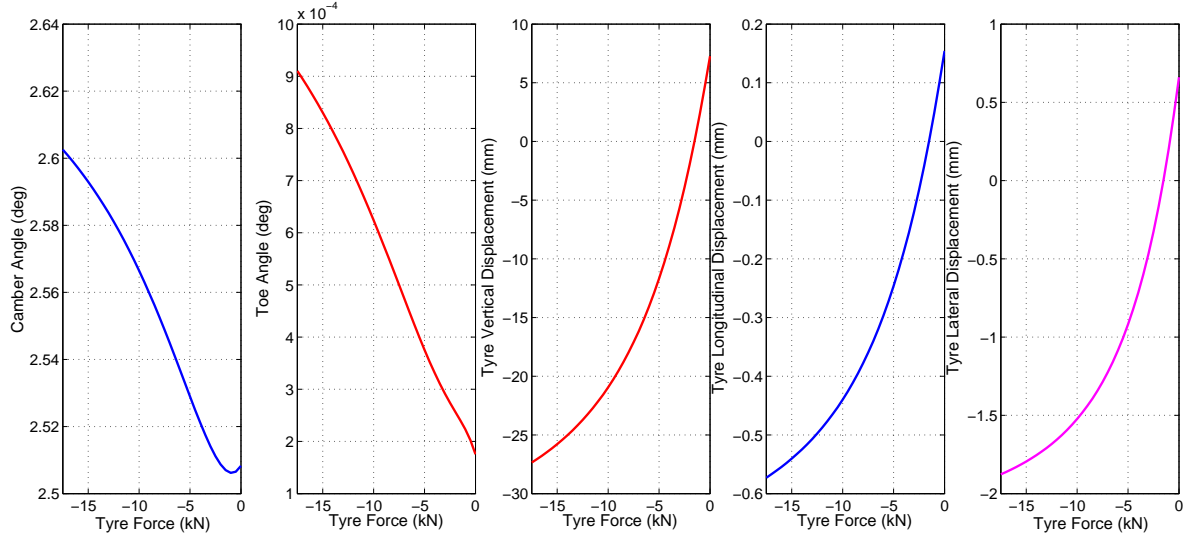


Fig. 4: Variations in the front-left wheel carrier camber angle, toe angle, and the ground contact point as a function of tyre vertical load. These results are expressed in an SAE coordinate system fixed in the car body - a right-hand rule is used for angles.

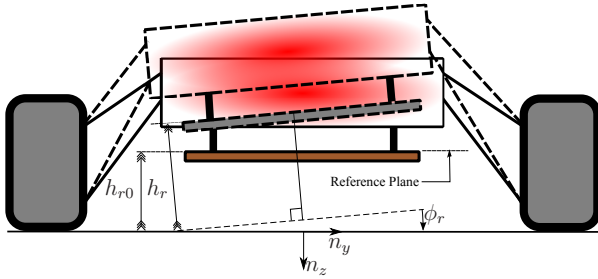


Fig. 5: Suspension and tire squash induced roll angle ϕ_r and ride height h_r at the rear axle; the car is driving into the page and into a right-hand bend. The rear-axle ride height in the nominal configuration is h_{r0} and is measured relative to the reference plane (fixed in the car); see Drawing 7 in [16]. The front axle is treated in the same way with ϕ_r , h_r and h_{r0} replaced with ϕ_f , h_f and h_{f0} . The ride height is expressed in the inertial coordinate system given by n_x and n_y .

from zero to -8 kN at the rear axle, the resulting ride height and roll angle variations are shown in Figure 6. A similar calculation was carried out for the front axle with the normal tire loads swept from zero to -6 kN. Broadly similar results were obtained, but are not shown, in the interests of reducing length.

Figure 6 shows that the rear ride height varies between -88 mm and -27 mm, with the rear axle roll angle ranging over $\pm 5^\circ$. The front ride height varies between -24 mm and 7 mm, while the front axle roll angle varies over the $\pm 2.4^\circ$ range. Under cornering conditions the car rolls out of the bend and so right-hand corners result in negative roll angles; this is the situation illustrated in Figure 5. As the tire loads increase, the car moves towards the road (in the positive n_z -axis direction;

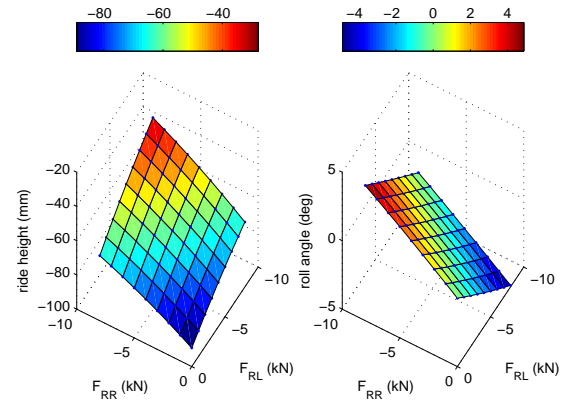


Fig. 6: Rear axle ride height and roll angle variations as a function of right- and left-hand normal tire load changes. The left-hand plot shows the rear axle ride height, while the right-hand plot shows the rear axle roll angle.

see Figure 5).²

In optimal lap time simulations it will be necessary to perform suspension deflection calculations many times over and so it is necessary to perform these calculations fast. To that end we will fit multivariate polynomials to pre-computed data of the type shown in Figure 6 using the least squares algorithm describe in the Appendix. The smooth (and almost linear) nature of the ride height and roll angle data shown in Figure 6 suggests that this data can be accurately represented by relatively low-order multivariate polynomials. Table I gives

²The fact that the reference frame can pitch relative to the road means that ride heights can take on positive and negative values within the suspension's operating range.

-	1	F_l	F_r	F^2	$F_l F_r$	F_r^2
h_f	-21.52	-2.513	-2.513	-0.08324	0.1126	-0.08324
ϕ_f	0	0.00851	-0.00851	0.000256	0	-0.0002562
h_r	-89.04	-4.1305	-4.1305	-0.09450	0.1119	-0.09450
ϕ_r	0	0.0112	-0.0112	6.1×10^{-05}	0	-6.1×10^{-05}

TABLE I: Multivariate polynomial coefficients of the front- and rear-axle ride height (in mm) and roll angle (in rad) in *Graded Reverse Lexicographic Order*. The left- and right-wheel normal loads F_l and F_r , respectively, are given in (kN).

the coefficients associated with the multivariate polynomial descriptions of the rear- and front-axle ride height and roll angles. Despite the relatively low order of these polynomial descriptions, the largest rear axle ride height error across all mesh points is only 0.03 mm with a corresponding maximum roll angle error of 0.027° . The largest front axle ride height error across all mesh points is 0.147 mm with a corresponding maximum roll angle error of 0.038° .

The external forces acting on the car come from the tires and from aerodynamic influences. The dynamic car model used in the optimal control calculations compute the normal tire loads $F_{zrr} \cdots F_{zfl}$, and the steering and side slip angles δ and β , respectively, for each track location on an optimal lap. As shown in Figure 7 these data are supplied to the suspension and aero models to produce the aerodynamic coefficients, which are then updated continuously. The front and rear axle down force coefficients are given by C_L^f and C_L^r respectively, while the drag coefficient is C_D . The resulting aerodynamic down

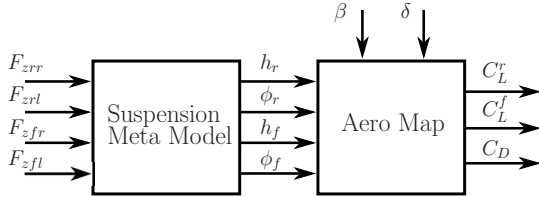


Fig. 7: Quasi-static aero-suspension model. The dynamic model generates the normal tire loads, and the steering and side-slip angles, while the quasi-static meta model generated the aero dynamic drag and down force coefficients.

forces are computed using

$$F_{az}^f = \frac{1}{2} C_L^f \rho A u^2, \quad (20)$$

$$F_{az}^r = \frac{1}{2} C_L^r \rho A u^2, \quad (21)$$

which are applied at the centres of the front and rear axles, while the drag is given by

$$F_{ax} = -\frac{1}{2} C_D \rho A u^2, \quad (22)$$

which is applied at the car's mass centre. The drag has a negative sign, since it acts in the negative x-axis direction. The drag and down-force coefficients used in this study derive from track and wind tunnel measurements and are given in Table II. These coefficients are responsive to the front and rear ride heights, the side-slip angle (β); see Figure 10, and the steer (δ) and roll (ϕ) angles.

