

Minimizing Emission for Timely Heavy-Duty Truck Transportation

Junyan Su, Runzhi Zhou, Qingyu Liu, Wenjie Xu, Minghua Chen, *Fellow, IEEE*, Haibo Zeng, *Member, IEEE*

Abstract—We consider the problem of minimizing emission of a heavy-duty truck transporting freight between two locations subject to a hard deadline constraint. The truck is equipped with a multi-speed transmission and a modern combustion engine that intelligently switches among multiple fuel injection strategies at certain engine speeds (called *switching speeds*) to achieve lower emission profiles. Our objective is to minimize the emission by optimizing both path and speed planning for heavy-duty trucks with multi-speed transmission and multiple injection strategies in the engine. This emission minimization problem, while pervasive in practice, has two challenges: i) the emission rate function is discontinuous and non-convex due to switching of the fuel injections and gear ratios, which makes the common practice of driving at a constant speed on a road segment not eco-friendly; ii) the problem is NP-hard due to the combinatorial nature of the simultaneous path and speed planning. We tackle the first challenge by considering the case where the truck can travel at a heterogeneous speed profile over a road segment and then formulate the speed planning problem as a convex problem. We further identify special structures in this problem and provide an efficient method for computing the optimal speed profile. We then tackle the second challenge by developing an efficient heuristic for both path planning and speed planning to solve the emission minimization problem on the scale of national highway systems. Our extensive simulations on the US highway system show that our solution reduces up to 46% NOx emission as compared to the commonly-adopted fastest path approach. We also find that optimizing heterogeneous speed profiles reduce up to 32% emission as compared to their homogeneous counterpart, thus are necessary to be considered in eco-friendly truck operations.

Index Terms—Energy-efficient transportation, timely transportation, engine fuel injection strategy, emission

I. INTRODUCTION

In 2021, 11.0 billion tons of freight were transported by heavy-duty trucks in the United States, representing 72.2% of total domestic tonnage shipped [2]. The trucking industry generated \$875.5 billion in revenue in 2021 [2]. This number

Part of this work has been presented at Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (ACM BuildSys), November 18–20, 2020, Virtual Event, Japan [1]. This work is supported in part by a General Research Fund from Research Grants Council, Hong Kong (Project No. 11206821), an InnoHK initiative, The Government of the HKSAR, Laboratory for AI-Powered Financial Technologies, and a Shenzhen-Hong Kong-Macau Science & Technology Project (Category C, Project No. SGDX2022053011203026). The authors would also like to thank the anonymous reviewers for their helpful comments.

Junyan Su and Minghua Chen are with City University of Hong Kong (email: junyan.su@my.cityu.edu.hk; minghua.chen@cityu.edu.hk). Ruizhi Zhou is with University of Pennsylvania (email: lawrencezhou9@gmail.com). Qingyu Liu is with Peking University (email: qy.liu@pku.edu.cn). Haibo Zeng is with Virginia Tech (email: hbzeng@vt.edu). Wenjie Xu is with Swiss Federal Institute of Technology Lausanne (EPFL) (email: wenjie.xu@epfl.ch). Corresponding authors: Minghua Chen (minghua.chen@cityu.edu.hk) and Haibo Zeng (hbzeng@vt.edu)

would rank 18th in the world if measured against countries' GDPs. This trend will likely continue, as the global freight activity is predicted to increase by a factor of 2.4 by 2050 [3].

Despite their importance to the economy, heavy-duty vehicles are a significant source of emissions, including Carbon Dioxide (CO₂), Nitrogen Oxides (NO_x), and fine particle matter (PM 2.5). With only 4% of the total vehicle population, heavy-duty trucks produce more than one-third of the CO₂ (the primary greenhouse gas causing global warming) emitted in the transportation sector around the world [3]. In the US, about 16–18% of NO_x is emitted by heavy-duty trucks [4]. In California, one of the most polluted areas in the US, heavy-duty trucks contribute to over 70% of the NO_x emissions from on-road vehicles [5]. Therefore, it is critical to reduce their exhaust emissions for a cleaner environment.

A recent effort to reduce emissions in the transportation sector is the introduction of multiple injection strategies for internal combustion engines, which also improves fuel economy and reduces combustion noise [8]. More specifically, in each engine revolution, the engine control system determines the timing and amount of fuel for several possible injections: (i) the main injection, which provides the bulk of the fuel; (ii) several optional injections (called pilot injections or pre-injections) before the main injection to heat the combustion chamber and ensure a more uniform fuel-air mixture; (iii) one or two optional injections (called post-injections) after the main injection to burn the residual and decrease the amount of pollutants. Fig. 1a illustrates the emission of NO_x as a function of the engine speed for different injection strategies [6], where the emission rate function for each strategy is approximated by a convex function. In addition, these functions are monotonically decreasing with the number of injections: for any engine speed, a higher number of injections means less emission.

However, multiple injections come with the cost of a higher computational load. In addition, when a truck drives at high speed, the time interval for one engine revolution becomes small. Therefore, the limited computational resources and tight real-time requirements may not allow sophisticated control strategies (and multiple injections). Engine control systems are thus often designed to be *self-adaptive* in that they switch to simplified control strategies (e.g., single-injection) at high engine speeds [7]. Fig. 1b shows a typical engine control software realized as a sequence of *if* statements [6], [7]. The control strategy at a speed higher than ω_4 (i.e., $\omega > \omega_4$) only executes function *f1()* for single-injection, compared to (*f1()*–*f5()*) for quintuple-injection at ω_1 . Such engine control software, in general, makes the cost model discontinuous at the switching speeds (thus non-convex). The thick black

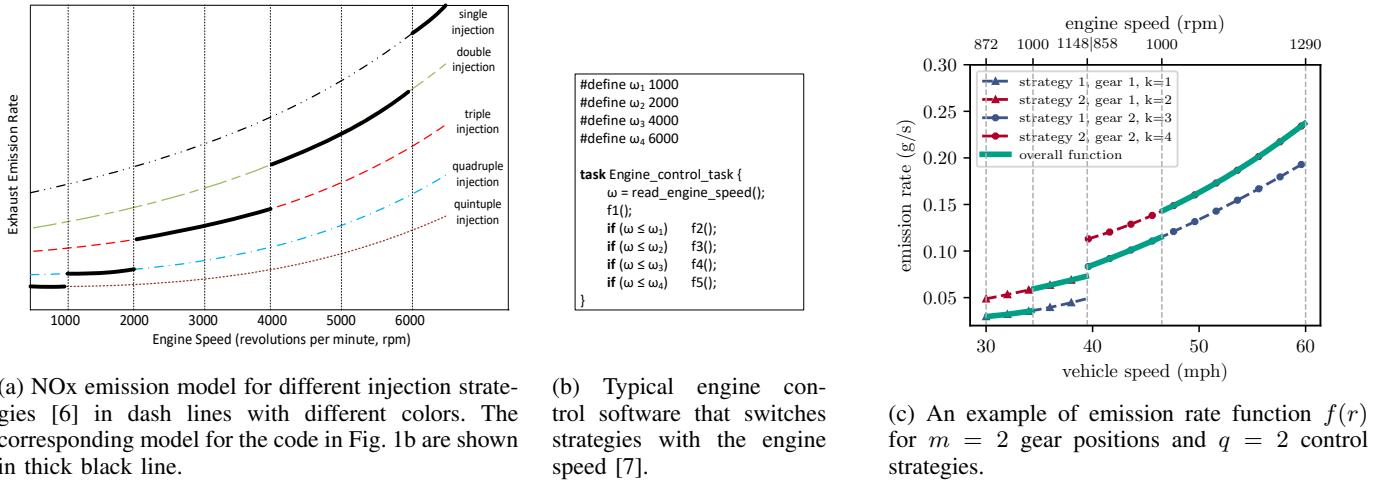


Fig. 1: Illustrations for the discontinuity of the emission rate function.

line in Fig. 1a gives an example of the overall NO_x emission model corresponding to the code in Fig. 1b.

Another source of discontinuity in the emission model is gear shifting, i.e., switching the gear so that the engine speed is within the fuel-economic range (see Fig. 2). When the gear is shifted, e.g., from position 9 to 10 at 40 mph, the engine speed falls from 1150 rpm to 850 rpm, which leads to a large difference of engine performance (e.g., emission rate) near the vehicle speed 40 mph. Therefore, the gear-shifting strategy also introduces points of discontinuity to the emission rate function with respect to vehicle speed. An example of the overall emission rate function is presented in Fig. 1c.

Timely transportation is a common requirement in the trucking industry, for three reasons [9]: (i) the nature of the goods (such as fresh food) [10]; (ii) service-level agreements to guarantee delivery delay such as those in Amazon¹, uShip² and Uber Freight³; and (iii) ease of scheduling and operation in the logistics [11]. For instance, mobile applications like uShip and Uber Freight collect freight transportation requests for truck operators, often associated with pickup and delivery time requirements.

In this work, we consider a common truck operation scenario where a long-haul truck drives across a national highway system. Our objective is to minimize the total emission subject to a hard deadline constraint by optimizing path planning and speed planning and leveraging adaptive engine control strategies and multi-speed transmission.

Our problem optimizes path planning and speed planning as in several previous studies [9], [15], [16]. However, we also consider adaptive engine control strategies and multi-speed transmission, which makes our problem uniquely challenging, since the overall emission is a discontinuous and non-convex function with respect to the vehicle speed due to the switching among these control strategies and gear positions at runtime. Table I summarizes the comparison between our work and

related studies. To our best knowledge, we are the first to study the problem of optimizing the operation of a long-haul heavy-duty truck subject to a hard deadline constraint, where the truck is equipped with multiple injection strategies and multi-speed transmission. The **contributions** are summarized as follows.

▷ We prove that our problem is NP-hard and show that switching among multiple engine control strategies and gear ratios makes the emission rate function discontinuous and non-convex, hence imposing a unique challenge compared to existing studies, which all deal with convex cost functions.

▷ We find that driving at a constant speed on one road segment is not optimal, and it is necessary to consider *heterogeneous speed profile* at one road segment. We explore the structures of this speed planning problem and formulate it into a convex programming form. We then further leverage the properties of the emission rate functions and derive an efficient method for the optimal speed profile that achieves three orders of magnitude runtime improvement than the convex program solver. We develop an efficient dual-based heuristic for both path planning and speed planning. Moreover, we give a sufficient condition under which the solution of our approach is optimal and an upper bound of the optimality gap when the condition is not satisfied.

▷ We conduct extensive simulations over the US highway network with 1,000 origin-destination pairs. The results show that our scheme reduces up to 46% emissions compared to the fastest path. We also find that heterogeneous speed profiles reduce up to 32% emission as compared to its homogeneous counterpart, thus are necessary to be considered in environmentally friendly truck operations.

II. RELATED WORK

Restricted Shortest Path (RSP). In an RSP problem, the cost and travel time associated with each road segment are both *fixed*. The objective is to find a path with the minimum total cost such that the total travel time is within a deadline constraint. However, existing approaches and results for RSP

¹Place an Order with Guaranteed Delivery, Amazon, <http://amazon.com>

²uShip, <https://www.uship.com/>

³Uber Freight, <https://freight.uber.com/>

Work	Path Planning	Speed Planning	Hard Deadline	Adaptive Fuel Injection	Transmission	Vehicle Type	Cost	Cost Model
RSP [12], [13], [14]	✓	✗	✓	✗	✗	Truck	Any	Constant
PASO [9], [15], [16]	✓	✓	✓	✗	✗	Truck	Fuel	Convex
Others, e.g., [17], [18], [19]	✗	✓	✗	✗	✓	Truck	Fuel	Nonlinear
Other, e.g., [20]	✗	✓	✓	✗	✓	Truck	Fuel	Nonlinear
Other, e.g., [21]	✓	✗	✗	✗	✗	Truck	Fuel	Constant
Others, e.g., [22], [23]	✗	✓	✗	✗	✓	General	Fuel	Nonlinear
Other, e.g., [24]	✗	✓	✓	✗	✓	General	Fuel	Nonlinear
Other, e.g., [25]	✓	✗	✗	✗	✓	General	Emission	Constant
VRP [26], [27]	✓	✓	✓	✗	✗	General	Emission	Nonlinear
This work	✓	✓	✓	✓	✓	Truck	Emission	Discontinuous

TABLE I: Comparison of our work and existing studies.

are not applicable to our problem, as the fixed cost and travel time for each road segment in RSP disregard the challenging design space of speed planning.

