An Economic Nonlinear Model Predictive Control Strategy for SI Engines: Model-Based Design and Real-Time Experimental Validation

Qilun Zhu, *Member, IEEE*, Simona Onori, *Senior Member, IEEE*, and Robert Prucka

***Abstract*— This paper proposes a model predictive torque control strategy for spark-ignition engines with external exhaust gas recirculation. The proposed economic nonlinear model predictive controller (E-NMPC) tries to minimize fuel consumption, given an indicated mean effective pressure (IMEP) tracking reference and abnormal combustion constraints such as knock and combustion variabilities. A nonlinear optimization problem is formulated and solved in real time using sequential quadratic programming (SQP) to obtain the desired control actions. The SQP utilizes active set quadratic programming (QP) algorithms, with warm-start techniques that exploit the structural similarities between successive sub-QPs along the SQP sequence. This process reduces QP iterations by approximately 60% for each SQP update. Simulation results demonstrate that the proposed model predictive controller can track the IMEP reference with an rms error of 1.1% for engine cycles without active combustion constraints. When the IMEP reference conflicts with constraints, the SQP E-NMPC can efficiently find close-to-optimal control actions that are similar to those from off-line feed-forward calibration. The proposed algorithm is validated on an engine dynamometer. The algorithm executes in a prototype engine controller with a mean computation time of 1.07 ms, proving its feasibility for future engine control unit implementation.**

***Index Terms*— Combustion constraints, economic nonlinear model predictive control (MPC), exhaust gas recirculation (EGR), experimental results, indicated mean effective pressure (IMEP) tracking, sequential quadratic programming (SQP), spark-ignition (SI) engine, suboptimal solution.**

NOMENCLATURE

A_e	Effective area.
A_{flow}	Effective flow area during valve overlap.
AFR	Air-to-fuel ratio.
$CA50$	Crank angle of 50% mass fraction burned.
C_D	Discharge coefficient.
c_{eg}	Cylinder exhaust gas fraction.
c_p	Constant pressure heat capacity.
COV	Coefficient of variation.
COV_{ub}	Upper bound of COV.
e	Slope factor.
ECU	Engine control unit.
EGR	Exhaust gas recirculation.
$E-MPC$	Economic model predictive control.
$E-NMPC$	Economic nonlinear model predictive control.
EVC	Exhaust valve closing.
η_V	Volumetric efficiency.
γ	Ratio of heat capacity.
IC	Internal combustion.
$IMEP$	Indicated mean effective pressure.
IVO	Intake valve opening.
KI	Knock intensity.
KI_{ub}	Upper bound of KI.
λ	Lagrange multiplier.
LHV	Low heating value.
LPV	Linear parameter variant.
m_α	Total air mass flowing through the throttle per engine cycle.
$m_{\beta_{air}}$	Total air flow into the cylinder per engine cycle.
$m_{\beta_{eg}}$	Total exhaust flow into the cylinder per engine cycle.
m_ε	Total flow through the EGR valve per engine cycle.
MAF	Mass air flow.
MAP	Manifold absolute pressure.
MPC	Model predictive control.
m_m	Intake manifold mass.
NLP	Nonlinear programming.
$NMPC$	Nonlinear MPC.
ODE	Ordinary differential equation.
OLV	Overlap volume between EVC and IVO.
ω_e	Engine speed.
P_0	Pumping effective pressure.
P_a	Ambient pressure.
P_{exh}	Exhaust pressure.
P_f	Fuel effective pressure.
P_i	Orifice input pressure.
P_m	Intake manifold pressure.
P_o	Orifice output pressure.
QP	Quadratic programming.
R	Gas constant.
RGM	Residual gas mass.
SI	Spark ignition.
σ_0	Stoich AFR.

Manuscript received April 15, 2017; revised September 13, 2017; accepted October 3, 2017. Manuscript received in final form October 30, 2017. Recommended by Associate Editor Y. Wang. *(Corresponding author: Qilun Zhu.)*

The authors are with the Clemson University International Center for Automotive Research, Greenville, SC 29607 USA (e-mail: qilun@g.clemson.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCST.2017.2769660

$SPKT$	Spark timing.
SQP	Sequential QP.
T_a	Ambient temperature.
τ	Time constant.
TDC	Top dead center.
T_{exh}	Exhaust temperature.
T_i	Orifice input temperature.
θ_{EGR}	EGR valve angle.
θ_T	Throttle angle.
ϑ	Ratio of heat transfer to coolant.
T_m	Intake manifold temperature.
V_c	Cylinder clearance volume.
V_d	Engine displacement.
V_m	Intake manifold volume.

I. Introduction

THE control objectives of an IC engine management system are to deliver demanded engine torque while minimizing fuel consumption and preventing abnormal combustion phenomena. These control objectives favor the application of model-based optimal control strategies. While many articles discuss the possibility of applying MPC to regulate engine torque output [1]–[4], minimal focus has been given to the minimization of fuel consumption and meeting combustion constraints. The most important factor causing this dilemma is that the engine models used for MPC torque control are heavily simplified to allow reasonable computation time of the optimal control actions. However, this model simplification process affects optimality of the control actions obtained by the MPC. This paper proposes an E-NMPC strategy for SI engines with external EGR. The strategy is able to utilize complex engine models to find the optimal control actions while achieving target IMEP, reducing fuel consumption, and meeting abnormal combustion constraints.

Control-oriented engine air path and torque generation models are well established (see [5]). Most of these models are constructed in the time domain, making them favorable for controllers with fixed sampling time. The most important drawback of this approach is that the IC engines are inherently discrete event systems with cyclic operational characteristics. The SI engine system is modeled and controlled in the engine cycle domain for this paper. While this approach agrees with the discrete nature of both MPC and IC engines, it also benefits from the fact that most control-oriented combustion models were constructed in the engine cycle domain [6]–[8]. This makes it convenient to impose abnormal combustion constraints such as knock and combustion stabilities during the calculation of optimal control actions.

This paper reveals that the nonlinear optimization problem is not convex for IMEP control of SI engines with external EGR, leading to the challenge of dealing with multiple local minima issues. Global NLP solvers, such as dynamic programming and particle swarm, can be employed for MPC applications [9], [10]. Stability of tracking and E-MPC with global optimal solutions was shown in [11] and [12]. However, these global NLP solvers require numerous evaluations of the system model, which are not feasible for engine control applications with fast update frequencies. Most model predictive engine control researchers have selected suboptimal strategies to reduce the computational demand [13]–[16]. LPV MPC is a widely adopted suboptimal predictive controller for nonlinear systems tracking problems [17], [18]. The validity of LPV MPC is based upon the assumption that the system behavior remains linear like if the system states are in the neighborhood of the nominal point of linearization. For SI engines, the COV of IMEP and knock constraints are highly nonlinear, which causes significant error if approximated linearly. Finding the nominal point of linearization requires high calibration effort to ensure that they are equilibrium points and economically optimal. Although closed-loop stability was established for some types of E-MPC [19], [20], a general stability criterion for suboptimal E-NMPC has not yet been developed. This paper identifies a Lyapunov function for the proposed suboptimal E-NMPC application, proving the closed-loop stability of the investigated engine control system.

SQP is a continuous NLP algorithm based on Newton's method. Previous research investigated the possibility of applying SQP to NMPC [15], [21], [22]. The most important advantage of SQP is that it transforms complex NLP problems into a sequence of sublevel QP problems (hence the name). The sub-QP problems can be solved efficiently with algorithms based on active set methods. As a result, the original nonlinear objective and constraint functions are only evaluated before the sub-QP (to compute the Hessian and Jacobian of the NLP problem), saving significant computation time compared to other NLP solvers. This characteristic is advantageous for engine control applications since the evaluation of most high-fidelity models requires a significant amount of time due to their complex structure with multiple calibration maps and ODEs.

Conventional SQP algorithms designed for general optimization purposes (see [27]) may not be favorable for specific real-time MPC applications due to the heavy computational load in computing the Hessian matrix for complex or implicit system models. In this case, numerical differentiation methods are necessary (e.g., algorithmic differentiation and finite difference). In practice, the Hessian is often approximated with the first-order derivatives to reduce computational burden. Quasi–Newton methods are well discussed in [23] and [25]. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) rank-two update method is widely used in SQP applications [26]–[28]. The computational load required to approximate the Hessian prevents the BFGS method from being utilized for MPC with fast sampling. Quirynen *et al.* [15] proposed an approach based on algorithmic differentiation to calculate the exact Hessian. This method significantly reduces the computation effort compared with BFGS Hessian approximation. However, the exact Hessian of the system is not necessarily semipositive definite. This paper proposes to use the mirrored Hessian instead of the exact Hessian for sub-QP formulation [15]. The effect of using the mirrored Hessian on SQP convergence is not theoretically addressed nor numerically discussed. The use of nonconvex QP algorithms with nonpositive definite Hessian SQP is discussed in [28], where it is shown to be not efficient enough for most MPC applications. This paper exploits the Gauss–Newton-like structure of the investigated E-NMPC to

simplify the computation of Hessian and Jacobian matrices. Although numerical real-time linearization of the nonlinear engine model induces some round-off error, the SQP preserves most of the nonlinearities of the system.

For NLP problems with multiple local optimal solutions, the search step is computed by a standard SQP iteration that can be very aggressive, making the SQP converge to an undesirable local optimal solution. In some extreme cases, the SQP cycles between local optima without converging (see [27, pp. 292, 293]). For MPC applications, this behavior increases computational load and causes undesirable control chattering. A merit function-based approach can be applied to enhance local convergence of the SQP [27], [29], [32]. This paper exploits this technique to ensure the SQP converges to the first local optimal solution along the search direction. This method also guarantees that the merit function is strictly decreasing as the SQP progresses, which can be utilized to prove the closed-loop stability of the proposed E-NMPC.

This paper expands the scope and details of [35]. The two-layer supervision control structure is explained. Modeling of IMEP generation is further discussed with some additional details to enhance the derivation of the final nonlinear state-space engine model. Convex analysis of the E-NMPC problem shows the existence of multiple local optimal solutions of the real-time NLP problem. While global optimization algorithms are not considered feasible for the real-time application, this paper utilizes a computationally efficient SQP algorithm, which finds the local optimal solution of the NLP problems. However, the existence of multiple local minima leads to potential issues with control chattering and stability. Warm start and merit function techniques are introduced to ensure closed-loop stability and eliminate control chattering. Finally, the experimental setup and validation results are provided to substantiate the performance and computational efficiency of the proposed E-NMPC engine controller. Effects of different horizon lengths on performance and stability are investigated and demonstrated in this paper. Discussion of E-NMPC performance for a range of engine speeds is also presented.

This paper is organized as follows. Section II introduces the control-oriented engine model. Section III formulates and analyzes the NLP solver for desired control actions. Section IV discusses the proposed SQP MPC strategy. Section V investigates two active set QP algorithms with warm-start techniques and Section VI provides simulation and experimental results. Finally, Section VII concludes the contribution of this paper highlighting possible future extensions.

