Near Optimal Control of a Ride-Hailing Platform via Mirror Backpressure

(Preliminary version, including Section 8. Feedback welcome.)

Yash Kanoria[*] Pengyu Qian[†]

March 31, 2019

Abstract

Ride-hailing platforms need to match supply and demand so as to maximize the rate of payoff generation while respecting the geographical flow constraints, namely, that the rate at which vehicles arrive at a location must equal the rate at which they leave that location. The platform's control levers include: (i) *entry control*, i.e. it can choose not to serve some customers, (ii) *dynamic pricing*, i.e. it can set prices for each ride depending on its origin and destination, and (iii) *assignment rule*, i.e. it can choose from which neighboring location to dispatch a car for pickup.

We consider two settings depending on whether pricing is an available lever to the platform. In joint-entry-assignment (JEA) setting, the platform can only use levers (i) and (iii) (for example, Didi in China does not use dynamic pricing). In joint-pricing-assignment (JPA) setting, the platform can use levers (ii) and (iii) (e.g. most ride-hailing platforms in North America). We introduce a novel family of *Mirror Backpressure* (MBP) platform control policies which are simple, practical, and do not require prior knowledge of the demand arrival rates. Key challenges include that serving a customer reduces supply availability at the dispatch location but increases it at the dropoff location ("supply externalities"), and that the number of vehicles at any location can never be negative ("no-underflow constraints"). Mirror Backpressure generalizes the backpressure policy such that it executes mirror descent, allowing us to address these challenges. MBP loses at most an $O(K/T + 1/K)$ fraction of the achievable payoff in the JEA setting and $O(\sqrt{K/T + 1/K})$ fraction in the JPA setting, where K is the number of vehicles and T is the horizon. Simulation results in a realistic environment support our theoretical findings.

Keywords: queueing network; optimal control; backpressure; maximum weight; mirror descent; dynamic pricing; ride-hailing.

[*] Graduate School of Business, Columbia University. Email: ykanoria@columbia.edu

[†] Graduate School of Business, Columbia University. Email: PQian20@gsb.columbia.edu

1 Introduction

We consider the control problem of shared transportation platforms for ride-hailing such as Uber, Lyft, DiDi and Ola. These platforms can be modeled as dynamic two-sided markets where demand units (customers) with different origin-destination pairs arrive stochastically over time. The system dynamically decides which customers to serve (or sets prices) and assigns supply units (vehicles) to customers based on the current system state. (For example, Didi in China does not use dynamic pricing to manage demand-supply imbalances. Instead it serves customers selectively, e.g., by showing a long wait time to customers in under-supplied locations, causing many such customers to abandon. In contrast, dynamic pricing is the most popular control lever in North America, Europe, etc.) Each assignment decision has two effects: it generates a certain *payoff* (this may include the platform's net revenue, the satisfaction of the customer, the driver, etc.), and *relocates* the assigned supply units from the dispatch location to the destination of the customer. The goal of the system is to maximize the collected payoff over a period of time.

The main challenges of designing a good control policy for the platform are the following:

1. *No-underflow constraints.* Because of the queueing network setting, each assignment decision needs to be backed by an available supply unit at the dispatch location. We refer to these constraints as *no-underflow constraints.* The no-underflow constraints capture the tension between the payoff effect and queueing effect of assignment decisions. It is one of the main difficulties in dynamic scheduling of queueing systems.

2. *Supply externalities.* Unlike traditional revenue management problems (e.g., see Gallego and Van Ryzin 1994) and dynamic bipartite matching problems (e.g., Caldentey et al. 2009) which study *one-hop* systems where the supply unit leaves the system once matched, our model is *multi-hop* with supply units circulating in the system. As a result, the service and assignment decisions need to consider both upstream (dispatch) and downstream (destination) locations.

In solving this problem, we introduce a novel class of joint-entry-assignment/joint-pricing-assignment control policies that we name *Mirror Backpressure (MBP).* MBP generalizes the celebrated backpressure (maximum weight) policy (Tassiulas and Ephremides 1992, Dai and Lin 2005) from the queueing literature in such a way that it executes mirror descent (Nemirovsky and Yudin 1983, Beck and Teboulle 2003) on the controller's optimization problem. The tight connection with mirror descent allows problem geometry (no-underflow constraints) to be incorporated systematically in the design of the control policy. We instantiate our general approach to provide an elegant solution to the ride-hailing control problem at hand; we are optimistic this will be one of many problems in the control of queueing networks that MBP will help solve.

