On the Equilibrium and Stability Properties of a Macroscopic Model for Ride-Hailing Services with Limited Fleet Size

Gustav Nilsson and Nikolas Geroliminis

Abstract— **With the growing popularity of ride-hailing services and the desire to operate those services efficiently, ride-hailing companies need to ensure a sufficiently large fleet size and an appropriate rebalancing of empty vehicles. Due to the complexity of city traffic dynamics, macroscopic modeling approaches are often required. In this work, we present a macroscopic compartment model for ride-hailing services and characterize its equilibrium properties. If the service is only operating in one region, we provide both sufficient and necessary conditions for the system to converge to a unique equilibrium. If the service is operating over a couple of regions, we provide the necessary conditions for the request queues to stay bounded. When operating over more than one region, there is a need for a rebalancing controller for sending idling vehicles to another region. Hence, we present a Model Predictive Control (MPC) approach to solve the rebalancing problem and compare its performance with some simpler myopic controllers.**

I. Introduction

Ride-hailing services, i.e., services where users can request a ride through different smartphone applications, have increased in popularity over the previous years. For those services to operate efficiently and provide a good user experience, the availability of ride-hailing vehicles is crucial. Hence there is a need for operators of ride-hailing services to ensure that the fleet size is large enough to satisfy the demand and that idling ride-hailing vehicles are positioned in areas where the demand is high to ensure efficient matching with the waiting passengers.

The rebalancing problem has previously been extensively studied on the microscopic level. For this level of granularity, control solutions such as MPC [1], [2] and coverage control [3] have been proposed.

However, in an urban traffic environment where most ride-hailing services operate, the traffic dynamics is very complex. Because of the complexity, macroscopic models that capture the main characteristics of the traffic dynamic have become popular, both due to their ability to capture some of the most relevant congestion effects [4], [5], and also allowing for more tractable analysis of urban traffic control problems such as routing [6] and perimeter control [7]. Macroscopic urban traffic models have also been utilized for modeling and control of the dispatching of taxi services [8]. While there are similarities between the model proposed in [8] and the model proposed in this paper, we in this work focus our analysis on equilibrium properties and theoretical stability guarantees. In particular, we will investigate the implications of having a constant fleet size, which imposes further limitations on the set of feasible ride-hailing demands.

In this paper, we will introduce and analyze a macroscopic model for the rebalancing problem. The model we are studying fits into the broad class of compartmental models, which have gained attention in modeling different dynamical systems such as epidemics [9] and chemical reactions [10], [11]. Unlike most control problems for compartmental models, the controller's task in our model is to move a finite and constant mass between the separate compartments to improve the overall throughput of passengers. The rebalancing problem ultimately is about allocating a finite set of resources to serve queues, in our case, the waiting passengers. However, different from similar problems in, e.g., communication networks [12] and traffic networks with signalized junctions [13], [14], where the service allocation to the queues can be decided upon instantaneously, the rebalancing of empty ride-hailing vehicles is a slower process since the vehicles need to travel in the traffic network and are hence affected by the traffic dynamics. While ride-hailing vehicles are affected by traffic dynamics, we will in this work assume that the number of ride-hailing vehicles is relatively small compared to the overall number of vehicles in the system, and hence will have a negligible impact on the overall traffic dynamics themselves.

Apart from introducing and analyzing the properties of the model, we also propose a model predictive control (MPC) solution to solve the rebalancing problem. While the dynamical model itself does not yield convex constraints, we propose a convex relaxation of the discretized dynamics by transforming non-linear equality constraints into inequality constraints. Similar techniques have previously been utilized to make convex relaxations for optimal control of non-linear compartmental models for highway traffic [15].

The paper is outlined as follows. In the following section, Section II, we introduce the fluid model for one region and investigate its stability properties. We then, in Section III, extend the model to cover two regions, characterize the equilibrium properties of the model, and provide necessary conditions on the ride-hailing demands for the system to be stable. In Section IV, we present an MPC solution for the re-balancing problem, and in Section V, we illustrate the performance of MPC through numerical examples and also compare its performance to some simpler myopic controllers. The paper is concluded with some suggested directions for future research.

G. Nilsson and N. Geroliminis are with the School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. {gustav.nilsson, nikolas.geroliminis}@epfl.ch.

This work was supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545.

II. Single Region Dynamics

For a single region, we consider a fluid model with three continuous non-negative states: the number of idling ride-hailing vehicles n^I, the number of occupied ride-hailing vehicles n^O, and the user request queue q. Occupied vehicles become idling with a possible time-varying trip completion rate $\gamma > 0$ which represents the number of completed trips per time unit, and an idling vehicle becomes occupied with a matching rate given by a matching function $f(q, n^I)$ that is both dependent on the size of the request queue and the availability of idling vehicles. Throughout the paper, we will make the following assumptions about the matching function:

Assumption 1: The matching function is suppose to be such that $f(0, n^I) = 0$ for all $n^I \geq 0$ and $f(q, 0) = 0$ for all $q \geq 0$.

The assumption ensures that the number of empty vehicles and the request queue can never become non-negative.