II. Control-Oriented Nonlinear Engine Model

This paper focuses on IMEP control of SI engines with external EGR. The fuel injection control is assumed to maintain a stoichiometric AFR, maximizing catalyst efficiency. The manifold temperature is assumed to be constant since the EGR is cooled with a heat exchanger. Finally, the air mass in the air-path system is considered uncompressible. Fig. 1 shows the engine configuration with labels of sensors, control, and modeling variables.

In the engine cycle domain, cylinder air charge per cycle can be computed according to intake manifold air density and

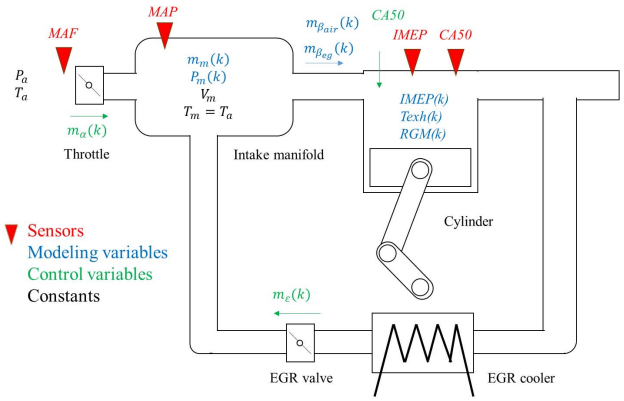


Fig. 1. Block diagram of the SI engine system with external EGR.

volumetric efficiency (for a given engine manifold pressure and engine speed) as

$$m_{\beta_{\text{air}}} = \frac{m_{m\text{air}}}{V_m} V_d \eta_V(P_m, \omega_e) \tag{1}$$

where

$m_{m_{\text{air}}}$	air mass in the intake manifold;
$\eta_V(P_m, \omega_e)$	volumetric efficiency;
V_d	engine displacement;
V_m	intake manifold volume;
P_m	intake manifold pressure;
ω_e	engine speed, treated as a varying parameter in this paper.

The air mass balance of the intake manifold can be expressed as

$$m_{m\text{air}}(k+1) = m_{m\text{air}}(k) + m_\alpha(k) - m_{\beta_{\text{air}}}(k). \tag{2}$$

Rearranging (1), and substituting it into (2) yields

$$m_{\beta_{\text{air}}}(k+1) = \frac{1}{K+1} m_{\beta_{\text{air}}}(k) + \frac{K}{K+1} m_\alpha(k) \tag{3}$$

where

$$K = \frac{\eta_V(P_m(k), \omega_e) V_d}{V_m}.$$

The cylinder EGR flow can be modeled similar to the air mass flow, as shown in the following:

$$m_{\beta_{\text{eg}}}(k+1) = \frac{1}{K+1} m_{\beta_{\text{eg}}}(k) + \frac{K}{K+1} m_\varepsilon(k). \tag{4}$$

The manifold pressure can be computed by reversing the speed density equation

$$P_m(k) = \frac{(m_{\beta_{\text{air}}}(k) + m_{\beta_{\text{eg}}}(k)) R T_m}{\eta_V(P_m(k), \omega_e) V_d}. \tag{5}$$

It can be observed from (3) to (5) that manifold dynamics are independent of engine speed (excluding the slowly varying volumetric efficiency η_V), unlike most time-domain models. In-cylinder gas composition includes air, exhaust gas, fuel, and other minor species that are neglected in this paper. The amount of air and fuel can be determined by $m_{\beta_{\text{air}}}$, assuming

stoichiometric AFR. The fuel mean effective pressure can be computed as

$$P_f(k) = \mathrm{LHV} \cdot \frac{m_{\beta_{\mathrm{air}}}(k)}{\sigma_0 V_d} \tag{6}$$

where σ_0 is stoichiometric AFR and LHV is the low heating value of the fuel.

The amount of in-cylinder exhaust gas is the summation of $m_{\beta_{\mathrm{eg}}}$ and RGM. This paper adopts the semiempirical model proposed in [37], which expands the RGM model proposed in [36]. This model separates the RGM into two parts: 1) from trapped residual at EVC due to unswept cylinder volume and 2) exhaust gas backflow into the cylinder and intake runner during the valve overlap period. After adding terms ΔP_{exh} and ΔP_m to account for wave tuning dynamics to the original Fox model, the RGM for each engine cycle can be calculated according to

$$\begin{aligned}\mathrm{RGM}(k) &= C_1 \frac{P_{\mathrm{exh}}}{RT_{\mathrm{exh}}(k-1)} V_c \\ &+ C_2 \sqrt{\frac{P_{\mathrm{exh}}}{RT_{\mathrm{exh}}(k-1)}((P_{\mathrm{exh}} + \Delta P_{\mathrm{exh}}) - (P_m + \Delta P_m))} A_{\mathrm{flow}} \\ &\times \frac{\mathrm{OLV}}{\omega_e}\end{aligned} \tag{7}$$

where

$P_{\mathrm{exh}} \approx 110$ kPa	exhaust pressure;
T_{exh}	exhaust temperature;
R	gas constant;
V_c	cylinder clearance volume;
A_{flow}	effective flow area during valve overlap period;
OLV	overlap volume which is the cylinder volume difference between EVC and IVO;
C_1 and C_2	calibration factors.

The total fraction of in-cylinder exhaust gas c_{eg} can be generated as

$$c_{\mathrm{eg}}(k) = \frac{m_{\beta_{\mathrm{eg}}}(k) + \mathrm{RGM}(k)}{m_{\beta_{\mathrm{air}}}(k) + m_{\beta_{\mathrm{eg}}}(k) + \mathrm{RGM}(k)}. \tag{8}$$

With the information of cylinder composition, the IMEP can be modeled using Willians approximation method [5]

$$\mathrm{IMEP}(k) = e(k) P_f(k) - P_0 \tag{9}$$

where e is the "slope" factor and $P_0 = P_{\mathrm{exh}} - P_m$ is pumping effective pressure.

The slope factor e is related to engine speed ω_e, $CA50$, and c_{eg}. It is the product of multiple slope factors with dependence of one or two inputs

$$e(k) = e_\omega(\omega_e) e_\zeta(CA50(k)) e_{\mathrm{eg}}(c_{\mathrm{eg}}(k), \omega_e). \tag{10}$$

Each slope factor in (10) can be approximated with a low-order polynomial function, as demonstrated by Figs. 2–4 regarding the engine used in this paper. Fig. 4 shows the contour plot of the computed e_{eg} with respect to in-cylinder exhaust gas fraction and engine speed for the calibration data. Clear correlation can be observed and modeled using 2-D polynomial functions (not shown on the plot).

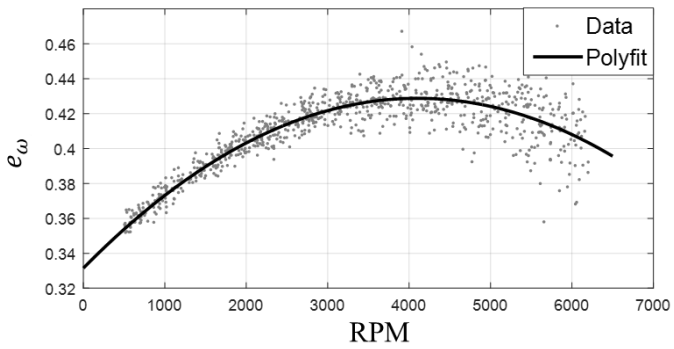


Fig. 2. Relationship between e_ω and RPM.

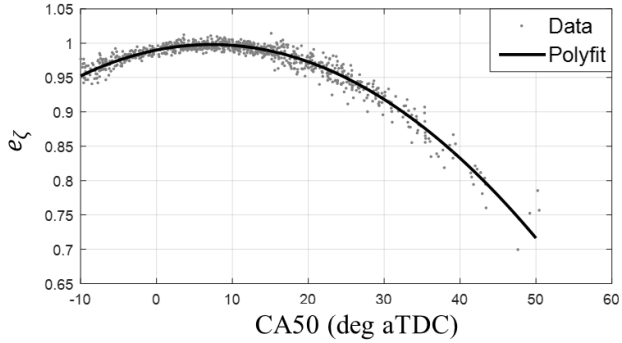


Fig. 3. Relationship between e_ζ and CA50.

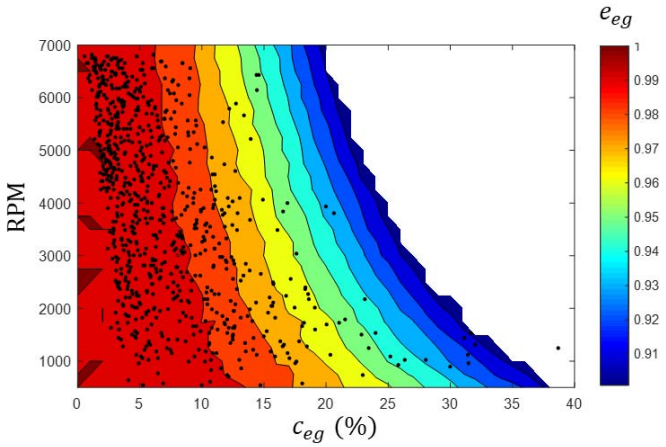


Fig. 4. Relationship between e_{eg} and c_{eg}.

An energy balance approach is utilized to calculate exhaust temperature T_{exh} for the RGM model. The IC engine transforms the chemical energy of the injected fuel into mechanical work (IMEP) and rejected heat, which is the summation of heat transfer to coolant and exhaust enthalpy. Thus the exhaust gas temperature can be calculated according to

$$T_{\mathrm{exh}}(k) = \frac{V_d(P_f(k) - \mathrm{IMEP}(k))(1-\vartheta)}{c_p m_{\beta_{\mathrm{air}}}(k)(1 + 1/\sigma_0)} \tag{11}$$

where c_p is the constant pressure gas heat capacity and ϑ is the ratio of transferred heat to coolant (in terms of the total

rejected heat). It can be estimated with engine speed, CA50, and load [39], [40].

COV of IMEP is utilized as an indication of cycle-by-cycle combustion variability. The proposed model predictive IMEP control should maintain the COV of IMEP below a certain value. The COV of IMEP is correlated with the cylinder air mass flow $m_{\beta_{\text{air}}}$ and CA90 [42]. CA90 is computed with an artificial neural network (one hidden layer and ten neurons) with CA50, RPM, and $m_{\beta_{\text{air}}}$ as inputs. KI is computed by integrating the end gas energy modeled by an Arrhenius type function in the crank angle domain [38]. The output of this model is the normalized KI, which indicates engine knock likely if KI $\geq$ 1. This model cannot be directly integrated in the E-NMPC due to its need for a crank angle-resolved combustion model. A fully empirical KI map is employed to approximate the crank angle-resolved physics-based KI model in the engine cycle domain. The inputs to this KI map are RPM, CA50, $m_{\beta_{\text{air}}}$, and RGM.

