Spark ignition engine control strategies for minimising cold start fuel consumption under cumulative tailpipe emissions constraints

D.J. Andrianov*, C. Manzie, M.J. Brear

Department of Mechanical Engineering, The University of Melbourne, Victoria 3010, Australia

**Abstract**

This paper proposes a methodology for minimising the fuel consumption of a gasoline fuelled vehicle during cold starting. It first takes a validated dynamic model of an engine and its aftertreatment reported in a previous study (Andrianov, Brear, & Manzie, 2012) to identify optimised engine control strategies using iterative dynamic programming. This is demonstrated on a family of optimisation problems, in which fuel consumption is minimised subject to different tailpipe emissions constraints and exhaust system designs. Potential benefits of using multi-parameter optimisation, involving spark timing, air–fuel ratio and cam timing, are quantified. Single switching control policies are then proposed that perform close to the optimised strategies obtained from the dynamic programming but which require far less computational effort.

1. Introduction

Road vehicles with internal combustion engines are a significant source of air pollution (Seinfeld, 2004). The pollutants found in the exhaust include carbon monoxide (CO), nitrogen oxides (NO and NO2, also referred to as NOx), unburned hydrocarbons (HC) and particulates. These substances present significant environmental and health risks, and are therefore regulated. To improve the air quality and account for the increasing number of road vehicles, the allowable emissions limits are continually tightened.

In most vehicles with gasoline fuelled engines, these emissions limits are achieved with the use of three-way catalysts in the exhaust system, designed to convert engine-out CO, NOx and HC emissions to CO2, H2O and N2. However, the chemical processes involved depend strongly on the catalyst temperature. Whilst the conversion efficiency of a hot catalyst can be high, a cool catalyst performs poorly. Consequently, cold start emissions play a critical role in meeting emissions standards.

Engine control is a cost effective approach to limit cold start emissions, whilst avoiding the need for additional or upgraded hardware. A common strategy is to retard the spark timing. This enables more heat to be rejected into the exhaust, which heats the catalyst more quickly. Another approach is to raise the engine’s idle speed to produce an increased number of combustion events, and thus higher enthalpy input to the catalyst. Both of these approaches, however, result in increased fuel consumption. More generally, maximising vehicle fuel economy whilst meeting emissions standards is a key and ongoing problem faced by all car manufacturers.

Manufacturers have traditionally relied heavily on experimentation to identify engine control set-points during cold starting (the work of Dohner, 1978 is an early example). However, since every cold start test must be followed by a cooling period, these approaches are both time consuming and costly. To reduce both the amount and duration of the testing required, a variety of model-based methods have been proposed. Some of these do not directly consider tailpipe emissions (Benz, Hehn, Onder, & Guzzella, 2011; Keynejad & Manzie, 2011b; Sanketi, Zavala, & Hedrick, 2004; Shaw & Hedrick, 2003; Sun & Sivashankar, 1997), whilst other methodologies make use of black-box (e.g. Cohen, Randall, Tether, VanVoorhis, & Tennant, 1984) or phenomenological (e.g. Kang, Kolmanovsky, & Grizzle, 2001; Kolmanovsky, Siverguina, & Lygoe, 2002; Kum, Peng, & Bucknor, 2011) models. However, indirect consideration of tailpipe emissions in the optimisation can yield inaccurate or misleading results. Furthermore, use of black-box and phenomenological models can be impractical, as a significant amount of engine testing can be required for their calibration.

This paper takes a different approach. The minimisation of fuel consumption under cumulative tailpipe emissions constraints is viewed as a dynamic optimisation problem, involving a computationally practical and validated cold start model of a spark ignition engine, an exhaust system and a three-way catalyst (Andrianov et al., 2012). In contrast to other numerical optimisation approaches, which use black-box or phenomenological models of similar functionality (e.g. Bérard, Cotta, Stokes, Thring, & Wheals,
2000; Cohen et al., 1984; Fiengo, Glielmo, Santini, & Serra, 2002; Fussey, Goodfellow, Oversby, Porter, & Wheals, 2001; Sanketi et al., 2006), this work takes advantage of physics-based modelling to significantly reduce the amount of engine testing required. In the following sections the methodology for obtaining optimised engine control strategies is demonstrated on several examples, where spark timing, air–fuel ratio and cam timing are subject to optimisation. The control policies found are then compared and validated experimentally when possible.

2. Problem formulation

To enable the dynamic optimisation, a computationally practical model capable of simulating fuel consumption and legislated tailpipe emissions as a function of the engine control setpoints is required. Throughout this work it is assumed that perfect setpoint controllers are in place to deliver the developed trajectories. The validated model of Andrianov et al. (2012) includes the appropriate control setpoint to fuel consumption and tailpipe emissions functionality. The accuracy in simulating cumulative fuel consumption and tailpipe emissions under transient driving conditions is of order 2% and 10% respectively with respect to experimental results. In this section this model is first briefly described, and then the optimal engine control problem is formulated.