Backpressure is very attractive as a simple and practical policy for state dependent control of queueing networks that does not require knowledge of arrival and service rates (it is a perfect candidate to handle supply externalities in isolation). However, it fails to systematically incorporate no-underflow constraints, and the workarounds that have been proposed have significant limitations that make them unsuitable in many settings, including ours (see Section 1.1). MBP generalizes backpressure, with the crucial feature that the MBP policy executes stochastic mirror descent on the operator's optimization problem with flow constraints dualized (this generalizes the known property that backpressure executes stochastic gradient descent with the queue lengths serving as the dual variables, see Georgiadis et al. 2006). The rough idea is as follows: The queue congestion cost variables $\mathbf{y}$ (dual variables to the flow constraints, one per queue) are the ones being optimized, via gradient steps in the queue length vector $\mathbf{q}$ (the mirror point). As is typical in applications of mirror descent (e.g., exponentiated gradient descent, see Kivinen and Warmuth 1997), the *mirror map* $\mathbf{y} \mapsto \mathbf{q}$ is chosen based on problem geometry, with the twist that the relevant constraints (no-underflow constraints) are on $\mathbf{q}$, i.e., in the mirror space (in typical applications of mirror descent, the constraints are on $\mathbf{y}$, in the original space). The inverse mirror map $\mathbf{q} \mapsto \mathbf{y}$ then defines the right MBP policy. Convergence of mirror descent to the optimum $\mathbf{y}^*$ leads to near optimality of MBP; the technical machinery of mirror descent is applicable. In our ride-hailing control problem on a closed network, the sum of queue lengths is constant, so Kullback-Leibler divergence emerges as the right notion of "distance" in the mirror space of normalized queue lengths, which leads to the inverse mirror map $y_i \propto \log(q_i)$ for each location i in MBP, and provably near optimal performance.

We now return to the ride-hailing control problem. We model the system as a closed queueing network with m *supply nodes* (locations) and n *demand nodes* (locations, which may be identical to the supply locations), and K supply units that circulate in the system. When a supply unit is assigned to a customer, it will drop off the customer at the destination and become available again; meanwhile the assignment generates a payoff depending on the supply unit's original location, the customer's origin and destination. This model is appropriate for ride-hailing systems where supply units are typically long-lived (see Banerjee et al. 2016, Waserhole and Jost 2016, Banerjee et al. 2018). For each demand location j' there are several *compatible supply locations* that are close enough, from where the system can assign supply units to pick up demand at j'. Customers arrive stochastically over slotted time. At each period the system uses entry control (JEA setting) or pricing (JPA setting) to modulate demand, and makes assignment decisions. We consider the JEA setting from Section 2 to Section 7 and use it as a primary example to illustrate the idea of MBP poilcy. In Section 8 we consider the JPA setting and show that most of the intuition in JEA setting carry through.

The platform's goal is to maximize the collected payoff. We first derive an upper bound for the average per period payoff over any time horizon through the system's fluid limit, and then evaluate the performance of policies through *regret* with respect to this upper bound. Our result is *non-asymptotic*, i.e. the regret bound is valid for finite number of supply units and over a finite horizon, thus covering both transient and stationary performance. See Section 2 for the formal definition of the JEA setting, and Section 8 for the JPA setting. The MBP policy does not require knowledge of the demand arrival rates, making it promising for applications.

To keep the state space manageable, we make the key simplification that assigned supply units relocate to the customer's destination in the next time slot, i.e. pickup and service of demand are both *instantaneous*. This allows us to avoid the complexity of tracking the positions of all in-transit supply units, while retaining the essence of our main challenge, i.e., maximizing payoff in the face of the no-underflow constraints and supply externalities. Realistic simulations based on NYC yellow cab data show that our theoretical findings strikingly retain their power even with transit times, achieving excellent performance over both short and long time horizons (Section 7).

Main Contributions. To summarize, we make two main contributions:

1. **A near optimal state dependent control policy for a ride-hailing platform.** In both JEA and JPA setting, we propose a family of dynamic control policies, the Mirror Backpressure policies, that addresses both the no-underflow constraints and supply externalities in a natural way. We show that the MBP policy loses at most an $O(K/T + 1/K)$ fraction of the achievable payoff in the JEA setting and an $O(\sqrt{K/T + 1/K})$ fraction in the JPA setting, where K is the number of vehicles and T is the horizon. The MBP policy is simple and practical, and does not require knowledge of demand arrival rates, making it promising for applications.

2. **Novel Mirror Backpressure control policy to systematically handle no-underflow constraints.** MBP generalizes backpressure in such a way that it executes mirror descent on the controller's optimization problem, which allows problem geometry (no-underflow constraints) to be incorporated systematically in the design of the control policy. We are optimistic this will be the first of many problems in the control of queueing networks for which MBP will provide provably near optimal policies.

1.1 Literature Review

MaxWeight/backpressure scheduling. MaxWeight and backpressure (Tassiulas and Ephremides 1992, Georgiadis et al. 2006) are well-studied scheduling policies for workload minimization (Stolyar 2004, Dai and Lin 2008), queue length minimization (Eryilmaz and Srikant 2012) and utility maximization (Eryilmaz and Srikant 2007), etc. in constrained queueing networks. One of the main attractive features of MaxWeight/backpressue policies is that they can achieve provably good performance without requiring any statistical knowledge of the system randomness. Most of the literature considers the open queueing networks setting, and there is much less work on closed networks. An exception is a recent paper on state dependent control of ride-hailing platforms (Banerjee et al. 2018), which shows large deviations optimality of a non-idling scaled MaxWeight policy in a setting which assumes equal pickup costs as well as Hall's matching condition (the present paper makes neither of these assumptions; in the absence of Hall's condition the near optimal policy we obtain is an idling policy). Another key difference between the existing works and our paper is that they use a power of queue lengths (Stolyar 2004) for decision making, while our analysis allows the use of a general nonlinear function (e.g., logarithm) of queue lengths.