In summary, the engine is modeled as a continuous fourth-order nonlinear state-space model. Lookup tables and artificial neural networks are applied in the state and output functions to improve model accuracy. We use subscript k to replace step indicators for the convenience of discussion. The engine model can be written as

$$\begin{aligned} x_{k+1} &= f_x(x_k, u_k) \\ y_k &= f_y(x_k) \\ z_k &= f_z(x_k) \end{aligned} \tag{12}$$

where $x \in \mathbb{R}^4$, $x = [m_{\beta_{\text{air}}}, m_{\beta_{\text{eg}}}, \text{RGM}, x_{CA50}]^T$; $u \in \mathbb{R}^3, u = [m_\alpha, m_\varepsilon, CA50]^T$; $y \in \mathbb{R}^1$, $y =$ IMEP; $z \in \mathbb{R}^3$, $z = [P_m, \text{COV}, \text{KI}]^T$; and x_{CA50} is an augmented state that delays CA50 by one step.

The CA50 output of the MPC is the target value for the next engine cycle, which induces a unit step delay.

Remark: The slowest dynamics of the investigated engine system are the intake manifold filling dynamics. Therefore, the characteristic "time" τ (in engine cycles of 4π crank angle degrees) is calculated as

$$\tau \approx \frac{V_m}{\eta_V V_d} = 2 \text{ cycles with } \eta_V \approx 1. \tag{13}$$

Hence, the control and preview horizons of the MPC can be as short as two steps without significant loss of optimality and stability [2].

III. Two-Layer Supervision Control Structure

The proposed E-NMPC is designed in the engine cycle domain. The E-NMPC manipulates throttle m_α (air mass flow per cycle), EGR valve m_ε (EGR mass flow per cycle), and combustion phasing $CA50$ (crank angle at which 50% of total heat release occurs). These variables are sent to lower level controllers as tracking references. The lower level controllers have faster update frequencies than the E-NMPC to account for fast- and low-order nonlinear dynamics of the actuators. The two-layer supervision control structure (shown in Fig. 5) exploits the frequency separation between engine and actuator dynamics. The lower level controllers remove nonlinearities

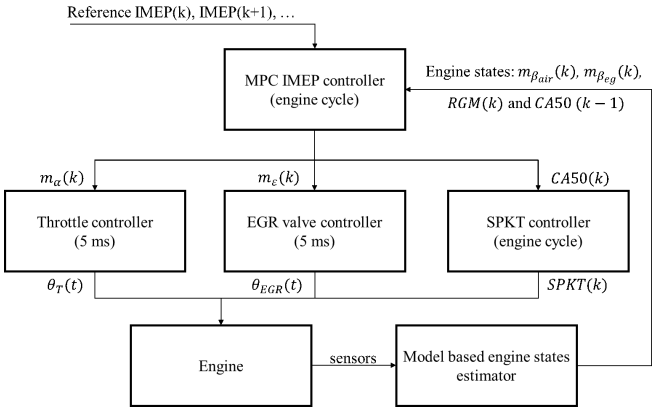


Fig. 5. Two-layer supervision engine control structure.

from the E-NMPC, significantly reducing the complexity of the NLP problem.

CA50 is an important input to many control-oriented engine models. In this paper, it is used to compute IMEP, exhaust temperature, COV of IMEP, and KI. SI engine control systems manipulate SPKT to provide desired CA50. However, the correlation between CA50 and SPKT varies with engine operating conditions. To capture this correlation, the current SI engine control systems utilize maps and combustion models [6], [7]. Both options are too computationally expensive to be implemented in the MPC. Therefore, we employed a separate SPKT controller so that the MPC can manipulate CA50 directly. This controller updates before every firing, and directly inverts factory calibrated combustion phasing maps.

Using the orifice flow model, the flow through the throttle and EGR valves can be computed as

$$\dot{m}_\alpha = C_D A_e \frac{P_i}{\sqrt{RT_i}} \begin{cases} \sqrt{\gamma \left(\dfrac{2}{\gamma+1}\right)^{\frac{\gamma+1}{\gamma-1}}}, & P_o < P_{\text{cr}} \\ \left(\dfrac{P_o}{P_i}\right)^{\frac{1}{\gamma}} \sqrt{\dfrac{2\gamma}{\gamma-1}\left[1-\left(\dfrac{P_o}{P_i}\right)^{\frac{\gamma-1}{\gamma}}\right]} & \\ & P_o \geq P_{\text{cr}} \end{cases} \tag{14}$$

and

$$P_{\text{cr}} = P_i \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma}{\gamma-1}}$$

where

$\dot{m}$	air mass flow rate through the valve;
C_D	discharge coefficient;
A_e	effective area;
P_i and P_o	pressure on input and output sides of the valve;
R	gas constant;
T_i	temperature of the input side;
γ	heat capacity ratio.

The desired valve opening is computed by inverting the above orifice flow model and sending it to the valve actuation controller every 5 ms. The time constant of valve position dynamics is approximately 10–100 ms for common automotive

butterfly valves. Frequency analysis of the throttle on the investigated engine shows 180° phase lag at approximately 40 Hz. Considering the engine cycle duration of 200–20 ms for engine speeds of 600–6000 r/min, the frequency separation assumption is valid for low to medium speed operation. For high-speed engine speed operation, the frequency difference in engine and valve dynamics diminishes. Engine cycle-based MPC should integrate the valve dynamics into the prediction process, but that is outside of the scope of this paper. As a result, the experimental validation of the proposed E-NMPC was carried out for engine speeds less than 3000 r/min.

Remark: The proposed engine IMEP control hierarchy transfers the nonlinear orifice valve flow and combustion phasing models to the fast sublevel controllers. It reduces modeling complexity of the MPC and in turn lowers computational load. Another advantage of this control structure is that feedback control and lookup tables can be applied to the sublevel controllers since they only deal with low-order nonlinear dynamics. These controllers can have faster update frequencies than the MPC has, exploiting the bandwidth of actuators. They also provide the ability to fine tune the control response to compensate for dynamics ignored by the MPC loop.

IV. Optimization Problem Formulation

The objective of the proposed model predictive IMEP control is to track an IMEP reference while minimizing fuel consumption. This determines that the stage cost of the objective function $g(x, u)$ should penalize the least-squares error of IMEP tracking and fuel consumption. The fuel consumption penalty term makes the proposed MPC problem an E-MPC (in contrast with the conventional tracking MPC). Assuming the engine operates under stoichiometric AFR, the fuel consumption can be calculated with throttle air mass flow m_α, which is the first element of u

$$g\ (x_k, u_k) = \frac{1}{2} q (y_k - (y_{\text{ref}})_k)^2 + u_k^T r \tag{15}$$

where $r = \tilde{r}[1, 0, 0]^T, \tilde{r} \in \mathbb{R}^1_{>0}$, and $q \in \mathbb{R}^1_{>0}$.

The objective function of the NLP is the summation of the stage cost over a horizon window of length N

$$J(x_k, U_k) = \sum_{i=1}^{N} \frac{1}{2} q (y_{k|k+i} - (y_{\text{ref}})_{k|k+i})^2 + u_{k|k+i}^T r \tag{16}$$

where subscript $k|k+i$ represents the prediction of $k+i$ step, while the system is at step k, and the vector U_k is

$$U_k = [u_{k|k+1}, u_{k|k+2}, \ldots u_{k|k+N}]^T \in R^{3N}.$$

After substituting the state-space equation (12) into (16), cost J becomes a function of current state x_k and future control sequence U_k. The weighting q and r can be tuned to balance between IMEP tracking performance and fuel consumption. If the model is normalized, then q and r are in the range of 0–1. Otherwise, Bryson's rule can be applied to compute default values of q and r, considering the range of y and u. It is noticed that the fuel consumption penalty makes the MPC an economic optimal controller rather than a conventional tracking MPC. However, the objective function of this specific control application still preserves the least-squares-like structure that favors Gauss–Newton methods. Section V discusses exploiting this property to reduce computation load.

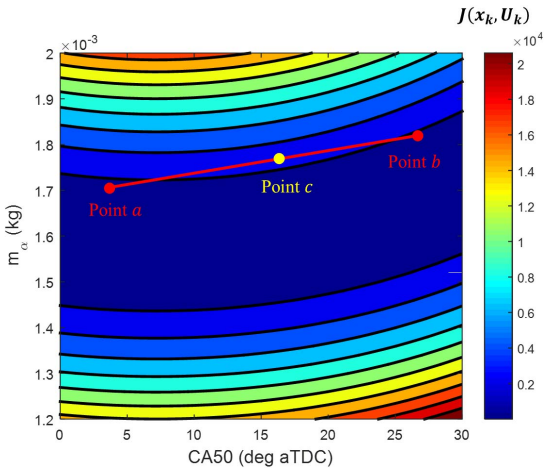


Fig. 6. Contour plot of the 2-D cost function surface, with CA50 (degrees after compression TDC) and m_α of the first step as decision variables.

The proposed model predictive engine control strategy has upper bounds on COV of IMEP (denoted as COV_{ub}) and KI (denoted as KI_{ub}). The manifold pressure has to be constrained to be less than the ambient pressure P_a since the engine is naturally aspirated. The air mass flowthrough the throttle, m_α, and EGR valve, m_ε, are also nonnegative to be physically reasonable. The CA50 is referenced to MBT and required to be later than MBT. Finally, the following equation shows the complete NLP that needs to be solved every engine cycle to obtain the optimal control sequence for the N steps of the future horizon:

$$\begin{aligned} &\min_{U(k)} \ J(x_k, U_k) \\ &\text{s.t.} \begin{cases} f_z(x_{k|k+i}) - b_z \le 0 \\ -u_{k|k+i} - b_u \le 0 \end{cases} \end{aligned} \tag{17}$$

where $i = 1, 2, \ldots N$; $b_z = [P_a, COV_{\text{ub}}, KI_{\text{ub}}]^T$ is the upper bounds on z; and $b_u = [0, 0, 0]^T$ is the lower bounds on u.

After lumping the constraint functions together

$$\begin{aligned} &\min_{U(k)} \ J(x_k, U_k) \\ &\text{s.t. } l(x_k, U_k) \le \mathbf{0} \end{aligned} \tag{18}$$

where $l : \mathbb{R}^{4+3N} \to \mathbb{R}^{5N}$ after eliminating the infinite lower bound on CA50.

This NLP problem is nonconvex with multiple local minima. In order to prove this property, *Lemma* 1 (refer to the Appendix) can be used. For any combination of given system states x_k the objective function J maps the control sequence, U, belonging to the admissible set, $\mathbb{U}^N$, into a scalar value. Without loss of generality, the control component m_ε is fixed (for example, set to zero), and the first components (the values at time k) of CA50 and m_α are considered as free decision variables. In this case, the objective function J is a 3-D surface, whose contour plots on the CA50 and m_α domain are shown in Fig. 6. One can see that the contours are nonconvex lines. In fact, the point c on the straight

line connecting the point a and b (within the same contour) belongs to a different contour with higher level. Therefore, the objective function J with all elements in U as decision variables is nonconvex with multiple local minima according to *Lemma* 1.