C_L^f	C_L^r	C_D	Input
1.2000	0.9494	0.9455	Nominal Value
10.0000	1.0000	0.1000	h_f (m)
-5.0000	-22.2222	-0.1000	h_r (m)
0	-222.2222	0	h_r^2 (m ²)
-5.1294	-10.2588	0	β^2 (rad ²)
-0.1641	-0.1641	0	δ^2 (rad ²)
-8.2070	-32.8281	0	ϕ^2 (rad ²)

TABLE II: Down force and drag coefficients and sensitivities.

III. CAR AND TRACK MODEL

The car model used in this paper is an extension of one that has been used elsewhere [12]–[14], and so will be described only briefly. A similar model described in [17], although the hyper-static structure associated with the under-determined normal force system at the wheels is dealt with in a different way. Relative to prior art models, the key upgrades relate to the introduction of a suspension system, and the introduction of an aerodynamic model that is responsive to the car's disposition relative to the road and its direction of travel.

In order to study suspension-related influences, one might consider augmenting existing vehicle models with dynamic suspension models, and main-body heave, roll and pitch freedoms, which would inform the aerodynamic maps given in Table II with the appropriate ride height and body-angle information. A model of this type would be complex, since it would have to include three new main-body freedoms and the multi-link suspension system described in Section II. Another complication, from an optimal control perspective, related to the fact that the suspension dynamics would contain 'fast' dynamics requiring a fine integration mesh. Rather than pursuing this line of investigation, we will use the quasi-static meta model given in Table I to compute the front- and rear-axle ride heights, and the front and rear vehicular roll angles. These suspension-related variables are then fed directly into the aerodynamic map — see Figure 7

If one wishes to model the car body as 'rigid', the front- and rear-axle roll angles can be constrained in the optimal control problem to be equal. The remaining modelling elements are now described briefly, with further details widely available in the literature.

As has been explained elsewhere [12], [13], [18], the track and car kinematics are modelled using ideas from classical differential geometry. 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 tire description given in [19] in combination with the upgraded aerodynamic model given in Table II. The important geometric modelling quantities are shown in plan view in Figure 8.

A. Track Model

We will model the track using a curvilinear coordinate system that follows the vehicle using the track *spine* (centre line) position as the curvilinear abscissa [20]. Referring to Figure 9, we describe the location of the mass centre of the

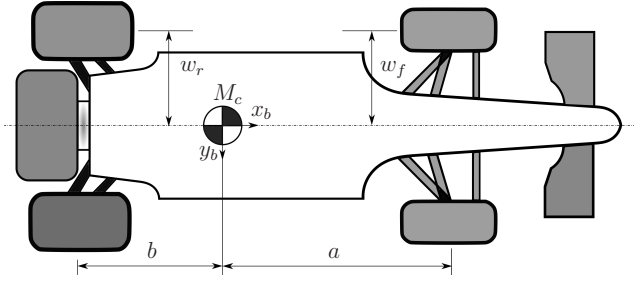


Fig. 8: Plan view of a Formula One car with its basic geometric parameters. The car's mass centre is at M_c , with the body-fixed axes x_b and y_b in the ground plane beneath M_c .

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 spine 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' are alternate independent variables. 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} . The track spine tangent vector \mathbf{t} will be described in terms of the track orientation angle θ , with the track's (possibly unequal) half-widths given by \mathcal{N}_r and \mathcal{N}_l . 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} . The car and track kinematics have been extended to three dimensions in [14] and [18].

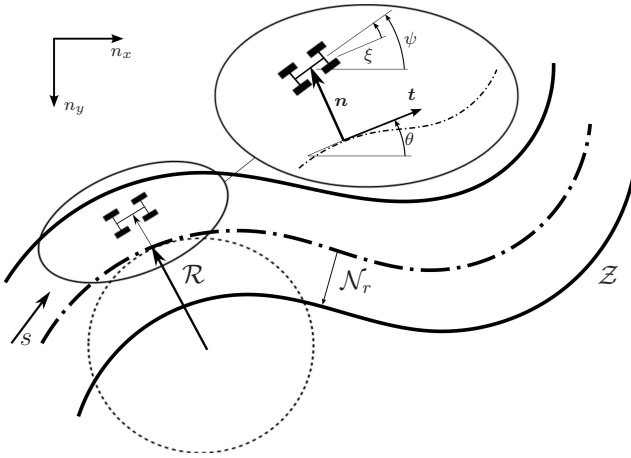


Fig. 9: Curvilinear-coordinate-based description of a track segment Z . The independent variable s represents the elapsed centre-line distance travelled, with $\mathcal{R}(s)$ the radius of curvature and $\mathcal{N}_r(s)$ is the the track's right-hand half-width; n_x and n_y represent an inertial reference frame.

It follows by routine calculation [12] that

$$\dot{s} = \frac{u \cos \xi - v \sin \xi}{1 - n\mathcal{C}}, \quad (23)$$

in which u and v are the longitudinal and lateral components of the car's velocity. This equation shows how the car 'drags' s along the track's spine. The rate of change of n is given by

$$\dot{n} = u \sin \xi + v \cos \xi. \quad (24)$$

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

$$\dot{\xi} = \dot{\psi} - \mathcal{C}\dot{s}. \quad (25)$$

1) *Change of Independent Variable:* The 'distance travelled', rather than time, 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 requisite number of state variables. Suppose

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

where S_f comes from (23) as follows

$$S_f = \left(\frac{ds}{dt} \right)^{-1} = \frac{1 - n\mathcal{C}}{u \cos \xi - v \sin \xi}. \quad (26)$$

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

$$\frac{dn}{ds} = S_f (u \sin \xi + v \cos \xi) \quad (27)$$

from (24), and

$$\frac{d\xi}{ds} = S_f \omega - \mathcal{C} \quad (28)$$

from (25); $\omega = \dot{\psi}$ is the vehicle yaw rate.

B. Car Model

Each tire produces longitudinal and lateral forces that are responsive to the tires' slip [21]. These forces together with the steer and yaw angle definitions are illustrated in Figure 10.

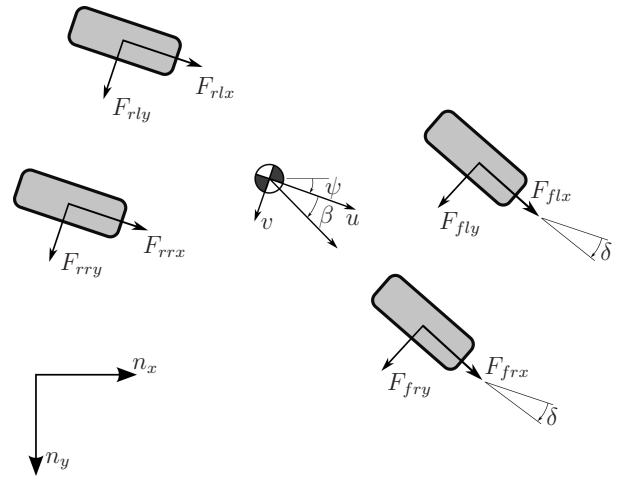


Fig. 10: Tyre force system. The car's yaw angle is given by ψ relative to the inertial reference frame n_x and n_y . The body side-slip angle is given by $\beta = \tan^{-1} \left(\frac{v}{u} \right)$, where u and v are the longitudinal and lateral components, respectively, of the car's mass centre velocity.