PAth selection and Speed Optimization (PASO) and its extensions. PASO [9] generalizes RSP with the additional design space of speed planning. An FPTAS and a dual-based heuristic have been developed to solve the challenging PASO problem [9]. An extension to a multi-task setting has been studied [15] to fulfill multiple transportation tasks under task pickup and delivery time window constraints. The design space of opportunistic driving [16] has been explored to leverage dynamic traffic conditions, wait for benign traffic conditions and reduce the cost. However, all these studies assume the fuel consumption model is convex, while in our problem the cost function is discontinuous and non-convex due to multiple injection strategies and multiple gear positions.

Other related studies. Reducing emission for vehicles has been studied extensively [26], [28], [25], [27]. Bektaş *et al.* [26] propose an extension of the Vehicle Routing Problem (VRP) called Pollution Routing Problem (PRP) with various optimization objectives including fuel consumption, CO₂ emission and travel time. Guo *et al.* [25] collect a number of real-world vehicle data and optimize the objectives of fuel consumption and emissions. Hellström *et al.* [17] embed the cost-speed tradeoff in the objective, a weighted sum of travel time and fuel consumption. They also use look-ahead information such as estimated road grade [29] to control the truck's speed profile under a given path. Boriboonsomsin *et al.* [21] present an eco-routing navigation system that determines the most fuel-economic path. Alam *et al.* [30] observe that improved fuel efficiency can be obtained by maintaining the platoon of trucks throughout a hill, motivating subsequent studies, e.g., [31], [32], [33], which focus on developing control strategies for truck platooning to save fuel.

Meanwhile, a continuous effort has been put into engine emission control from both the industry and the regulatory authorities; see a recent review in [34]. In particular, heavy-duty engines are improving at a much slower pace than light-duty ones [34]. However, all these studies focus on the optimization of the design and operation of various parts in the engine (e.g., fuel system and injection strategy [35], [36], [6], [37], exhaust after-treatment system [38]), and they do not consider the planning of path and/or speed profile. For example, Biondi *et al.* [6] present methods to optimize the switching speeds of multiple injection strategies at design time

against standard driving cycles (i.e., with fixed path and speed profile), while Peng *et al.* [39], [37] propose to adjust the switching speeds at runtime using predicted driving cycles.

We summarize the comparison of our work and related studies in Tab. I. To our best knowledge, we are the first to study the problem of minimizing emissions for a long-haul heavy truck equipped with multiple engine control strategies and multi-speed transmission. Compared to existing studies that also simultaneously optimize path planning and speed planning under a hard deadline constraint [9], [15], [16], the consideration of switching among multiple injection strategies and gear ratios makes our problem uniquely challenging due to the discontinuity and non-convexity of the emission rate function.

III. MODEL AND PROBLEM FORMULATION

We model a national highway network as a directed graph $G \triangleq (V, E)$ where an edge $e \in E$ represents a road segment, and a node $v \in V$ represents a point of junction for connected road segments. We denote the distance of an edge $e \in E$ by $D^e > 0$. We also denote the minimum (resp. maximum) speed of an edge e by r_l^e (resp. r_u^e). We assume the driving conditions to be homogeneous on each road segment, i.e., the grade and surface resistance are the same on a road segment (otherwise, we can break it into multiple segments). We consider the scenario where a truck travels from an origin $o \in V$ to a destination $d \in V$ across the highway network G within a hard deadline requirement T .

A. Emission Rate function

For a truck and its i -th engine control strategy, we denote the engine performance function by $\hat{f}_i(\omega, p) : \mathbb{R}^2 \rightarrow \mathbb{R}$ where ω is the engine speed in revolutions per minute (rpm), and p is the output power in kW. The output $\hat{f}_i(\omega, p)$ is the (instantaneous) emission rate (unit: g/s).

For a fixed gear position j , the engine speed is linear with respect to the vehicle speed (see Fig. 2). Therefore, we model the mapping from the vehicle speed r to the engine speed ω

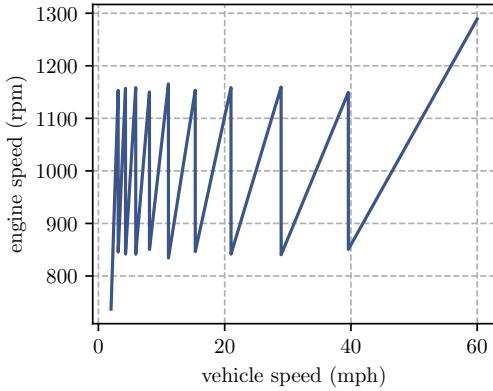


Fig. 2: A typical 10-speed transmission plot from the vehicle speed to the engine speed [40].

as a piece-wise linear function:⁴

$$\omega(r) = a_j r + b_j \quad \text{if } r \in (r_{j-1}^\omega, r_j^\omega] \quad (1)$$

for all gear positions $j \in \{1, \dots, m\}$, where m is the number of gear positions. Moreover, for a truck driving at a constant speed r on a road segment e , we model the output power $p(r)$ as a third-degree polynomial function [43], [44]:

$$p^e(r) = a_3^e r^3 + a_2^e r^2 + a_1^e r + a_0^e \quad (2)$$

Thus the emission rate function of a truck driving at i -th control strategy and j -th gear position over a segment e is

$$\tilde{f}_{ij}^e(r) = \hat{f}_i(a_j r + b_j, a_3^e r^3 + a_2^e r^2 + a_1^e r + a_0^e) \quad (3)$$

The overall emission rate function is given by

$$f^e(r) = \tilde{f}_{ij}^e(r), \text{ if } r \in (r_{j-1}^\omega, r_j^\omega] \text{ and } \omega(r) \in (\omega_{i-1}, \omega_i], \quad (4)$$

where ω_i is the switching speed from strategy i to $i+1$, $i \in \{1, \dots, q\}$ and q is the number of control strategies. Note that we can write the function in a standard piece-wise form as follows, since the domain for each piece is disjoint

$$f^e(r) = f_k^e(r) = \tilde{f}_{ij}^e(r) \quad \text{if } r \in (s_{k-1}, s_k] \quad (5)$$

for $k = \{1, \dots, n\}$. Here $n \leq mq$ is the number of pieces. We set the boundary points to the speed limits on edge e : $s_0 = r_l^e$, $s_n = r_u^e$. We then derive each s_k , $k = \{1, \dots, n-1\}$ as:

$$s_{k-1} = \frac{\omega_{i-1} - b_j}{a_j}, \quad s_k = \frac{\omega_i - b_j}{a_j}$$

$$i = k - q \left\lfloor \frac{k-1}{q} \right\rfloor, \quad j = \left\lfloor \frac{k-1}{q} \right\rfloor + 1$$

Fig. 1c gives an illustrative example of the emission rate function $f^e(r)$ with $n = 4$ pieces. There are three points of discontinuity in total: two of them are due to switching of

⁴ In practice, drivers often shift the gear position to keep the engine speed within a “sweet spot” range [41], [42] to ensure smooth driving and engine performance. In this study, we consider that the driver follows such a gear-shifting strategy during the whole trip. Consequently, this paper does not optimize the vehicle-to-engine speed map. That is, given the vehicle speed r , the engine speed is determined by a fixed piece-wise linear function $\omega(r)$ like Fig. 2.

control strategies at 1,000 rpm and one of them is due to switching of gear positions at 40 mph.

In this paper we make the following assumptions on the high-level emission rate $f^e(r)$:

- **piece-wise convex:** each function f_k^e is convex over the interval $[r_l^e, r_u^e]$;
- **staircase-shaped:** they satisfy that

$$\forall k_1 < k_2 \in [n], \forall r \in [r_l^e, r_u^e], \quad f_{k_1}^e(r) < f_{k_2}^e(r) \quad (6)$$

where $[n]$ is the set of positive integers no larger than n .

We remark that the above assumptions on the cost model are realistic for a number of metrics related to eco-friendly truck operations. For example, the fuel rate can be described as an increasing polynomial function of the vehicle speed, verified with both theoretical and empirical studies [9]. The amount of CO₂ emission is roughly proportional to the amount of fuel consumption (e.g., about 10.18kg per gallon of diesel, or 8.887kg per gallon of gasoline) [45]. Hence it exhibits the same characteristics as the fuel-rate function. Similarly, the NOx emission can be modeled as an exponential function that remains convex in the engine speed range based on extensive simulation data [6]. The rationale behind the second assumption (6) is that a more complicated and time-consuming fuel injection strategy makes sense only if it provides some benefit such as reduced emission, but its complexity makes it only feasible at lower speeds [6].

While our approach generally applies to any cost function that satisfies the above assumptions, we focus on minimizing the emission cost for concreteness in the rest of this paper. Those assumptions are also verified later in our simulations.

B. Total Emission Function

In contrast to convex cost rate functions discussed in existing literature [9], [15], [16], the emission rate function $f^e(r)$ we study is discontinuous and non-convex which makes it non-trivial to optimize. In particular, while it is optimal to drive at a constant speed to travel through one road segment with a convex cost rate function [9], this approach is sub-optimal for discontinuous cost functions. In such cases, operating within two speed ranges can yield lower costs, as demonstrated in Example 1 in Appendix VIII-A. Therefore, it is necessary to consider a *heterogeneous speed profile* on one road segment to minimize emission costs. In the following, we shall study how to optimize the heterogeneous speed profile and compute the total emission function $c^e(t)$ that gives the minimum emission cost (in grams) for a truck to traverse the edge e with the travel time of t . We first present a similar result to [9, Lemma 1].

Lemma 1. *If the total travel time t_i following the speed in the range $(s_{i-1}, s_i]$ of the i -th piece $f_i(r)$ is given, then the optimal speed profile is to maintain some constant speed r_i for the whole duration t_i .*

Proof. Similar to [9, Lemma 1], it is proven by applying the continuous Jensen’s inequality to the convex function $f_i(r)$. \square

Given Lemma 1, we are now ready to formulate the problem of optimizing the speed profile to pass an edge e , with the

length of the edge D^e , and the total travel time t^e . We denote r_k and t_k as the selected speed and travel time for the k -th piece, respectively. We introduce the auxiliary variable $d_k = r_k \cdot t_k$ to denote the driving distance for the k -th piece and formulate the problem as follows⁵

$$c^e(t^e) = \min_{t_k \geq 0, d_k \geq 0, \forall k \in [n]} \sum_{k \in [n]} t_k \cdot f_k\left(\frac{d_k}{t_k}\right) \quad (7a)$$

$$\text{s.t. } \sum_{k \in [n]} t_k = t^e, \quad \sum_{k \in [n]} d_k = D^e \quad (7b)$$

$$s_{k-1} t_k \leq d_k \leq s_k t_k, \forall k \in [n] \quad (7c)$$

The objective (7a) is to minimize the total emission on the road segment. The constraint (7b) makes sure that the total driving distance is equal to the length D^e of the road segment and that the total travel time is equal to t^e . The constraint (7c) is to make sure the driving speed $\frac{d_k}{t_k}$ is within the corresponding speed range for each piece k .

Lemma 2. *The speed planning problem in (7) is a convex optimization problem.*

Proof. In the objective, each summand $t_k \cdot f_k\left(\frac{d_k}{t_k}\right)$ is the perspective of the convex function f_k , thus is also convex [47] in d_k and t_k . Therefore, the objective function is the sum of convex functions and thus is convex. Then convexity of the objective combined with the observation that all the constraints (7b)-(7c) are linear, giving the desired result. \square

Therefore, we can efficiently compute the total emission function $c^e(t^e)$ by solving problem (7) as a standard convex program. This approach, however, only exploits the piece-wise convexity of the emission rate function $f^e(r)$. In Section IV, we will delve deeper into the unique structures of this problem and introduce a more efficient method. Notably, the proposed method offers a runtime that is three orders of magnitude faster than directly solving the convex program (7), as detailed in Table III in Section VI.

C. Problem Formulation

In this paper, we consider the problem of minimizing the total emission cost for a truck to travel from an origin o to a destination d within a hard deadline T . We consider the problem from the driver's perspective, where the truck parameters and engine parameters are predefined. The problem inputs include the graph $G = (V, E)$, the emission rate function $f^e(r)$ for each road segment e that incorporates the truck's and its engine's characteristics, the vehicle speed to engine speed map $\omega(r)$, the speed limits r_l^e and r_u^e for each edge e , the origin o , the destination d , and the deadline T .