2.1. The integrated model (Andrianov et al., 2012)

The structure of the model used is shown in Fig. 1. The engine is represented by a second order mean value model similar to that of Keynejad and Manzie (2011a) and considers fluid, thermal and mechanical domains. It calculates the intake manifold pressure $p_{in}$, exhaust port gas temperature $T_{cyl}$ and fuel mass flow rate $m_{fuel}$ as a function of throttle angle $\alpha$, engine speed $N$, normalised

![Fig. 1. Structure of the combined engine, emissions, exhaust and aftertreatment system model.](image-url)
air–fuel ratio setpoint \( \lambda \), spark timing \( \theta \), and cam timing \( \theta_{int} \) and \( \theta_{avlp} \). To simulate driving conditions using representative engine speed and torque trajectories, the dynamometer control system model is used. This model is implemented as a PI controller, adjusting the engine’s throttle angle \( \alpha \) to track the reference torque \( \tau_{\text{ref}}^{\text{brake}} \) for some prescribed engine speed \( N_{\text{ref}}^{\text{ref}} \). Engine-out CO, NO and HC emissions are approximated by static functions of the engine state and control setpoints. Nitrogen dioxide (NO\(_2\)) emissions from spark ignition engines are negligible (Heywood, 1988) and are therefore not considered. Other emissions (O\(_2\) and H\(_2\)) required by the catalyst model are estimated based on chemical equilibrium calculations, whilst ensuring consistency between the exhaust gas composition modelled and the \( \lambda \) setpoint.

The exhaust manifold and the pre-catalyst connecting pipe are both modelled using a lumped parameter approach. Warm-up behaviour and heat transfer between the exhaust gas and inner surfaces, and between outer surfaces and the ambient air are considered. These models estimate the gas temperature drop from the exhaust port to the catalyst inlet. The three-way catalyst is modelled using a variation of the one-dimensional physics-based approach of Pontikakis and Stamates (2004). The model considers substrate warm-up dynamics, heat and mass transfer between the exhaust gas and the washcoat, thermal conduction in the substrate, heat release from exothermic reactions, oxygen storage and a reduced order chemical kinetic scheme with 10 reactions.

The overall model is represented by a system of differential algebraic DAEs

\[
\begin{align*}
&x(t) = F_{\text{integ}}(x(t), z(t), u(t)), \\
&0 = G_{\text{integ}}(x(t), z(t), u(t)),
\end{align*}
\]

where functions \( F_{\text{integ}} \) and \( G_{\text{integ}} \) enclose the model equations, \( x(t) \in \mathbb{R}^{35}, z(t) \in \mathbb{R}^{7} \) and \( u(t) \in \mathbb{R}^{6} \) are state, algebraic and input vectors respectively. Whilst this model encloses a significant number of equations, it has been shown (Andrianov et al., 2012), nonetheless, that such formulation provides a reasonable compromise between the model’s accuracy and complexity. An alternative integrated model, for a closely coupled catalyst, is formulated such that the outlet boundary conditions of the exhaust manifold specify the inlet conditions for the catalyst. This model is represented by

\[
\begin{align*}
&x(t) = F_{\text{integ,ccc}}(x(t), z(t), u(t)), \\
&0 = G_{\text{integ,ccc}}(x(t), z(t), u(t)),
\end{align*}
\]

where functions \( F_{\text{integ,ccc}} \) and \( G_{\text{integ,ccc}} \) differ from \( F_{\text{integ}} \) and \( G_{\text{integ}} \) with the exception of the connecting pipe dynamics. Consequently, the state and algebraic vectors become \( x(t) \in \mathbb{R}^{34} \) and \( z(t) \in \mathbb{R}^{7} \). In both model formulations the input vector \( u(t) \) is specified by

\[
u(t) = [\tau_{\text{ref}}^{\text{brake}}, N_{\text{ref}}^{\text{ref}}, \lambda, \theta, \theta_{int}, \theta_{avlp}]^T,
\]

and the outputs are given by

\[
m_{\text{fuel}}, m_{\text{CO, out}}, m_{\text{NO, out}}, m_{\text{HC, out}}^\text{int} = H_{\text{integ}}(x(t), u(t)),
\]

where \( m_{\text{X, out}} \) is the mass flow rate of tailpipe emissions \( \text{X} \). Importantly, only fully warm engine maps and the results from a single transient test are required for complete calibration of the integrated model. For further details on the modelling, model calibration and validation see Andrianov (2011) and Andrianov et al. (2012).