V. SQP Strategy With Gaussian Hessian Approximation

Finding the global optimal solution in real time is considered infeasible for engine control due to the fast update frequency. Applying suboptimal NLP algorithms that find local optimal solutions is more practical. However, closed-loop stability of the suboptimal E-NMPC needs to be proved. SQP is a computationally efficient numerical algorithm searching for the local optimal solution of problem (16). This paper tailors the conventional SQP algorithm for the proposed E-NMPC to reduce computation effort, ensuring closed-loop stability and convergence to predictable local optima.

A. Hessian and Jacobian Calculation

With a given initial guess of U (represented by U_0), the SQP computes the search direction ΔU by solving a subquadratic programming problem as follows:

$$\begin{aligned} &\min_{\Delta U(j)} \ \frac{1}{2}\Delta U^T H J_{(x_k,U_0)} \Delta U + \Delta U^T \nabla J_{(x_k,U_0)} \\ &\text{s.t. } l|_{(x_k,U_0)} + \nabla l|_{(x_k,U_0)} \Delta U \le \mathbf{0}. \end{aligned} \tag{19}$$

It is noticed that the conventional SQP algorithm utilizes the Hessian of the Lagrangian equation to account for the curvature of constraints. While the second-order derivatives of the constraints are computationally expensive, the Hessian of the objective function can be computed from linearization of the state-space model. A finite-difference approach is applied to linearize the complex engine model that consists of neural networks, lookup tables, and algebraic equations. This approach provides robustness against discontinuities (e.g., from lookup tables) and exploits the fact that most of the model is linear or can be linearized analytically.

The Hessian $H J_{(x_k,U_0)}$ can be calculated by taking the second-order derivatives of $J(x_k, U_k)$ with respect to U_k

$$\begin{aligned} H J_{(x_k,U_0)} \ &= 2\left(\left.\frac{\partial Y}{\partial U_k}\right|_{x_k,U_0}\right)^T Q \left.\frac{\partial Y}{\partial U_k}\right|_{x_k,U_0} \\ &\quad + 2\left.\frac{\partial^2 Y}{\partial U_k^2}\right|_{x_k,U_0} Q(Y_0 - Y_{\text{ref}}) \end{aligned} \tag{20}$$

where

$$Y = [y_{k|k+1}, y_{k|k+2}, \dots y_{k|k+N}]^T$$

$$Y_0 = Y(x_k, U_0), \quad Q = \begin{bmatrix} q & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & q & 0 \\ 0 & \cdots & 0 & q \end{bmatrix} \in \mathbb{R}^{N\times} N.$$

The second term on the right-hand side of (20) can be neglected if $Y_0 \approx Y_{\text{ref}}$. This assumption is reasonable if the initial guess is not far from the optimal solution. This can be achieved with warm start techniques or additional calibration effort to improve the initial guess of the optimal solution. Furthermore, this assumption becomes more reasonable as the SQP converges to the optimal solution. It is noticed that the tracking performance weighting matrix Q is positive definite. Depending on the rank of $\partial Y/\partial U$ the Hessian of the proposed MPC objective function is inherentively semipositive definite. Some active set QP algorithms are able to handle problems with semipositive definite Hessian by adding a small quadratic penalty to the search step ΔU to potentially reduce QP iterations by making the Hessian positive definite. This penalty narrows the search of local optimal solutions close to the initial guess U_0 and improves the chance of finding the first local minima along the search direction. The quadratic penalty of ΔU also provides convenience for proving the closed-loop stability of the proposed suboptimal E-NMPC, which will be discussed later in this paper. However, this penalty reduces the size of search step ΔU and decreases the convergence rate of the SQP. Therefore, the weighting s should be small. The objective function augmented with quadratic search step penalty becomes

$$\begin{aligned} J(x_k, U_k) \ &= \sum_{i=1}^{N} \frac{1}{2} q (y_{k|k+i} - (y_{\text{ref}})_{k|k+i})^2 + u_{k|k+i}^T r \\ &\quad + \Delta u_{k|k+i}^T s \Delta u_{k|k+i} \end{aligned} \tag{21}$$

where $s \in \mathbb{R}_{>0}^{3\times 3}$.

It is noted that $u_{k|k+i} = \Delta u_{k|k+i} + u_0$. After dropping the constant term $u_0^T r$ that does not influence the optimal solution, the objective function becomes

$$\begin{aligned} J(x_k, U_k) \ &= \sum_{i=1}^{N} \frac{1}{2} q (y_{k|k+i} - (y_{\text{ref}})_{k|k+i})^2 \\ &\quad + \Delta u_{k|k+i}^T r + \Delta u_{k|k+i}^T s \Delta u_{k|k+i}. \end{aligned} \tag{22}$$

The following equation shows the Hessian calculation of the augmented objective function:

$$H J_{(x_k,U_0)} = 2\left(\left.\frac{\partial Y}{\partial U}\right|_{x_k,U_0}\right)^T Q \left.\frac{\partial Y}{\partial U}\right|_{x_k,U_0} + S \tag{23}$$

where

$$S = \begin{bmatrix} s & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s \end{bmatrix} \in \mathbb{R}_{>0}^{3N\times 3N}.$$

The Jacobian of the objective function J can be computed as

$$\nabla J_{(x_k,U_0)} = -2 \left.\frac{\partial Y}{\partial U}\right|_{x_k,U_0} Q(Y_{\text{ref}} - Y_0) + R \tag{24}$$

where

$$R = [r, r, \dots r]^T \in R^{3\times N}.$$

The Jacobian of the constraint function l can be generated in a similar fashion. $\partial Y/\partial U$ is calculated using a finite-difference linearization approach, which provides a robust solution to

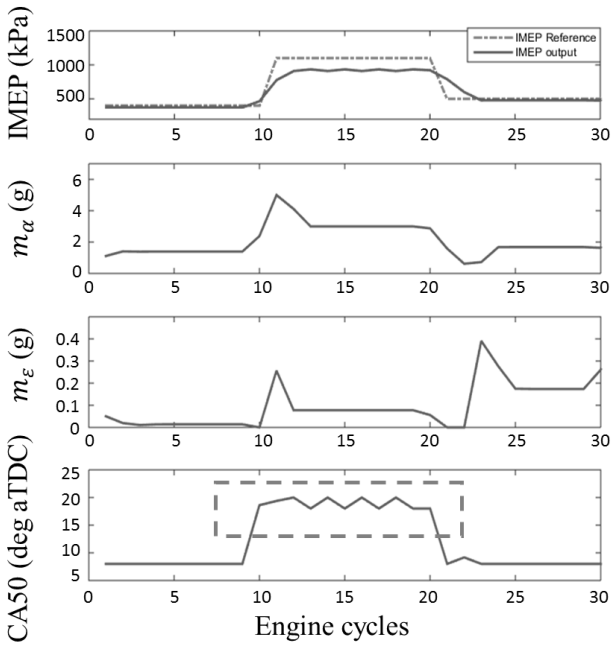


Fig. 7. Steady-state control chattering (CA50 of 10–20 engine cycles) of the SQP MPC without merit function step scaling.

differentiate general first principle models like lookup tables and nonphysics-based models. Furthermore, the computation load of the finite-difference approach is acceptable for the investigated system, since the manifold dynamics are linear and known. For the proposed control strategy $\partial Y/\partial U$ is fully computed with $4 \times N + 1$ number of engine model evaluations.

Remark: The objective function of the NLP can be considered as the summation of a nonlinear least square function and a linear function. The linear function, accounting for the economical penalty, does not affect the Hessian of the NLP. Therefore, the Hessian and Jacobian computations for the proposed E-NMPC are similar to the Gauss–Newton approach. Both the Hessian and the Jacobian can be generated by linearizing the system model at every step along the preview horizon. The computational load of this approach is significantly less than solving the exact Hessian numerically. The augmentation of the quadratic penalty of ΔU makes the proposed approach similar to the Levenberg–Marquardt (LM) method, with differences in the Jacobian calculation to account for the economical penalty. The LM method is widely used to solve least-squares problems in engineering due to its robustness and fast convergence rate.

B. Merit Function Technique

For systems with high nonlinearities, the full search step ΔU calculated from the sub-QP can be so large that the algorithm may miss some local minima along the search direction. This situation reduces the convergence rate of SQP, and can cause the "cycling" phenomenon (example can be found in [27, pp. 292–298]). From the perspective of MPC applications, this situation increases computation load and causes unnecessary control chattering (CA50 of 10–20 engine cycles in Fig. 7). The merit function approach is often applied to solve the global convergence issues of SQP [27], [32].

The search step size toward direction ΔU is scaled by a factor α which is generated by solving a line search problem of a merit function of the original NLP. For each major iteration, j (whereas the iterations solving the sub-QP problems are referred to as minor iterations), the updated solution is calculated as

$$U_k^* = U_0(j+1) = U_0(j) + \alpha \Delta U(j). \tag{25}$$

Let us define a merit function $V(x_k, U) : R^4 \times R^{3N} \rightarrow R^1$ such that

$$V(x_k, U_k) = J(x_k, U_k) + \sigma \sum_{i=1}^{q} \max(0, l_i(x_k, U_k)) \tag{26}$$

where q is the total number of inequality constraints and $\sigma \in \mathbb{R}^1$ is tuning parameter that penalizes the constraints violation.

The scaling of the search step is obtained by solving the following 1-D search problem

$$\begin{aligned} &\alpha(i) = \arg V(x_k, U_0 + \alpha \Delta U) \\ &\text{s.t. } 0 \le \alpha \le 1. \end{aligned} \tag{27}$$

While multiple line search algorithms can be applied to solve (27), this paper fixes the starting point of α as zero and the search direction as $\alpha \rightarrow 1$. If the search cannot find any $\alpha \in [0, 1]$ that reduces the merit function value the SQP converges to U_0 and $\alpha = 0$. This approach guarantees that the SQP converges to the local optimal solution that is closest to the starting point U_0 along the search direction. Furthermore, the value of the merit function g is strictly decreasing as SQP progresses.

The SQP is converged if the search step $\alpha \Delta U(j)$ is smaller than a certain threshold. In this situation, the algorithm terminates outputting U^* as the final solution. The rest of this section discusses the computation of Hessian $HJ_{(x_k, U_0)}$ and Jacobians $\nabla J_{(x_k, U_0)}$ in (19) and the initialization of the SQP and merit function techniques.

C. Warm-Start Strategy and Stability

This section will introduce a warm-start technique that is able to guarantee closed-loop stability of the proposed E-NMPC solved by the suboptimal SQP algorithm. Some notations are necessary for the convenience of this discussion:

An admissiable set for control sequences is defined as $\mathbb{U}^N \subseteq \mathbb{R}^{3N}$, while the admissible set for u is $\mathbb{U} \subseteq \mathbb{R}^3$. The feasible set of states is $\mathbb{X} \subseteq \mathbb{R}^4$. Let $\phi(i; x, U)$ denote the solution of $x_{k+1} = f_x(x_k, u_k)$ at time i if the initial state is x and the control sequence is U. Since the constraint function $l(\cdot)$ compiles with the constraints of N preview steps, the merit function V can be written as the summation of stage penalties

$$V(x_k, U_k) = \sum_{i=1}^{N} g(\phi(i; x_k, U_k), u_k). \tag{28}$$

After augmenting with the terminal states penalty $V_f(\cdot)$ and constraints $\mathbb{X}_f$ the NLP problem can be written as

$$\begin{aligned} &\min_{U_k \in \mathbb{U}^N} \ V_N(x_k, U_k) \\ &\text{s.t. } \phi(N+1; x_k, U_k) \in X_f \end{aligned} \tag{29}$$

where

$$V_N(x_k, U_k) = V(x_k, U_k) + V_f(\phi(N+1; x_k, U_k)).$$

Assumption 1: The open loop system (12) is stablizable $\forall x \in X$.