No-underflow constraint in queueing control. The non-negativity constraint of queue lengths (a.k.a. no-underflow constraint) is one of the main challenges in the control of queueing systems. The popular approaches in the literature to address this problem are: (i) imposing assumptions on the network primitives (Dai and Lin 2005) such that backpressure policy does not violate no-underflow constraints, which rules out many networks of interest including our model; (ii) perturbing the MaxWeight policy (Huang and Neely 2011), where the perturbation parameters need to be carefully chosen; (iii) using virtual queue based controls (Stolyar 2005, Eryilmaz and Srikant 2007), which is difficult to analyze theoretically for finite systems and may fail entirely in our closed network setting, since the number of fake packets (generated when dispatching from an empty queue) would grow without bound. Our approach is systematic and provides a transient bound for finite systems (i.e., before taking the asymptotic limit).

Shared transportation systems. Most of this literature studied the state independent control of ride-hailing platforms: Ozkan and Ward (2016) studied revenue maximizing assignment control in an open queueing network model, Braverman et al. (2016) derived the optimal state independent routing policy that sends empty vehicles to under-supplied locations, Banerjee et al. (2016) adopted the Gordon-Newell closed queueing network model and considered state independent pricing/repositioning/matching policies that maximize throughput, welfare or revenue. Banerjee et al. (2018) which assumes CRP and equal pickup costs may be the only one on state dependent control, despite the key advantage that state dependent control can yield near optimal performance even in the absence of knowledge about system parameters. Comparing with Banerjee et al. (2016) which obtains a steady state regret scaling of $O(1/K)$ (in the absence of travel times) assuming *perfect* knowledge of demand arrival rates, our control policy achieves the same steady state regret for joint-entry-assignment controls with *no* knowledge of demand arrival rates, and further achieves a transient regret scaling of $O(K/T + 1/K)$ for a finite horizon T.

Several recent works study pricing aspects of ride-hailing; see, e.g., Adelman (2007), Bimpikis et al. (2016), Waserhole and Jost (2016), Cachon et al. (2017), Hall et al. (2015).

Other related works. There is a related stream of research on online stochastic bipartite matching, see, e.g., Caldentey et al. (2009), Adan and Weiss (2012), Buŷić and Meyn (2015), Mairesse and Moyal (2016). The main difference between their setting and ours is that we study a *closed* system where supply units never enter or leave the system. Jordan and Graves (1995), Désir et al. (2016), Shi et al. (2015) and others study how process flexibility can facilitate improved performance, analogous to our use of assignment control to maximize payoff (when all pickup costs are equal). Along similar lines, network revenue management is a classical dynamic resource allocation problem, see, e.g., Gallego and Van Ryzin (1994), Talluri and Van Ryzin (2006), and recent works, e.g., Bumpensanti and Wang (2018). Again, this setting is "open" in that each service token or supply unit can be used only once, in contrast to our setting.

1.2 Organization of the Paper

The remainder of our paper is organized as follows. From Section 2 to Section 7 we focus on the JEA setting as a primary example of our approach. Section 2 presents our model and the platform objective. Section 3 introduces the static fluid problem and reviews the connection between the "vanilla" backpressure policy and stochastic gradient decent. In Section 4 we introduce the Mirror Backpressure policy. Section 5 presents our main theoretical result, i.e., a performance guarantee for the MBP policy. Section 6 outlines the proof of our main result. In Section 7, we describe our simulation study of MBP policies using NYC yellow cab data. In Section 8, we design the MBP policy for the JPA setting, demonstrating the versatility of our approach. We conclude in Section 9 and discuss the future directions.

2 Network Model

2.1 Notation

All vectors are column vectors if not specified otherwise. We use $\mathbf{e}_i$ to denote the i-th unit column vector with the i-th coordinate being 1 and all other coordinates being 0, and $\mathbf{1}$ to denote the all 1 column vector, where the dimension of the vector will be indicated in the superscript when it is not clear from the context, e.g., $\mathbf{e}_i^n$. For two vectors $\mathbf{a} = (a_1, a_2, \dots)$ and $\mathbf{b} = (b_1, b_2, \dots)$ of the same dimension, their Hadamard (element-wise) product $\mathbf{a} \circ \mathbf{b}$ has the same dimension as the operands with elements given by $(\mathbf{a} \circ \mathbf{b})_i = a_i b_i$.