Example 1: Examples of matching functions are variants of the Cobb-Douglas meeting function $f(q, n^I) = a_0 (n^I)^{\alpha_e} q^{\alpha_c}$, where $a_0 > 0$, $\alpha_e > 0$, and $\alpha_c > 0$ are parameters [16], [17]. For this kind of matching function, the frequency of which matches occur increases both with the queue length and the number of idle vehicles. Another example is $f(q, nI) = \beta(1 - \exp(-q))n^I$ with $\beta > 0$. In this case, the number of matches saturates with queue length, such that it will at maximum, be βn^I.

Moreover, we will, throughout the paper, assume that the fleet size of the ride-hailing vehicles will be relatively small compared to the total number of vehicles in the system:

Assumption 2: The number of ride-sharing vehicles is relatively small compared to the total number of vehicles in each region and will hence have a negligible effect on the trip completion rates.

The assumption above implies that the trip completion rate will be an exogenous variable in our model.

To complete our model, we introduce the exogenous, possibly time-varying, request rate $r \geq 0$ of users joining the request queue. The dynamics for one region can then be stated as:

$$\dot{n}^I = \gamma(t) n^O - f(q, n^I)\,, \tag{1}$$
$$\dot{n}^O = f(q, n^I) - \gamma(t) n^O\,, \tag{2}$$
$$\dot{q} = r(t) - f(q, n^I)\,. \tag{3}$$

Under the assumption that the system has an equilibrium point, the following theorem shows that all trajectories will converge to this equilibrium.

Theorem 1: Consider the system (1)–(3). Assume that the matching function is strictly increasing in both the request queue lengths and the number of idling vehicles, i.e., $\frac{\partial f}{\partial q} > 0$ and $\frac{\partial f}{\partial n^I} > 0$. If the request rate r and trip completion rate γ is constant and such that $n^{O*} = \frac{r}{\gamma} < n(0)$ and there exists a q^* such that $r = f(q^*, n^{O*})$, then for every initial state $q(0) \geq 0$, $n(0) > 0$ and $0 \leq n^O(0) \leq n(0)$ it holds that $\lim_{t \to +\infty} (n^O(t), n^I(t), q(t)) = (n^{O*}, n(0) - n^{O*}, q^*)$.

Proof Since the total number of ride-hailing vehicles is constant, it holds that $n^I = n(0) - n^O$ the system dynamics can equivalently be expressed as

$$\dot{n}^O = f(q, n(0) - n^O) - \gamma(t) n^O := g_{n^O}(q, n^O)\,, \tag{4}$$
$$\dot{q} = r(t) - f(q, n(0) - n^O) := g_q(q, n^O)\,. \tag{5}$$

Let $(q^*(t), n^{O*}(t))$ denote the solution to (4)–(5) with $(q^*(0), n^{O*}(0)) = (q^*, n^{O*})$, i.e., the trajectory that will stay at the equilibrium for all $t \geq 0$. Let $(q(t), n^{O*}(t))$ denote any solution to the system (4)–(5). Introduce the Lyapunov candidate $V(t) = |q(t) - q^*(t)| + |n^O(t) - n^{O*}(t)|$. Then it holds that $\frac{\mathrm{d}}{\mathrm{d}t} V(t) = \mathrm{sign}(q(t) - q^*(t)) \cdot (g_q(q(t), n^O(t)) - g_q(q^*(t), n^{O*}(t))) + \mathrm{sign}(n^O(t) - n^{O*}(t)) \cdot (g_{n^O}(q(t), n^O(t)) - g_{n^O}(q^*(t), n^{O*}(t)))$. Now, since $\frac{\partial g_{n^O}}{\partial q} > 0$ and $\frac{\partial g_q}{\partial n^O} > 0$, and also $\frac{\partial (g_{n^O} + g_q)}{\partial q} = 0$ and $\frac{\partial (g_{n^O} + g_q)}{\partial n^I} < 0$, Lemma 1 in [18] can be applied to guarantee that $\frac{d}{dt} V(t) < 0$ apart from when $q(t) = q^*(t)$ and $n^{O*}(t) = n^{O*}(t)$, which proves the theorem. ∎

While we in this section have analyzed the dynamics for one region, we will extend the model to two regions in the next section and introduce the need for a rebalancing controller.

III. Two Region Dynamic with Rebalacing

When extending our model to two regions, we need to introduce states for both the trips within and between the regions. We will refer to the regions as Region 1 and Region 2 and let $q_{1,1}, q_{1,2}, q_{2,1}, q_{2,2}$ denote the different request queues, where, e.g., $q_{1,2}$ are the users that want to travel between Region 1 and Region 2. In the same manner, we introduce the states $n^O_{1,1}, n^O_{1,2}, n^O_{2,1}, n^O_{2,2}$ for the occupied vehicles. Moreover, we let n^I_1, n^I_2 denote the number of idling vehicles in each region. We also allow the trip completion rates γ_1, γ_2 to be different for the different regions and the matching functions $f_{1,1}, f_{1,2}, f_{2,1}, f_{2,2}$ to be different for the different queues, all satisfying Assumption 1.

To rebalance the idling vehicles between the two regions, we introduce the control actions $u_{1,2}, u_{2,1} \in [0, 1]$ that tells how large a fraction of the idling vehicles that should go from one region to another. A graphical overview of the whole problem setting is shown in Figure 1.