Assumption 2: The admissible control set U is a compact set containing the nominal control actions. The terminal states set $\mathbb{X}_f$ is a closed set and contains the nominal states $\mathbb{X}_f \subseteq \mathbb{X}$.

Assumption 2 is adapted from [11, Assumption 2.3, p. 97]. For the investigated engine system, this suggests that the final IMEP reference is achievable with feasible control actions.

Assumption 3: For any $x \in X_f$, there exists a terminal control law

$$u := K_1 x - (s^{-1})^T r \tag{30}$$

that makes $V_f(f_x(x,u)) + g(x,u) \leq V_f(x)$, where $V_f(x) := (1/2)x^T P_f x$ and $P_f \in \mathbb{R}^{4\times 4}_{>0}$.

The proof of the system satisfying this assumption can be found in the Appendix. The terminal control law (30) has a state dependent term which regulates the states to the nominal condition. The second term of (30) is a constant that is related to the weighting of the control variation and fuel economy penalty. Therefore, the system converges to a nonzero steady-state x_{ss} which can be obtained by solving the following linear system:

$$[A + BK_1 - I \;\; B]\begin{bmatrix} x_{ss} \\ -(s^{-1})^T r \end{bmatrix} = 0. \tag{31}$$

A solution of the above linear system exists since the system is stablizable (according to *Assumption* 1). It is observed that $x_{ss} \to 0$ as $r \to 0$, and the E-NMPC becomes a tracking NMPC. If $r > 0$, the nominal condition must be shifted to x_{ss}. This process is automatically completed by the proposed SQP algorithm with real-time linearization, which updates the nominal conditions according to U_0 of each iteration.

The terminal state penalty is not necessary if N is sufficiently large to ensure the origin is asymptotically stable for the proposed controller, (see [2], [11, pp. 147–153]). Due to the complex nature of the system model, it is difficult to find the exact control horizon N that makes the terminal states penalty unnecessary. Simulation and experimental results show that the optimal control actions converge for N as small as two steps. Therefore, the terminal state penalty is removed from the proposed E-NMPC.

Based on *Assumptions* 1–3, a starting point U_0 can be selected to warm start the SQP algorithm for receding horizon MPC applications. At step k, the E-NMPC computes an optimal control sequence using the suboptimal SQP algorithm

$$U^*(k) = \left[u^*_{k|k+1}, u^*_{k|k+2}, \ldots u^*_{k|k+N}\right]^T. \tag{32}$$

For the next step, $+1$, the starting point $U_0(k+1)$ can be generated from the current optimal control sequence $U^*(k)$ by removing the first element (which is applied to the system a $k+1$ step) and adding a new control action u^+ for the final step of the control horizon

$$U_0(k+1) = \left[u^*_{k|k+2}, u^*_{k|k+2}, \ldots u^*_{k|k+N-1}, u^+\right]^T \tag{33}$$

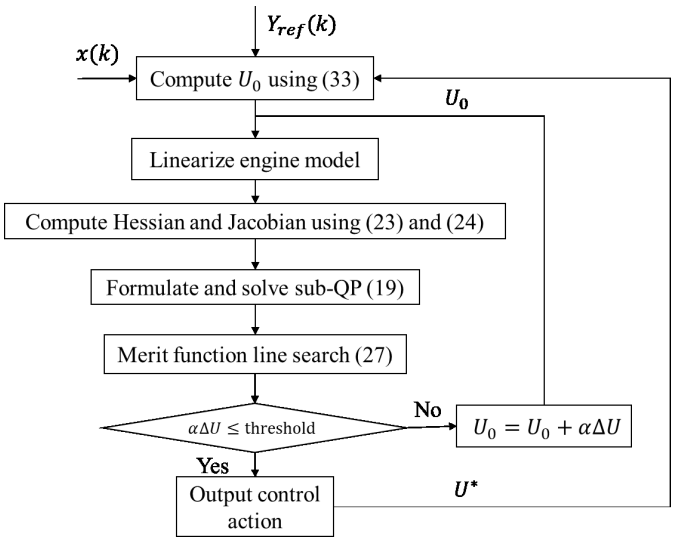


Fig. 8. Flowchart of the proposed SQP model predictive IMEP control.

where control action u^+ is obtained according to (30).

Assumption 4: The stage cost $g(\cdot)$ and the terminal cost $V_f(\cdot)$ satisfy

$$g(x,u) \geq \alpha_1(|x|) \quad \forall x \in X, \; \forall u \in U \tag{34}$$

$$V_f(x) \leq \alpha_2(|x|) \quad \forall x \in X_f \tag{35}$$

where $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are in K_∞ class.

Equation (34) is satisfied since $g(x,u) \geq (1/2)x^T C^T q C x = \alpha_1(|x|)$. Equation (35) is also satisfied. We can select a matrix $P_c \in \mathbb{R}^{4\times 4}_{>0}$, $P_c \geq P_f$. Then, $V_f(x) = (1/2)x^T P_f x \leq (1/2)x^T P_c x = \alpha_2(|x|)$.

It is noticed that the terminal states penalty is quadratic, making integration into the SQP algorithm previously discussed straightforward. Using similar reasoning from [11], it can be proved that *Assumptions* 1–4 result in upper and lower bounds of $V_N(x,U)$, where U is computed by the proposed E-NMPC

$$\beta_1(|x|) \leq V_N(x) \leq \beta_2(|x|) \quad \forall x \in \mathbb{X} \tag{36}$$

where $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are K_∞ functions.

Let $\beta_3(|x|) := V_N(x) - V_N(x^+)$, where x^+ are the successor states generated by the E-NMPC control law. The merit function search process discussed previously ensures that $\{V_N(x_i)|i \in I_{>0}\}$ decreases as the SQP progresses (or $i \to \infty$). Thus

$$\beta_3(|x|) \geq 0. \tag{37}$$

Furthermore, $V_N(x_i)|i \in I_{>0}$ is nonincreasing (as $i \to \infty$) and bounded below by zero (every terms in V_N is greater than or equal to zero). Therefore, as $i \to \infty$, $V_N(x_i, U_i) \to V_N^*$ (a constant), $x_i \to 0$, and $\beta_3(|x|) \to 0$. This means $\beta_3(|x|)$ is in PD class, proving that $V_N(x)$ is a Lyapunov function of the engine system controlled by the proposed E-NMPC.

Figure 8 shows the flow chart that summaries the entire proposed SQP model predictive IMEP control strategy.

TABLE I
OPERATION CONDITION RANGE OF SQP MPC VALIDATION

Engine parameters	Min	Max
RPM	800	6000
Manifold pressure (kPa)	30	100
IMEP reference (kPa)	250	1200

TABLE II
ITERATION NUMBER AND COMPUTATIONAL TIME FOR SQP MPC SIMULATION

Solver	Mean iter.	Max iter.	Min iter.	Mean time (ms)
Cold dual	14	42	3	0.23
Warm dual	8	46	1	0.16
Cold parametric	11	16	8	0.25
Warm parametric	3	7	1	0.09

VI. ACTIVE SET QP ALGORITHMS

Two active set QP algorithms for solving the sub-QP problems are considered in this paper: 1) a dimensional search dual active set (see [44], [45]) and 2) parametric active set methods (e.g. *qpOASES* [33]). The first method does not require matrix inversion during each QP iteration, making it more numerically robust and efficient. However, it has unsatisfactory performance against constraint degeneracy (a situation where the constraint gradient $\nabla l_{(x_k,U_0)}$ is row rank insufficient). The parametric active set method, on the other hand, was refined by Ferreau *et al.* [34] to improve convergence and degeneracy handling performance. The parametric active set QP requires solving a linear system during each iteration resulting in longer computation time per iteration compared to the dual active set method. However, since the parametric method uses less iterations both active set methods have similar computational time for the proposed MPC. Finally, both active set methods can be warm started, exploiting the similarities between successive QPs along the sequence. This characteristic makes both algorithms favorable for the current E-NMPC application, although each method has its own unique advantages.

Both dual and parametric active set QP algorithms are implemented in the E-NMPC and evaluated in simulation for 10^6 consecutive engine cycles. The IMEP tracking reference and engine speed traces are generated randomly according to a uniform distribution. Table I shows the range of operating conditions used for validation.

The *qpOASES* software is used in this paper to implement the parametric active set QP algorithm in MATLAB. Table II summarizes the number of iterations and mean computational time (per SQP iteration) of both QP algorithms. It can be observed that both methods can benefit from "warm" start strategies that exploit similar active constraint sets between successive SQP iterations. The parametric method requires fewer number of iterations in both "cold" and "warm" start situations. The simulation is carried out on a desktop computer with a 3.2-GHz 64-b CPU and 16 GB of RAM. Although the parametric method has obvious advantages in terms of iteration number the difference in computation time between the two algorithms is not very significant. The dual QP requires scalar computation within each iteration, making it faster per iteration compared to the parametric method. There were compatibility issues between *qpOASES* and the prototype engine controller of this paper. Therefore, the experimental validation of the proposed SQP MPC was carried out with the dual QP algorithm.

TABLE III
ENGINE PARAMETERS

Fuel	Gasoline (87 Pump Octane)
Compression Ratio	10.2
Bore	96 mm
Stroke	83 mm
Connecting Rod Length	156.5 mm
Intake Valve Diameter	39 mm
Exhaust Valve Diameter	30 mm

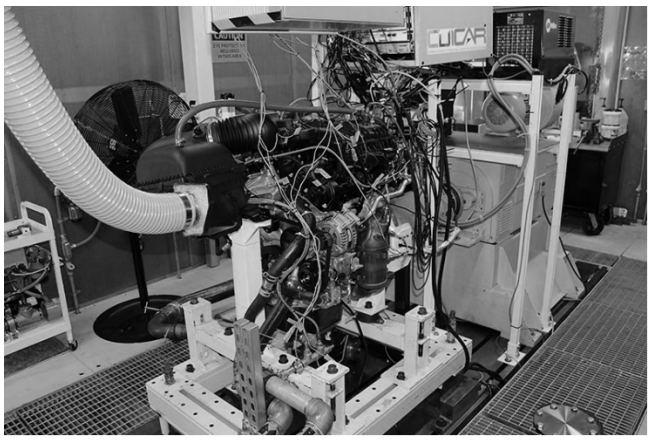


Fig. 9. Dynamometer test setup for the investigated engine located at CU-ICAR, Clemson University.

VII. SIMULATION AND EXPERIMENTAL RESULTS

The spark ignited V6 engine investigated in this paper has port fuel injection and a displacement of 3.6 L. The EGR system has an intercooler and postthrottle setup, as shown in Fig. 1. A summary of engine geometry is given in Table III.