2.2 Setting

Supply/demand types and compatibility graph. We consider a finite-state Markov chain model with slotted time $t = 0, 1, 2, \dots$, where a fixed number (denoted by K) of identical *supply units* (vehicles) circulate among m *supply nodes* (locations), indexed by $i \in V_S$. (Our theoretical development will ignore travel times. As a result, in the context of a real system with travel times, an appropriate interpretation of K is the number of *free* cars in the system; see Section 7.) There are n *demand nodes* (locations, which may be identical to the supply locations) indexed by $j' \in V_D$. Each demand type is specified by an origin-destination pair (j', k) where $j' \in V_D$, $k \in V_S$, representing demand units (customers) going from j' to k. Demand (customers) arrive over time. To serve an arriving customer of type (j', k), the system must immediately assign (dispatch) a supply unit from one of the *pickup-compatible* supply nodes of j', and the assigned supply unit will subsequently relocate to supply node k. This flexibility structure is modeled by an undirected bipartite graph $G = (V_S \cup V_D, E)$, called the *compatibility graph* and we denote the neighborhood of any node $k \in V_S \cup V_D$ by $\mathcal{N}(k)$.

Demand arrival process. The demand in period t is denoted by $\mathbf{a}(t) \in \mathbb{N}^{n \times m}$, where $a_{j'k}(t)$ is the number of customers in period t who want to go from $j' \in V_D$ to $k \in V_S$. We assume $\mathbf{a}(t)$ to be i.i.d. across time with distribution $\mathcal{F}$, where $\mathcal{F}$ can be correlated across different origin-destination pairs. We make the following assumption on the distribution of demand arrivals at each time.

Condition 1 (Bounded support of demand arrival.)**.** *We assume that there exists $B < \infty$ such that for all $i \in V_S$ we have $\sum_{j' \in \mathcal{N}(i)} \sum_{k \in V_S} a_{j'k}(1) \leq B$ almost surely, i.e., in any time period, the number of customers requesting a pickup in the neighbourhood of i is at most B.*

Note that the total number of customers who arrive in any time period is upper bounded by mB. Let $\boldsymbol{\phi} \triangleq \mathbb{E}[\mathbf{a}(1)] \in \mathbb{R}_+^{n\times m}$ be the demand arrival rate vector.

Assignment payoff. Assigning a supply unit from location $i \in \mathcal{N}(j')$ to serve a type (j', k) customer produces a payoff $w_{ij'k} = w^c_{ij'} + w^r_{j'k}$, where $w^c_{ij'}$ is the cost of assignment (thus we would expect it to take on a negative value, though our results do not depend on this), and $w^r_{j'k}$ is the reward of serving a customer of type (j', k). Informally, a ride-hailing platform operator may set the cost $w^c_{ij'}$ based on criteria like the distance, pickup time, etc. between i and j', and the reward $w^r_{j'k}$ based on the fare the customer pays, the amount payable to the driver, the estimated surplus of the customer and of the driver, the duration of the trip, etc. (We treat the price as exogenous for now. This already captures the current environment for Didi in China, for instance, where dynamic pricing is *not* used. We will consider the pricing case in Section 8.) We define

$$w_{\max} \triangleq \max\{|w_{ij'k}|\colon j' \in V_D, k \in V_S, \phi_{j'k} > 0, i \in \mathcal{N}(j')\}\,. \tag{1}$$

2.3 Assignment Policies and System Dynamics

Assignment policies. An assignment policy $\pi \in \mathcal{U}$ picks an assignment decision $\mathbf{x}^\pi(t)$ in every time slot t based on the current (supply) queue lengths $\mathbf{q}(t)$. (Because the model considered is a finite state Markov decision process, there exists a Markov policy that is optimal (Derman 1970).) The assignment decision $\mathbf{x}^\pi(t)$ for each t is constrained by the compatibility graph G. Specifically, letting $r \triangleq m(\sum_{j'\in V_D} |\mathcal{N}(j')|)$, we have

$$\mathbf{x}^\pi(t) \in \mathcal{X} \triangleq \left\{ (x_{ij'k})_{j'\in V_D, k\in V_S, i\in\mathcal{N}(j')} \in \{0,1\}^r : \sum_{i\in\mathcal{N}(j')} x_{ij'k} \le 1, \forall j' \in V_D, k \in V_S \right\}. \tag{2}$$

Note that $\mathcal{X}$ admits the following decomposition:

$$\begin{aligned} &\mathcal{X} = \Pi_{j'\in V_D, k\in V_S} \mathcal{X}_{j'k}\,, \\ &\text{where } \mathcal{X}_{j'k} \triangleq \left\{ \mathbf{x}_{j'k} \in \{0,1\}^{|\mathcal{N}(j')|} : \mathbf{x}_{j'k} = \mathbf{e}_i^{|\mathcal{N}(j')|} \text{ for some } i \in \mathcal{N}(j') \right\} \cup \{\mathbf{0}\}\,, \end{aligned} \tag{3}$$

where Π over sets denotes the Cartesian product set. Here $x_{ij'k} = 1$ if and only if the system assigns from node i to all the type (j', k) demand arriving during this period. The assignment decision includes *whether* to serve each customer type; $\mathbf{x}_{j'k} = \mathbf{0}$ corresponds to customers of type (j', k) not being served (we say such demand is "dropped"). The convex hull of $\mathcal{X}_{j'k}$ is $\mathrm{conv}(\mathcal{X}_{j'k}) = \left\{ \mathbf{x}_{j'k} \in \mathbb{R}^{|\mathcal{N}(j')|} : \mathbf{1}^{\mathrm{T}}\mathbf{x}_{j'k} \le 1,\ \mathbf{x}_{j'k} \ge \mathbf{0} \right\}$. Elements of $\mathrm{conv}(\mathcal{X}_{j'k})$ can be interpreted as randomized or mixed assignment decisions.