The whole dynamical system can now be described as:

$$\begin{aligned} \dot{n}^I_1 = &\gamma_1(t) n^O_{1,1} + \gamma_2(t) n^I_2 u_{2,1} - \gamma_1(t) n^I_1 u_{1,2} \\ &- f_{1,1}(q_{1,1}, n^I_1) - f_{1,2}(q_{1,2}, n^I_1)\,, \end{aligned} \tag{6}$$
$$\begin{aligned} \dot{n}^I_2 = &\gamma_2(t) n^O_{2,2} + \gamma_1(t) n^I_1 u_{1,2} - \gamma_2(t) n^I_2 u_{2,1} \\ &- f_{2,2}(q_{2,2}, n^I_2) - f_{2,1}(q_{2,1}, n^I_2)\,, \end{aligned} \tag{7}$$
$$\dot{n}^O_{1,1} = f_{1,1}(q_{1,1}, n^I_1) + \gamma_2(t) n^O_{2,1} - \gamma_1(t) n^O_{1,1}\,, \tag{8}$$
$$\dot{n}^O_{1,2} = f_{1,2}(q_{1,2}, n^I_1) - \gamma_1(t) n^O_{1,2}\,, \tag{9}$$
$$\dot{n}^O_{2,1} = f_{2,1}(q_{2,1}, n^I_2) - \gamma_2(t) n^O_{2,1}\,, \tag{10}$$
$$\dot{n}^O_{2,2} = f_{2,2}(q_{2,2}, n^I_2) + \gamma_1(t) n^O_{1,2} - \gamma_2(t) n^O_{2,2}\,, \tag{11}$$

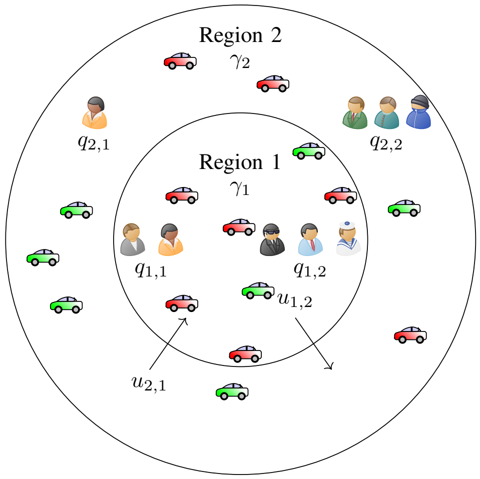

Fig. 1. A schematic sketch of the rebalancing problem studied in this paper. The area is divided into two regions, where Region 1 is the central region and Region 2 is the suburban region. Each region has its own trip completion rate, γ_1 and γ_2, respectively. In each region, users queue up for service, either for a trip within the region or a trip to the other region, where $q_{1,1}$ and $q_{2,2}$ denotes demands within each region, while $q_{1,2}$ and $q_{2,1}$ denotes trips from region 1 to region 2 and vice versa. The controller's, $u_{1,2}$ and $u_{2,1}$, task is to send a fraction of empty ride-hailing vehicles from one region to another.

$$\dot{q}_{1,1} = r_{1,1}(t) - f_{1,1}(q_{1,1}, n_1^I)\,, \tag{12}$$
$$\dot{q}_{1,2} = r_{1,2}(t) - f_{1,2}(q_{1,2}, n_1^I)\,, \tag{13}$$
$$\dot{q}_{2,1} = r_{2,1}(t) - f_{2,1}(q_{2,1}, n_2^I)\,, \tag{14}$$
$$\dot{q}_{2,2} = r_{2,2}(t) - f_{2,2}(q_{2,2}, n_2^I)\,. \tag{15}$$

In the dynamics above (6) and (7) describes the amount of idle ride-hailing vehicles in each region. Looking specifically at Region 1, the amount increases when trips within the region finish, which happens with a rate $\gamma_1(t)n_{1,1}^O$ and decreases when idle vehicles get occupied, both for trips within the region $f_{1,1}(q_{1,1}, n_1^I)$ and trips to the neighboring region $f_{1,2}(q_{1,2}, n_1^I)$. Mover, idle vehicles can be sent from Region 2 to Region 1 through the term $\gamma_2(t)n_2^I u_{2,1}$ which will increase the number of idle vehicles in Region 1, or in the opposite direction, $\gamma_1(t)n_1^I u_{1,2}$, which will decrease the number of vehicles in Region 1.

Equations (8)–(11) describe the dynamics of the occupied vehicles. The number of occupied vehicles increases with the matching and decreases when a trip is completed. It should be noted that for a trip between regions, e.g., from Region 1 to Region 2, once the vehicle reaches Region 2, the trip becomes equivalent to a Region 2 trip. In other words, a vehicle can not go from the states/compartments $n_{1,2}^O$ or $n_{2,1}^O$ to an idle state without first passing through the states/compartments $n_{1,1}^O$ or $n_{2,2}^O$.