2.2. Optimal control problem formulation

If \( t_{\text{cyc}} \) is the duration of the drive cycle, specified by engine brake torque \( \tau_{\text{ref}}^{\text{brake}} \), and engine speed \( N_{\text{ref}}^{\text{ref}} \) inputs, and vector \( u^*_c(t) \) is the optimal trajectory of the engine control setpoints, with variables in \( u_c \) constrained within physically reasonable ranges, then the dynamic optimisation problem is formulated as

\[
f(u_c) = \int_0^{t_{\text{cyc}}} \dot{m}_{\text{fuel}}(u_c, t) \, dt,
\]

\[
u_c^*(t) = \arg \min_{u_c(t)} f(u_c(t)),
\]

such that

\[
\begin{align*}
\dot{\psi}(t) &= \arg \min_{\psi(t)} f(\psi(t)), \\
\theta &= \arg \min_{\theta(t)} f(\theta(t)), \\
\theta_{int} &= \arg \min_{\theta_{int}(t)} f(\theta_{int}(t)), \\
\theta_{avlp} &= \arg \min_{\theta_{avlp}(t)} f(\theta_{avlp}(t)),
\end{align*}
\]

\[
\dot{\psi} \leq \psi \leq \dot{\psi} \quad \text{and} \quad \theta \leq \theta \leq \theta
\]

\[
\theta_{int} \leq \theta_{int} \leq \theta_{int}
\]

\[
\theta_{avlp} \leq \theta_{avlp} \leq \theta_{avlp}
\]

\[
\int_0^{t_{\text{cyc}}} m_{\text{out}}(u_c(t), t) \, dt < m_{\text{out, max}}^\text{int}
\]

Eq. (5f) constrains the ignition timing within a range supported by an ECU, whilst (5g) sets acceptable drivability characteristics and limits engine knock. Maximum brake torque spark timing \( \theta_{\text{MBT}}(t) \) is specified as a static function of the engine state and control setpoints. Cumulative tailpipe emissions are limited by \( m_{\text{out}} = [m_{\text{CO, out}}, m_{\text{NO, out}}, m_{\text{HC, out}}] \). Depending on whether an under-floor or a closely coupled catalyst configuration is required, either (1) and (4) or (2) and (4) form the equality constraints for (5).

2.3. Modified control problem formulation

The development of solutions to (5) over the full duration of the NEDC cycle \( t_{\text{cyc}} = 1180 \) s is computationally very intensive. Thus, an approximation for this optimisation problem is considered.

To reduce the computational requirements and improve the time resolution of the optimised control policies, the cycle is limited to the first 400 s, which includes catalyst light-off and covers engine warm-up from cold to fully warm operating conditions. Furthermore, the number of control input variables in \( u_c \), considered for optimisation is reduced. To distinguish the variables to be optimised, \( u_c \), is partitioned into vectors \( u_{\text{co}} \) and \( u_{\text{avlp}} \), containing engine control setpoints that are optimised and unchanged (preset to a production ECU strategy) respectively,

\[
u_c = [u_{\text{co}}, u_{\text{avlp}}]^T.
\]

If (5j) are implemented using barrier functions

\[
b_b(u_{\text{co}}) = \begin{cases} 0 & \text{if } \int_0^{t_j} m_{\text{out}}(u_{\text{co}}, t) \, dt < m_{\text{out, max}}^\text{int} \\
1 & \text{if } \int_0^{t_j} m_{\text{out}}(u_{\text{co}}, t) \, dt \geq m_{\text{out, max}}^\text{int} \end{cases}
\]

which are 0, if the constraints (5j) are satisfied, and are greater than or equal to 1, when they are violated, the optimal control problem (5) is approximated as

\[
f(u_{\text{co}}) = \int_0^{t_j} \dot{m}_{\text{fuel}}(u_{\text{co}}, t) \, dt + \sum_{i = \text{CO,NO,HC}} w_i \dot{b}_i(u_{\text{co}}),
\]

\[
u_{\text{co}}^*(t) = \arg \min_{u_{\text{co}}(t)} f(u_{\text{co}}(t)),
\]

\[
\text{where } w_i \text{ is a weighting factor and } b_i(u_{\text{co}}) \text{ is a barrier function.}
\]
such that
\[ u = [r_{\text{torque}}^\text{ref}, N^\text{ref}, u_{\text{in}}, u_{\text{out}}]^T, \] (7d)
\[ \lambda_{\text{min}} \leq \lambda(t) \leq \lambda_{\text{max}}, \] (7e)
\[ \theta_{\text{min}} \leq \theta(t) \leq \theta_{\text{max}}, \] (7f)
\[ \theta_{\text{MBT}} + \theta_{\text{adv,min}} \leq \theta(t) \leq \theta_{\text{MBT}} + \theta_{\text{adv,max}}, \] (7g)
\[ \theta_{\text{int,min}} \leq \theta_{\text{int}}(t) \leq \theta_{\text{int,max}}. \] (7h)
\[ \theta_{\text{ovlp,min}} \leq \theta_{\text{ovlp}}(t) \leq \theta_{\text{ovlp,max}}. \] (7i)

In this formulation \( t_f = 400 \) s. In order to simulate hard constraints, the value for \( w_{\text{con}} \) selected is several orders of magnitude larger than the overall fuel consumption.