System dynamics. In the t-th time slot, the following events occur in sequence.

- At the beginning of the time slot, based on the current (supply) queue lengths $\mathbf{q}(t) \in \mathbb{R}^m$ and time t, the system makes an assignment decision $\mathbf{x}(t) \in \mathcal{X}$.
- The demand arrival $\mathbf{a}(t) \in \mathbb{N}^{n \times m}$ reveals itself.
- The system evolves according to the following dynamics:

$$q_i(t+1) = \; q_i(t) \; - \sum_{j' \in \mathcal{N}(i)} \sum_{k \in V_S} a_{j'k}(t) x_{ij'k}(t) + \sum_{j' \in V_D} \sum_{l \in \mathcal{N}(j')} a_{j'i}(t) x_{lj'i}(t) \,, \quad \forall i \in V_S \,. \tag{4}$$

To make the notation more compact, we introduce the *input-output* matrix $\mathbf{R} \in \mathbb{R}^{m \times r}$, where

$$R_{l,(ij'k)} \triangleq \mathbb{1}\{l = i\} - \mathbb{1}\{l = k\} \,.$$

In other words, $R_{l,(ij'k)}$ is the number of type l supply units *consumed* as the result of assigning a type i supply unit to a type (j', k) customer. Let $\tilde{\mathbf{a}}(t) \in \mathbb{R}^r$ be the augmented demand arrival vector where $\tilde{a}_{ij'k}(t) \triangleq a_{j'k}(t)$ for all $j' \in V_D, k \in V_S, i \in \mathcal{N}(j')$. Then we can rewrite (4) in terms of the element-wise product $\tilde{\mathbf{a}}(t) \circ \mathbf{x}(t)$ as

$$\mathbf{q}(t+1) = \mathbf{q}(t) - \mathbf{R}(\tilde{\mathbf{a}}(t) \circ \mathbf{x}(t)) \,. \tag{5}$$

Since the operator cannot dispatch more supply units than the number available at that location at that time, the assignment $\mathbf{x}(t)$ at each time needs to satisfy the *no-underflow constraints*:

$$\sum_{j' \in \mathcal{N}(i)} \sum_{k \in V_S} a_{j'k}(t) x_{ij'k}(t) \le q_i(t) \quad \forall i \in V_S \,. \tag{6}$$

Analogous to $\tilde{\mathbf{a}}(t)$, we find it convenient to define the augmented demand arrival rate vector $\tilde{\boldsymbol{\phi}} \in \mathbb{R}^r$ by $\tilde{\phi}_{ij'k} \triangleq \phi_{j'k}$. In particular, $\mathbb{E}[\tilde{\mathbf{a}}(t)] = \tilde{\boldsymbol{\phi}}$.

2.4 Platform Objective

Let $v^\pi(t) \triangleq \mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}(t) \circ \mathbf{x}(t))$ be the payoff obtained in the t-th time slot under policy π. The platform's goal is to maximize the average payoff over a finite horizon of length $T < \infty$, i.e.,

$$\text{maximize}_\pi \; \bar{v}^\pi(T) \quad \text{for } \bar{v}^\pi(T) \triangleq \frac{1}{T} \mathbb{E}\left[\sum_{t=1}^{T} v^\pi(t)\right] \,. \tag{7}$$

3 The Static Problem and the Vanilla Backpressure Policy

In this section, we first define a deterministic optimization problem (the "fluid limit"), whose value we use to upper bound the optimal (average) payoff the system can generate in the first T time slots. We then review the interpretation of the celebrated Vanilla Backpressure policy as a stochastic gradient descent algorithm on the dual of the deterministic problem as a precursor to the generalization of backpressure we will introduce in the next section.

3.1 The (Primal) Static Fluid Problem

Consider the following "static fluid problem":

$$
\begin{aligned}
&\text{maximize}_{\mathbf{x}} \sum_{j' \in V_D} \sum_{k \in V_S} \phi_{j'k} \sum_{i \in \mathcal{N}(j')} w_{ij'k} x_{ij'k} && (8)\\
&\text{s.t.} \sum_{j' \in \mathcal{N}(i)} \sum_{k \in V_S} \phi_{j'k} x_{ij'k} - \sum_{j' \in V_D} \sum_{l \in \mathcal{N}(j)} \phi_{j'i} x_{lj'i} = 0 \quad \forall i \in V_S && \text{(flow balance)}\,,\\
&\sum_{i \in \mathcal{N}(j')} x_{ij'k} \le 1 \quad \forall j' \in V_D,\ k \in V_S, && \text{(demand constraint)}\,,\\
&x_{ij'k} \ge 0 \quad \forall i,k \in V_S,\ j' \in V_D\,.
\end{aligned}
$$

In matrix form:

$$
\begin{aligned}
&\text{maximize}_{\mathbf{x}}\ \mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) && (9)\\
&\text{s.t.}\ \mathbf{R}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) = \mathbf{0}\,, && (10)\\
&\mathbf{x}_{j'k} \in \operatorname{conv}(\mathcal{X}_{j'k}) \qquad \forall j' \in V_D\,, k \in V_S\,. && (11)
\end{aligned}
$$

The idea behind the above problem is that, ignoring stochastic fluctuations, for a static policy to be feasible in the long run, the inflow of supplies to each node in V_S must equal the outflow in order to avoid underflow, and the policy must employ a feasible (possibly mixed/randomized) assignment rule $\mathbf{x}_{j'k}$ for each demand type (j', k). Subject to these constraints, the policy is chosen to maximize the per period payoff generated.