The following proposition is a direct consequence of standard equilibrium analysis:

Proposition 1: Suppose that the request rates $r_{1,1}, r_{1,2}, r_{2,1}, r_{2,2}$ and trip completion factors γ_1, γ_2 are constant. Then, if the system (6)–(15) has an equilibrium, the number of occupied vehicles at the equilibrium is given by

$$n_{1,2}^{O*} = \frac{r_{1,2}}{\gamma_1}\,, \quad n_{2,1}^{O*} = \frac{r_{2,1}}{\gamma_2}\,,$$
$$n_{1,1}^{O*} = \frac{r_{1,1} + r_{2,1}}{\gamma_1}\,, \quad n_{2,2}^{O*} = \frac{r_{2,2} + r_{1,2}}{\gamma_2}\,. \tag{16}$$

Moreover, the amount of idle vehicles at the equilibrium satisfies

$$\begin{bmatrix} -\gamma_1 u_{1,2} & \gamma_2 u_{2,1} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} n_1^{I^*} \\ n_2^{I^*} \end{bmatrix} = \begin{bmatrix} r_{1,2} - r_{2,1} \\ n(0) - n_{1,2}^{O*} + n_{2,1}^{O*} + n_{1,1}^{O*} + n_{2,2}^{O*} \end{bmatrix}, \tag{17}$$

and the request queues at equilibrium will be such that $r_{1,1} = f_{1,1}(q_{1,1}^*, n_1^{I^*})$, $r_{1,2} = f_{1,2}(q_{1,2}^*, n_1^{I^*})$, $r_{2,1} = f_{2,1}(q_{2,1}^*, n_2^{I^*})$, and $r_{2,2} = f_{2,2}(q_{2,2}^*, n_2^{I^*})$.

Remark 1: It should be noted that there are several possible values of $u_{1,2}, u_{2,1}$ satisfying (17). If we suppose that $r_{1,2} > r_{2,1}$, then given that $\frac{r_{1,2} - r_{2,1}}{n_2^{I^*}\gamma_2} \leq 1$, one possible choice of control action that has an equilibrium is $(\tilde{u}_{1,2}, \tilde{u}_{2,1}) = (0, \frac{r_{1,2} - r_{2,1}}{n_2^{I^*}\gamma_2})$ but another choice is $(\tilde{u}_{1,2}, \tilde{u}_{2,1}) = (k, \frac{r_{1,2} - r_{2,1} + k\gamma_1}{n_2^{I^*}\gamma_2})$, where $0 \leq k \leq \min(1, \frac{n_2^{I*}\gamma_2 + r_{2,1} - r_{1,2}}{\gamma_1})$. In the latter case, there is, however, an unnecessary rebalancing of empty ride-hailing vehicles in both directions.

From (16) in the proposition, we can see that a slower trip completion rate, i.e., smaller values of γ will lead to more occupied vehicles at equilibrium. Moreover, as expected, as long as the matching function can provide a sufficiently large matching rate to meet the demand, the choice of matching functions will not influence the equilibrium.

From Proposition 1, it follows that a necessary condition for the existence of a non-negative equilibrium point is that $n_{1,2}^{O*} + n_{2,1}^{O*} + n_{1,1}^{O*} + n_{2,2}^{O*} < n(0)$, which can equivalently be expressed as

$$\frac{r_{1,2} + r_{2,1} + r_{1,1}}{\gamma_1} + \frac{r_{2,2} + r_{1,2} + r_{2,1}}{\gamma_2} < n(0)\,. \tag{18}$$

However, since $0 \leq u_{1,2}, u_{2,1} \leq 1$, (17) also provides another necessary condition for the existence of a non-negative equilibrium point, namely

$$\frac{r_{1,2} - r_{2,1}}{\gamma_2 n_2^{I*}} \leq 1\,, \quad \text{if } r_{1,2} \geq r_{2,1}\,, \tag{19}$$
$$\frac{r_{2,1} - r_{1,2}}{\gamma_1 n_1^{I*}} \leq 1\,, \quad \text{if } r_{2,1} > r_{1,2}\,. \tag{20}$$

By combining (19) and (20) with (18), we obtain the following necessary conditions for a non-negative equilibrium point:

$$\frac{r_{1,1} + r_{1,2} + r_{2,1}}{\gamma_1} + \frac{2r_{1,2} + r_{2,2}}{\gamma_2} \leq n(0)\,, \quad \text{if } r_{1,2} \geq r_{2,1}\,, \tag{21}$$

and

$$\frac{r_{1,1} + 2r_{2,1}}{\gamma_1} + \frac{r_{1,2} + r_{2,1} + r_{2,2}}{\gamma_2} \leq n(0)\,, \quad \text{if } r_{1,2} < r_{2,1}\,. \tag{22}$$

We will next show that the conditions (21) and (22) are necessary for the request queues to stay bounded.

Proposition 2: Suppose that the request rates $r_{1,1}, r_{1,2}, r_{2,1}, r_{2,2}$ and trip completion factors γ_1, γ_2 are constant. Then the conditions (21) and (22) are necessary for the system (6)–(15) have bounded request queues, i.e., the conditions have to be satisfied for the existence of a constant $D > 0$ such that $q_{1,1}(t) + q_{1,2}(t) + q_{2,1}(t) + q_{2,2}(t) < D$ for all $t \geq 0$.