As previously, the equality constraints are specified either by (1) and (4) or (2) and (4), depending on whether an under-floor or a close-coupled catalyst configuration is required. The brake torque \( e_{\text{brake}}^\text{ref} \) and engine speed \( N^\text{ref} \) prescribe the driving conditions, and the unoptimised control variables \( u_{\text{in}} \) are assigned trajectories previously obtained from an NEDC test with the production control strategy. This paper considers several compositions of \( u_{\text{in}} \), two exhaust system configurations and tailpipe emissions limits based on Euro-3 and Euro-4. These cases are presented in Table 1.

Note that the valve overlap \( \theta_{\text{ovlp}} \) in Case 6 is part of \( u_{\text{in}} \), which is prescribed. The intake valve closing angle \( \theta_{\text{int}} \) can therefore be limited by the physical constraints in the cam timing mechanism, in addition to (7h). To consider a wider range of possible \( \theta_{\text{int}} \) in the optimisation, the valve overlap is specified in terms of \( \theta_{\text{int}} \) when the exhaust cam-shaft is in the fully advanced state. Consequently, the following applies:
\[ \theta_{\text{ovlp}} = \begin{cases} g_{\text{ovlp}}^\text{ref} & \text{if } \theta_{\text{int}} \text{--ID} + g_{\text{ovlp}}^\text{ref} > \theta_{\text{enh,min}} \text{,} \\ \theta_{\text{enh,min}} - (\theta_{\text{int}} \text{--ID}) & \text{if } \theta_{\text{int}} \text{--ID} + g_{\text{ovlp}}^\text{ref} \leq \theta_{\text{enh,min}} \text{.} \end{cases} \] (8)
where ID is the intake duration and \( \theta_{\text{enh,min}} \) is the most advanced exhaust valve closing angle. For the engine used in this work \( \theta_{\text{enh,min}} = -4.5^\circ \) ATDC and ID = 256°.

3. Proposed solution approach

The integrated models (1) and (2) are of relatively higher order. Consequently, indirect methods (involving calculus of variations or Pontryagin’s minimum principle), direct methods (where a dynamic optimisation problem is converted to a static problem) and dynamic programming (Bellman, 1954) are impractical for solving the dynamic optimisation problem (to convert a dynamic optimisation problem into a static problem) in reasonable time frames.

In this paper a simplified variation of dynamic programming called iterative dynamic programming (Luus, 2000) is used. However, whilst this procedure possesses good speed, flexibility and convergence properties, it does not guarantee convergence to a global optima, but is nonetheless needed to make the optimalisation computationally feasible.

The state-space is initially allocated by solving the model equations subject to some nominal engine control policy \( u_{\text{in}}(t) \), then by trajectories perturbed from this policy. For the optimisation cases considered in this work, 3–5 trajectory perturbations were typically used at each iteration. The inputs tested at each of the discretised values of the state vector \( x \) are chosen to lie in the