We show the following upper bound on the expected payoff of any policy over a finite horizon, in terms of W^{OPT}.

Proposition 1. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}(0)$, the expected payoff generated by any policy $\pi \in \mathcal{U}$ is upper bounded as*

$$\bar{v}^{\pi}(T) \le W^{\mathrm{OPT}} + \frac{(m-1)w_{\max}K}{T}\,, \tag{12}$$

where W^{OPT} is the value of the static fluid problem. In particular, the long run payoff under any policy $\pi \in \mathcal{U}$ is bounded above as $\limsup_{T\to\infty} \bar{v}^{\pi}(T) \le W^{\mathrm{OPT}}$.

The proof is in Appendix A.

The idea behind Proposition 1 is as follows. As is typical in such settings, W^{OPT} is an upper bound to the long run expected payoff. We obtain a *finite horizon* upper bound in addition by slightly relaxing the flow constraint (10) in the static fluid problem to

$$|\mathbf{1}_S^T \mathbf{R}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x})| \leq \frac{K}{T} \qquad \forall\, S \subset V_S\,, \tag{13}$$

where $\mathbf{1}_S$ is the vector with 1s at nodes in S and 0s at all other nodes. In words, constraint (13) requires that for any subset S of supply nodes, the magnitude of net per period flow into S should not exceed K/T, i.e., over T periods the net movement of supply into or out of S should not exceed K. Clearly this constraint must hold since there are only K supply units circulating in the system. Call the optimization problem defined by (9), (13) and (11) the *finite horizon fluid problem*, and denote its value by W_T^{OPT}. The proof of Proposition 1 has to two key components. First we show that W_T^{OPT} is an upper bound on the expected payoff over horizon T under any policy. Second, we show that $W_T^{\mathrm{OPT}} \leq W^{\mathrm{OPT}} + \frac{(m-1)w_{\max}K}{T}$ by decomposing any feasible $\mathbf{x}$ for the finite horizon fluid problem into a circulation (feasible for the static fluid problem) plus a directed acyclic flow with at most K/T flow crossing any cut. Combining we get the proposition.

3.2 Dual Subgradient Descent and Vanilla Backpressure Policy

Note that the demand constraint (11) is satisfied at all time slots under any policy, but the flow balance constraint (10) is a long-term constraint and can be violated at an individual time slot. To satisfy the flow balance in the long run, we introduce Lagrange multipliers $\mathbf{y} \in \mathbb{R}^m$ for (10) and derive the partial dual problem of (9)-(11):

$$\text{minimize}_{\mathbf{y}}\ g(\mathbf{y})\,,\ \text{ for } g(\mathbf{y}) \triangleq \max_{\mathbf{x}\in\mathcal{X}}\ (\mathbf{w}^{\mathrm{T}} + \mathbf{y}^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) = \sum_{j'\in V_D, k\in V_S} \phi_{j'k} \left[\max_{i\in\mathcal{N}(j')}\ \left(w_{ij'k} + y_i - y_k\right)\right]^+\,, \tag{14}$$

where $[\ell]^+ \triangleq \max(0, \ell)$.

The dual variables to the flow balance constraints can be interpreted as the "congestion costs" (see, e.g., Srikant 2012), i.e., y_i can be thought of as the price of having one extra supply unit at node i. Rich dividends have been obtained by treating the current queue lengths $\mathbf{q}$ as the dual variables $\mathbf{y}$ to the flow balance constraints, resulting in the celebrated maximum weight and backpressure control policies, introduced in the thesis of Tassiulas (Tassiulas 1992). If we take the current normalized queue lengths $\bar{\mathbf{q}}(t) \triangleq \mathbf{q}(t)/K$ as the current value of $\mathbf{y}$, and greedily maximize the quantity (assignment payoff + congestion cost at dispatch location - congestion cost at destination), we end up with the backpressure policy (also called primal-dual policy, see, e.g., Stolyar 2005, Eryilmaz and Srikant 2007). We henceforth refer to it as the *Vanilla Backpressure (VBP)* policy

$$x_{ij'k}^{\mathrm{VBP}}(t) \triangleq \begin{cases} 1 & \text{if } i = \mathrm{argmax}_{l \in \mathcal{N}(j')} \ \left\{ w_{lj'k} + \bar{q}_l(t) - \bar{q}_k(t) \right\} \text{ and } w_{ij'k} + \bar{q}_l(t) - \bar{q}_k(t) \geq 0\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{15}$$

with ties broken arbitrarily.