The experiment validation was carried out on an ac dynamometer. The test engine was controlled using an ETAS ES910 system overriding the stock ECU. AVL Concerto software was used to process the data that were acquired during each experiment. Each cylinder of one engine bank was instrumented with passage-mounted AVL GH12D piezoelectric cylinder pressure sensors. The sensors were located in the cylinder head to maximize accuracy according to Patterson and Davis [41], and were equipped with flame guards to minimize

Fig. 10. Prototype control system setup.

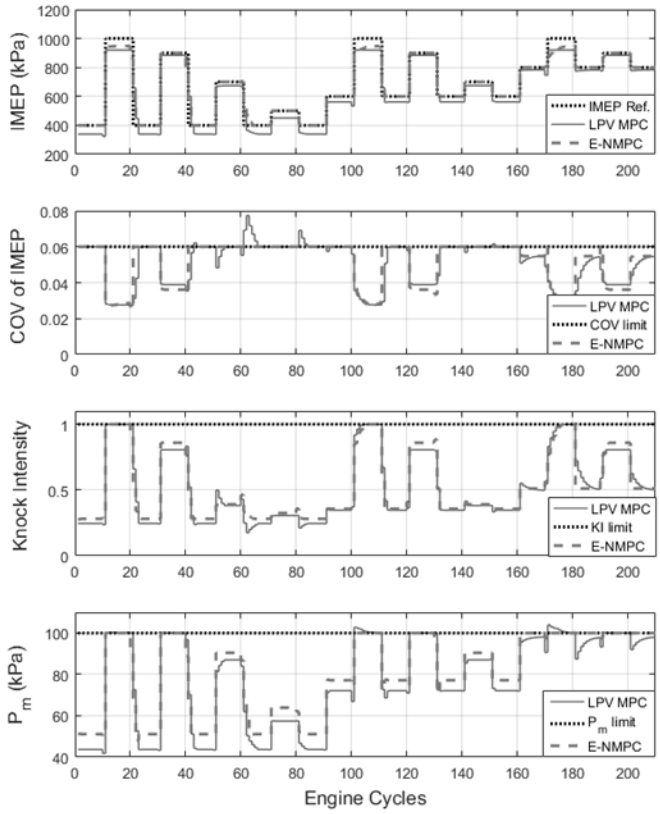


Fig. 11. Engine performance comparison between LPV E-MPC and E-NMPC IMEP control.

thermal shock errors. The cylinder pressure measurement was sent to a cylinder pressure development controller (CPDC) unit to compute CA50 and IMEP [43]. Communication between the CPDC and ES910 was established via a CAN communications link. Figs. 9 and 10 show the hardware and prototype control system setup of the experimental validation.

Figs. 11 and 12 show the engine performance and control actuation of the proposed SQP model predictive IMEP controller in simulation with comparison to an LPV E-MPC. The LPV E-MPC shares the same finite-difference real-time linearization and warm-start techniques as the proposed SQP

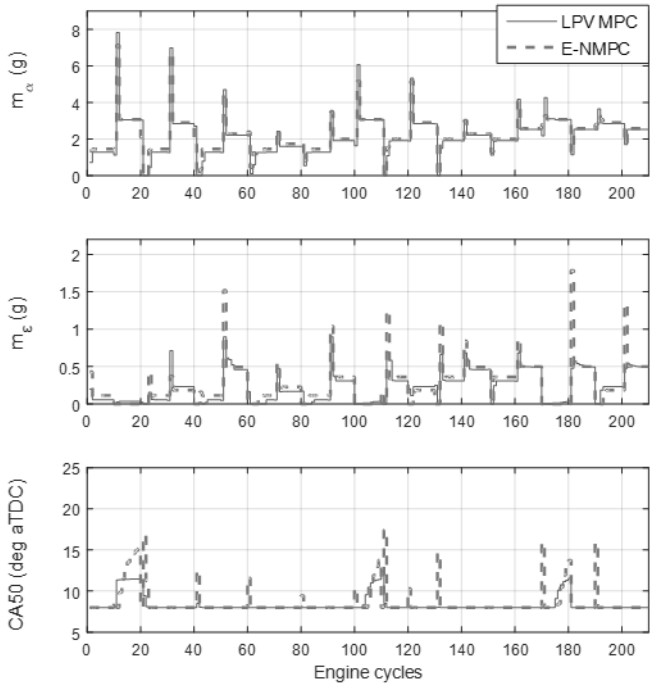


Fig. 12. Control actuation comparison between LPV MPC and SQP MPC IMEP control.

E-NMPC. The engine speed is fixed at 1500 r/min during this simulation. The COV of IMEP limit is selected as 6%.

It can be observed from Fig. 11 that the proposed SQP E-NMPC is able to track the IMEP reference without violating knock, COV of IMEP and intake manifold pressure constraints. The normalized root-mean-square error (NRMSE) of IMEP tracking is in the neighborhood of 1%. In comparison, the NRMSE of IMEP for the LPV E-MPC is 9.8%. The inferior IMEP tracking performance of the LPV E-MPC can be observed in Fig. 11, with steady-state IMEP tracking error caused by model linearization errors. It can also be observed from Fig. 11 that the LPV E-MPC cannot guarantee the COV of IMEP constraints during tip-out situations (e.g., 40th, 60th, and 80th engine cycles). The EGR valve actuation from Fig. 12 reveals that the LPV E-MPC over dilutes the cylinder charge during these transient situations due to errors caused by linearizing the COV of IMEP model. The LPV MPC also violates the manifold pressure constraint at the 100th and 170th engine cycles, resulting in a false improvement of IMEP tracking performance. The large IMEP tracking error at high load conditions (1000-kPa IMEP reference) is due to the knock constraint, which causes the IMEP reference to be infeasible. This is a situation where *Assumption* 2 is invalid. However, the E-NMPC (and the LPV MPC) is still able to find a reasonable control action to mitigate the knock constraints without inducing instability. Both LPV MPC and the E-NMPC mitigate engine knock by postponing combustion phasing. Close examination of these high load situations reveals that the E-NMPC generates slightly higher IMEP output than the LPV MPC by using larger throttle air mass flow and less EGR. Therefore, the E-NMPC needs to retard CA50 further than the LPV MPC. The large IMEP tracking error during these high

load situations makes the fuel economy penalty insignificant, causing the E-NMPC to find the actuator setpoints that generates the maximum IMEP possible.

Control actuation of the proposed SQP E-NMPC is demonstrated in Fig. 12. It can be observed between engine cycles 10–20 that the control chattering issue is eliminated with the merit function technique (compared to Fig. 7). For a "tip-in" situation (around 10th, 30th, 50th, etc., engine cycles), the throttle air mass flow m_α overshoots during IMEP reference steps. This maneuver is to compensate for the manifold delay and quickly increase the IMEP output. During "tip-out" situations (around 20th, 40th, 60th, etc., engine cycles), the throttle air mass is reduced to zero initially to compensate for manifold delay. Then, it converges to a steady-state value without oscillation. For a "heavy" tip-out situation where IMEP reference reduces significantly (40th, 60th, 80th, etc., engine cycles), the E-NMPC closes the EGR valve 1–2 cycles earlier than the throttle in order to prevent excessive dilution and meet the COV of IMEP constraint. The LPV MPC does not find this maneuver most of the time, resulting in the COV limit violation for "heavy" tip-out situations. During part load conditions, the E-NMPC maximizes EGR flow to the limit of COV of IMEP or manifold pressure, reducing the pumping losses. In comparison, the EGR flow allowed by the LPV MPC during part load conditions is much less. If the IMEP demand is very high (10th–20th, 100th–110th, etc., engine cycles), both E-NMPC and LPV MPC reduce EGR to maximize engine air mass flow. The optimal transient maneuvers generated by the proposed E-NMPC are in agreement with the production engine calibration when the future IMEP demand is known (for example, during the gear shift phase). However, the calibration effort of the E-NMPC is much less than conventional map-based engine control strategies.

Fig. 13 compares E-NMPCs with different preview and control horizons. It can be observed that the E-NMPC with longer horizon closes the EGR valve earlier if EGR purging is required to avoid high COV of IMEP. The CA50 output is also modified to accommodate the EGR flow to avoid knock and COV of IMEP constraints. However, the differences in IMEP tracking performance and throttle air mass flow are not distinguishable after $N \geq 2$. This suggests that further increasing horizon length will only provide marginal improvement of optimality, suggesting the possibility of decreasing horizon length to reduce computational load of the E-NMPC.

In Fig. 14, the E-NMPC is evaluated for different engine speeds. The two-step preview E-NMPC is provided with the same IMEP reference in the time domain. While the accelerated engine speed reduces characteristic time of manifold filling and combustion, it also reduces the change rate of IMEP reference in the engine cycle domain. Therefore, the optimal control actions found by the MPC become less "aggressive" as engine speed increases. The throttle actuation almost becomes a slow transition between two steady states, and the CA50 is maintained at MBT. This suggests fast throttle position control is not necessary for high speed operation. However, the E-NMPC demands fast EGR actuation to purge the EGR out of the system during high speed tip-out situations. This behavior requires fast EGR valve position tracking to maintain the engine-actuator frequency separation assumption. Considering the mechanical design of current EGR valves, this requirement is difficult to realize. In this case, the E-NMPC formulation should be integrated with the EGR valve position dynamics, which will be addressed in a future extension of this paper.

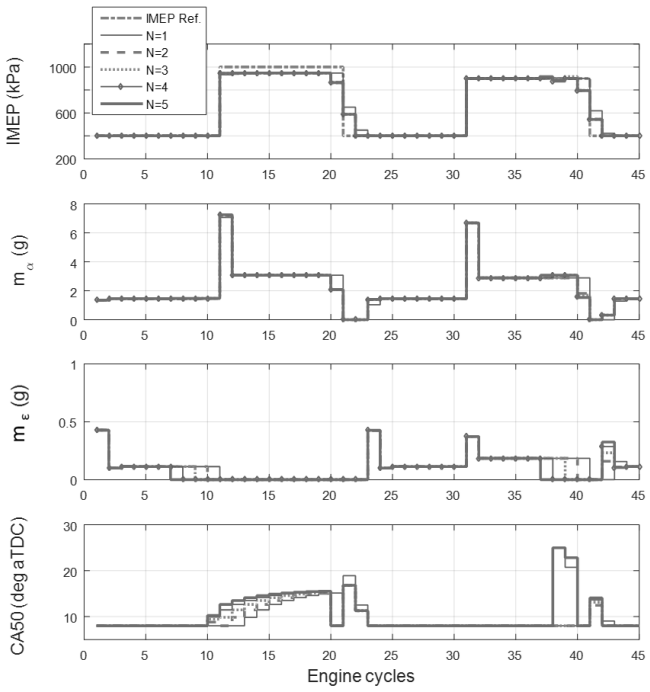


Fig. 13. Comparison between E-NMPCs with different preview/control horizons.