<table>
<thead>
<tr>
<th>Case</th>
<th>Optimisation vector</th>
<th>Control constraints</th>
<th>Emission constraints</th>
<th>Model equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( u_{\text{in}} = [\theta(t)] )</td>
<td>( \theta_{\text{min}} = -10^\circ )</td>
<td>( m_{\text{CO, out}} = 22.8 \text{ g} )</td>
<td>(1) and (4) under-floor catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{max}} = 50^\circ )</td>
<td>( m_{\text{NOx, out}} = 1.29 \text{ g} )</td>
<td>(based on Euro-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{adv,min}} = 30^\circ )</td>
<td>( m_{\text{PM, out}} = 1.58 \text{ g} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{adv,max}} = 5^\circ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( u_{\text{in}} = [\theta(t)] )</td>
<td>( \theta_{\text{min}} = -10^\circ )</td>
<td>( m_{\text{CO, out}} = 22.8 \text{ g} )</td>
<td>(2) and (4) close-coupled catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{max}} = 50^\circ )</td>
<td>( m_{\text{NOx, out}} = 1.29 \text{ g} )</td>
<td>(based on Euro-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{adv,min}} = 35^\circ )</td>
<td>( m_{\text{PM, out}} = 1.58 \text{ g} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \theta_{\text{adv,max}} = 10^\circ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( u_{\text{in}} = [\theta(t), \lambda(t)] )</td>
<td>( \lambda_{\text{min}} = 0.9 )</td>
<td>( m_{\text{CO, out}} = 22.8 \text{ g} )</td>
<td>(1) and (4) under-floor catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \lambda_{\text{max}} = 1.1 )</td>
<td>( m_{\text{NOx, out}} = 1.29 \text{ g} )</td>
<td>(based on Euro-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( m_{\text{PM, out}} = 1.58 \text{ g} )</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( u_{\text{in}} = [\theta(t), \lambda(t)] )</td>
<td>( \lambda_{\text{min}} = 0.9 )</td>
<td>( m_{\text{CO, out}} = 22.8 \text{ g} )</td>
<td>(2) and (4) close-coupled catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \lambda_{\text{max}} = 1.1 )</td>
<td>( m_{\text{NOx, out}} = 1.29 \text{ g} )</td>
<td>(based on Euro-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( m_{\text{PM, out}} = 1.58 \text{ g} )</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( u_{\text{in}} = [\theta(t), \lambda(t)] )</td>
<td>( \lambda_{\text{min}} = 0.9 )</td>
<td>( m_{\text{CO, out}} = 9.9 \text{ g} )</td>
<td>(2) and (4) close-coupled catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \lambda_{\text{max}} = 1.1 )</td>
<td>( m_{\text{NOx, out}} = 0.69 \text{ g} )</td>
<td>(based on Euro-4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( m_{\text{PM, out}} = 0.79 \text{ g} )</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( u_{\text{in}} = [\theta(t), \theta_{\text{int}}(t)] )</td>
<td></td>
<td>( m_{\text{CO, out}} = 22.8 \text{ g} )</td>
<td>(2) and (4) close-coupled catalyst</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( m_{\text{NOx, out}} = 1.29 \text{ g} )</td>
<td>(based on Euro-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( m_{\text{PM, out}} = 1.58 \text{ g} )</td>
<td></td>
</tr>
</tbody>
</table>

\( m_{\text{CO, out}} \) and \( m_{\text{NOx, out}} \) are the emissions of CO and NOx, respectively. \( m_{\text{PM, out}} \) is the PM emission.
proximity of the recorded best control policy from the previous
iteration. The range of input values tested is reduced with each
iteration, making the discretisation of the inputs more refined. A
typical input range contraction factor used was 0.85.

Whilst these modifications to dynamic programming can sig-
nificantly reduce its computational requirements, use of finely
resolved time grids can require a large number of simulations,
which can be time demanding. Use of rough grids, however, limits
the frequency at which the optimised control inputs can be varied,
and as a result, some higher frequency characteristics of optimal
trajectories, such as rapid spark timing switching or air–fuel ratio
pulsation, might be missed. In this paper temporal resolutions of
\( \Delta t = 20 \text{ s} \) and \( \Delta t = 10 \text{ s} \) are used across the 400 s drive cycle
duration considered. Whilst these resolutions limit the ability of
the solution to capture higher frequency information, they also
enable more general trends in the engine control setpoints to be
established.

4. Experimental methods

All model calibration and validation work, as well as imple-
mentation of control schemes were carried out on a 4 l Ford BF
engine and a Horiba-Schenck Titan 460 kW transient dynam-
ometer. The exhaust system comprised of a cast iron exhaust
manifold, a connecting pipe and an aged catalyst (75 h Ford 4-
mode schedule, equivalent to roughly 80,000 km of road driving).
A photo of the test rig is presented in Fig. 2.

Normalised air–fuel ratio measurements were taken using
Bosch LSU 4.9 wide-band sensor. If the engine was required to
operate at a non-stoichiometric air–fuel ratio, the sensor was
utilised in the feedback loop of a PI controller, adjusting the
injection duration. Fuel flow measurement was performed using
AVL KMA 4000. Exhaust gas composition results presented were
obtained from a Horiba vehicle certification grade emissions
bench.

5. Simulation and experimental results

Solutions to (7) are developed for each of the cases listed in
Table 1. The results are explicitly validated using experimental
data whenever possible, whilst trends in the optimised trajectories
are identified and examined.

5.1. Case 1: spark timing solution for an under-floor catalyst

5.1.1. \( \Delta t = 20 \text{ s} \) grid

The dynamic optimisation problem (7) was solved using an
evenly spaced temporal grid of \( \Delta t = 20 \text{ s} \) resolution, requiring
roughly a day of computation on a typical desktop PC. The solution
was specified in terms of spark timing offset relative to a produc-
tion control strategy previously recorded over the same driving
conditions, which enabled to consider this production strategy as a
reasonable initial guess of the control solution in the first iteration
of iterative dynamic programming. The resulting optimised spark
timing is presented in Fig. 3 along with the constraints (7g).

---

**Fig. 2. Test rig.**

**Fig. 3.** Spark timing strategy optimised using \( \Delta t = 20 \text{ s} \) grid under Euro-3 constraints for the under-floor catalyst.
This trajectory was implemented on the engine. Simulated and measured fuel consumption and tailpipe emissions are shown in Fig. 4, demonstrating reasonable agreement. Note that the observed differences are not only the result of the modelling assumptions used, but also the consequence of imprecise control policy implementation, caused primarily by phasing errors between the spark timing strategy and the prescribed engine torque and speed trajectories.