It is worth noting that $(\mathbf{w} + \mathbf{R}^{\mathrm{T}}\mathbf{y}) \circ \tilde{\boldsymbol{\phi}}$ is the *reduced cost* in linear programming (see, e.g., Bertsimas and Tsitsiklis 1997). We will describe the main attractive feature of this policy and then point out several problems with it in our setting, motivating the novel generalization of backpressure that we introduce in the next section.

The VBP policy can be viewed as a *stochastic subgradient descent (SGD)* algorithm on the dual problem (14) in the *interior* of the state space, i.e., when all the supply nodes have sufficient supply units (see, e.g., Stolyar 2003, Huang and Neely 2009). First, note that the definition of VBP (15) written in matrix form is

$$\mathbf{x}^{\mathrm{VBP}}(t) = \arg\max_{\mathbf{x} \in \mathcal{X}} \ (\mathbf{w}^{\mathrm{T}} + \bar{\mathbf{q}}(t)^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x})\,. \tag{16}$$

Denote the subdifferential (set of subgradients) of function $g(\mathbf{y})$ as $\partial g(\mathbf{y})$. Define the dual variable at the t-th time slot by $\mathbf{y}(t) \triangleq \bar{\mathbf{q}}(t)$ corresponding to the definition of VBP (16). Recalling $\mathbb{E}[\tilde{\mathbf{a}}] = \tilde{\boldsymbol{\phi}}$, one can verify that *the expected change in queue lengths is proportional to the negative of a subgradient of* $g(\cdot)$ *at* $\mathbf{y} = \mathbf{y}(t) = \bar{\mathbf{q}}(t)$, in particular

$$\mathbb{E}[\mathbf{R}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\mathrm{VBP}}(t))] = \ \mathbf{R}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{VBP}}(t)) \in \partial g(\bar{\mathbf{q}}(t)) = \partial g(\mathbf{y}(t))\,, \tag{17}$$

$$\mathbf{y}(t+1) = \ \mathbf{y}(t) - \frac{1}{K}\mathbf{R}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\mathrm{VBP}}(t)) \tag{18}$$

using (5). (18) is exactly an iteration of stochastic subgradient descent with step size $\frac{1}{K}$.

This interpretation of VBP as stochastic subgradient descent leads to desirable properties in open networks including stability, minimization of delay/workload, and payoff maximization (see, e.g., Georgiadis et al. 2006, Eryilmaz and Srikant 2007, etc.). However, the success does not translate to our closed queuing network setting for the following reasons.

Problem 1: The dual optimum may lie outside the positive orthant. The first problem is that the relevant dual optimum may not be in the positive orthant, and so $\bar{\mathbf{q}}(t)$ cannot get close to it. Note that one of our flow constraints is redundant (flow balance at any $m-1$ nodes automatically implies flow balance at the remaining node), and correspondingly, $g(\mathbf{y})$ is unchanged if all the y_is are translated by the same amount, i.e.,

$$g(\mathbf{y} + \xi\mathbf{1}) = g(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbb{R}^m, \xi \in \mathbb{R}\,. \tag{19}$$

As a result the dual problem (14) has a *family* of optima resulting from translating any optimum. Since for closed queueing networks we have $\sum_{i\in V_S} \bar{q}_i = 1$, the proxy dual variables $\bar{\mathbf{q}}(t)$ will converge to a member of the family of optimal $\mathbf{y}$ which satisfies $\sum_{i\in V_S} y_i = 1$. However, this particular optimum lies *outside* the positive orthant in a large class of instances. See Appendix B for such an example. In any such instance, the proxy dual variables $\bar{\mathbf{q}}(t)$ under the VBP policy cannot approach a dual optimum and hence the payoff under the VBP policy does not achieve the upper bound even asymptotically (the long run per period payoff is bounded away from W^{OPT} even for $K \to \infty$). A possible solution is to ensure that $\bar{\mathbf{q}}$ corresponding to the dual optimum lies in the positive orthant by interpreting a suitable translation $\bar{q}_i(t) - \zeta_i$ as the corresponding dual variable (Huang and Neely 2011), but then choosing an appropriate translation $\boldsymbol{\zeta}$ would depend upon adequate knowledge of the demand arrival rates $\boldsymbol{\phi}$. Even if such knowledge is available, analysis of the VBP policy is complicated by another issue, which we describe next.

Problem 2: Supply externalities and underflow. The analysis is complicated by the non-negativity constraint of queue lengths: the number of supply units assigned from a location cannot exceed the number of supply units available there. As a result, the intended assignment (18) in each time slot might only be partially fulfilled, and the realized update is

$$\mathbf{y}(t+1) = \mathbf{y}(t) - \frac{1}{K}\mathbf{R}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\mathrm{VBP}}(t)) + \mathbf{u}(t)\,, \tag{20}$$

where $\mathbf{u}(t)$ accounts for: (i) supplies that would be dispatched if the dispatched queue was non-empty; (ii) supplies that would arrive if the dispatched upstream queue was non-empty.