The solution to our problem include a path from o to d and the vehicle speed(s) on each road segment along the path. The

⁵In this paper, we ignore the emission cost due to acceleration and deceleration during the speed transition phase, because this phase usually spans over only several hundred feet [46] while the length of one road segment is several miles or longer. Meanwhile, as we will show in Lemma 3, it is sufficient to drive with at most two different speeds to achieve the minimum emission cost at one road segment and our method suggests to travel for more than a mile every speed change on average (see Fig. 6).

constraints include the speed limits on each road segment, and a hard deadline that requires the total travel time to be no more than T . In particular, we formulate our problem as follows:

$$\min_{x \in \mathcal{X}, t \in \mathcal{T}} \sum_{e \in E} x^e \cdot c^e(t^e) \quad (8a)$$

$$\text{s.t. } \sum_{e \in E} x^e \cdot t^e \leq T, \quad (8b)$$

where \mathcal{X} defines a simple path from o to d

$$\mathcal{X} \triangleq \left\{ x : x^e \in \{0, 1\}, \forall e \in E, \text{ and} \sum_{e \in \text{out}(v)} x^e - \sum_{e \in \text{in}(v)} x^e = 1_{\{v=o\}} - 1_{\{v=d\}}, \forall v \in V \right\}.$$

Here $1_{\{\cdot\}}$ is the indicator function, $\text{in}(v) \triangleq \{(u, v) : (u, v) \in E\}$ is the set of incoming edges of node v , $\text{out}(v) \triangleq \{(v, u) : (v, u) \in E\}$ is the set of outgoing edges of node v . The set \mathcal{T} captures the speed limits of all roads, which is defined as

$$\mathcal{T} \triangleq \{t : t_l^e \leq t^e \leq t_u^e, \forall e \in E\},$$

where $t_l^e = \frac{D^e}{r_u^e}$ and $t_u^e = \frac{D^e}{r_l^e}$ are the minimum and maximum travel times of traversing the edge e , respectively.

We remark that there are two challenges to solve the problem (8). The first challenge comes from the discontinuity and the non-convexity of the emission rate function $f^e(r)$, which makes it more difficult to compute $c^e(t)$ than previous work [9], [15], [16]. The second challenge comes from the combinatorial nature of the problem 8, which makes it NP-hard by the following theorem.

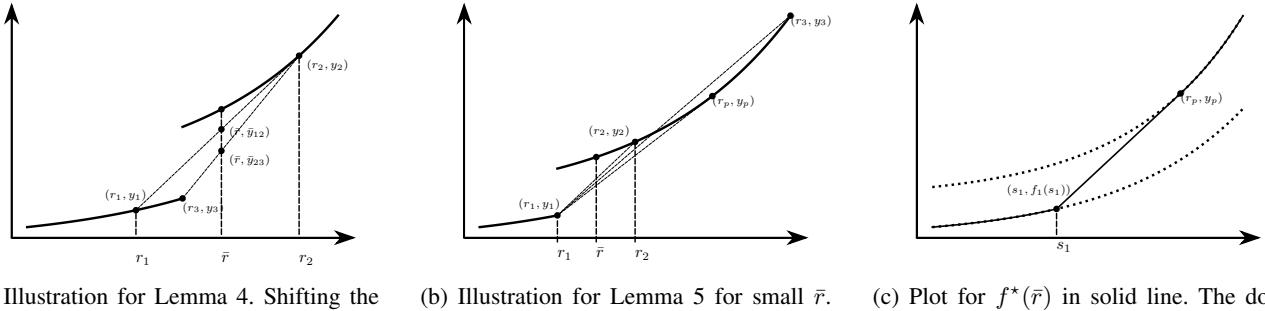
Theorem 1. *The problem (8) is NP-hard.*

Proof. This directly follows the fact that the NP-complete problem PASO [9] is a special case of our problem, where PASO only considers one control strategy (i.e., $n = 1$). \square

In the following two sections, we shall tackle those two challenges. We shall provide an efficient method for solving the speed planning problem (7) by indentifying special structures of the problem in Sec. IV. In Sec. V, we shall provide an efficient dual based method that solves the overall path planning and speed planning problem (8).

IV. SPEED PLANNING

In this section, we shall present an efficient method for computing $c^e(t)$ by exploring the properties of the emission rate function $f^e(r)$. For ease of presentation, we omit the superscript e from the notations in the rest of the section, e.g.,



(a) Illustration for Lemma 4. Shifting the left point (r_1, y_1) to right point (r_3, y_3) decreases the cost \bar{y} .

(b) Illustration for Lemma 5 for small \bar{r} . The objective is minimized at the generalized right point of tangency.

(c) Plot for $f^*(\bar{r})$ in solid line. The dotted lines are functions $f_1(r_1)$ and $f_2(r_2)$ respectively.

Fig. 3: Illustrations for ideas of solving the speed planning problem.

$f^e(r)$ is simplified as $f(r)$. We now rewrite problem (7) in the following equivalent form to better illustrate our method

$$f^*(\bar{r}) = \min_{\beta_i \geq 0, r_i \geq 0, \forall i \in [n]} \sum_{i \in [n]} \beta_i \cdot f_i(r_i) \quad (9a)$$

$$\text{s.t. } \sum_{i \in [n]} \beta_i = 1, \quad (9b)$$

$$\sum_{i \in [n]} \beta_i \cdot r_i = \bar{r} \quad (9c)$$

$$r_i \in (s_{i-1}, s_i], \forall i \in [n] \quad (9d)$$

where $\beta_i = \frac{t_i}{t}$, and $\bar{r} = \frac{D}{t}$. Essentially, problem (9) is to find a set of speeds, one r_i for each piece i , such that their convex combination is equal to the average speed $\bar{r} = \frac{D}{t}$ over the edge, and the weighted average of the emission rate is minimized (denoted as $f^*(\bar{r})$). Note that this reformulation is not convex anymore due to the nonlinear equality constraint (9c) and the non-convex objective. However, problem (9) has special structures that allow us to design an efficient method for solving it. The proposed method is three orders of magnitude faster than directly solving the convex program (7) (cf. Tab. III in Sec. VI). In the following, we shall present this method for the function f with $n > 1$ pieces, because when $n = 1$, the optimal solution is simply $r_1 = \bar{r}$ and $\beta_1 = 1$ by Lemma 1. We outline the derivation steps as follows:

- We first show by Lemma 3 that we only need to focus on at most two out of n pieces of function f to obtain the optimal solution.
- Given i -th piece and j -th piece of function f_i and f_j with $1 \leq i < j \leq n$, we show by Lemma 4 that how to choose the speed r_i for i -th piece and show by Lemma 5 that how to choose the speed r_j for j -th piece. Therefore, we can obtain the minimum-cost solution on two specific pieces f_i, f_j .
- Finally, we present our efficient method by searching over $O(n^2)$ pairs of candidate pieces and identifying the optimal solution.

Next, we present the details on how to derive our efficient method for solving the speed planning problem (9). Suppose the average speed \bar{r} lies in the i_r -th piece of the function f , i.e., $\bar{r} \in (s_{i_r-1}, s_{i_r})$. Let $\bar{r}^* = [\bar{r}_1^*, \dots, \bar{r}_n^*]$ and $\vec{\beta}^* = [\beta_1^*, \dots, \beta_n^*]$

be an optimal solution to problem (9). Then we have the following lemma.

Lemma 3. *There exists an optimal solution $\vec{\beta}^*$ that has at most two non-zero entries.*

Proof. See Appendix VIII-B. \square

Lemma 3 suggests that we can focus on finding only two points on the curve of the function f whose convex combination is minimized. Fig. 3a gives a geometric illustration of our goal. We want to find two points (r_1, y_1) and (r_2, y_2) on two pieces of function f such that the interpolated point (\bar{r}, \bar{y}_{12}) has the minimum function value \bar{y}_{12} . To investigate how to choose two points on two given pieces. We first fix the right point (r_2, y_2) and shift the left point (r_1, y_1) to reduce the cost, which yields the following lemma.

Lemma 4 (Illustrated by Fig. 3a). *Consider a convex function $g(r)$, a fixed point $\bar{r} \leq r_2$, and a point (r_2, y_2) above the curve of $y = g(x)$, i.e., $y_2 > g(r_2)$. For every $r_1 < \bar{r}$, let (\bar{r}, \bar{y}) be the convex combination of two points $(r_1, y_1 = g(r_1))$ and (r_2, y_2) , i.e.,*

$$\bar{y} = \frac{y_2 - y_1}{r_2 - r_1} (\bar{r} - r_1) + y_1 \quad (10)$$

Then for every fixed \bar{r} and $r_2 \geq \bar{r}$, \bar{y} is monotonically non-increasing with r_1 for $r_1 < \bar{r}$. In addition, $\bar{y} \geq g(\bar{r})$.

Proof. See Appendix VIII-C. \square

Fig. 3a illustrates the geometric interpretation of Lemma 4. It suggests that given fixed right point (r_2, y_2) , we should always choose larger point (r_3, y_3) on the right boundary of the same piece to decrease the objective. Next, we consider the case where the left point (r_1, y_1) is fixed and we want to shift the right point (r_2, y_2) to reduce the cost, which is summarized in the following lemma.

Lemma 5 (Illustrated by Fig. 3b). *Consider a convex function $g(r)$ and a fixed point (r_1, y_1) that is below the curve of $y = g(r)$, i.e., it satisfies $y_1 < g(r_1)$. Define $(p, y_p = g(p))$ as the generalized right point of tangency from (r_1, y_1) to its right hand side of the curve $y = g(r)$, i.e., $p > r_1$ satisfies the following equation*

$$\partial_- g(p) \leq \frac{g(p) - y_1}{p - r_1} \leq \partial_+ g(p) \quad (11)$$

where $\partial_- g(p)$ and $\partial_+ g(p)$ are the left and right derivatives of g at p .

Given $\bar{r} > r_1$, for every $r_2 \geq \bar{r}$, let (\bar{r}, \bar{y}) be the convex combination between (r_1, y_1) and $(r_2, y_2 = g(r_2))$, that is,

$$\bar{y} = \frac{y_2 - y_1}{r_2 - r_1} (\bar{r} - r_1) + y_1 \quad (12)$$

Then \bar{y} has the following properties:

- If $\bar{r} \leq p$, then \bar{y} is monotonically non-increasing with r_2 until p , and monotonically non-decreasing with r_2 afterwards. In addition, $\forall \bar{r} \leq r_2 \leq p, \bar{y} \leq g(\bar{r})$.
- If $\bar{r} \geq p$, then \bar{y} is monotonically non-decreasing with r_2 . In addition, $\forall r_2 \geq \bar{r}, \bar{y} \geq g(\bar{r})$.

Proof. See Appendix VIII-D \square

Fig. 3b illustrates the geometric interpretation of Lemma 5 when \bar{r} is small. It suggests that given fixed left point (r_1, y_1) , we want the right point (r_2, y_2) as close to (r_p, y_p) as possible to reduce to cost. Meanwhile, when \bar{r} is larger than a threshold p , we can simply drive at the constant speed \bar{r} .

The efficient method. With Lemma 4 and Lemma 5, we can directly obtain the optimal solution if the optimal piece(s) are known. Combined with Lemma 3, we notice that it is sufficient to check $O(n^2)$ pairs of points to obtain the optimal solution. In particular, we summarize our efficient method in the following:

Theorem 2. For any two pieces i, j such that $1 \leq i < i_r \leq j \leq n$, we denote by p_{ij} the generalized right point of tangency from point $(s_i, f_i(s_i))$ to the curve $y = f_j(r)$. We also denote the clamped p_{ij} as

$$\tilde{p}_{ij} = \begin{cases} s_{j-1}, & \text{if } p_{ij} < s_{j-1}, \\ p_{ij}, & \text{if } p_{ij} \in [s_{j-1}, s_j], \\ s_j, & \text{if } p_{ij} > s_j \end{cases} \quad (13)$$

Then by Lemma 4 and Lemma 5, the minimum objective and the corresponding solution on pieces i, j are given by:

$$\begin{aligned} r_i &= s_i, \quad r_j = \tilde{p}_{ij} \\ \beta_i &= \frac{\bar{r} - r_i}{r_j - r_i}, \quad \beta_j = \frac{r_j - \bar{r}}{r_j - r_i} \\ y_{ij} &= \beta_i f_i(s_i) + \beta_j f_j(\tilde{p}_{ij}) \end{aligned}$$

We additionally define the case $i = j = i_r$:

$$r_{i_r} = \bar{r}, \quad \beta_{i_r} = 1, \quad y_{i_r i_r} = f_{i_r}(\bar{r}).$$

Then the optimal solution and objective to the problem (9) are identified by:

$$\begin{aligned} (i^*, j^*) &= \arg \max_{1 \leq i \leq i_r, i_r \leq j \leq n} y_{ij} \\ y^* &= \max_{1 \leq i \leq i_r, i_r \leq j \leq n} y_{ij} \end{aligned}$$

Proof. The proof directly follows from combining Lemma 3, Lemma 4, and Lemma 5. \square

Therefore, we only need to check $i_r(n - i_r + 1) \leq n^2$ number of candidate points to solve the problem (9). There are other techniques to reduce the number of candidates by further exploring the relationship between different pairs of pieces.