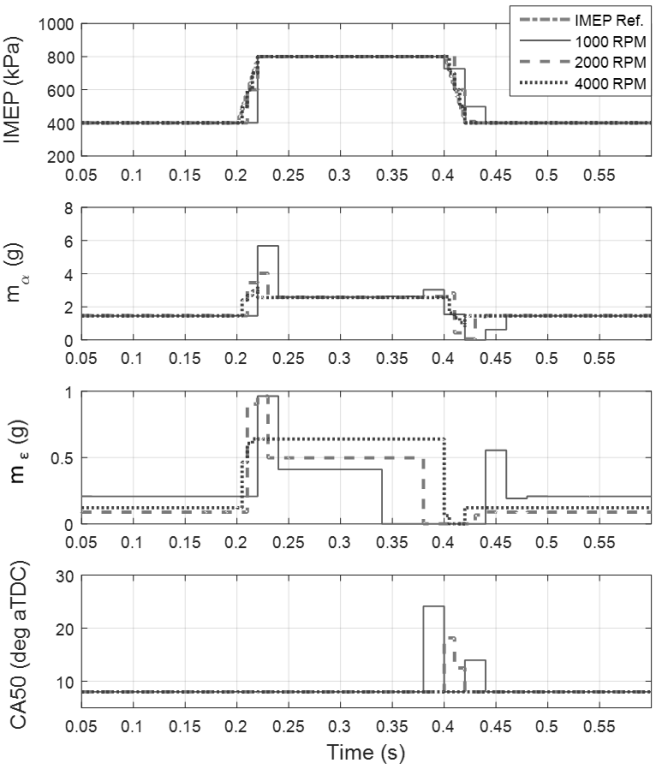


Fig. 14. E-NMPC performance comparison between different engine speeds.

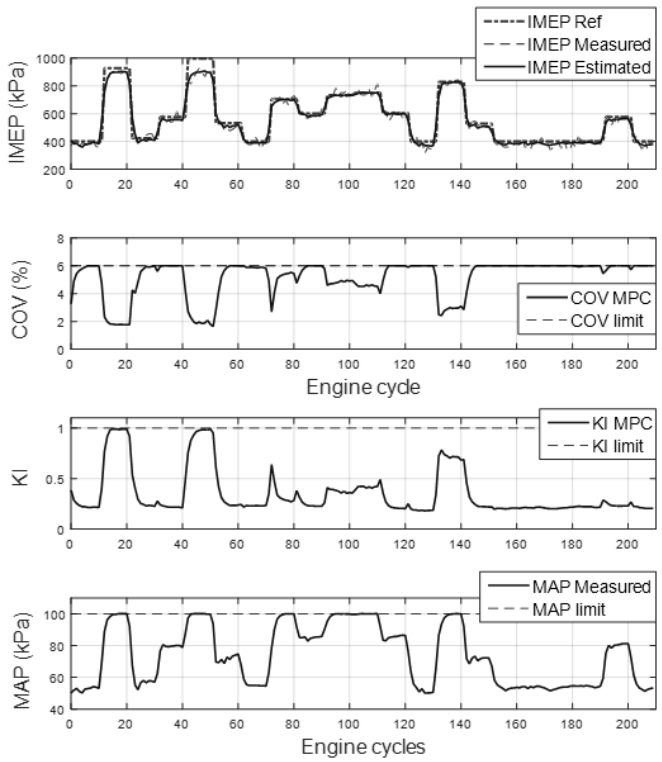


Fig. 15. Engine performance during experiment validation at 2000 r/min.

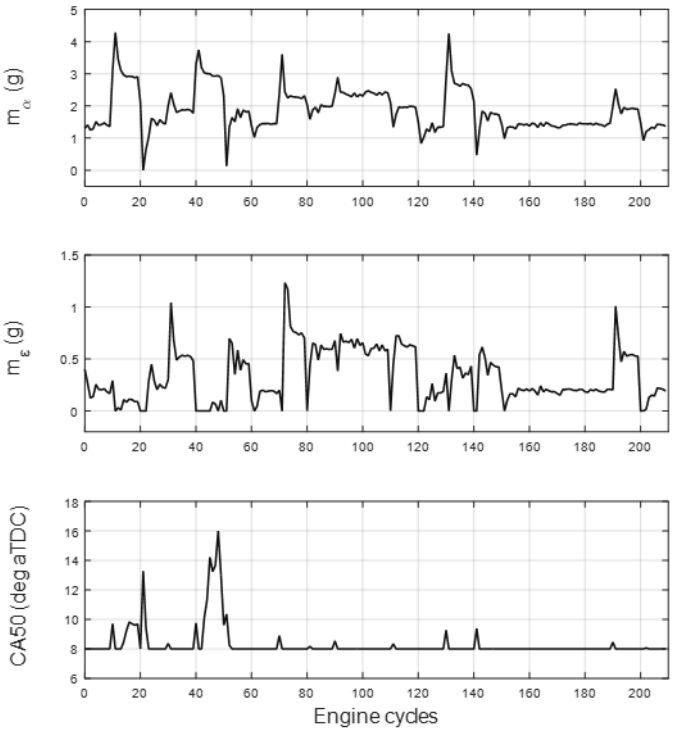


Fig. 16. Actuator output of E-NMPC during experimental validation at 2000 r/min.

Experimental validation of the proposed IMEP controller was carried out for engine speeds less than 3000 r/min. Fig. 15 shows the engine performance results and Fig. 16 shows the control actions found by the proposed E-NMPC. The experimental results exhibit similar transient behaviors as the simulation results. The engine states are estimated using an extended Kalman filter. The estimated IMEP is shown in the first plot of Fig. 15, but engine state(s) estimation is not the focus of this paper.

TABLE IV
STATISTICS OF THE PROPOSED SQP MODEL PREDICTIVE IMEP CONTROLLER

	Mean	Max	Min
Number of major iterations	5	15	1
Number of QP iterations	8	50	1
Execution time per engine cycle (ms)	1.07	9	0.11
Execution time per major iteration (ms)	0.21	0.76	0.09
Time for model evaluation per major iteration (ms)	0.15	0.42	0.07
Time for QP per major iteration (ms)	0.06	0.32	0.02

Table IV summarizes execution time statistics of the proposed SQP MPC during experimental validation. The ETAS ES910 prototype engine controller has a double precision floating CPU with 800-MHz clock. The memory is DDR2-RAM with 512 MB of space and a 400-MHz clock. It can be observed that the control algorithm is computationally efficient for cyclic engine control considering the duration of every engine cycle is 200–20 ms for engine speeds from 600 to 6000 r/min. Table IV also reveals that QP computation takes approximately 1/3 of the overall computational time, while the rest of the execution time is spent on evaluating the engine model. It can be concluded that the simplification of the engine models can significantly reduce the computational time.

VIII. CONCLUSION

This paper proposes an E-NMPC strategy for managing the IMEP control of spark-ignition engines. The E-NMPC is solved by an SQP algorithm. The control objective is to track IMEP reference while minimizing fuel consumption. This E-NMPC also guarantees abnormal combustion constraints during the search for optimal control actions. The proposed SQP E-NMPC is designed to work in the engine cycle domain, which reduces the engine speed dependence of air-path dynamics. The application of fast low-order sublevel controllers reduces the modeling complexity needed in the NMPC formulation. These fast controllers can also maximize the bandwidth of actuators.

The SQP algorithm is tailored for this application for high computational efficiency. It exploits the Gauss–Newton-like structure of the NLP formulated for MPC to simplify computation of the Hessian matrix. Warm-start and merit function techniques are applied to ensure closed-loop stability on top of the conventional terminal state penalty technique. Merit function

step scaling also improves global convergence performance and eliminates steady-state control chattering issues. Two active set QP algorithms were investigated for implementation with the SQP MPC. Both algorithms can be warm started and exploit similar active constraint sets between successive QP problems of each SQP iteration. Finally, the finite-difference linearization technique makes the SQP strategy adaptive to different types of engine models with complex structures.

Simulation and experimental results demonstrate that the proposed model predictive IMEP control strategy achieves its design objectives, in terms of tracking torque reference, minimizing fuel consumption, and respecting combustion constraints. Compared to LPV MPC, the E-NMPC demonstrates superior ability to guarantee constraints and reducing fuel consumption. Although it is possible to calibrate conventional map-based engine control strategies to make control decisions similar to E-NMPC, the calibration effort is much less with E-NMPC. This advantage of E-NMPC is expected to increase as more actuators are integrated into engines for the purpose of improving fuel economy. Finally, the computational time analysis of the proposed SQP MPC with prototype engine controllers demonstrates high potential for real-time implementation with future production ECUs.

APPENDIX

This Lemma is used to prove the nonconvexity of the NLP problem that is solved in every MPC update step. The existence of multiple local optimal solutions makes it impractical to apply global optimization algorithms in real-time implementation. Therefore, this paper considers the suboptimal SQP algorithm for the proposed MPC problem.

Lemma 1: A function $f(x) : X \in \mathbb{R}^m \rightarrow Y \in \mathbb{R}^n$ is nonconvex, if $\exists \mathbb{Z} \subseteq \mathbb{X}$ such that $f(x)$ is nonconvex for $x \in \mathbb{Z}$.

Proof of Lemma 1: By the definition of convexity, if $\exists \mathbb{Z} \subseteq \mathbb{X}$ such that $f(x)$ is nonconvex for $x \in \mathbb{Z}$, $\exists x_1, x_2 \in \mathbb{Z}, \exists \alpha \in [0, 1] : f[\alpha x_1 + (1-\alpha)x_2] > \alpha f(x_1) + (1-\alpha) f(x_2)$. Since $\mathbb{Z} \subseteq \mathbb{X}, x_1, x_2 \in \mathbb{X}$. Therefore $f(x)$ is nonconvex for $x \in \mathbb{X}$

Proof of Assumption 3: After linearizing the system model at $\phi\ (N+1; x_k, U_0)$ and $(u_0)_{k|k+N}$, we have linear state-space model $x^+ = Ax + Bu$, and $y = Cx$. The stage cost is $g(x,u) = (1/2)(x^T C^T q C x + u^T s u + u^T r)$. The constraint penalty is neglected according to *Assumption* 2. Design the control law as $u := K_1 x + K_2$. Let $A_k := A + BK_1$, and let $V_f : R^4 \rightarrow R_{\geq 0}$ be defined by $V_f(x) := (1/2)x^T P_f x$, $P_f \in \mathbb{R}^{4\times 4}_{>0}$. Then

$$\begin{aligned} &V_f(x^+) - V_f(x) + g(x,u) \\ &\quad = x^T \left(A_k^T P_f A_k + Q_k - P_f\right)x + \left(K_2^T s - r^T\right)K_1 x \\ &\qquad + K_2^T (sK_2 - r) \end{aligned}$$

where $Q_k = C^T q C + K_1^T s K_1$

We can choose $K_2 = -(s^{-1})^T r$, while K_1 and P_f solve the Lyapunov equation

$$A_k^T P_f A_k + Q_k - P_f = 0.$$

Then we have

$$V_f(x^+) - V_f(x) + g(x,u) = 0.$$

REFERENCES

[1] A. Ali and J. P. Blath, "Application of modern techniques to SI-engine torque control," in *Proc. IEEE Int. Conf. Control Appl.*, Munich, Germany, Oct. 2006, pp. 2405–2410.