The spark timing policy is characterised by a period of initially significant retard relative to MBT timing, where it is in close proximity of the retard constraint. This is followed by a transition to near MBT timing. The maximum indicated thermal efficiency is observed at the MBT spark advance. Thus, retarded ignition leads to a reduction in the efficiency. Under such conditions, more heat is rejected with the exhaust, which causes the exhaust enthalpy to increase and enables the catalyst warm-up time to be reduced. This strategy also tends to reduce NO and HC emissions from the engine, which provides additional benefits.

During warm idle events the ignition is again retarded with the intention to operate within the calibrated region of the engine model, whilst minimising fuel use. It is therefore not clear at this stage whether use of MBT ignition is beneficial during these periods. However, as warm idle events are marked by low fuel usage, optimisation results involving an exhaustively calibrated engine model are expected to be comparable.

Simulated and measured fuel consumption, as well as precatalyst (feedgas) and post-catalyst (tailpipe) emissions are presented in Fig. 4. While catalyst light-off is observed roughly 60 s after engine start, Fig. 3 shows that the optimised spark timing remains retarded from MBT until approximately 130 s. The additional influx of heat appears to be required to raise the catalyst temperature further, in order to achieve higher pollutant conversion efficiencies during the acceleration event at roughly 120 s. Near MBT operation later in the cycle facilitates better fuel economy at the expense of reduced exhaust enthalpy input into the catalyst. However, by that stage the catalyst is sufficiently warm to maintain a working temperature from the exothermic reactions taking place in the washcoat.

By the end of 400 s, cumulative HC emissions are almost at the Euro-3 limit, suggesting that further reduction in fuel consumption may be limited by the emissions constraints. This tradeoff therefore emphasises the opportunity to improve fuel economy if the emissions are below the limits.

5.1.2. $\Delta t = 10$ s grid

To test whether the solution developed in Section 5.1.1 is time grid independent, the spark timing strategy was reproduced on a grid with intervals $\Delta t$ fixed at 10 s. This required roughly 3 days...
of computation on a typical desktop PC, approximately 3 times longer than with $\Delta t$ set to 20 s. The resulting spark timing policy, demonstrated in Fig. 5, is characterised by a more significant retard in the initial phase of the cycle, and a much sharper and earlier transition from highly retarded to near MBT spark advance. The benefit in terms of the overall fuel consumption is roughly 1.4% with respect to the strategy with $\Delta t = 20$ s.

5.1.3. Validation of local optimality

To test the spark timing solution, the optimised spark timing was offset by $-10^\circ$, $-5^\circ$, $5^\circ$ and $10^\circ$. The perturbed control strategies were implemented on an engine and the results are shown in Fig. 6, where the offset of $0^\circ$ corresponds to the optimised trajectory. While the trajectories with retarded spark timing appear to satisfy the emissions constraints, they result in an overall fuel consumption increase. Conversely, trajectories with more advanced spark timing improve the fuel economy, but violate the hydrocarbon limits. The minimum achievable fuel consumption, that allows all the emissions constraints to be satisfied, corresponds to an offset of $0^\circ$. Therefore, the experiments confirm the local optimality of the control policy. Of course, this is not an exhaustive evaluation of optimality, which would require an impractically large number of experiments.

5.1.4. Switching result

As the results of Sections 5.1.1 and 5.1.2 appear to be converging towards a bang–bang type policy for spark timing, there is interest in investigating the merit of such a policy directly. The potential benefits of exclusively searching for a switched policy is that the search space is significantly reduced, with only the switching times required. This results in an obvious speed up in developing the solution.

The maximum retard from MBT is prescribed by drivability constraints, leading to a policy defined exclusively in terms of the switching time $t_{sw}$ in the following form:

$$\theta_{sw}(t_{sw}, t) = \begin{cases} \max(-10^\circ, \theta_{MBT}(t)-30^\circ) & \text{for } t < t_{sw}, \\ \min(50^\circ, \theta_{MBT}(t)) & \text{for } t \geq t_{sw}. \end{cases} \quad (9)$$

Note that such spark timing automatically satisfies ECU and drivability constraints (7f) and (7g). Prior to time $t_{sw}$, the strategy is specified by the most retarded spark timing possible under these constraints. It is then switched to approach MBT timing. The optimisation problem is reformulated as

$$u_{sw} = [\theta_{sw}(t_{sw}, t)], \quad (10)$$

$$t_{sw}^* = \arg \min_{t_{sw}} J(t_{sw}, u_{sw}), \quad (11)$$

and solved, yielding

$$t_{sw}^* = 61.3 \text{ s}. \quad (12)$$

Fig. 7 shows the effect of the switching time $t_{sw}$ on the overall fuel economy and tailpipe emissions as simulated by the model. Whilst earlier switching tends to improve the fuel economy, it also appears to violate cumulative tailpipe emissions limits, as less heat is made available for the catalyst. Later switching causes more heat to be rejected into the exhaust, which generally reduces overall tailpipe emissions, but increases fuel consumption. The switching time $t_{sw}^*$ is therefore a balanced compromise between cumulative tailpipe emissions and fuel economy.