Proof We start by proving the second case, i.e., when $r_{1,2} < r_{2,1}$, and will then see that the second case follows analogously. Assume that

$$\frac{r_{1,1} + 2r_{2,1}}{\gamma_1} + \frac{r_{1,2} + r_{2,1} + r_{2,2}}{\gamma_2} > n(0)\,. \quad (23)$$

Next, assume without loss of generality that all queues are initiated empty and observe that

$$\begin{aligned} A := \frac{q_{1,1}(t) + 2q_{2,1}(t)}{\gamma_1} + \frac{q_{1,2}(t) + q_{2,1}(t) + q_{2,1}(t)}{\gamma_2} = \\ \frac{1}{\gamma_1}\int_0^t \big(r_{1,1} + 2r_{1,2} - f_{1,1}(q_{1,1}(s), n_1^I(s)) \\ -2f_{2,1}(q_{2,1}(s), n_2^I(s)\big)\,\mathrm{d}s \\ + \frac{1}{\gamma_2}\int_0^t \big(r_{1,2} + r_{2,1} + r_{2,2} - f_{1,2}(q_{1,2}(s), n_1^I(s)) \\ -f_{2,1}(q_{2,1}(s), n_2^I(s) - f_{2,2}(q_{2,2}(s), n_2^I(s)\big)\,\mathrm{d}s\,. \end{aligned}$$

Due to (23) it holds that

$$\begin{aligned} A \geq (n(0) + \epsilon)t \\ - \int_0^t \left(\frac{f_{1,1} + 2f_{2,1}}{\gamma_1} + \frac{f_{1,2} + f_{2,1} + f_{2,2}}{\gamma_2}\right)\mathrm{d}s \quad (24) \end{aligned}$$

where $\epsilon > 0$, and we have dropped the dependencies of some functions for clarity. Now, by utilizing (9), (10), and (11) we get that

$$\frac{f_{1,2} + f_{2,1} + f_{2,2}}{\gamma_2} = n_{1,2}^O + n_{2,2}^O + \frac{\dot{n}_{1,2}^O + \dot{n}_{2,1}^O + \dot{n}_{2,2}^O}{\gamma_2}\,, \quad (25)$$

and by utilizing (7), (8),(10) and (11) we get that

$$\begin{aligned} \frac{f_{1,1} + 2f_{2,1}}{\gamma_1} = n_{1,1}^O + n_{2,1}^O \\ + \frac{\gamma_1 n_1^I u_{1,2} - \gamma_2 n_2^I u_{2,1}}{\gamma_1} + \frac{\dot{n}_{1,1}^O + \dot{n}_{1,2}^O - \dot{n}_1^I - \dot{n}_{2,2}^O}{\gamma_1} \\ \leq n_{1,1}^O + n_{2,1}^O + n_1^I + \frac{\dot{n}_{1,1}^O + \dot{n}_{1,2}^O - \dot{n}_1^I - \dot{n}_{2,2}^O}{\gamma_1}\,, \quad (26) \end{aligned}$$

where the last inequality follows from the fact that $u_{1,2}, u_{2,1} \in [0, 1]$. By inserting (25) and (26) into (24), we obtain $A \geq (n(0) + \epsilon)t - \int_0^t (n_{1,1}^O(s) + n_{1,2}^O(s) + n_{2,1}^O(s) + n_{2,2}^O(s) + n_1^I(s))\mathrm{d}s - \int_0^t (\frac{\dot{n}_{1,1}^O + \dot{n}_{1,2}^O - \dot{n}_1^I - \dot{n}_{2,2}^O}{\gamma_1} + \frac{\dot{n}_{1,2}^O + \dot{n}_{2,1}^O + \dot{n}_{2,2}^O}{\gamma_2})\mathrm{d}s$. The last integral term has to be bounded for all $t \geq 0$, since the states for the ride-hailing fleet are bounded between 0 and $n(0)$. Moreover, $\int_0^t (n_{1,1}^O(s) + n_{1,2}^O(s) + n_{2,1}^O(s) + n_{2,2}^O(s) + n_1^I(s))\mathrm{d}s \leq n(0)t$ and hence $A \geq \epsilon t \to +\infty$ when $t \to +\infty$, which combined with the fact that all the queues are non-negative, proves the statement. ■

Remark 2: In the case of a single region, as analyzed in Section II, the necessary condition above simplifies to $\frac{r}{\gamma} \leq n(0)$ which is arbitrarily close to the sufficient condition in Theorem 1. Hence the condition is both a necessary and sufficient condition for the single region dynamics.

While we in this section only have provided necessary conditions for the request queues to stay bounded, we will later numerically investigate if there exists rebalancing controllers that can keep the request queues bounded, i.e., if the necessary conditions could be sufficient as well. Before doing so, we will in the next section introduce a convex model predictive controller.

IV. Convex Relaxation of the Discretized MPC problem

In this section, we will design a model predictive control (MPC) solution that will determine the rebalancing control actions with the objective to minimize the total request queue length over time.