However, we remark that the proposed method is efficient enough in practice because the number of pieces n is relatively small. In fact, since we are considering a truck travelling across the highway system with vehicle speed larger than 20 mph on most road segments, there are only $m = 3$ gear positions involved in our scenario (cf. Fig. 2). If we consider an engine with $q = 3$ control strategies, then the number of pieces $n \leq mq = 9$ is indeed small. Therefore, Theorem 2 provides an efficient method for computing $c(t) = t \cdot f^*(D/t)$. It allows us to efficiently solve the overall emission minimization problem (8) in the next section. Fig. 3c gives an illustration of $f^*(\bar{r})$. It can be seen as the continuous version of the original discontinuous function $f(r)$ by drawing a line between the break point $(s_1, f_1(s_1))$ and (r_p, y_p) . As one may observe from Fig. 3c, we can show that $f^*(r)$ is convex with respect to r and thus $c(t) = t \cdot f^*(D/t)$ is also convex, as stated in the following proposition.

Proposition 1. $c(t)$ is convex over $[t_l, t_u]$.

Proof. See Appendix VIII-E. \square

In the next section, we shall leverage the efficient computation method and convexity of $c(t)$ to provide an efficient dual-based method that solves the overall emission minimization problem (8).

V. AN EFFICIENT DUAL-SUBGRADIENT ALGORITHM WITH PERFORMANCE GUARANTEE

With the efficient method to the speed planning for a given travel time t on an edge e and hence a fast computation for $c^e(t)$, we consider the overall path planning and speed planning problem (8) which now amounts to finding a path and assign a travel time to each edge on the path, such that the emission is minimized and the subject to the hard deadline constraint T . We shall design an efficient heuristic based on Lagrangian relaxation and derive a theoretical condition under which our heuristic outputs the optimal solution. As illustrated in the simulations, our heuristic quickly finds close-to-optimal solutions for the scale of the US national highway network.

A. Lagrangian Relaxation and Dual Problem

We introduce a Lagrangian dual variable $\lambda \geq 0$, and derive the Lagrangian relaxation for problem (8) as

$$\begin{aligned} L(x, t, \lambda) &\triangleq \sum_{e \in E} x^e \cdot c^e(t^e) + \lambda \cdot \left(\sum_{e \in E} x^e \cdot t^e - T \right) \\ &= \sum_{e \in E} x^e \cdot (c^e(t^e) + \lambda \cdot t^e) - \lambda \cdot T. \end{aligned}$$

The corresponding dual function is given by

$$D(\lambda) \triangleq \min_{x \in \mathcal{X}, t \in \mathcal{T}} L(x, t, \lambda).$$

and the dual problem of the original problem (8) is

$$\max_{\lambda \geq 0} D(\lambda)$$

Given λ , we can obtain the value of $D(\lambda)$ by computing the shortest path with easily computed weights for each edge. In particular, we have

$$\begin{aligned} D(\lambda) &= -\lambda T + \min_{x \in \mathcal{X}} \left[\min_{t \in \mathcal{T}} \sum_{e \in E} x^e \cdot (c^e(t^e) + \lambda t^e) \right] \\ &= -\lambda T + \min_{x \in \mathcal{X}} \sum_{e \in E} x^e \cdot \min_{t_l^e \leq t^e \leq t_u^e} (c^e(t^e) + \lambda t^e) \\ &\stackrel{(a)}{=} -\lambda T + \min_{x \in \mathcal{X}} \sum_{e \in E} x^e \cdot [c^e(t^{e*}(\lambda)) + \lambda \cdot t^{e*}(\lambda)] \\ &\stackrel{(b)}{=} -\lambda T + \min_{x \in \mathcal{X}} \sum_{e \in E} x^e \cdot w^e(\lambda) \\ &\stackrel{(c)}{=} -\lambda T + \sum_{e \in p^*(\lambda)} w^e(\lambda), \end{aligned} \quad (14)$$

Here $t^{e*}(\lambda)$ in (a) is defined as

$$t^{e*}(\lambda) \triangleq \arg \min_{t^e \leq t^e \leq t_u^e} (c^e(t^e) + \lambda t^e). \quad (15)$$

That is, $t^{e*}(\lambda)$ is the optimal travel time that minimizes $c^e(t^e) + \lambda t^e$ for edge $e \in E$, $w^e(\lambda)$ in (b) is the corresponding optimal cost

$$w^e(\lambda) \triangleq c^e(t^{e*}(\lambda)) + \lambda \cdot t^{e*}(\lambda), \quad (16)$$

and $p^*(\lambda)$ in (c) is the resulting minimum-cost path where each edge is associated with an edge cost of $w^e(\lambda)$. Given a value to the dual variable λ , Equation (14) suggests that we can figure out $D(\lambda)$, the value of the dual function, by finding a shortest path with each edge e assigned an edge cost of $w^e(\lambda)$. In the following, we first derive an analytical solution to $t^{e*}(\lambda)$ and hence $w^e(\lambda)$ for each edge $e \in E$, then in Section V-B we propose an iterative procedure to find an appropriate value for λ .

From Section IV, we can efficiently compute the value of $c^e(t)$ by checking a small number of candidates. Moreover, Proposition 1 shows that $c^e(t)$ is convex (but potential piecewise), hence it allows left and right derivatives. The following lemma provides an analytical solution to $t^{e*}(\lambda)$ and hence $w^e(\lambda)$.

Lemma 6. *In case that $\partial_+ c^e(t_l^e) \leq -\lambda \leq \partial_- c^e(t_u^e)$, define t^* as any t such that $\partial_- c^e(t) \leq -\lambda \leq \partial_+ c^e(t)$, which is well-defined since the derivatives of the convex function $c^e(t)$ are non-decreasing. Then $t^{e*}(\lambda)$ is given as*

$$t^{e*}(\lambda) = \begin{cases} t_l^e, & \text{if } \lambda + \partial_+ c^e(t_l^e) > 0 \\ t^*, & \text{if } \partial_+ c^e(t_l^e) \leq -\lambda \leq \partial_- c^e(t_u^e) \\ t_u^e, & \text{if } \lambda + \partial_- c^e(t_u^e) < 0 \end{cases} \quad (17)$$

Proof. Observe that $c^e(t) + \lambda t$ is also convex with respect to t . Hence its derivatives are non-decreasing.

If $\lambda + \partial_+ c^e(t_l^e) > 0$, then $c^e(t) + \lambda t$ is non-decreasing for $t \geq t_l^e$, hence its minimum is achieved at the lower bound of t , i.e., t_l^e . If $\lambda + \partial_- c^e(t_u^e) < 0$, then $c^e(t) + \lambda t$ is non-increasing for $t \leq t_u^e$, hence its minimum is achieved at the upper bound of t , i.e., t_u^e .

If $\partial_+ c^e(t_l^e) \leq -\lambda \leq \partial_- c^e(t_u^e)$, then the derivatives of $c^e(t) + \lambda t$ remain to be non-positive for $t < t^*$, and then are

always non-positive for $t > t^*$, hence its minimum is achieved at t^* . \square

When $\partial_+ c^e(t_l^e) \leq -\lambda \leq \partial_- c^e(t_u^e)$, we can find t^* with a binary search scheme, because the derivatives of the convex function c^e is non-decreasing. The complexity is thus $O(n^2 \log \lceil \frac{t_u^e - t_l^e}{\epsilon_t} \rceil)$ where n is the number of control strategies and ϵ_t is the error tolerance for t .

B. Our Heuristic Algorithm

For a given λ , we define the total travel time of the minimum-cost path $p^*(\lambda)$ as

$$\delta(\lambda) \triangleq \sum_{e \in p^*(\lambda)} t^{e*}(\lambda), \quad (18)$$

We introduce an important observation on $\delta(\lambda)$ below.

Lemma 7. *$\delta(\lambda)$ is non-increasing over $\lambda \in [0, +\infty)$.*

Proof. Refer to [9, Thm. 3], which is still applicable to our problem since it only uses the facts that $t^{e*}(\lambda)$ minimizes $c^e(t) + \lambda t$ and $p^*(\lambda)$ is the minimum-cost path. \square

By the Lagrangian dual relaxation, the value of $D(\lambda)$ as calculated in (14) is always a lower bound of the minimized emission to the original problem (8). Hence, we observe that the Lagrangian dual variable λ^* with $\delta(\lambda^*) = T$ defines the optimal solution $\mathcal{P}^*(\lambda^*)$. By Lemma 7, our heuristic suggests to use a binary-search scheme to update λ to approach λ^* , by comparing $\delta(\lambda)$ with T . The details of our heuristic are described in Algorithm 1, where ϵ_λ is the error tolerance for λ . λ can be interpreted as a price on the delay. Hence λ_{\max} can be set as the upper bound on the emission rate.

In the algorithm, whenever we find a λ such that $\delta(\lambda) = T$ (Line 8), it must be the optimal value λ^* . If $\delta(\lambda) > T$ (Line 10), then the deadline constraint is violated, and we set λ as the new lower bound λ_l . Otherwise (Line 12), we set λ as the new upper bound λ_u , and update the current best solution p^* and t^* .

Algorithm 1 Our Heuristic Approach

```

1: procedure
2:   Set  $\lambda_l = 0$  and  $\lambda_u = \lambda_{\max}$ 
3:   while  $\lambda_u - \lambda_l > \epsilon_\lambda$  do
4:     Set  $\lambda = \frac{\lambda_l + \lambda_u}{2}$ 
5:     Obtain  $t^{e*}(\lambda)$  according to Lemma 6 for all  $e \in E$ 
6:     Set  $w^e(\lambda)$  according to Equation (16) for all  $e \in E$ 
7:     Get the shortest path  $p^*(\lambda)$  from  $o$  to  $d$  in  $G$ 
8:     if  $\delta(\lambda) = T$  then
9:       return  $p^*(\lambda)$  and  $\{t^{e*}(\lambda), \forall e \in E\}$ 
10:    else if  $\delta(\lambda) > T$  then
11:      Set  $\lambda_l = \lambda$ 
12:    else
13:      Set  $\lambda_u = \lambda$ ,  $p^* = p^*(\lambda)$ , and  $\{t^{e*} = t^{e*}(\lambda), \forall e \in E\}$ 
14:    return  $p^*$  and  $\{t^{e*}, \forall e \in E\}$ 

```

We remark that our heuristic has a strong theoretical performance guarantee. That is, Algorithm 1 always returns a

feasible solution as long as the problem is feasible. Moreover, we give a sufficient condition under which the solution of our approach is optimal and an upper bound of the optimality gap when the condition is not satisfied.

Theorem 3. *If Algorithm 1 returns in Line 9, then the returned solution is optimal to our problem. Otherwise if Algorithm 1 returns in Line 14, the returned solution $\mathcal{S} = (p^* \text{ and } \{t^{e*}, \forall e \in E\})$ satisfies the deadline constraint T and hence is feasible. Furthermore, it has the following theoretical performance guarantee:*

$$C(\mathcal{S}) - OPT \leq \lambda^\times \cdot (T - \delta(\lambda^\times)), \quad (19)$$

where $C(\mathcal{S})$ is the total emission of the solution \mathcal{S} , OPT is the optimal emission of our problem, and λ^\times is the value of the dual variable corresponding to the returned solution \mathcal{S} .