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

$$\begin{aligned} M \frac{d}{dt} u(t) &= M\omega v + F_x \\ M \frac{d}{dt} v(t) &= -M\omega u + F_y \\ I_z \frac{d}{dt} \omega(t) &= a(\cos \delta(F_{fry} + F_{fly}) + \sin \delta(F_{frx} + F_{flx})) \\ &\quad + w_f(\sin \delta F_{fry} - \cos \delta F_{frx}) - w_r F_{rrx} + \\ &\quad w_f(\cos \delta F_{flx} - \sin \delta F_{fly}) + w_r F_{rlx} - \\ &\quad b(F_{rry} + F_{rly}), \end{aligned} \quad (29)$$

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

$$F_x = \cos \delta(F_{frx} + F_{flx}) - \sin \delta(F_{fry} + F_{fly}) + (F_{rrx} + F_{rlx}) + F_{ax} \quad (30)$$

$$F_y = \cos \delta(F_{fry} + F_{fly}) + \sin \delta(F_{frx} + F_{flx}) + (F_{rry} + F_{rly}) \quad (31)$$

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

$$\frac{du}{ds} = S_f(s)\dot{u} \quad (32)$$

$$\frac{dv}{ds} = S_f(s)\dot{v} \quad (33)$$

$$\frac{d\omega}{ds} = S_f(s)\dot{\omega}. \quad (34)$$

C. 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 10. In each case these forces are a function of the normal load and the tyre's longitudinal slip coefficient κ and a slip angle α ; further details are available in [12].

1) *Load Transfer*: In our optimal control codes the tyre normal forces are treated as time-varying non-positive inputs normal to the ground plane that must be constrained by the laws of mechanics. To this end 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 8. Balancing the vertical forces gives

$$0 = F_{rrz} + F_{rlz} + F_{frz} + F_{flz} + Mg + F_{az}^f + F_{az}^r, \quad (35)$$

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}^f and F_{az}^r are the front and rear, respectively, aerodynamic down forces acting on the car. Balancing moments around the car's body-fixed x_b -axis gives

$$0 = w_r(F_{rlz} - F_{rrz}) + w_f(F_{flz} - F_{frz}) + hF_y, \quad (36)$$

in which F_y is the lateral inertial force acting on the car's mass centre; see (31). Balancing moments around the car's body-fixed y_b -axis gives

$$0 = b(F_{rrz} + F_{rlz} + F_{az}^r) - a(F_{frz} + F_{flz} + F_{az}^f) + hF_x, \quad (37)$$

where F_x is the longitudinal inertial force acting on the car's mass centre (see (30)).

Equations (35), (36) and (37) are a set of linear equations in four unknowns. A unique solution for the tyre loads is obtained by including the suspension-related roll constraint

$$\phi_f = \phi_r, \quad (38)$$

which ensures that the front- and rear-axle roll angles are the same; see Figure 5. Equation (38) enforces a rigid vehicular main body, which is not the same as using a suspension roll stiffness parameter [12]–[14].

D. Wheel Torque Distribution

In order to optimise the vehicle's performance, one needs to control the torques applied to the 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, which is stipulated by the Formula One technical regulations [16]. The drive torques applied to the rear wheels are controlled by a differential mechanism.

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

$$0 = \max(\omega_{fr}, 0) \max(\omega_{fl}, 0)(F_{frx} - F_{flx}), \quad (39)$$

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 (39) becomes inactive.

2) *Differential*: The drive torque is delivered to the rear wheels through a limited-slip differential, which is modelled by

$$R(F_{lrx} - F_{rrx}) = -k_d(\omega_{lr} - \omega_{rr}), \quad (40)$$

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.

IV. OPTIMAL CONTROL

General purpose numerical optimal control problem solvers are now widely available. When solving new problems for the first time, it is prudent to check the solutions from multiple initial conditions, and if possible against measured data. One should also verify that parameter changes produce the trends expected, and if possible, consistency against alternative optimisation algorithms should be checked. Other important

issues include: (A) the selection of a suitable discretisation technique; (B) scaling the problem; (C) the approximation of non-smooth features; (D) regularization when the problem when the solution contains singular arcs; (E) the efficient computation of gradients and (F) mesh refinement strategies. The reader may wish to consult [12]–[14], [18] for further information on the practical implementation of numerical optimal control solvers in the motor sport context.

We will use a direct numerical method based on an orthogonal collocation technique to solve the optimal control problems studied here.

A. Continuous-Time Problem

A wide class of optimal control problems can be posed in the *Mayer-Pontryagin* form. The more familiar *Bolza* formulation can be converted into this form if so desired [22].

Suppose the system to be controlled is described by the state-space model:

$$\frac{dx}{d\tau} = f(x(\tau), u(\tau), \tau) \quad x(-1) = x_0, \quad (41)$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, with the initial value of the state-vector x given by x_0 ; the controls are given by u . The general optimisation interval $[t_0, t_f]$ can be transformed into the interval $[-1, 1]$, which is particularly well suited to pseudospectral methods, with the affine transformation $t = (t_f - t_0)\tau/2 + (t_f + t_0)/2$. The control vector u may be subject to control constraints with $u \in \mathcal{U}$, where \mathcal{U} is some region in the control space. The problem may also have path constraints, but these are omitted for clarity of exposition. The optimization problem is to find a trajectory (i.e., a solution of (41) consistent with the boundary and control constraints) and that minimizes

$$\Phi(x(1)) \quad (42)$$

for some scalar function $\Phi(\cdot)$ of the terminal state.

In accordance with any of the standard texts such as [23]–[25], first-order necessary conditions for optimality are given by the Pontryagin minimum principle (PMP)

$$u^* = \arg \min_u \mathcal{H}(x, \lambda, u, \tau), \quad (43)$$

in which \mathcal{H} is the control Hamiltonian

$$\mathcal{H}(x, \lambda, u, \tau) = \langle \lambda, f \rangle, \quad (44)$$

where $\langle \cdot, \cdot \rangle$ is the inner product. The control u^* is an extremal control with λ an n -dimensional costate vector that is a continuous function of time. In this notation (41) and λ are given by

$$\dot{x} = \nabla_\lambda \mathcal{H} \quad x(-1) = x_0 \quad (45)$$

$$\dot{\lambda} = -\nabla_x \mathcal{H} \quad \lambda(1) = \nabla \Phi(x(1)). \quad (46)$$

Equations (45) and (46), in combination, form a two-point boundary value problem in which the initial state and the terminal costate are specified. In the case of arcs for which u is in the interior of \mathcal{U} , assuming the existence and continuity of the relevant partial derivatives, (43) implies that

$$\nabla_u \mathcal{H} = 0 \quad (47)$$

with the weak Legendre-Clebsch condition

$$\nabla_{uu} \mathcal{H} \geq 0 \quad (48)$$

also satisfied.