When implementing the MPC, we discretize the dynamics by step size $h > 0$ and let the index $k \in \mathcal{K} := \{0, \ldots, k^{\max}\}$ denote the kth time step, i.e., at time kh. We will be using the same state-space as our continuous time model in (6)–(15), but to simplify the notation, we introduce the vectors $q(k) = [q_{1,1}(k), q_{1,2}(k), q_{2,1}(k), q_{2,2}(k)]$, $f(k) = [f_{1,1}(q_k), f_{1,2}(k), f_{2,1}(k), f_{2,2}(k)]$, $r(k) = [r_{1,1}(k), r_{1,2}(k), r_{2,1}(k), r_{2,2}(k)]$, $n^O(k) = [n_{1,1}^O(k), n_{1,2}^O(k), n_{2,1}^O(k), n_{2,2}^O(k)]$, $n^I(k) = [n_1^I(k), n_2^I(k)]$, and $\tilde{u}(k) = [\tilde{u}_{1,2}(k), \tilde{u}_{2,1}(k)]$, where $\tilde{u}$ corresponds to the number of vehicles rebalanced, i.e., $0 \leq \tilde{u}_{1,2}(k) \leq n_1^I(k)$ and $0 \leq \tilde{u}_{2,1}(k) \leq n_2^I(k)$, instead of the fraction previously used as the control input. Moreover, we will introduce additional variables that will model the actual matching rate, $\alpha(k) = [\alpha_{1,1}(k), \alpha_{1,2}(k), \alpha_{2,1}(k), \alpha_{2,2}(k)]$, which we will allow to be lower than the matching rates given by $f(k)$. The reason for introducing $\alpha(k)$ is two-fold. First, by introducing them, we will later see how the MPC problem becomes a convex problem. Second, by naively discretizing the system (6)–(15), the introduction of α prevents some of the states from going below zero. However, as we will see in the numerical studies in the next section, for the problem we are studying, $\alpha(k)$ will equal the true matching rates.

With the vector $\alpha(k)$, the optimal control problem in discrete time can be formulated as

$$\text{minimize} \quad \sum_{k \in \mathcal{K}} q_{1,1}(k) + q_{1,2}(k) + q_{2,1}(k) + q_{2,2}(k)$$

subject to

$$q(k+1) = q(k) + h(r(k) - \alpha(k))\,, \quad \forall k \in \bar{\mathcal{K}}\,, \quad (27)$$

$$n^O(k+1) = n^O(k) + h\alpha(k) - h\begin{bmatrix} \gamma_1 & 0 & \gamma_2 & 0 \\ 0 & \gamma_1 & 0 & 0 \\ 0 & 0 & \gamma_2 & 0 \\ 0 & \gamma_1 & 0 & \gamma_2 \end{bmatrix} n^O(k)\,, \forall k \in \bar{\mathcal{K}}\,, \quad (28)$$

$$
\begin{aligned}
n^I(k+1) =& n^I(k) + h \begin{bmatrix} \gamma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_2 \end{bmatrix} n^O(k) \\
&- h \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \alpha(k) \\
&+ h \begin{bmatrix} -\gamma_1 & \gamma_2 \\ \gamma_1 & -\gamma_2 \end{bmatrix} \tilde{u}(k)\,, \quad \forall k \in \bar{\mathcal{K}}\,, \quad (29)
\end{aligned}
$$

$$0 \le \tilde{u}_{1,2}(k) \le n_1^I(k)\,, \quad \forall k \in \mathcal{K}\,, \quad (30)$$

$$0 \le \tilde{u}_{2,1}(k) \le n_2^I(k)\,, \quad \forall k \in \mathcal{K}\,, \quad (31)$$

$$n^I(k) \ge 0\,, n^O(k) \ge 0\,, n^I(k) \ge 0\,, \forall k \in \mathcal{K}\,, \quad (32)$$

$$0 \le \alpha(k) \le f(k)\,, \quad \forall k \in \mathcal{K}\,, \quad (33)$$

where $\bar{\mathcal{K}} := \mathcal{K} \setminus \{k^{\max}\}$ and the inequalities in (32) and (33) applies element-wise.

Proposition 3: If the matching functions $f(k)$ are concave, then the optimal rebalancing problem is convex.

Proof The objective function is linear. All the equality constraints, i.e., (27), (28), (29), are affine in the decision variables. The inequality constraints (30), (31) are all concave, and hence the problem is convex. ■

V. NUMERICAL EXAMPLE

To demonstrate the dynamics and investigate the controller, we study the two regions setting with $\gamma_1 = 3$ trips/hour, $\gamma_2 = 2$ trips/hour. We start the simulations with no occupied vehicles, and let $n_1^I(0) = 170$ vehicles and $n_2^I(0) = 50$ vehicles. Moreover, we let each of the four request queues be initiated with 10 users, i.e., $q_{1,1}(0) = q_{1,2}(0) = q_{2,1}(0) = q_{2,2} = 10$. We run the simulations with a step size $h = 0.01$ for 800 steps, i.e., the total simulation horizon is 8 hours. For simplicity, we let the matching functions for all queues be $f_{i,j}(q_{i,j}, n_i^I) = \sqrt{q_{i,j} n_i^I}$.

In our first simulation, we let the request rates be $r_{1,1} = 30, r_{1,2} = 10, r_{2,1} = 50$, and $r_{2,2} = 10$ requests/hour, which we refer to as the low demand scenario. For those requests rates, the necessary condition in Proposition 2 is clearly satisfied, since $\frac{r_{1,1}+2r_{2,1}}{\gamma_1} + \frac{r_{1,2}+r_{2,1}+r_{2,2}}{\gamma_2} \approx 78 < n(0) = 220$.