Proof. By Equation (14), for a given dual variable λ , the duality gap of our problem is:

$$\text{Duality Gap} = \lambda \cdot (T - \delta(\lambda)), \quad (20)$$

hence the theorem holds. \square

Time Complexity. We now analyze the time complexity of Algorithm 1. The total number of iterations is $O(\log \frac{\lambda_{\max}}{\epsilon_\lambda})$. Within each iteration, the calculation of the optimal $t^{e*}(\lambda)$ and $w^e(\lambda)$ for all edges takes time $O(n^2 M \cdot \log \lceil \frac{t_{\max} - t_{\min}}{\epsilon_t} \rceil)$ where M is the number of edges in the highway network, n is the total number of control strategies of the engine, t_{\max} (resp. t_{\min}) is the maximum (resp. minimum) travel time among all edges. The step of finding the shortest path takes time $O(M + N \cdot \log N)$, where N is the number of vertices. Overall, Algorithm 1 has a time complexity of $O\left(\log \frac{\lambda_{\max}}{\epsilon_\lambda} \left(n^2 M \cdot \log \lceil \frac{t_{\max} - t_{\min}}{\epsilon_t} \rceil + M + N \cdot \log N \right)\right)$.

VI. PERFORMANCE EVALUATION

In this section, we present simulation results with real-world traces to evaluate the performance of our algorithm. Our objectives are i) to study the impact of heterogeneous speed over one road segment; ii) to study the impact of the deadline on the performance of different approaches. iii) to study the performance of the proposed approach as compared to the conceivable alternatives;

A. Simulation Setup

Transportation Network. We collect the highway network data from the Map-based Educational Tools for Algorithm Learning (METAL) project [48]. The constructed network consists of 84504 nodes and 178238 directed edges. The grade of each road segment is derived from the elevations of its end nodes provided by the Shuttle Radar Topography Mission (SRTM) [49] project.

Traffic Data. We set the maximum speed r_u^e of a road segment e as the historical average speed by collecting real-time speed data from HERE map [50] for two weeks. The minimum speed r_l^e is manually set to be $r_l^e = \min\{30\text{mph}, r_u^e\}$.

Origin	Destination	Distance (miles)	Value (billion USD)
Los Angeles CA	Columbus OH	1977	17.725
Los Angeles CA	Dallas-Fort Worth TX	1240	12.247
Los Angeles CA	Chicago IL	1745	11.293
Los Angeles CA	Nashville TN	1780	10.718
Los Angeles CA	Houston TX	1373	7.837

TABLE II: Five popular origin-destination pairs from Los Angeles. Here the distance means the straight-line distance between the origin and destination.

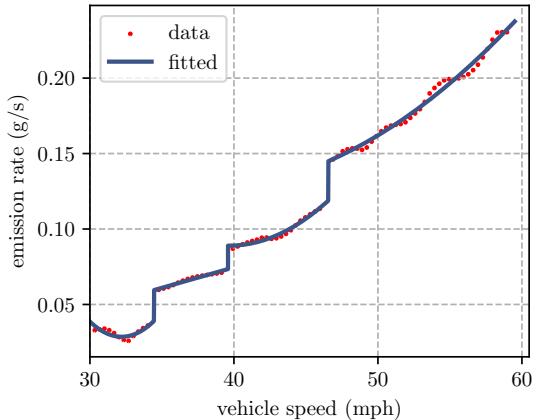


Fig. 4: The fitted emission rate function with respect to the vehicle speed with switching speed $\omega_1 = 1000$ rpm.

Origin-destination pair. We collect origin-destination pairs from the Freight Analysis Framework (FAF) [51]. We select 1,000 origin-destination pairs with distances longer than 1,000 miles. Those pairs represented 950 billion dollars of freight by trucks in 2017. Tab. II illustrates a subset of the selected origin-destination pairs in the US starting from Los Angeles.

Emission Model. We consider a class-8 heavy-duty truck Kenworth T800 with a 36-ton full load [52]. We collect the emission data for a typical four-stroke six-cylinder diesel engine [41] using the engine simulation software Diesel-RK [53]. We consider two fuel injection strategies: the traditional single injection and the cost-effective triple injection. We use the 10-speed transmission data from [40]. We consider the case where the driver follows the gear-shifting strategy that maintains the engine speed at around 1,000 RPM. The resulting engine speed to vehicle speed mapping is illustrated in Fig. 2. We then fit each piece of fuel rate function by a three-order polynomial. As shown in Fig. 4, the overall emission rate function satisfies the two assumptions in Sec. III.

Baseline Comparison. In the simulations, we implement our algorithm in Julia [54]. We run all the simulations on a desktop with 13th Gen Intel(R) Core(TM) i5-13400F processor and 64 GB RAM. We implement and compare the following approaches:

- **FAST:** The fastest path driving at its maximum speed. Given a graph G , we set the weight of each edge e as its minimum travel time D^e/r_u^e . We then find the fastest path from the origin to the destination in G using the Dijkstra algorithm.
- **MFI:** Our heuristic algorithm with both single-injection

and triple-injection control strategies. By default, we set the switching speed to 1,000 rpm. We set ϵ_λ and ϵ_t to be both 10^{-6} .

- **MFI-NS:** Same as MFI but with a constant speed over one road segment. That is, we use $c^e(t) = t \cdot f^e(D^e/t^e)$ in Algorithm 1. We use the golden-section search method to obtain $t^{e*}(\lambda)$ in (15).
- **SFI:** Our heuristic algorithm with only the single-injection strategy. That is, the emission rate function $f^e(r)$ only reflects the single-injection strategy. Note that the $f^e(r)$ is still non-convex due to different gear positions at different driving speeds.
- **TFI:** Our heuristic algorithm with only the triple injection strategy. That is, the emission rate function $f^e(r)$ only reflects the triple-injection strategy. This approach gives a performance upper bound for MFI, but it is not computationally feasible for the engine's software due to the high load of the triple-injection strategy.

Deadline Selection and Performance Evaluation. Given the origin and destination, we denote T^f as the minimum traveling time by FAST. We then set the deadline T as $T = \rho \cdot T^f$, where the ρ ranges from 1.05 to 2.0 in the simulation. We call ρ the *delay factor*. For two algorithms \mathcal{A}_1 and \mathcal{A}_2 , with emission cost C_1 and C_2 , the *relative emission reduction* of \mathcal{A}_1 as compared to \mathcal{A}_2 is given by $(C_2 - C_1)/C_2 \times 100\%$.

B. A Case Study

We first present a case study from Los Angeles, CA, to Columbus, OH, which represents the first row in Tab. II. We set the switching speed to 1,000 rpm and the deadline to be $T = 1.5T_f$ to better illustrate the benefit of heterogeneous speed planning. Using our MFI approach, we determine an emission-efficient solution that includes both a path and a speed profile, depicted by the blue lines in Figure 5. For comparison, we also present a modified speed profile that remains constant over each road segment, shown by the red lines in the same figure. To provide a clearer view, we zoom into a section of the trip in Figure 5, showcasing the truck's operational states. In the first two road segments, our approach MFI suggests driving at heterogeneous speed profiles, maintaining the engine speed just below the switching speed (1,000 rpm) to leverage the benefit of the triple-injection control strategy. Such a heterogeneous speed profile indeed achieves a smaller emission cost than the homogeneous speed profile, as evidenced in the last row of Figure 5.

To assess computational efficiency, we compare the runtime of calculating the emission function $c(t)$ using the convex program (7) and our efficient method introduced in Theorem 2. The convex problem is solved using the nonlinear programming solver MadNLP [55] integrated with JuMP [56]. The results, presented in Table III, reveal that our method is three orders of magnitude faster than directly solving the convex problem. Such improvement makes our method practical for computing the emission-efficient timely transportation plan on the national scale highway network.

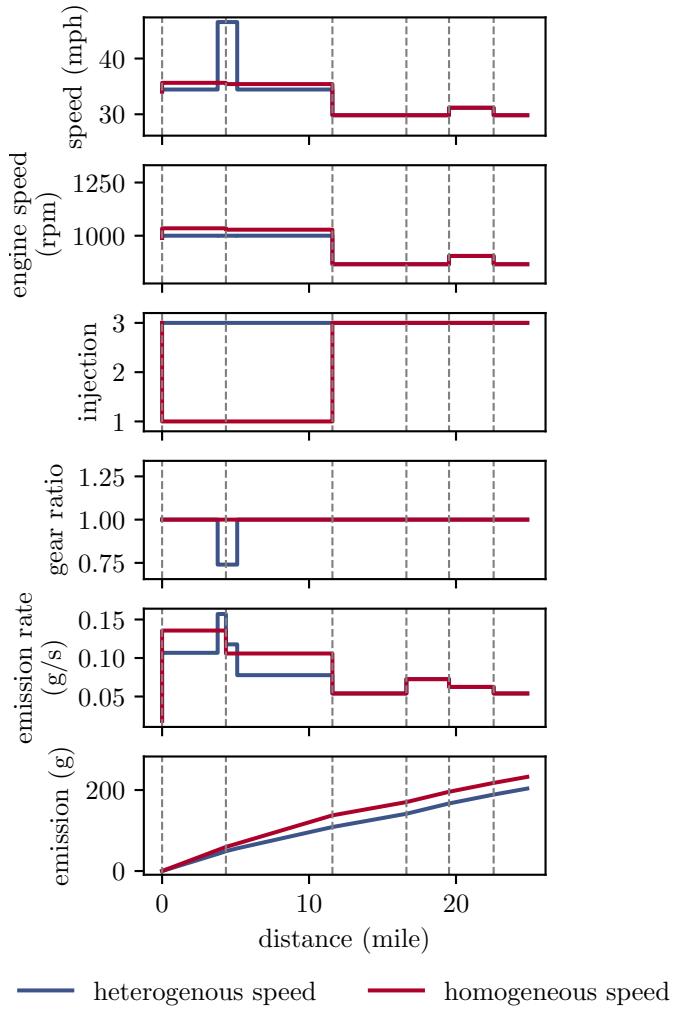


Fig. 5: A demonstration of the operational states of a truck traveling from Los Angeles, CA to Columbus, OH, using our MFI approach. The figure zooms into a specific section of the journey for enhanced visibility. The switching speed is 1,000 rpm, and the delay factor is 1.5. Each road segment is delineated by two vertical grey dashed lines.

	Convex program (7)	Our efficient method for problem (9)
Total runtime (seconds)	23,271	11
Number of solves	5.1×10^6	5.1×10^6
Average runtime per solve	4.5×10^{-3}	2.2×10^{-6}

TABLE III: Runtime comparison of the convex program (7) and our efficient method for problem (9) for solving the case study with MFI.

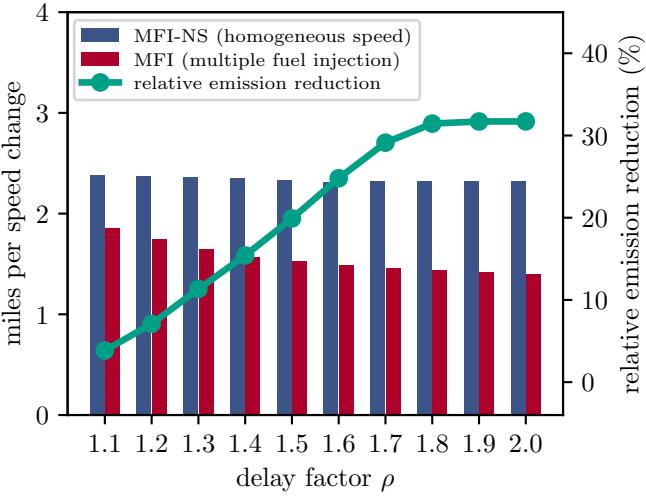


Fig. 6: Impact of heterogeneous speed over one road segment. The relative emission reduction of **MFI** as compared to **MFI-NS** is shown as green line.

C. Impact of heterogeneous speed over one road segment

We found in Sec. IV that the driver can drive with heterogeneous speed profile on one road segment to achieve emission reduction. In this subsection, we further explore this interesting observation by s. We set the engine switching speed to 1000 rpm and conduct the simulations with different delay factors over 1,000 origin-destination pairs. The results are presented in Fig. 6. When the deadline is tight (e.g. $\rho = 1.1$), the design space of speed planning is small, so heterogeneous speed profile **MFI** only saves 4% emission as compared to its homogeneous counterpart **MFI-NS**. As the deadline gets relaxed, **MFI** can leverage more design space from heterogeneous speed planning over one road segment and thus can achieve up to 32% emission reduction on average. Therefore, it is necessary to consider heterogeneous speed profiles in environmentally friendly truck operations.