[2] T.-K. Lee and Z. S. Filipi, "Nonlinear model predictive control of a dual-independent variable valve timing engine with electronic throttle control," *Proc. Inst. Mech. Eng. D, Transp. Eng.*, vol. 225, no. 9, pp. 1221–1234, Sep. 2011.

[3] E. Feru, M. Lazar, R. H. Gielen, I. V. Kolmanovsky, and S. Di Cairano, "Lyapunov-based constrained engine torque control using electronic throttle and variable cam timing," in *Proc. Amer. Control Conf.*, Montreal, QC, Canada, Jun. 2012, pp. 2866–2871.

[4] M. Kang and T. Shen, "Nonlinear model predictive torque control for IC engines," in *Proc. 11th World Congr. Intell. Control Autom.*, Shenyang, China, Jun./Jul. 2014, pp. 804–809.

[5] L. Guzzella and C. H. Onder, *Introduction to Modeling and Control of Internal Combustion Engine Systems*, 2nd ed. Heidelberg, Germany: Springer, 2009.

[6] C. M. Hall, G. M. Shaver, J. Chauvin, and N. Petit, "Combustion phasing model for control of a gasoline-ethanol fueled SI engine with variable valve timing," in *Proc. Amer. Control Conf.*, Montreal, QC, Canada, Jun. 2012, pp. 1271–1277.

[7] S. Wang, R. Prucka, Q. Zhu, M. Prucka, and H. Dourra, "A real-time model for spark ignition engine combustion phasing prediction," *SAE Int. J. Eng.*, vol. 9, no. 2, pp. 1180–1190, 2016.

[8] K.-H. Lee and K. Kim, "Influence of initial combustion in SI engine on following combustion stage and cycle-by-cycle variations in combustion process," *Int. J. Autom. Technol.*, vol. 2, no. 1, pp. 25–31, 2001.

[9] M. Diehl and J. Bjornberg, "Robust dynamic programming for min-max model predictive control of constrained uncertain systems," *IEEE Trans. Autom. Control*, vol. 49, no. 12, pp. 2253–2257, Dec. 2004.

[10] E. S. Meadows, "Dynamic programming and model predictive control," in *Proc. Amer. Control Conf.*, Albuquerque, NM, USA, Jun. 1997, pp. 1635–1639.

[11] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control Theory and Design*. Madison, WI, USA: Nob Hill Publishing, 2009.

[12] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[13] T. Ohtsuka, "A continuation/GMRES method for fast computation of nonlinear receding horizon control," *Automatica*, vol. 40, no. 4, pp. 563–574, 2004.

[14] M. Lazar, "Flexible control Lyapunov functions," in *Proc. Amer. Control Conf.*, St. Louis, MO, USA, Jun. 2009, pp. 102–107.

[15] R. Quirynen *et al.*, "Symmetric algorithmic differentiation based exact Hessian SQP method and software for economic MPC," in *Proc. IEEE 53rd Annu. Conf. Decision Control*, Dec. 2014, pp. 2752–2757.

[16] T. Broomhead, C. Manzie, P. Hield, R. Shekhar, and M. Brear, "Economic model predictive control and applications for diesel generators," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 2, pp. 388–400, Mar. 2017.

[17] Q. Zhu, S. Onori, and R. Prucka, "Pattern recognition technique based active set QP strategy applied to MPC for a driving cycle test," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, Jul. 2015, pp. 4935–4940.

[18] R. Sharma, D. Nešić, and C. Manzie, "Idle speed control using linear time varying model predictive control and discrete time approximations," in *Proc. IEEE Int. Conf. Control Appl.*, Yokohama, Japan, Sep. 2010, pp. 1140–1145.

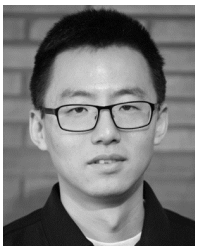
[19] M. Diehl, R. Amrit, and J. B. Rawlings, "A Lyapunov function for economic optimizing model predictive control," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 703–707, Mar. 2011.

[20] T. J. Broomhead, C. Manzie, R. C. Shekhar, and P. Hield, "Robust periodic economic MPC for linear systems," *Automatica*, vol. 60, pp. 30–37, Oct. 2015.

[21] M. J. Tenny, S. J. Wright, and J. B. Rawlings, "Nonlinear model predictive control via feasibility-perturbed sequential quadratic programming," *Comput. Optim. Appl.*, vol. 28, no. 1, pp. 87–121, Apr. 2004.

[22] B. Houska, H. J. Ferreau, and M. Diehl, "An auto-generated real-time iteration algorithm for nonlinear MPC in the microsecond range," *Automatica*, vol. 47, no. 10, pp. 2279–2285, 2011.

[23] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.
[24] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999, pp. 431–448.
[25] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York, NY, USA: Wiley, 1987.
[26] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms: 2. The new algorithm," *IMA J. Inst. Math. Appl.*, vol. 6, no. 3, pp. 222–236, 1970.
[27] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 2000.
[28] M. J. Goldsmith, "Sequential quadratic programming methods based on indefinite hessian approximations," Ph.D. dissertation, Dept. Oper. Res., Stanford Univ., Stanford, CA, USA, 1999.
[29] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 1982.
[30] R. Fletcher, "Methods related to Lagrangian functions," in *Numerical Methods for Constrained Optimization*, P. E. Gill and W. Murray, Eds. New York, NY, USA: Academic, 1974, pp. 228–234.
[31] J. Goodman, "Newton's method for constrained optimization," *Math. Programm.*, vol. 33, no. 2, pp. 162–171, 1985.
[32] S. P. Han, "A globally convergent method for nonlinear programming," *J. Optim. Theory Appl.*, vol. 22, no. 3, pp. 297–309, 1977.
[33] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, "qpOASES: A parametric active-set algorithm for quadratic programming," *Math. Program. Comput.*, vol. 6, no. 4, pp. 327–363, 2014.
[34] H. J. Ferreau, H. G. Bock, and M. Diehl, "An online active set strategy to overcome the limitations of explicit MPC," *Int. J Robust Nonlinear Control*, vol. 18, no. 8, pp. 816–830, 2007.
[35] Q. Zhu, S. Onori, and R. Prucka, "Nonlinear economic model predictive control for SI engines based on sequential quadratic programming," in *Proc. Amer. Control Conf.*, Boston, MA, USA, Jul. 2016, pp. 1802–1807.
[36] J. Fox, W. Cheng, and J. Heywood, "A model for predicting residual gas fraction in spark-ignition engines," SAE Tech. Paper 931025, 1993.
[37] S. Wang, R. Prucka, M. Prucka, and H. Dourra, "Control-oriented residual gas mass prediction for spark ignition engines," *Int. J. Eng. Res.*, vol. 226, no. 6, pp. 828–839, Jun. 2012.
[38] J. C. Livengood and P. C. Wu, "Correlation of autoignition phenomena in internal combustion engines and rapid compression machines," *Symp. Int. Combustion*, vol. 5, no. 1, pp. 347–355, 1955.
[39] O. Arici, J. H. Johnson, and A. J. Kulkarni, "The vehicle engine cooling system simulation part 1—Model development," SAE Tech. Paper 1999-01-0240, 1999.
[40] E. Cortona, H. C. Onder, and L. Guzzella, "Engine thermal management with components for fuel consumption reduction," *Int. J. Eng. Res. IMechE*, vol. 3, no. 3, pp. 157–170, 2002.
[41] G. Patterson and R. Davis, "Geometric and topological considerations to maximize remotely mounted cylinder pressure transducer data quality," SAE Tech. Paper 2009-01-0644, 2009.
[42] T. K. Lee, R. G. Prucka, and Z. S. Filipi, "Real-time estimation of combustion variability for model-based control and optimal calibration of spark ignition engines," *Proc. Inst. Mech. Eng. D, Transp. Eng.*, vol. 223, no. D11, pp. 1361–1372, Nov. 2009.
[43] K. Schten, G. Ripley, A. Punater, and C. Erickson, "Design of an automotive grade controller for in-cylinder pressure based engine control development," SAE Tech. Paper 2007-01-0774, 2007.
[44] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY, USA: Wiley, 1969.
[45] L. Wang, *Model Predictive Control System Design and Implementation Using MATLAB*. London, U.K.: Springer, 2009, pp. 63–68.



Qilun Zhu (M'14) was born in Changchun, Jilin, China, in 1988. He received the B.S. degree from Jilin University, Changchun, in 2010, the M.Eng. degree from Cornell University, Ithaca, NY, USA, in 2011, and the Ph.D. degree from Clemson University, Greenville, SC, USA, in 2015.

From 2015 to 2017, he was a Post-Doctoral Research Fellow with the Department of Automotive Engineering, Clemson University, where he is currently an Assistant Research Professor. He is also appointed as a fellow of Robert H. Brooks Sports Science Institute, Clemson University. His current research interests include real-time optimization with estimation, identification, learning, and control.

Simona Onori (SM'15) received the Laurea degree (*summa cum laude*) in computer science engineering from the University of Rome Tor Vergata, Rome, Italy, in 2003, the M.S. degree in electrical and computer engineering from the University of New Mexico, Albuquerque, NM, USA, in 2004, and the Ph.D. degree in control engineering from the University of Rome Tor Vergata, in 2007.

Since 2013, she has been an Assistant Professor with Clemson University International Center for Automotive Research, Greenville, SC, USA, where she also held a joint apportionment with the Electrical and Computer Engineering Department. She held visiting professor positions at the University of Trento, Trento, Italy, in 2014, the Beijing Institute of Technology, Beijing, China, in 2015, and the University of Orleans, Orleans, France, in 2016. Since 2017, she has been an Assistant Professor with the Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA. She is a PSG Distinguished Visiting Professor awarded by the Managing Trustee of the PSG College of Technology, Coimbatore, India, in 2017.

Dr. Onori was a recipient of the 2017 NSF CAREER award, the 2017 Clemson University College of Engineering and Science Dean's Faculty Fellows Award, the 2017 Clemson University Esin Gulari Leadership and Service Award, the 2016 Energy Leadership Award in the category Emerging Leader (for the Carolinas), the 2015 Innovision Award (South Carolina), the 2012 Lumley Interdisciplinary Research Award, and the 2011 Outstanding Technology Team Award, TechColumbus.

Robert Prucka received the B.S.E, M.S.E, and Ph.D. degrees in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2000, 2004, and 2008 respectively.

He is currently an Associate Professor with the Department of Automotive Engineering, Clemson University International Center for Automotive Research, Greenville, SC, USA, where he is also the Team Leader for one of the current Deep Orange Prototype Vehicle Programs, the Faculty Advisor for Racing Student Team, and is active in the motorsports engineering initiatives of the Brooks Sports Science Institute. His current research interests include the design, performance, control, calibration, and emissions of advanced internal combustion engines.

Dr. Prucka was a recipient of the Clemson University's Murray Sokely Award in 2015, given in recognition for his outstanding contributions to engineering education.