The trajectories for this simulation are shown in Figure 2. First, we observe that the α values in the relaxations are so close to the matching rates that we can conclude that the convex relaxation is tight in this setting. As expected, the queues stay stable. More interestingly, the controller sends empty vehicles in both directions. A question for further research is if it is possible to rescale the control action such that this behavior is avoided while still keeping the same overall performance.

To further test the MPC's ability when the necessary condition is barely satisfied, we increase the demands to be $r_{1,1} = 78, r_{1,2} = 26, r_{2,1} = 130$, and $r_{2,2} = 26$, which we refer to as the high demand scenario. Now, $\frac{r_{1,1}+2r_{2,1}}{\gamma_1} + \frac{r_{1,2}+r_{2,1}+r_{2,2}}{\gamma_2} \approx 204 < n(0) = 220$.

To better capture the convergence of the queues, we now let the simulation run for 2000 steps with $h = 0.1$ instead, although this horizon is superficial from a practical viewpoint. Figure 3 shows the trajectories for this demand profile. Since the demands barely satisfy the necessary conditions, it

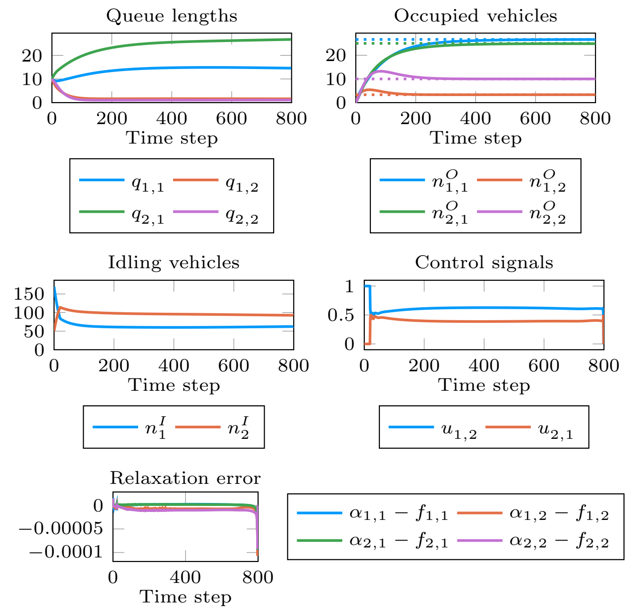

Fig. 2. The trajectories with the MPC for the low demand scenario. For the occupied vehicles, the dotted lines represent the equilibrium.

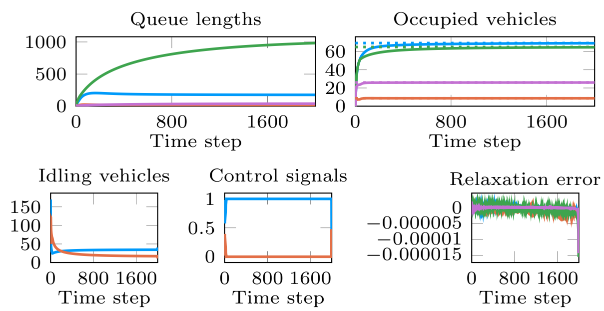

Fig. 3. The trajectories with the MPC for the high demand scenario. For the occupied vehicles, the dotted lines represent the equilibrium. For legend description, see Fig. 2.

takes longer for the system to converge, but it still converges. Therefore, we conjecture that the necessary conditions may also be sufficient when the control signal is optimal, but further theoretical investigations are needed in this regard.

For comparison, we also do simulations with two myopic feedback controllers. The first one rebalances the vehicles in proportion to the difference in the aggregate queue lengths in each region, such that

$$u_{1,2}^{\mathrm{P}} = \begin{cases} \frac{q_2 - q_1}{q_2} & \text{if } q_2 \ge q_1 \\ 0 & \text{otherwise} \end{cases}, \quad u_{2,1}^{\mathrm{P}} = \begin{cases} \frac{q_1 - q_2}{q_1} & \text{if } q_1 > q_2 \\ 0 & \text{otherwise,} \end{cases}$$

where $q_1 = q_{1,1} + q_{1,2}$ and $q_2 = q_{2,1} + q_{2,2}$. We also try with a bang-bang-like controller,

$$u_{1,2}^{\mathrm{BB}} = \begin{cases} 1 & \text{if } q_2 > q_1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad u_{2,1}^{\mathrm{BB}} = \begin{cases} 1 & \text{if } q_1 > q_2 \\ 0 & \text{otherwise.} \end{cases}$$

The trajectories for both myopic controllers in the low-demand scenario are shown in Figure 4 and for the high-demand scenario in Figure 5.

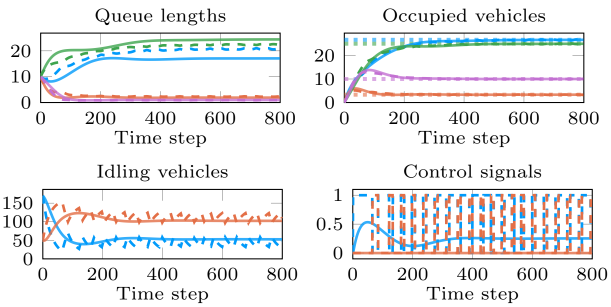

Fig. 4. The trajectories with the proportional controller (solid) and the bang-bang controller (dashed) for the low-demand scenario. For the occupied vehicles, the dotted lines represent the equilibrium. For legend description, see Fig. 2.