However, this emission reduction comes with side effects. The driver has to change the speed more frequently to achieve this further emission reduction. In particular, when the delay factor ρ is 1.1, our approach **MFI** suggests changing speed every 1.85 miles on average while **MFI-NS** suggest changing speed every 2.38 miles. When the ρ increases to 2.0, **MFI** increases the speed change frequency to 1.4 miles per speed change to further leverage the design space of speed planning.

D. Impact of Deadline

We now study the impact of timely delivery requirements on different approaches. We set the switching speed to 1,000 rpm and vary the delay factor from 1.05 to 2.0. Fig. 7 illustrates the relative emission reductions of different approaches as compared to **FAST**. We observe that as the delay factor increases, all compared approaches reduce more emissions because a relaxed deadline allows for slower travel speeds and more space for speed planning, resulting in more emission reduction. Meanwhile, when the delay factor exceeds 1.8,

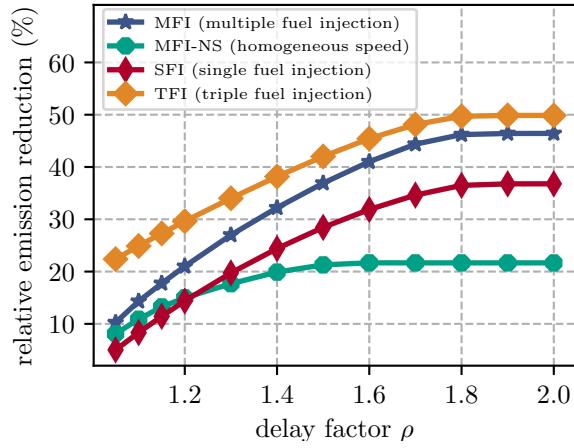


Fig. 7: Average relative emission reduction of different algorithms as compared to **FAST** w.r.t. delay factor. The switching speed is 1,000 rpm.

the travel speeds reach the lower bounds, and there is no further space for speed planning. As a result, the emission reductions saturate when the delay factor is higher than 1.8 for all alternatives.

Among all alternatives, the **TFI** has the biggest emission reduction. However, **TFI** is the ideal case that requires the engine to use the computationally expensive complicated triple injection at all engine speed ranges, which is infeasible since the engine control software task may miss its deadline. The second best is **MFI** with the heterogeneous speed profile. We find that it outperforms **MFI-NS**, which coincides with our observations in Sec. VI-C. **MFI** also outperforms **SFI** because **SFI** only considers single fuel injection and has less design space to optimize. Meanwhile, as the delay factor increases to 2.0, the cost reduction of **MFI** (46%) becomes close to the cost reduction of the ideal case **TFI** (50%). This is because a relaxed deadline allows **MFI** to optimize the speed profile to use the cost-effective triple injection control strategy at most of the road segments, achieving a close-to-ideal performance.

Another interesting observation comes from the comparison between **MFI-NS** and **SFI**. Those two methods consider two separate ways to reduce the emission cost: **MFI-NS** has cost-effective triple-injection but only drives at a constant speed over one road segment, while **SFI** considers heterogeneous speed profile but does not have the triple injection strategy. When the deadline is tight (e.g., $\rho = 1.1$), **SFI** is slightly worse than **MFI-NS** because there is little design space for heterogeneous speed profile so the benefit of multiple control strategies is larger. When the deadline gets relaxed (e.g., ρ from 1.2 to 2.0), the emission reduction of **SFI** outperforms the emission reduction of **MFI-NS**. This is because, with a relaxed deadline, a heterogeneous speed profile has larger design space and its benefit outperforms the benefit of multiple control strategies. It thus also justifies the necessity of heterogeneous speed profiles in environmentally related truck operations.

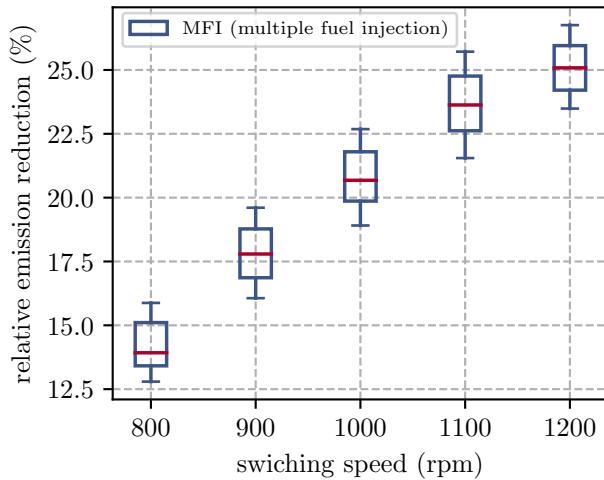


Fig. 8: Relative emission reduction of MFI as compared to FAST with different switching speeds. The delay factor is 1.2.

E. Impact of Switching Speed

In this subsection, we study the impact of switching speed for MFI. We set the delay factor ρ to 1.2 for all 1,000 origin-destination pairs and vary the switching speed from 800 rpm to 1,200 rpm. Note that a lower switching speed leads to a smaller feasible rpm range for the cost-effective triple injection. The results are presented in Fig. 8. We find that as switching speed increases, the relative emission reduction of MFI decreases because higher switching speeds allow larger engine speed ranges and thus larger vehicle speed ranges for the cost-effective triple injection strategy. Meanwhile, for all switching speeds from 800 to 1,200, the boxes in Fig. 8 have small lengths. This means that the variance of relative emission reduction of MFI is small and MFI has uniformly good performance for all 1,000 origin-destination pairs.

VII. CONCLUSION

In this paper, we consider a scenario where a heavy-duty truck hauls freight across a national highway network subject to a hard deadline. We ride on the recent advancement in engine control that adaptively selects the fuel injection strategy to effectively reduce the emission. We show that the problem is NP-hard, and the adaptive fuel injection strategy imposes a unique challenge compared to existing studies due to the non-convexity of the emission rate function. We reformulate the speed planning problem as a convex problem and leverage the problem structure to compute the optimal solution by checking a small number of candidates. We then propose an efficient heuristic for the overall problem with both path planning and speed planning, and derive an upper bound of the performance gap for the heuristic.

We evaluate the performance of our approach using real-world traces over the US national highway system. The results show that our approach reduces up to 46% emission as compared to the fastest path, which is commonly adopted in practice. We also study the impact of heterogeneous speed profile on one road segment and find that it reduces up to 32% as compared to its counterpart that uses constant speed on one

road segment. It is thus necessary to consider heterogeneous speed profiles in environment-friendly truck operations. An interesting future direction is to study the multiple objectives by jointly considering fuel consumption and emissions. It would also be interesting to jointly optimize the gear position shifting strategy, path planning and speed planning.

REFERENCES

- [1] R. Zhou, Q. Liu, W. Xu, M. Chen, and H. Zeng, "Minimizing emission for timely truck transportation with adaptive fuel injection," in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2020, pp. 240–249.
- [2] A. T. Associations, "Economics and industry data," <https://trucking.org/economics-and-industry-data>, 2022, accessed: 2022-11-22.
- [3] International Energy Agency, *The Future of Trucks*. <https://www.oecd-ilibrary.org/content/publication/9789264279452-en>, 2017.
- [4] California Air Resources Board, "NOx Emissions of In-Use Trucks." <https://ww2.arb.ca.gov/resources/documents/nox-emissions-of-in-use-trucks>.
- [5] ———, "California Air Resources Board staff current assessment of the technical feasibility of lower NOx standards-White paper," https://ww3.arb.ca.gov/msprog/hdlnownox/white_\paper_04182019a.pdf.
- [6] A. Biondi, M. D. Natale, G. C. Buttazzo, and P. Pazzaglia, "Selecting the transition speeds of engine control tasks to optimize the performance," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 1, pp. 1–26, 2018.
- [7] D. Buttle, "Keynote speech: Real-time in the prime-time," in *Euromicro Conference on Real-Time Systems*, 2012.
- [8] Y. Hotta, M. Inayoshi, K. Nakakita, K. Fujiwara, and I. Sakata, "Achieving lower exhaust emissions and better performance in an hsdiesel engine with multiple injection," in *SAE Technical Paper*, 2005.
- [9] L. Deng, M. H. Hajiesmaili, M. Chen, and H. Zeng, "Energy-efficient timely transportation of long-haul heavy-duty trucks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2099–2113, 2018.
- [10] H. Ashby, *Protecting Perishable Foods during Transport by Truck*. U.S. Department of Agriculture, 2006.
- [11] W. Mallett, *Freight Performance Measurement: Travel Time in Freight-Significant Corridors*. U.S. Federal Highway Administration, 2006.
- [12] R. Hassin, "Approximation schemes for the restricted shortest path problem," *Mathematics of Operations research*, vol. 17, no. 1, pp. 36–42, 1992.
- [13] D. Lorenz and D. Raz, "A simple efficient approximation scheme for the restricted shortest path problem," *Operations Research Letters*, vol. 28, no. 5, pp. 213–219, 2001.
- [14] A. Juttner, B. Sziyatovski, I. Mécs, and Z. Rajkó, "Lagrange relaxation based method for the qos routing problem," in *IEEE Conf. Computer Communications*, 2001.
- [15] Q. Liu, H. Zeng, and M. Chen, "Energy-efficient timely truck transportation for geographically-dispersed tasks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, 2020.
- [16] W. Xu, Q. Liu, M. Chen, and H. Zeng, "Ride the tide of traffic conditions: Opportunistic driving improves energy efficiency of timely truck transportation," in *ACM Conf. Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2019.
- [17] E. Hellström, M. Ivarsson, J. Åslund, and L. Nielsen, "Look-ahead control for heavy trucks to minimize trip time and fuel consumption," vol. 17, no. 2, pp. 245–254. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0967066108001251>
- [18] C. Sundström, A. Voronov, O. Lindgärde, and A. Lagerberg, "Optimal speed and gear shift control of long-haulage trucks," vol. 52, no. 5, pp. 471–477. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2405896319306986>
- [19] B. Wingelaar, G. R. G. da Silva, and M. Lazar, "Model predictive eco-driving control for heavy-duty trucks using branch and bound optimization." [Online]. Available: <http://arxiv.org/abs/2206.02447>
- [20] E. Hellström, J. Åslund, and L. Nielsen, "Design of an efficient algorithm for fuel-optimal look-ahead control," vol. 18, no. 11, pp. 1318–1327. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0967066109002366>
- [21] K. Boriboonsomsin, M. J. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," vol. 13, no. 4, pp. 1694–1704, conference Name: IEEE Transactions on Intelligent Transportation Systems.