B. Discretisation

The optimal control problems studied here will be solved using GPOPS-II [26], which is a direct pseudospectral method based on Legendre-Gauss-Radau (LGR) collocation and Radau's integration formula; see page 103 in [27]. In this framework the state and controls are treated as parameters on a finite collocation, or implicit integration mesh that is determined by N Radau points (τ_0, \dots, τ_N) . These points are the roots of $P_{N+1}(-\tau) + P_N(-\tau)$, in which $P_N(\tau)$ is the N th-degree Legendre polynomial; these roots lie in $[-1, 1]$ with one root at $\tau_0 = -1$ always. The Radau integration formula is exact for polynomials of degree $2N - 2$ and lower. In this discretisation process the state is approximated in a basis of Lagrange polynomials

$$\mathcal{L}_i(\tau) = \prod_{\substack{j=0 \\ j \neq i}}^N \frac{\tau - \tau_j}{\tau_i - \tau_j} \quad i = 0, \dots, N. \quad (49)$$

The point $\tau_0 = -1$ is not collocated and the time derivative of the state is not approximated there. The state in (41) is approximated by

$$x(\tau) \approx \begin{bmatrix} \mathcal{L}_0(\tau) & \dots & \mathcal{L}_N(\tau) \end{bmatrix} \begin{bmatrix} x(\tau_0) \\ \vdots \\ x(\tau_N) \end{bmatrix} \quad (50)$$

in which the state vector is treated as a row vector $\begin{bmatrix} x_1(\tau) & \dots & x_n(\tau) \end{bmatrix} \in \mathbb{R}^n$. This means that

$$\begin{bmatrix} \dot{x}(\tau_1) \\ \vdots \\ \dot{x}(\tau_N) \end{bmatrix} \approx \begin{bmatrix} \dot{\mathcal{L}}_0(\tau_1) & \dots & \dot{\mathcal{L}}_N(\tau_1) \\ \vdots & \ddots & \vdots \\ \dot{\mathcal{L}}_0(\tau_N) & \dots & \dot{\mathcal{L}}_N(\tau_N) \end{bmatrix} \begin{bmatrix} x(\tau_0) \\ \vdots \\ x(\tau_N) \end{bmatrix}. \quad (51)$$

The $N \times (N + 1)$ matrix on the right-hand side of (51) is a constant differentiation matrix. If the first column of the differentiation matrix in (51) is removed, the resulting $N \times N$ matrix is non-singular with its inverse an implicit Legendre-Gauss-Radau integration matrix. Using this discretisation, the optimal control problem given in Section IV-A can be transformed into a nonlinear programming problem (NLP).

In some problems, including those studied here, the state equation (41) and performance index (42) may be functions of adjustable static parameters. In these cases the static parameters can be included as decision variables in the NLP decision vector and optimised alongside the state and controls. These details are explained in [26], which also provides a detailed description of the general-purpose optimal control problem solver GPOPS-II that is used to perform all the calculations presented in this paper. The solution of similar problems, using GPOPS-II, are described in [13], [14], where additional references are provided.

C. Scaling

Scaling can have a significant influence on the performance of optimisation algorithms. One scaling methodology is to transform variables from their original physical quantities into dimensionless variables that have desirable optimisation properties. In our work we take the length of the car to be the fundamental unit of length — after scaling the car has a dimensionless length of one. In the same way the fundamental unit of mass is the mass of the car. To scale time we choose a non-dimensional gravitational constant of unity and so a time scale of $\sqrt{g/l_0}$ is used in which l_0 is the car's length. This scaling of the basic length, mass and time variables induces a scaling on all the other physical quantities involving acceleration, velocity, force, torque, power and so on. This scaling scheme forces the NLP's decision variables into a 'more spherical' space than that associated with the original physical units.

D. Non-smooth features

Interior point nonlinear programming algorithms such as IPOPT require 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 [14]. One may also make use of $\max(a, b) = a + \max(0, b - a)$. In the case of $|x|$, one might use

$$|x| \approx \sqrt{x^2 + \epsilon}$$

in this case

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

with the derivative approximation at $x = 0$ again well defined. These approximations become more accurate as the value of ϵ is reduced, with values in the range $10^{-5} \leq \epsilon \leq 10^{-2}$ typical for the results given here.

E. Regularisation

When the controls enter the system dynamics and performance index linearly, the possibility of singular arcs exists. If the control Hamiltonian ceases to be an explicit function of the control variable u , no information about the optimal control can be gleaned from the PMP; these problems are referred to as 'singular'. In other words, if \mathcal{H} does not depend upon u explicitly, the usual procedure of selecting u^* so as to maximize \mathcal{H} breaks down and the PMP cannot determine directly a unique optimal control as a function of the state and costate variables. Instead, the optimal control must be determined by a new set of necessary conditions [28]–[31]. All the general-purpose optimal control software known to the authors assume the strong form of the Legendre-Clebsch necessary condition (ie., $\nabla_{uu} \mathcal{H} > 0$), and so these codes cannot be used to solve problems containing singular arcs.

In order to avoid oscillatory solutions and non-convergent NLP behaviour in these problems, it is often useful to introduce into the performance index small terms that are quadratic in the controls. These terms ensure, at least in theory, that the Pontryagin Minimum Principle can be used to find the optimal controls (for the perturbed problem). If these terms are 'small' relative to the elapsed time, they do not change the problem in a significant way. In our case we used an index of the form

$$J = \int_0^T (1 + \sum_{i=1}^m \epsilon_i u_i^2) dt \quad (52)$$

in which the ϵ_i 's are small constants with the u_i 's the controls. This is clearly no longer a pure minimum-time problem, but it is close to one if the ϵ 's are chosen sufficiently small. Instead of using J as a measure of the minimum lap time, the 'true' lap time is computed as a separate auxiliary state. In our case the minimum value of J is of the order 80 s, while the computed 'true' lap time is typically 0.05 s smaller; these differences are comparable with the changes obtained with variable numbers of mesh refinement steps.

F. Computing gradients

Gradient-based methods for solving NLPs require the derivatives of the objective function and constraints with respect to the states and the controls. The most obvious way to computing these derivatives is analytically, either by hand, or by using a computer algebra package. While the exactness of this approach is appealing, and generally results in faster rates of convergence, it is burdensome to compute derivatives this way. Another approach is to compute derivatives numerically using finite differencing. In the case of forward differencing one uses

$$\frac{df}{dx} \approx \frac{f(x + \Delta) - f(x)}{\Delta}$$

in which Δ is the step length. The difficulty with this approach is that Δ must be chosen small enough to provide a good approximation to the derivative, but not so small that round-off errors occur when calculating the difference $f(x + \Delta) - f(x)$. Another approach is *complex-step differentiation*, which is both accurate and efficient [32], [33]. In the case of higher order derivatives, a more general version of the complex-step derivative approximation, called the multi-complex-step derivative approximation, can be used as a way of computing Hessians without truncation errors [34]. Another powerful approach is automatic differentiation that computes function derivatives to the precision that one would achieve using analytic or symbolic differentiation. While any of these approaches can be used, in our experience automatic or symbolic differentiation produce the most reliable results.

G. Mesh Refinement

The numerical solution of optimal control problems using direct transcription methods involves three key steps: (A) Transcribing the optimal control problem into a nonlinear programming (NLP) problem using a numerical integration scheme; (B) Solving the NLP and (C) Reviewing the accuracy

of the solution and, if necessary, refining the mesh and resolving the (refined) problem. The accuracy and efficiency of this process can be influenced by many things including the choice of integration scheme, the solution mesh, the mesh refinement strategy (if one is used), and the problem characteristics themselves (cf. singular arcs, stiff and/or non-smooth dynamics and/or rapidly varying path constraints).

In the case of global orthogonal collocation methods the state and costate are approximated by polynomials over the entire optimisation interval, with increased accuracy coming from increasing the degree of the approximating polynomials used. For problems with smooth solutions pseudospectral (or orthogonal collocation) methods converge rapidly. In the case of non-smooth problems, however, the convergence rate of global pseudospectral methods will be slow (as the degree of the approximating polynomial increases), and a poor approximation may result even if high-degree polynomials are used.