The performance of the switching strategy appears to approach that of the policy from Section 5.1.2, with the minimum value of the cost function $J$ falling within 0.4% of the result produced using iterative dynamic programming. However, only minutes of computation were required to solve this static optimisation problem, as opposed to days in the case of Section 5.1.2. Furthermore, the approximate solution might lead to real-time implementable policies, with the switching time specified, for example, by representative temperatures or species in the exhaust.
5.2. Case 2: spark timing solution for a close-coupled catalyst

In Section 5.1 it was demonstrated that as far as the cost function $J_s$ is concerned, there are small differences between the results of optimisation using $\Delta t = 20$ s and $\Delta t = 10$ s. Therefore, to reduce excessive computational requirements and to ensure consistency between all of the results, all further studies are limited to $\Delta t = 20$ s grids. For this case roughly 1 day was required for the development of the solution on a typical desktop PC.

Fig. 8 demonstrates the spark timing solution for the closely coupled catalyst. The duration of the initial spark timing retard is reduced relative to the optimised control policy for the under-floor catalyst. This is because close-coupling avoids some of the heat losses that occur for the under-floor catalyst, in the exhaust system between the engine manifold and the catalyst, and so can heat the catalyst with less spark retard.

As seen in Fig. 9(a), the spark timing is heavily retarded during the first idle event. This increases the mass flow rate of the exhaust and the catalyst inlet gas temperature, which helps to bring the catalyst to its operating temperature quicker.

Fig. 9(b)–(d) indicates that while the new control strategy results in lower overall engine-out CO emissions, engine-out NO and HC emissions are significantly increased. However, as the catalyst is exposed to higher inlet exhaust enthalpy, its conversion efficiency is enhanced, which results in very similar cumulative
tailpipe emissions to the optimisation case with the under-floor catalyst. The time to catalyst light-off remains almost fixed at 60 s for the two exhaust systems.

5.2.1. Switching result
As in Section 5.1.4, approximating the spark timing solution by a switching strategy and optimising the switching time yielded
t_sw = 32.5 s.

The minimum value of the cost function J_s agreed closely with the iterative dynamic programming result, which is consistent with earlier findings. As previously, only minutes of computation were required.

5.3. Case 3: spark timing and $\lambda$ solution for an under-floor catalyst

The solution to Case 3 was developed using $\Delta t = 20$ s time discretisation, requiring roughly 3 days of computation on a typical desktop PC. It is presented in Fig. 10. Simulated emissions are compared against single spark timing optimisation results of Section 5.1.1 in Fig. 11.

The $\lambda$ control policy is initially lean and spark timing is retarded to help reduce engine-out CO and HC emissions, whilst maintaining similar levels of NO. In later parts of the cycle, however, $\lambda$ is characterised by multiple switches, and the engine operates lean during most of the time. This compromises the NO conversion efficiency in the hot catalyst in favour of the fuel economy and lower tailpipe CO and HC emissions, such that both NO and HC emissions fall onto the Euro-3 limits by the end of the 400 s of NEDC conditions considered.

The trends in the spark timing strategy closely resemble those of Section 5.1. It is characterised by an initial period of significant retard, transition towards MBT, and then near MBT operation. With this strategy catalyst light-off occurs roughly at 60 s, whilst the spark timing is retarded significantly from the MBT setpoint for another 70 s. This provides additional heat for the catalyst, increasing its conversion efficiency during some of the later higher power events.

The overall fuel consumption saving relative to the single parameter spark timing optimisation of Section 5.1.1 is approximately 6.1%, which is achieved at the cost of increased cumulative NO tailpipe emissions.

5.3.1. Switching result
Fig. 10 suggests that a bang–bang strategy with a single switch is inappropriate for approximating the $\lambda$ trajectory. Approximating $\lambda$ by a trajectory with multiple switches is computationally more involved and is not considered in this paper. Consequently, if only the spark timing solution is approximated by a switching strategy, whilst the $\lambda$ trajectory is fixed to the iterative dynamic programming solution, the optimised switching time evaluates to
t_sw = 57.3 s.

The minimum value for the cost function $J_s$ is then within 2% of the iterative dynamic programming result, which is consistent with previous observations.