- [22] S. Xu, S. E. Li, B. Cheng, and K. Li, "Instantaneous feedback control for a fuel-prioritized vehicle cruising system on highways with a varying slope," vol. 18, no. 5, pp. 1210–1220, conference Name: IEEE Transactions on Intelligent Transportation Systems.
- [23] S. Xu, S. E. Li, X. Zhang, B. Cheng, and H. Peng, "Fuel-optimal cruising strategy for road vehicles with step-gear mechanical transmission," vol. 16, no. 6, pp. 3496–3507, conference Name: IEEE Transactions on Intelligent Transportation Systems.
- [24] A. Hamednia, N. K. Sharma, N. Murgovski, and J. Fredriksson, "Computationally efficient algorithm for eco-driving over long look-ahead horizons," vol. 23, no. 7, pp. 6556–6570, conference Name: IEEE Transactions on Intelligent Transportation Systems.
- [25] D. Guo, J. Wang, J. B. Zhao, F. Sun, S. Gao, C. D. Li, M. H. Li, and C. C. Li, "A vehicle path planning method based on a dynamic traffic network that considers fuel consumption and emissions," vol. 663, pp. 935–943. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0048969719303067>
- [26] T. Bektas and G. Laporte, "The Pollution-Routing Problem," *Transportation Research Part B: Methodological*, vol. 45, pp. 1232–1250, 2011.
- [27] O. Jabali, T. Van Woensel, and A. de Kok, "Analysis of Travel Times and CO₂ Emissions in Time-Dependent Vehicle Routing," *Production and Operations Management*, vol. 21, pp. 1060–1074, 2012.
- [28] M. S. Gajanand and T. T. Narendran, "Green route planning to reduce the environmental impact of distribution," vol. 16, no. 5, pp. 410–432. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/13675567.2013.831400>
- [29] P. Sahlholm and K. H. Johansson, "Road grade estimation for look-ahead vehicle control using multiple measurement runs," *Control Engineering Practice*, vol. 18, no. 11, pp. 1328–1341, 2010.
- [30] A. Alam, J. Mårtensson, and K. Johansson, "Look-ahead cruise control for heavy duty vehicle platooning," in *IEEE International Conference on Intelligent Transportation Systems*, 2013.
- [31] ——, "Experimental evaluation of decentralized cooperative cruise control for heavy-duty vehicle platooning," *Control Engineering Practice*, vol. 38, pp. 11–25, 2015.
- [32] A. Alam, B. Besselink, V. Turri, J. Mårtensson, and K. Johansson, "Heavy-duty vehicle platooning for sustainable freight transportation: A cooperative method to enhance safety and efficiency," *IEEE Control Syst. Magazine*, vol. 35, no. 6, pp. 34–56, 2015.
- [33] W. Xu, T. Cui, and M. Chen, "Optimizing two-truck platooning with deadlines," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [34] T. Johnson and A. Joshi, "Review of vehicle engine efficiency and emissions," *SAE International Journal of Engines*, vol. 11, no. 6, pp. 1307–1330, 2018.
- [35] Y. Kim, Y. Kim, S. Jun, K. Lee, S. Rew, D. Lee, and S. Park, "Strategies for particle emissions reduction from gdi engines," in *SAE Technical Paper*, 2013.
- [36] J. Su, M. Xu, P. Yin, Y. Gao, and D. Hung, "Particle number emissions reduction using multiple injection strategies in a boosted spark-ignition direct-injection (sidi) gasoline engine," *SAE International Journal of Engines*, vol. 8, no. 1, pp. 20–29, 2015.
- [37] C. Peng, Y. Zhao, and H. Zeng, "Dynamic switching speed reconfiguration for engine performance optimization," in *56th Design Automation Conference*, 2019.
- [38] M.-F. Hsieh and J. Wang, "No and no₂ concentration modeling and observer-based estimation across a diesel engine aftertreatment system," *Journal of Dynamic Systems, Measurement, and Control*, vol. 133, no. 4, 2011.
- [39] C. Peng, Y. Zhao, and H. Zeng, "Schedulability analysis of adaptive variable-rate tasks with dynamic switching speeds," in *IEEE Real-Time Systems Symposium (RTSS)*, 2018, pp. 396–407.
- [40] *The FR Series 10-Speed ("B" ratio, "C" ratio and High Torque Models)*, Eaton Corporation. [Online]. Available: <https://www.eaton.com/content/dam/eaton/products/transmissions/vehicle-transmissions/fr-series-manual/fr-series-10-speed-brochure-trsl0261-en-us.pdf>
- [41] PACCAR, "THE PACCAR EPA MX-13 ENGINE," <https://paccarpowertrain.com/products/epa-mx-13/>.
- [42] J. Park, "Understand your truck's powertrain and improve fuel economy: Driver's Education," <https://paccarpowertrain.com/products/epa-mx-13/>.
- [43] J. Wang and H. A. Rakha, "Fuel consumption model for heavy duty diesel trucks: Model development and testing," *Transportation Research Part D: Transport and Environment*, vol. 55, pp. 127–141, 2017.
- [44] J. Y. Wong, *Theory of ground vehicles*. John Wiley & Sons, 2001.
- [45] United States Environmental Protection Agency, "Greenhouse Gases Equivalencies Calculator," <https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references>.
- [46] G. Yang, H. Xu, Z. Wang, and Z. Tian, "Truck acceleration behavior study and acceleration lane length recommendations for metered on-ramps," *International journal of transportation science and technology*, vol. 5, no. 2, pp. 93–102, 2016.
- [47] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge univ. press, 2004.
- [48] J. Teresco, Map-based Educational Tools for Algorithm Learning (METAL) project. <https://courses.teresco.org/metal/>.
- [49] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth *et al.*, "The shuttle radar topography mission," *Reviews of geophysics*, vol. 45, no. 2, 2007.
- [50] Traffic flow using corridor in HERE maps. <https://developer.here.com/api-explorer/rest/traffic/flow-using-corridor>.
- [51] H.-L. Hwang, S. Hargrove, S.-M. Chin, D. W. Wilson, H. Lim, J. Chen, R. Taylor, B. Peterson, and D. Davidson, "The freight analysis framework version 4 (faf4)-building thefaf4 regional database: Data sources and estimation methodologies," Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), Tech. Rep., 2016.
- [52] Kenworth, "Kenworth T800 vehicle," <http://www.kenworth.com/trucks/t800>.
- [53] A. Kuleshov, "Model for predicting air-fuel mixing, combustion and emissions in di diesel engines over whole operating range," *SAE paper*, no. 2005-01, p. 2119, 2005.
- [54] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017. [Online]. Available: <https://pubs.siam.org/doi/10.1137/141000671>
- [55] S. Shin, F. Pacaud, and M. Anitescu, "Accelerating optimal power flow with GPUs: SIMD abstraction of nonlinear programs and condensed-space interior-point methods," *arXiv preprint arXiv:2307.16830*, 2023.
- [56] M. Lubin, O. Dowson, J. Dias Garcia, J. Huchette, B. Legat, and J. P. Vielma, "JuMP 1.0: Recent improvements to a modeling language for mathematical optimization," *Mathematical Programming Computation*, vol. 15, p. 581–589, 2023.
- [57] H. L. Royden and P. Fitzpatrick, *Real analysis*. Macmillan New York, 1988, vol. 32.
- [58] A. V. Fiacco and J. Kyprasis, "Convexity and concavity properties of the optimal value function in parametric nonlinear programming," *Journal of optimization theory and applications*, vol. 48, no. 1, pp. 95–126, 1986.

VIII. APPENDIX

A. An example of heterogenous speed planning

Example 1. Let us consider the following emission rate function

$$f(r) = \begin{cases} (r - 30)^2 / 100 + 1 & \text{if } 30 \leq r \leq 50, \\ (r - 50)^2 / 100 + 10 & \text{if } 50 < r \leq 60 \end{cases}$$

We further set the length of the edge to be 110, and the total time for traversing this edge is $t = 2$.

(i) By following a constant speed of $110/2 = 55$, the total emission is

$$2 \cdot f(55) = 20.5$$

(ii) In comparison, consider another solution where we first drive at a speed of 40 for time 0.5, and then drive at a speed of 60 for time 1.5. This solution is feasible since it traverses the edge (with a length of 110) by a total travel time of 2. The incurred emission is

$$0.5 \cdot f(40) + 1.5 \cdot f(60) = 17.5$$

Therefore, driving at a constant speed of 55 incurs a larger emission than the solution (ii); the solution (i) is not optimal.

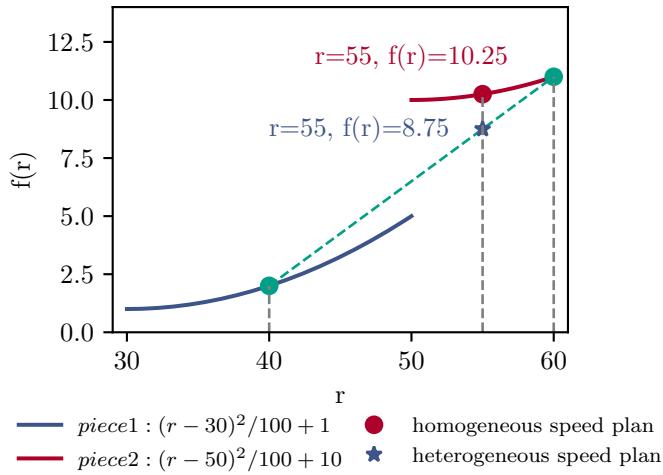


Fig. 9: An illustration of Example 1

B. Proof of Lemma 3

Proof. Suppose there is a solution $\vec{\beta}$ with at least three non-zero entries and its corresponding speed vector \vec{r} . We will show that we can decrease the non-zero entries by one without loss of optimality. Pick any three of nonzero entries of $\vec{\beta}$ and denote them by $\beta_{i_1}, \beta_{i_2}, \beta_{i_3} > 0$ with corresponding function values $y_1 = f_{i_1}(r_{i_1}), y_2 = f_{i_2}(r_{i_2}), y_3 = f_{i_3}(r_{i_3})$. Let

$$\hat{\beta} = \beta_{i_1} + \beta_{i_2} + \beta_{i_3}$$

$$\hat{r} = \frac{\beta_{i_1}r_{i_1} + \beta_{i_2}r_{i_2} + \beta_{i_3}r_{i_3}}{\hat{\beta}}$$

Without loss of generality, we assume the following conditions

$$r_{i_1} < \hat{r} < r_{i_2}, \hat{r} < r_{i_3} \quad (21)$$

$$\frac{y_2 - y_1}{r_{i_2} - r_{i_1}} \leq \frac{y_3 - y_1}{r_{i_3} - r_{i_1}} \quad (22)$$

Note that similar argument holds for the case when r_{i_1} and r_{i_2} is less than \bar{r} or when the slope order in (22) is swapped. Let $\tilde{\beta}_{i_1}, \tilde{\beta}_{i_2}$ to the following equations:

$$\begin{aligned} \hat{\beta} &= \tilde{\beta}_{i_1} + \tilde{\beta}_{i_2} \\ \hat{\beta}\hat{r} &= \tilde{\beta}_{i_1}r_{i_1} + \tilde{\beta}_{i_2}r_{i_2} \end{aligned} \quad (23)$$

Then we can construct a new solution

$$\tilde{\beta}_j = \begin{cases} \tilde{\beta}_{i_1}, & \text{if } j = i_1 \\ \tilde{\beta}_{i_2}, & \text{if } j = i_2 \\ 0, & \text{if } j = i_3 \\ \beta_j, & \text{otherwise.} \end{cases}$$

and the same speed profile \vec{r} . By construction, $\vec{\beta}$ and \vec{r} satisfy the constraints of problem (9). Meanwhile, we have

the following inequality.

$$\sum_{i=1}^n \beta_i f_i(r_i) - \sum_{i=1}^n \tilde{\beta}_i f_i(r_i) \quad (24)$$

$$= \beta_{i_1}y_1 + \beta_{i_2}y_2 + \beta_{i_3}y_3 - (\tilde{\beta}_{i_1}y_1 + \tilde{\beta}_{i_2}y_2) \quad (25)$$

$$= \frac{\beta_{i_3}}{r_{i_2} - r_{i_1}} ((y_3 - y_1)(r_{i_2} - r_{i_1}) - (y_2 - y_1)(r_{i_3} - r_{i_1})) \quad (26)$$

$$= \beta_{i_3}(r_{i_3} - r_{i_1}) \left(\frac{y_3 - y_1}{r_{i_3} - r_{i_1}} - \frac{y_2 - y_1}{r_{i_2} - r_{i_1}} \right) \quad (27)$$

$$\geq 0 \quad (28)$$

Here, equality (26) is derived by substituting the solution of equations (23). The last inequality is true because all of the product components in (27) is positive. The above inequality means that the new solution decrease the number of non-zero entries with no larger objective. Therefore, if there is an optimal solution with more than two nonzero entries, we can always follow the above procedure to construct another optimal solution with at most two non-zero entries. \square

C. Proof of Lemma 4

Our proof replies on the following lemma, which is also known as chordal slope lemma.