In order to deal with non-smooth problems segmentation is used with global 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 [35]. If the error across a particular segment has a uniform behaviour, the number of collocation points is increased. If the error at an isolated point within a segment is significantly larger than the errors at other points within the segment, the segment is subdivided. The interested reader can refer to Figures 5 and 6 in [13] for an example of adaptive grid refinement in a race car application.

V. RESULTS

As with earlier research [12], [13], the study presented in this paper is based on minimum-time optimal laps of a 2D model of the Circuit de Barcelona-Catalunya, which is shown in Figure 11 (A). This particular track has sixteen corners with their approximate distances from the start-finish given in Figure 11 (B). The track's corner distances are useful features when seeking to understand the detailed behaviour of the car.

We will begin our study by considering the case when the car's suspension is 'frozen' in its nominal configuration ($h_f = -14$ mm and $h_r = -74.1$ mm), and with an aero map that has no angular sensitivities; in this case the last three rows of Table II are set to zero. The optimal lap time is computed as 83.10 s as shown in case (A) in Table III. Introducing the steer angle sensitivity (row 6 in Table II) causes a marginal increase in the lap time that is too small to compute reliably. As shown in row (B), when the steer and sideslip sensitivities are recognised, the lap time increases to 83.45 s.

The baseline car, corresponding to row (C) Table III, includes the aerodynamic coefficients given in Table II and the suspension system described by Table I. Nominal values for the car and tire parameters can be found in [12]. The introduction of the suspension system means that the car has roll and pitch freedoms and all the rows in Table II come into effect. As compared with rows (A) and (B) in Table III, the car will now benefit from increased down force due to reduced ride heights at high speed, but it will also suffer the detrimental effects of roll motions on the downforce coefficients.

Figure 12 (A) shows the whole-track speed profile of the baseline car, with the optimal racing line for the baseline car between turns 9 and 16 illustrated in Figure 12 (B); this figure shows the trajectory of the car's mass centre. In the optimal

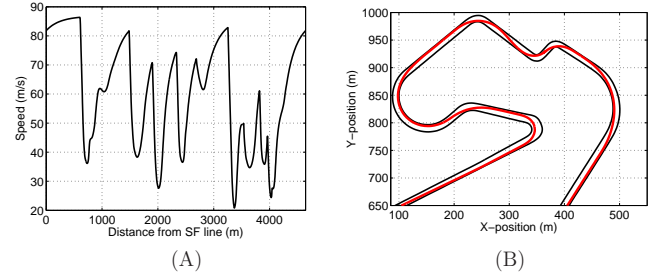


Fig. 12: Speed (A) and racing line (B) for one lap of the Circuit de Barcelona; the red trajectory shows the car's mass centre. The suspension and aerodynamic parameters are set to their baseline values; see row (C) in Table III.

control problem formulation used here the car's mass centre is constrained to remain within ± 4 m of the track centre. The track width is approximately 12 m on almost all parts of the track. This constraint ensures that the car's front wheels remain on the track whatever the vehicle's orientation (relative to the track). This explains why the racing line in turn 10 appears slightly conservative, where the track widens to approximately 20 m. A constraint set that allows for a variable width track, and which ensures that each of the four wheels remain on the track, is easy to implement, but takes longer to process. The differences between the solutions with these alternative problem formulations are not visible on plotted graphs and so the more complex constraint formulation was not used.

Figure 13 shows that as the car accelerates out of turn 16 onto the 1 km straight towards turn 1, the air flow over the front wing causes the front of the car to move closer to the road under the influence of increased down force.

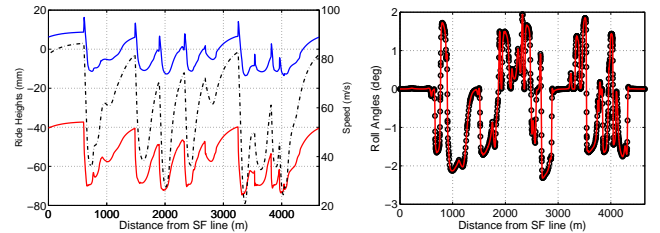
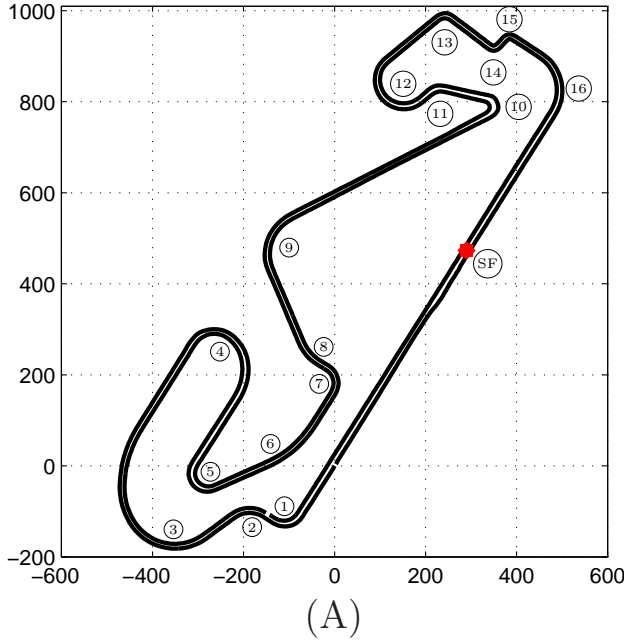


Fig. 13: Suspension ride heights and roll angles with the suspension and aerodynamic parameters set to their base-case values; see row (C) in Table III. The left-hand figure shows the vehicle speed (black dot-dash), the front ride height (blue) and the rear ride height (red). The right-hand diagram shows the front and rear axle roll angles, which are constrained to be equal in the optimisation process; see equation (38).

The car's increasing speed, and the increased airflow under the vehicle, also causes the car's rear axle to move closer to the road. The high-speed straight running car operates with a very small roll angle on this section of the track. Figure 13 shows these increased ride heights (in road-fixed SAE coordinates)



Corner	Distance (m)	Corner	Distance (m)
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

(B)

Fig. 11: Plan view of the Circuit de Barcelona-Catalunya with its corner numbers and the start-finish line (SF); the axes are in (m). Table (B) gives the approximate distances to the mid corner (in metres from the start-finish line).

between track positions 4400 m and 600 m, relative to the start-finish line, with an essentially zero roll angle over this track section. As the car's speed drops into the right-hand turn 1, the car's suspension is unloaded and the car moves away from the track surface, with the vehicle simultaneously rolling out of the corner; see the roll angle decrease at 725 m. Following on immediately after turn 1, is the left-hand corner 2, which causes the roll angle to switch from approximately -1.8° to 1.8° . As the car accelerates towards turn 3, the vehicle moves closer to the track with the roll angle remaining negative in the long, fast right-hand corner 3. Figure 13 illustrates the speed, roll angle and ride heights of the car as it traverses the rest of the track. The roll angle and ride heights are derived from the normal tire loads and the suspension meta model given in Table I. The variations in the suspension posture produce changes in the aerodynamic drag and down force coefficients in accordance with Table II. These changes, as well as the accompanying changes in the drag and down forces, are shown in Figure 14. It is evident that the drag Coefficient C_D is essentially constant (at approximately 0.95). To explain this, we observe that the aerodynamic model in Table II shows that the drag is linearly dependent on the difference between the front and rear ride heights; $C_D = 0.9455 + 0.1(h_f - h_r)$. Further, a pitch angle of 1° , which is large for a F1 race car, would produce a ride-height difference of about 59 mm. This difference represents a change of 0.006 to the drag coefficient. For that reason the drag coefficient is not significantly affected by suspension pitch movements, or differential ride height variations. In contrast the down force coefficients C_L^f and C_L^r vary significantly between approximately 1.3 and 1.6. It is also evident that the total down force reaches approximately 20 kN, which is roughly 3.5 times the weight of the car.