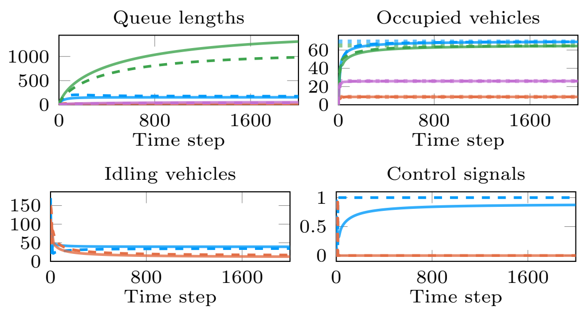

Fig. 5. The trajectories with the proportional controller (solid) and the bang-bang controller (dashed) for the high-demand scenario. For the occupied vehicles, the dotted lines represent the equilibrium. For legend description, see Fig. 2.

A summary of all controllers' performances is shown in Table I, where $q = q_1 + q_2$. The results show that the bang bang controller performs almost as well as the MPC for both the low-demand and high-demand scenarios. However, the MPC yields a smoother rebalancing strategy.

VI. Conclusion

This paper presents a macroscopic model for ride-hailing services with a fixed fleet size that allows for rebalancing empty vehicles between different regions. We analyzed the equilibrium properties and provided necessary conditions on the ride-hailing demand levels for the system to stay stable. Moreover, we showed how the rebalancing control action could be determined through a convex MPC problem.

In the future, we plan to analytically study the stability properties of the different controllers proposed in this paper. Moreover, we plan to integrate the current framework in a hierarchical framework, where at the upper level, i.e., regional level, rebalancing is done with a macroscopic approach, while on the lower level, each vehicle is controlled individually.

TABLE I

Comparison of the different controller's performance

Scenario	Controller	Bounded q	$\frac{h}{T}\sum_{k=0}^{T} q(k)$
Low-demand	MPC	Yes	0.41
Low-demand	u^{P}	Yes	0.42
Low-demand	u^{BB}	Yes	0.43
High-demand	MPC	Yes	101.51
High-demand	u^{P}	Yes[1]	121.99
High-demand	u^{BB}	Yes[1]	101.54

[1] Convergence observed when extending the simulation to 10000 time steps

References

[1] A. Carron, F. Seccamonte, C. Ruch, E. Frazzoli, and M. N. Zeilinger, "Scalable model predictive control for autonomous mobility-on-demand systems," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 2, pp. 635–644, 2021.

[2] M. Tsao, D. Milojevic, C. Ruch, M. Salazar, E. Frazzoli, and M. Pavone, "Model predictive control of ride-sharing autonomous mobility-on-demand systems," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6665–6671.

[3] P. Zhu, I. I. Sirmatel, G. F. Trecate, and N. Geroliminis, "Idle-vehicle rebalancing coverage control for ride-sourcing systems," in *2022 European Control Conference (ECC)*, 2022, pp. 1970–1975.

[4] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

[5] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, 2008.

[6] M. Hajiahmadi, V. L. Knoop, B. De Schutter, and H. Hellendoorn, "Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 1022–1028.

[7] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2013.

[8] M. Ramezani and M. Nourinejad, "Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach," *Transportation Research Part C: Emerging Technologies*, vol. 94, pp. 203–219, 2018, iSTTT22.

[9] T. Harko, F. S. Lobo, and M. Mak, "Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates," *Applied Mathematics and Computation*, vol. 236, pp. 184–194, 2014.

[10] F. Horn and R. Jackson, "General mass action kinetics," *Archive for Rational Mechanics and Analysis*, vol. 47, no. 2, pp. 81–116, 1972.

[11] F. Blanchini and G. Giordano, "Piecewise-linear Lyapunov functions for structural stability of biochemical networks," *Automatica*, vol. 50, no. 10, pp. 2482–2493, 2014.

[12] L. Massoulié, "Structural properties of proportional fairness: Stability and insensitivity," *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 809–839, 06 2007.

[13] G. Nilsson and G. Como, "Generalized proportional allocation policies for robust control of dynamical flow networks," *IEEE Transactions on Automatic Control*, vol. 67, no. 1, pp. 32–47, 2022.

[14] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177 – 195, 2013.

[15] G. Como, E. Lovisari, and K. Savla, "Convexity and robustness of dynamic traffic assignment and freeway network control," *Transportation Research Part B: Methodological*, vol. 91, pp. 446–465, 2016.

[16] H. Yang and T. Yang, "Equilibrium properties of taxi markets with search frictions," *Transportation Research Part B: Methodological*, vol. 45, no. 4, pp. 696–713, 2011.

[17] S. Wei, S. Feng, J. Ke, and H. Yang, "Calibration and validation of matching functions for ride-sourcing markets," *Communications in Transportation Research*, vol. 2, p. 100058, 2022.

[18] G. Como, E. Lovisari, and K. Savla, "Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 57–67, 2015.