Lemma 8 (Chordal Slope Lemma [57]). *Suppose $g(r)$ is a function of one real variable r . Consider the slope $S(r_1, r_2) = \frac{g(r_2) - g(r_1)}{r_2 - r_1}$ of the line connecting two points $(r_1, g(r_1))$ and $(r_2, g(r_2))$ on the curve. If g is convex, then $S(r_1, r_2)$ is monotonically non-decreasing in r_1 , for every fixed r_2 .*

Proof of Lemma 4. The intercept point \bar{y} can be written as

$$\begin{aligned} \bar{y} &= \frac{y_2 - g(r_1)}{r_2 - r_1} (\bar{r} - r_1) + g(r_1) \\ &= y_2 + \frac{g(r_2) - g(r_1)}{r_2 - r_1} (\bar{r} - r_2) + \frac{(r_2 - \bar{r})(y_2 - g(r_2))}{r_1 - r_2} \end{aligned}$$

By Lemma 8 and that $(\bar{r} - r_2)$ is non-positive, the second summand on the right hand side is monotonically non-increasing with r_1 . The third summand is also monotonically non-increasing with r_1 , since both $(r_2 - \bar{r})$ and $(y_2 - g(r_2))$ are non-negative. Hence, \bar{y} is also monotonically non-increasing with r_1 . \square

D. Proof of Lemma 5

We consider the slope of the line connecting (r_1, y_1) and $(r_2, g(r_2))$, which is the major component of \bar{y} .

$$T(r_2) = \frac{g(r_2) - y_1}{r_2 - r_1}$$

In the following, we will first prove the following two useful Lemmas before we prove Lemma 5.

Lemma 9. *The following two statements are true.*

- If $r_1 \leq r \leq p$, then $\partial_- g(r) \leq T(r)$;
- If $r \geq p \geq r_1$, then $\partial_+ g(r) \geq T(r)$.

Proof. We first consider a help function

$$h(r) = g(r) + \partial g(r)(r_1 - r)$$

where $\partial g(r)$ is the subgradient of function g at point r and both left derivative and right derivative lie in $\partial g(r)$. For any a, b such that $r_1 \leq a \leq b$, we have

$$\begin{aligned} & h(b) - h(a) \\ &= g(b) - g(a) + \partial g(b)(r_1 - b) - \partial g(a)(r_1 - a) \\ &\leq \partial g(b)(b - a) + \partial g(b)(r_1 - b) - \partial g(a)(r_1 - a) \\ &= (\partial g(b) - \partial g(a))(r_1 - a) \leq 0 \end{aligned} \quad (29)$$

The inequality directly follows from the convexity of function g . Therefore, function $h(x)$ is non-increasing when $x \geq r_1$. When $r_1 \leq r \leq p$, we have

$$\begin{aligned} & \partial g_-(r) - T(r) \\ &= \frac{y_1 - (g(r) + \partial g(r)(r_1 - r))}{r - r_1} \end{aligned} \quad (30)$$

$$\leq \frac{y_1 - (g(r) + \partial g(p)(r_1 - p))}{r - r_1} \quad (31)$$

$$= \frac{r_1 - p}{r - r_1} \cdot \left(\frac{y_1 - g(p)}{r_1 - p} - \partial g(p) \right) \quad (32)$$

$$\leq 0 \quad (33)$$

Here, inequality (31) follows from (29) and inequality (32) follows from the definition of the generalized right point of tangency p . Similarly, when $r \geq p \geq r_1$, we have

$$\begin{aligned} & \partial g_+(r) - T(r) \\ &= \frac{y_1 - (g(r) + \partial g(r)(r_1 - r))}{r - r_1} \\ &\geq \frac{y_1 - (g(r) + \partial g(p)(r_1 - p))}{r - r_1} \\ &= \frac{r_1 - p}{r - r_1} \cdot \left(\frac{y_1 - g(p)}{r_1 - p} - \partial g(p) \right) \\ &\geq 0 \end{aligned}$$

Then we complete our proof for Lemma 9. \square

The following Lemma characterize the monotonicity of $T(r)$.

Lemma 10. $T(r)$ is first monotonically non-increasing with r before p , then monotonically non-decreasing with r after p .

Proof. When $a \leq b \leq p$, we have

$$\begin{aligned} T(a) &= \frac{g(a) - y_1}{a - r_1} \\ &\stackrel{(I1)}{\geq} \frac{g(b) + \partial g(b)(a - b) - y_1}{a - r_1} \\ &\stackrel{(I2)}{\geq} \frac{g(b) + \frac{g(b) - y_1}{b - r_1}(a - b) - y_1}{a - r_1} \\ &= \frac{g(b) - y_1}{b - r_1} = T(b) \end{aligned}$$

Here inequality (I1) is because g is convex hence $g(a) - g(b) \geq \partial g(b)(a - b)$, and inequality (I2) is because b is no larger than the generalized right point of tangency p .

Similarly, when $p \leq a \leq b$, we have

$$\begin{aligned} T(b) &= \frac{g(b) - y_1}{b - r_1} \\ &\stackrel{(I3)}{\geq} \frac{g(a) + \partial g(a)(b - a) - y_1}{b - r_1} \\ &\stackrel{(I4)}{\geq} \frac{g(a) + \frac{g(a) - y_1}{a - r_1}(b - a) - y_1}{b - r_1} \\ &= \frac{g(a) - y_1}{a - r_1} = T(a) \end{aligned}$$

Then we complete our proof for Lemma 10. \square

Now we proceed to prove Lemma 5.

Proof of Lemma 5. Recall that we can write \bar{y} as

$$\bar{y} = \frac{g(r_2) - y_1}{r_2 - r_1}(\bar{r} - r_1) + y_1$$

By Lemma 10 and that $(\bar{r} - r_1)$ is non-negative, we have the monotonicity of \bar{y} follows the monotonicity of $T(r_2)$. That is, \bar{y} is monotonically non-increasing before p , then monotonically non-decreasing after p .

If $\bar{r} \leq p$, since \bar{y} is monotonically non-increasing with r_2 , $\forall r_2 \in [\bar{r}, p]$, thus we have

$$\begin{aligned} \bar{y} &= y_1 + \frac{g(r_2) - y_1}{r_2 - r_1}(\bar{r} - r_1) \\ &\leq y_1 + \frac{g(\bar{r}) - y_1}{\bar{r} - r_1}(\bar{r} - r_1) = g(\bar{r}) \end{aligned}$$

If $\bar{r} \geq p$, since \bar{y} is monotonically non-decreasing with r_2 , $\forall r_2 \geq \bar{r} \geq p$, thus we have

$$\begin{aligned} \bar{y} &= y_1 + \frac{g(r_2) - y_1}{r_2 - r_1}(\bar{r} - r_1) \\ &\geq y_1 + \frac{g(\bar{r}) - y_1}{\bar{r} - r_1}(\bar{r} - r_1) = g(\bar{r}) \end{aligned}$$

\square

E. Proof of Proposition 1

Proof. The proof idea follows from [58, Proposition 2.1]. We first rewrite problem in the following compact form

$$c(\tau) = \min_{\vec{z} \in Z(\tau)} F(\vec{z})$$

where $\vec{z} = [d_1, \dots, d_n, t_1, \dots, t_n]$ is the stacked variable, and

$$Z(\tau) = \left\{ \vec{z}: \sum_{i \in [n]} t_i = \tau, \sum_{i \in [n]} d_i = D, s_{i-1}t_i \leq d_i \leq s_i t_i, \forall i \in [n] \right\} \quad (34)$$

$$F(\vec{z}) = \sum_{i \in [n]} t_i \cdot f_i \left(\frac{d_i}{t_i} \right) \quad (35)$$

are the constraint set and the objective. By Lemma 2, we know that $F(\vec{z})$ is convex in \vec{z} . Note that for every $\tau \in [t_l, t_u]$,

$Z(\tau)$ is simply a polyhedron and it is linear in τ , thus for any $\tau_1, \tau_2 \in [t_l, t_u]$, $\lambda \in (0, 1)$, we have

$$\lambda Z(\tau_1) + (1 - \lambda)Z(\tau_2) \subset Z(\lambda\tau_1 + (1 - \lambda)\tau_2) \quad (36)$$

That is, for every $\bar{z}^1 \in Z(\tau_1)$ and $\bar{z}^2 \in Z(\tau_2)$, we have $\lambda\bar{z}^1 + (1 - \lambda)\bar{z}^2 \in Z(\lambda\tau_1 + (1 - \lambda)\tau_2)$. Therefore, we have

$$c(\lambda\tau_1 + (1 - \lambda)\tau_2) \quad (37)$$

$$= \min_{\bar{z} \in Z(\lambda\tau_1 + (1 - \lambda)\tau_2)} F(\bar{z}) \quad (38)$$

$$\leq \min_{\substack{\bar{z}^1 \in Z(\tau_1) \\ \bar{z}^2 \in Z(\tau_2)}} F(\lambda\bar{z}^1 + (1 - \lambda)\bar{z}^2) \quad (39)$$

$$\leq \min_{\substack{\bar{z}^1 \in Z(\tau_1) \\ \bar{z}^2 \in Z(\tau_2)}} \lambda F(\bar{z}^1) + (1 - \lambda)F(\bar{z}^2) \quad (40)$$

$$= \lambda \min_{\bar{z}^1 \in Z(\tau_1)} F(\bar{z}^1) + (1 - \lambda) \min_{\bar{z}^2 \in Z(\tau_2)} F(\bar{z}^2) \quad (41)$$

$$= \lambda c(\tau_1) + (1 - \lambda)c(\tau_2) \quad (42)$$

Here (39) follows from (36) and that a smaller constraint set leads to a bigger optimal value. (40) follows from the convexity of objective $F(z)$. Therefore, function $c(\tau)$ is a convex function. \square



Junyan Su received his B.Eng degree from the Department of Information Science and Technology at ShanghaiTech University in 2019. He is currently a Ph.D. student in the Department of Data Science, City University of Hong Kong. He received ACM e-Energy Best Paper Award in 2023. His research interests include online optimization and intelligent transportation systems.



Runzhi Zhou received his B.S. degrees in Computer Science and Electrical Engineering from Case Western Reserve University in 2021. He received his M.S. degrees in Computer Information Science and Electrical Engineering at University of Pennsylvania in 2024. His research interests include optimization, intelligent transportation systems and automation in warehouse systems.



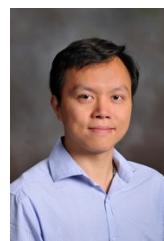
Qingyu Liu is currently an Assistant Professor with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen, China, which he joined in June 2023. Prior to joining Peking University, he was a Postdoc and then a Research Assistant Professor of Electrical and Computer Engineering with Virginia Tech, VA, USA, from September 2019 to May 2023. He received the Ph.D. degree in computer engineering from Virginia Tech in 2019. His research interests include wireless networking, Internet of Things, and edge AI. He has been serving on the TPC of IEEE INFOCOM since 2021, and was awarded Distinguished Member of the INFOCOM TPC in 2023.



Wenjie Xu is currently a doctoral student at Swiss Federal Institute of Technology Lausanne (EPFL). He received his MPhil degree in Information Engineering from The Chinese University of Hong Kong, and B.E. degree in Electronic Engineering from Tsinghua University in 2018. His research interests lie in the interface of optimization, control and machine learning, with applications to building control and intelligent transportation.



Minghua Chen (S'04 M'06 SM'13 F'22) received his B.Eng. and M.S. degrees from the Department of Electronic Engineering at Tsinghua University. He received his Ph.D. degree from the Department of Electrical Engineering and Computer Sciences at University of California Berkeley. He is currently a Professor in Department of Data Science, City University of Hong Kong. Minghua received the Eli Jury award from UC Berkeley in 2007 (presented to a graduate student or recent alumnus for outstanding achievement in the area of Systems, Communications, Control, or Signal Processing) and The Chinese University of Hong Kong Young Researcher Award in 2013. He also received IEEE ICME Best Paper Award in 2009, IEEE Transactions on Multimedia Prize Paper Award in 2009, ACM Multimedia Best Paper Award in 2012, IEEE INFOCOM Best Poster Award in 2021, and ACM e-Energy Best Paper Award in 2023. He receives the ACM Recognition of Service Award in 2017 and 2020 for the service contribution to the research community. He is currently a Senior Editor for IEEE Systems Journal, an Area Editor of ACM SIGEnergy Energy Informatics Review, and an Award Chair of ACM SIGEnergy. Minghua's recent research interests include online optimization and algorithms, machine learning for real-time optimization with hard constraints and its application in power systems, intelligent transportation systems, distributed optimization, and delay-critical networked systems. He is an ACM Distinguished Scientist and an IEEE Fellow.



Haibo Zeng is with Department of Electrical and Computer Engineering at Virginia Tech, USA. He received his Ph.D. in Electrical Engineering and Computer Sciences from University of California at Berkeley. He was a senior researcher at General Motors R&D until October 2011, and an assistant professor at McGill University until August 2014. His research interests are embedded systems, cyber-physical systems, and real-time systems. He received five paper awards in the above field.