The remainder of Table III shows the effect on the optimal

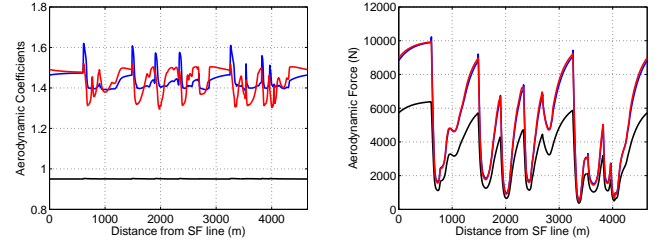


Fig. 14: Aerodynamic coefficients (left) and aerodynamic forces (right) with the suspension and aerodynamic parameters set to their base values; see row (C) in Table III. The drag force and drag coefficient are shown in black, the front down force and down force coefficient are shown in blue, and the rear down force and down force coefficient are shown in red.

lap time of a variety of suspension and aerodynamic set up changes, some of which act in combination. In row (D) the front and rear ride heights can be adjusted by ± 10 mm and ± 30 mm respectively. In order to minimise the lap time, the optimiser finds that the front ride height should be maximised, while the rear ride height should be adjusted by 12.7 mm. It appears that the rear ride height adjustment is effectively optimising the aerodynamic balance of the car (the position of the aerodynamic centre of pressure). These ride height adjustments reduce the predicted lap time by 0.50 s.

In the next three tests, rows (E-G) in Table III, the effect of changing the aerodynamic sensitivities to side slip, steer and roll are investigated. If the changes in the yaw, steer and roll sensitivities are ΔC_L^ψ , ΔC_L^δ and ΔC_L^ϕ , then changes in the nominal down force coefficients ΔC_L is constrained by

$$\Delta C_L = - \left(w_1 \Delta C_L^\psi + w_2 \Delta C_L^\delta + w_3 \Delta C_L^\phi \right) \quad w_i \geq 0,$$

—	ΔR_{rh}	ΔF_{rh}	$\frac{\partial C_L}{\partial \psi^2}$	$\frac{\partial C_L}{\partial \delta^2}$	$\frac{\partial C_L}{\partial \phi^2}$	Aero Balance ^a	Lap Time
(A) Rigid suspension (no aero sensitivities)	—	—	—	—	—	—	83.10
(B) Rigid suspension (with aero sensitivities)	—	—	—	—	—	—	83.54
(C) Suspension BaseLine	—	—	—	—	—	—	84.55
(D) RH offset ^b	0.0127	0.01	—	—	—	—	84.05
(E) Side slip sensitivity	—	—	-5 ^c -10	—	—	—	84.43
(F) Steer angle sensitivity	—	—	—	-0.15 -0.15	—	—	84.55
(G) Roll angle sensitivity	—	—	—	—	5 20	—	84.44
(H) Fixed aero balance	—	—	—	—	—	0.1936	83.66
(I) All-angle sensitivity RH and aero balance	0.0015	0.01	-5 -10	-0.15 -0.15	5 20	0.2200	82.75
(J) Point-wise aero balance	—	—	—	—	—	Continuous	82.49

^a Positive values move the aero-dynamic centre of pressure rearwards.

^b Optimised ride height offset in metres.

^c Front axle quantities are in the North-East corner.

TABLE III: Aero-dynamic properties of the car and suspension. All data is associated with a racing lap of Barcelona.

for weights w_i . The purpose of this constraint is to ensure that the optimisation process is not offered a ‘free lunch’, whereby the benefits of reducing the aerodynamic sensitivities can be introduced without reducing the nominal down force coefficient in compensation. One example of a constraint of this form is a water-bed constraint that preserves the “area under the aerodynamic down force curves”. In this case the change in the nominal down force coefficients are given by

$$0 = \int_{-\psi_{max}}^{\psi_{max}} \int_{-\delta_{max}}^{\delta_{max}} \int_{-\phi_{max}}^{\phi_{max}} \left(\Delta C_L + \Delta C_L^\psi \psi^2 + \Delta C_L^\delta \delta^2 + \Delta C_L^\phi \phi^2 \right) d\psi d\delta d\phi,$$

which means that

$$\Delta C_L = - \left(\frac{\psi_{max}^2}{3} \Delta C_L^\psi + \frac{\delta_{max}^2}{3} \Delta C_L^\delta + \frac{\phi_{max}^2}{3} \Delta C_L^\phi \right) \quad (53)$$

in which $\psi_{max} = \frac{4\pi}{45}$ rad, $\delta_{max} = \frac{\pi}{12}$ rad and $\phi_{max} = \frac{\pi}{90}$ rad are the ranges of validity of the aerodynamic maps.

In row (E) we allow the front and rear side slip aerodynamic sensitivities to be changed by ± 5 and ± 10 respectively, with the down force coefficients constrained by (53). The nominal values for these sensitivities are -5.1294 and -10.2588 respectively; see the fifth row in Table II. In this case the optimal strategy is to maximise the front and rear side slip sensitivities (more negative values) in order to increase the nominal front and rear down force coefficients. The associated lap-time reduction is 0.12 s. In row (F) the front and rear steer aerodynamic sensitivities are allowed to change by ± 0.15 . The nominal values for these sensitivities are -0.1641; see the sixth row in Table II. Again, the water bed constraint allows a higher nominal down force coefficient to be ‘bought’ with increased sensitivity. There is no discernible change in the predicted lap-time in this case. In row (G) the front and rear roll aerodynamic sensitivities are allowed to change by ± 5 and

± 20 respectively. The nominal values for these sensitivities are -8.2070 and -32.8281 respectively; see the seventh row in Table II. In this case the sensitivities are minimised thereby decreasing the nominal front and rear down force coefficients. The predicted lap-time reduction is 0.11 s.

In row (H) the aerodynamic balance of the car is adjusted by facilitating a relocation of the aerodynamic centre of pressure. The optimal control calculation is allowed to add a fixed aero offset $-1 \leq a_{off} \leq 1$ to the nominal rear down force coefficient of 0.9494; see Table II. If a_{off} is added to the rear, then it is subtracted in compensation from the front nominal down force coefficient of 1.20 in order to keep the total down force constant. In effect, positive values move the centre of pressure (CoP) rearwards, while negative values move the CoP forwards. In the case of the vehicle, tires and aero map used here, the optimal aero balance adjustment is to move the CoP rearwards by $a_{off} = 0.1936$, with a resulting lap time reduction of 0.89 s. Row (I) shows the effect of adjusting the front and rear ride heights in combination with the three aerodynamic sensitivities and the aero balance parameter a_{off} described above. The optimal aero balance coefficient is 0.220, with the front ride height dropped by the maximum 10 mm, while the rear ride height is dropped by only 1.5 mm. It appears that adjusting the aero balance coefficient renders rear ride height adjustments virtually unnecessary, because these changes produce the same effect. These changes, in combination, reduce the predicted lap time of the base line car by a significant 1.8 s.