5.4. Cases 4 and 5: spark timing and $\lambda$ solution for a close-coupled catalyst

Solutions to (7) for Cases 4 and 5 are presented in Fig. 12 and have taken each roughly 3 days to generate on a typical desktop PC. The spark timing strategies differ by almost a constant crank-angle during the first 60 s. Consequently, fuel consumption resulting from the Euro-4 strategy is 3.7% higher than of the Euro-3 strategy, as indicated in Fig. 13(a). The small differences in the $\lambda$ trajectories have minor effects on the fuel economy.

Fig. 13(b)–(d) demonstrates modelled cumulative emissions. As previously, whilst all of the cumulative tailpipe emissions constraints are fulfilled, both NO and HC emissions lie on the...
Apart from the reduced catalyst warm-up time and its enhanced conversion efficiency, another major contribution towards meeting the Euro-4 constraints comes from the reduction of engine-out emissions, partly caused by a more significant spark retard early in the cycle.

5.4.1. Switching result

Consideration of a switching spark timing trajectory, with the \( \lambda \) trajectory set to the solution of iterative dynamic programming, results in

\[
\tau_{sw}^* = 8.0 \text{ s} \quad \text{and} \quad \tau_{sw}^* = 35.5 \text{ s}
\]  

with respect to Euro-3 and Euro-4 constraints respectively. In both of these cases the minimum values of cost functions \( J_i \) are within 1% of iterative dynamic programming results.

---

**Fig. 10.** Spark timing and \( \lambda \) control strategies optimised under Euro-3 constraints for the under-floor catalyst.

**Fig. 11.** Modelled cumulative (a) fuel consumption, and (b) CO, (c) NO and (d) HC emissions using spark timing, and spark timing and \( \lambda \) strategies for an under-floor catalyst.
5.5. Case 6: spark and cam timing solution for a close-coupled catalyst

The study examines the benefits of including cam timing in the dynamic optimisation. The solutions to Cases 2 and 6 are compared in Fig. 14. The solution to Case 6 took roughly 3 days to generate on a typical desktop PC.

The optimised spark timing strategies from the single and two parameter optimisation studies agree closely during most of the drive cycle. Some of the most significant differences in the spark timing (and cam timing) are observed between 80 and 120 s. However, Fig. 15 shows that neither fuel economy nor emissions are greatly affected, as these instances are characterised by a low fuel flow. Simulated cumulative tailpipe NO emissions increase only slightly due to the increase in the respective engine-out emissions, but remain below the specified emissions limits.

The improvement in overall fuel economy of only 1.3% over the single parameter spark timing optimisation result and reasonable
agreement of the optimised and production cam timing trajectories suggest that the ECU cam control strategy in this engine is already near the expected optima.

The performance of the control policies developed are compared in Table 2 in terms of the fuel consumption saving relative to the Case 1 iterative dynamic programming result with $\Delta t = 20$ s. Subscripts $sw$ and $idp$ identify switching trajectories and those derived using iterative dynamic programming respectively. As a guide, the 400 s period considered in the optimisation accounts for approximately 30% of the fuel consumed during the entire NEDC drive cycle.

6. Conclusion

This paper presented a methodology for minimising the fuel consumption of a gasoline fuelled vehicle during cold starting and subject to cumulative emissions constraints. This problem was
solved numerically using a validated transient model of the engine and exhaust system reported in a previous study (Andrianov et al., 2012).

Including the model calibration, this methodology should generate optimised results significantly more quickly than many other approaches in the open literature. One and two parameter optimisations using iterative dynamic programming were achievable on a desktop PC within 1–3 days, which suggests that parallelisation would enable both higher dimensional optimisation and finer temporal resolution in practical time frames. Single switching control policies were also proposed, and these were found to perform close to the optimised strategies obtained from the iterative dynamic programming but with far less computational effort.

The methodology also quantified how the physics in this problem imposed limits on the achievable fuel consumption. These were as follows, all which are consistent with industrial practice.

- There was a trade-off between the lowest achievable fuel consumption and the emissions standard achievable with a given engine and exhaust system design. This is an important trade-off, since catalysts (and in particular their precious metal loading) are a significant cost that needs to be justified in terms of total vehicle performance.
- Close coupling a given catalyst reduces fuel consumption for a given emissions standard. This is because the reduced heat losses between the engine and the catalyst enables the engine to operate more efficiently for a greater part of the drive cycle, whilst maintaining high enough enthalpy exhaust for acceptable catalyst performance.
- Multiple parameter optimisation yielded significant fuel consumption savings when spark timing and air–fuel ratio were used. This primarily arose from frequent use of lean mixtures throughout the drive cycle conditions considered. The inclusion of cam timing did not yield significant benefits, which was likely due to the production cam control strategy being already well optimised.

**Acknowledgements**

This research was supported by the Advanced Centre for Automotive Research and Testing (ACART, www.acart.com.au), the Ford Motor Company of Australia and Australian Research Council Grants FT0991117 and FT100100538.

**References**