The study given in row (J) treats the aero balance coefficient as an idealised and continuously variable control input $-1 \leq u_{CoP}(t) \leq 1$. The front down force coefficient is given by $C_L^f - u_{CoP}(t)$, while the rear down force coefficient is given by $C_L^r + u_{CoP}(t)$ thereby keeping $C_L = C_L^r + C_L^f$ constant. Positive values of $u_{CoP}(t)$ move the CoP rearwards, while negative values move the CoP forwards. Figure 15 shows the

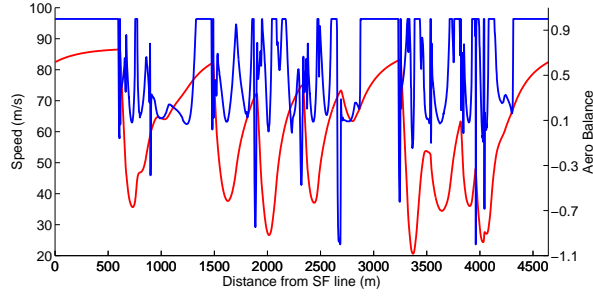


Fig. 15: Continuously adjustable aero-balance u_{CoP} (blue) and vehicle speed (red); see row (J) in Table III. The blue curve shows a continuously variable aero-balance control; positive values increase the rear down force coefficient.

aero balance control u_{CoP} (in blue) with the vehicle speed (red). The optimal strategy is to move the CoP rearwards under firm acceleration when high normal loads on the rear tires are required. The CoP is then moved forwards under braking and prior to turning into corners when high front tire side forces are needed. The influence of this hypothetical control is substantial and reduces the lap time by over 2 s, thereby emphasizing the importance of good aerodynamic balance.

By way of comparison, Figure 16 shows the changes in the

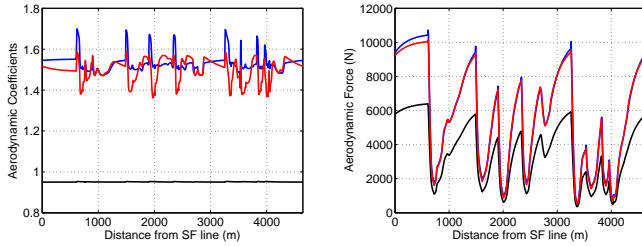


Fig. 16: Aerodynamic coefficients (left) and aerodynamic forces (right) with optimised suspension and aerodynamic parameters (see row (I) in Table II). The drag force and drag coefficient are shown in black, the front down force and down force coefficient are shown in blue, and the rear down force and down force coefficient are shown in red.

aerodynamic coefficients and aerodynamic forces produced by optimised suspension and aerodynamic parameters; compare with Figure 14. It is evident that the rear down force coefficient has been increased with its minimum value rarely less than 1.4, whereas this minimum is approximately 1.3 in the base case. In the optimised case the front down force coefficient is rarely less than 1.5, while this minimum is approximately 1.4 in the base case. The increased front and rear down force resulting from the optimisation process is self-evident, while there is virtually no change in the vehicle's drag force.

Figure 17 shows the effect of suspension and aerodynamic parameter optimisation on the optimal lap speed profile. The main optimisation gains appear to be on the long straight between turn 16 and turn 1, the long high-speed turn 3, the straight between turns 9 and 10, with some small benefit achieved almost everywhere else.

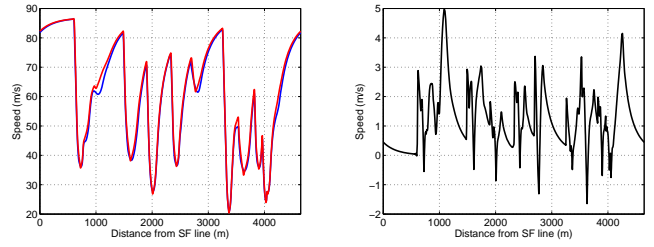


Fig. 17: Speed comparison between the baseline case (blue) and the study with optimised ride heights, aero-balance and aero-sensitivities (red); see rows (C) and (I) respectively in Table III. The right-hand plot shows the speed difference; positive values indicate that the optimised car is faster.

VI. CONCLUSION

Pseudospectral methods for solving numerical optimal control problems are a useful tool for the aerodynamic performance optimisation of high-performance race cars. The results presented in this paper focus on the interactions between the vehicle's aerodynamic performance and its suspension system. The kinematics of multi-link suspension systems that are typical of open-wheeled race cars involve the solution of a complex set of nonlinear algebraic equations. One source of complication are the torsion bars and heave spring rocker assemblies that produce coupling influences between the road wheels on each axle. It is not practical, or necessary, to include these equations in the vehicle dynamics model used for optimal control calculations. Instead, the solution to these equations can be accurately approximated by a multivariate polynomial, or 'meta' model, which is fitted off-line to the solution of the describing kinematic equations. In this way the interactions between the car's aerodynamics and its pull-rod suspension system can be accurately represented by smooth (almost linear) polynomials that can be quickly evaluated as part of an optimal control calculation. It is shown that parameter optimisation problems, involving the car's suspension and aerodynamics, are readily soluble. The results demonstrate that the track-specific optimisation of these systems can substantially reduce the lap time. The specific figures will depend on the track, the tire and vehicle properties, and the closeness of the base vehicle to its optimal set up. This work suggests that ever more complex vehicular optimisation problems may now become tractable.

VII. ACKNOWLEDGEMENT

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

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{bmatrix} - \underbrace{\begin{bmatrix} 1 & (x_1)_1 & (x_2)_1 & (x_3)_1 & \cdots & (x_n^d)_1 \\ 1 & (x_1)_2 & (x_2)_2 & (x_3)_2 & \cdots & (x_n^d)_2 \\ \vdots & & \ddots & \ddots & & \vdots \\ 1 & (x_1)_k & (x_2)_k & (x_3)_k & \cdots & (x_n^d)_k \end{bmatrix}}_{\Phi} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_q \end{bmatrix} \quad (55)$$

REFERENCES

- [1] J. Reimpell, H. Stoll, and J. W. Betzler, *The Automotive Chassis: Engineering Principles*, 2nd ed. Society of Automotive Engineers, 2001.
- [2] M. Raghavan, "Number and dimensional synthesis of independent suspension mechanisms," *Mech. Mach. Theory*, vol. 31, no. 8, pp. 1141–1153, 1996.
- [3] D. L. Cronin, "Macpherson strut kinematics," *Mechanism and Machine Theory*, vol. 16, no. 6, pp. 631–644, 1981.
- [4] K. Chen and D. G. Beale, "Base dynamic parameter estimation of a Macpherson suspension mechanism," *Vehicle System Dynamics*, vol. 39, no. 3, pp. 227–244, 2003.
- [5] D. E. Woods and B. A. Jawad, "Numerical design of racecar suspension parameters," in *SAE Technical Series Papers*, no. 1999-01-2257, 1999.
- [6] H. M. Hiller and S. Frik, "Road vehicle benchmark 2 five-link suspension," *Vehicle System Dynamics*, vol. 22, no. S1, pp. 254–262, 1993.
- [7] J. Knapczyk and S. Dzierżek, "Displacement and force analysis of five-rod suspension with flexible joints," *ASME J Mechanical Desgn*, vol. 117, no. 4, pp. 532–538, 1995.
- [8] P. Simionescu and D. Beale, "Synthesis and analysis of the five-link rear suspension system used in automobiles," *Mechanism and Machine Theory*, vol. 37, pp. 815–832, 2002.
- [9] R. Sancibrian, P. Garcia, F. Viadero, A. Fernandez, and A. De-Juan, "Kinematic design of double-wishbone suspension systems using a multiobjective optimisation approach," *Vehicle System Dynamics*, vol. 48, no. 7, pp. 793–813, 2010.
- [10] K. P. Balike, S. Rakheja, and I. Stiharu, "Development of kinetodynamic quarter-car model for synthesis of a double wishbone suspension," *Vehicle System Dynamics*, vol. 49, pp. 107–128, 2011.
- [11] Y. A. Papegay, J.-P. Merlet, and D. Daney, "Exact kinematics analysis of cars suspension mechanisms using symbolic computation and interval analysis," *Mechanism and Machine Theory*, vol. 40, no. 4, pp. 395–413, 2005.
- [12] G. Perantoni and D. J. Limebeer, "Optimal control for a formula one car with variable parameters," *Vehicle System Dynamics*, vol. 52, no. 5, pp. 653–678, 2014.
- [13] D. J. N. Limebeer, G. Perantoni, and A. V. Rao, "Optimal control of formula one car energy recovery systems," *International Journal of Control*, vol. 87, no. 19, pp. 2065–2080, 2014.
- [14] D. J. N. Limebeer and G. Perantoni, "Optimal control of a formula one car on a three-dimensional track part 2: Optimal control," *ASME J. Dynamic Systems, Measurement and Control*, vol. 137, pp. 051019–1 to 13, May 2015.
- [15] G. G. Wang and S. Shan, "Review of metamodeling techniques in support of engineering design optimization," *ASME J Mechanical Desgn*, vol. 129, no. 4, pp. 370–380, 2007.
- [16] *Formula One Technical Regulations*, Fédération Internationale de l'Automobile Std., 2015. [Online]. Available: <http://www.fia.com/sport/regulations>
- [17] A. Rucco, G. Notarstefano, and J. Hauser, "Development and numerical validation of a reduced-order two-track car model," *European Journal of Control*, vol. 20, pp. 163–171, 2014.
- [18] G. Perantoni and D. J. N. Limebeer, "Optimal control of a Formula One car on a three-dimensional track. Part 1: Track modelling and identification," *ASME J. Dynamic Systems, Measurement and Control*, vol. 137, pp. 051018–1 to 11, May 2015.
- [19] D. P. Kelly, *Lap Time Simulation with Transient Vehicle and Tyre Dynamics*. Cranfield University School of Engineering, 2008, PhD Thesis.
- [20] 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, no. 2, pp. 113–135, 1999.
- [21] H. B. Pacejka, *Tyre and Vehicle Dynamics*, second edition ed. Butterworth-Heinemann, 2008.
- [22] G. A. Bliss, "The problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 19, pp. 305–314, 1918.
- [23] A. E. J. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. Massachusetts: Blaisdell Publishing Company, 1969.
- [24] D. E. Kirk, *Optimal Control Theory: An Introduction*. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1970.
- [25] A. P. Sage and I. C. White, *Optimum Systems Control*. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1977.
- [26] M. A. Patterson and A. V. Rao, "Gpops - II: A matlab software for solving multiple-phase optimal control problems using hpadaptive gaussian quadrature collocation methods and sparse nonlinear programming," *ACM Transactions on Mathematical Software*, vol. 39, no. 1, pp. 1–37, October 2014.
- [27] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. Academic Press, Orlando, 1984.
- [28] H. J. Kelley, "A second variation test for singular extremals," *AIAA J.*, vol. 2, pp. 1380–1382, 1964.
- [29] B. S. Goh, "Necessary conditions for singular extremals involving multiple control variables," *SIAM J Control*, vol. 4, no. 4, pp. 716–731, 1966.
- [30] H. M. Robbins, "A generalized legendre-clebsch condition for the singular cases of optimal control," *IBM J. Res. Develop.*, vol. 11, pp. 361–372, 1967.
- [31] A. J. Krenner, "The high order maximal principle and its application to singular extremals," *SIAM J. Control*, vol. 15, no. 2, pp. 256–293, 1977.
- [32] W. Squire and G. Trapp, "Using complex variables to estimate derivatives of real functions," *SIAM Review*, vol. 40, p. 110112, 1998.
- [33] J. R. R. Martins, P. Sturda, and J. J. Alonso, "The complex-step derivative approximation," *ACM Transactions on Mathematical Software*, vol. 29, p. 245262, 2003.
- [34] G. Lantoine, R. P. Russell, and T. Dargent, "Using multicomplex variables for automatic computation of high-order derivatives," *ACM Transactions on Mathematical Software*, vol. 38, no. 3, p. 16:116:21, 2012.
- [35] 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.
- [36] D. A. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*, 2nd ed., S. Axler, F. Gehring, and K. Ribet, Eds. Springer, 2005.

APPENDIX

The set of all multivariate polynomials in n variables up to degree d over the field of complex numbers \mathbb{C} , together with addition and multiplication with a scalar, form a vector space \mathcal{C}_d^n . A basis for this vector space consists of all monomials from degree 0 up to d . Since the total number of monomials in n variables from degree 0 to degree d is given by $q = \frac{(n+d)!}{d!n!}$, it follows that $\dim(\mathcal{C}_d^n) = q$. The degree of a monomial $x^a = x_1^{a_1} \cdots x_n^{a_n}$ is defined as $|a| = \sum_{i=1}^n a_i$. The degree of a polynomial p is the degree of the monomial of p with highest degree. The terms of multivariate polynomials can be ordered in different ways. A formal definition of monomial orderings together with a detailed description of some orderings in computational algebraic geometry is given in [36]; we will use the *Graded Reverse Lexicographic Order*. In this ordering system, $a > b$ if $|a| = \sum_{i=1}^n a_i > |b| = \sum_{i=1}^n b_i$, or if $|a| = |b|$ and $a >_{xel} b$, if in the vector difference $a - b$ the leftmost nonzero entry is negative. The ordering is ‘graded’ because it first compares the degrees of the two monomials and then applies the $>_{xel}$ ordering when there is a tie.

Example $(2, 0, 0) > (0, 0, 1)$, because $|(2, 0, 0)| > |(0, 0, 1)|$, which implies that $x_1^2 > x_3$. Likewise, $(0, 1, 1) > (2, 0, 0)$, because $(0, 1, 1) >_{xel} (2, 0, 0)$ and so $x_2 x_3 > x_1^2$.

A monomial ordering allows a multivariate polynomial p to be represented by a vector of its coefficients in a consistent way. To do this one orders the coefficients in a row vector that is induced by the chosen ordering system. Suppose that $p = 2 + 3x_1 - 4x_2 + x_1x_2 - 8x_1x_3 - 7x_2^2 + 3x_3^2 \in \mathcal{C}_3^3$; since $\frac{5!}{3!2!} = 10$, the associated basis is 10 dimensional. As a result p can be represented by the vector of coefficients in graded reverse lexicographic order

$$\begin{array}{cccccccccc} 1 & x_1 & x_2 & x_3 & x_1^2 & x_1x_2 & x_1x_3 & x_2^2 & x_2x_3 & x_3^2 \\ 2 & 3 & -4 & 0 & 0 & 1 & -8 & -7 & 0 & 3 \end{array}$$

The least-squares fitting of multivariate polynomials to given data is an easy extension of the univariate case. Suppose that we wish to fit a multivariate polynomial of the form

$$\begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_n^d \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_q \end{bmatrix} \quad (54)$$

to given data, in which the terms are listed in *graded reverse lexicographic order*. Since the dimension of \mathcal{C}_d^n is q , we will suppose from now on that we have $k \geq q$ independent experimental outcomes, which can be stacked as (55) in which e is a k -vector of errors, f is a k -vector of experiment outcomes, and p is a q -vector of polynomial coefficients. More compactly we have

$$e = f - \Phi p. \quad (56)$$

Optimum values for the parameter vector p can be found by completing the square

$$\begin{aligned} e^T e &= (p - (\Phi^T \Phi)^{-1} \Phi^T f)^T (\Phi^T \Phi) (p - (\Phi^T \Phi)^{-1} \Phi^T f) \\ &\quad + f^T (I_k - \Phi (\Phi^T \Phi)^{-1} \Phi^T) f. \end{aligned} \quad (57)$$

Since the first term on the right-hand side of (57) is non-

negative for any choice of p , the optimal polynomial coefficients come from making this term zero by setting

$$p_{opt} = (\Phi^T \Phi)^{-1} \Phi^T f \quad (58)$$

with the resulting optimal estimation error

$$\|e\|_2^2 = f^T (I_k - \Phi (\Phi^T \Phi)^{-1} \Phi^T) f. \quad (59)$$