In some applications, VBP has provably good performance despite such boundary behavior. For example, when the system is a one-hop network (i.e., the input-output matrix $\mathbf{R}$ does not have negative entries), iteration (20) is a *projected SGD* step

$$\mathbf{y}(t+1) = \mathcal{P}_{\mathbb{R}^m_+}\left(\mathbf{y}(t) - \frac{1}{K}\mathbf{R}(\mathbf{a} \circ \mathbf{x}^{\mathrm{VBP}}(t))\right)\,,$$

here $\mathcal{P}_{\mathcal{S}}(\cdot)$ is the Euclidean projection onto set $\mathcal{S}$.

In multi-hop settings such as ride-hailing, iteration (20) is *not* a projected SGD step due to the existence of *supply externalities*: *starvation of upstream supply nodes (the dispatch location in ride-hailing) can affect supply availability at downstream supply nodes (the customer destination in ride-hailing).* Many works in the literature indirectly approach this problem by sending "null units" when the assigned (dispatch) location runs out of supply units (see, e.g., Huang and Neely 2009). In our closed queueing network setting, however, this approach is not an option because "null supplies" will gradually replace regular supplies since they never leave the system. The virtual queues approach (see, e.g., Stolyar 2005) would seem to suffer from the same issue. There are a small number of works that directly address no-underflow constraints, e.g., by considering a "perturbed" backpressure policy that keeps the queue lengths away from zero (e.g., Huang and Neely 2011). The main drawback of this approach is that its parameters need to carefully chosen. In the next section we present an elegant and systematic approach to address the underflow problem via the mirror descent algorithm.

4 The Mirror Backpressure Policy

4.1 Mirror Descent Algorithm for Convex Optimization

The origin of mirror descent (MD) algorithm for convex optimization dates back to the 1980s (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003). It is a generalization of the subgradient descent algorithm. Consider the optimization problem

$$\text{minimize}_{\mathbf{y}\in\mathcal{M}_1}\ g(\mathbf{y})\,, \tag{21}$$

where $\mathcal{M}_1$ is a convex set in a metric space and the objective function $g(\cdot)$ is convex. A subgradient descent update with step-size η is

$$\mathbf{y}(t+1) = \mathbf{y}(t) - \eta\mathbf{z}\,, \quad \text{where } \mathbf{z}\in\partial g(\mathbf{y}(t))\,.$$

When applying mirror descent, we perform the descent step in the *mirror space* $\mathcal{M}_2$ defined via a *mirror map*: The mirror map is usually defined as the gradient of a continuously differentiable and strictly convex function Φ, as $\nabla\Phi : \mathcal{M}_1 \to \mathcal{M}_2$. An iteration of mirror descent is then

$$\nabla\Phi(\mathbf{y}(t+1)) = \nabla\Phi(\mathbf{y}(t)) - \eta\mathbf{z}\,, \quad \text{where } \mathbf{z}\in\partial g(\mathbf{y}(t))\,. \tag{22}$$

Note that when choosing $\Phi(\mathbf{y}) = ||\mathbf{y}||_2^2$, the mirror map becomes the identical map $\nabla\Phi(\mathbf{y}) = \mathbf{y}$, and mirror descent reduces to subgradient descent. Separately, note that if we have the weaker guarantee $\mathbb{E}[\mathbf{z}]\in\partial g(\mathbf{y}(t))$, then (22) is an iteration of *stochastic* mirror descent.

Let $\tilde{\mathbf{q}}(t) \triangleq \nabla\Phi(\mathbf{y}(t))$ denote the mirror point of the current dual variables $\mathbf{y}(t)$. Denote by Φ^* the convex conjugate (Legendre-Fenchel transformation) of Φ. Let $\mathbf{y}^*$ be an optimal solution of problem (21), and let $\tilde{\mathbf{q}}^*$ be its mirror. The typical Lyapunov function used to analyze the mirror descent algorithm is the *Bregman divergence*

$$L(\tilde{\mathbf{q}}(t)) \triangleq D_{\Phi^*}\left(\tilde{\mathbf{q}}(t), \tilde{\mathbf{q}}^*\right) = \Phi^*(\tilde{\mathbf{q}}(t)) - \Phi^*(\tilde{\mathbf{q}}^*) - \langle\nabla\Phi^*(\tilde{\mathbf{q}}^*), \tilde{\mathbf{q}}(t) - \tilde{\mathbf{q}}^*\rangle\,. \tag{23}$$

Below is a high level (and informal) convergence argument of mirror descent. In the continuous time "fluid limit" of the system, we have

$$\nabla_{\tilde{\mathbf{q}}(t)} L(\tilde{\mathbf{q}}) = \nabla_{\tilde{\mathbf{q}}(t)} D_{\Phi^*}\left(\tilde{\mathbf{q}}(t), \tilde{\mathbf{q}}^*\right) = \nabla\Phi^*(\tilde{\mathbf{q}}(t)) - \nabla\Phi^*(\tilde{\mathbf{q}}^*) = \mathbf{y}(t) - \mathbf{y}^*\,.$$

It follows that for a small step size η we have

$$\begin{aligned} L(\tilde{\mathbf{q}}(t+1)) - L(\tilde{\mathbf{q}}(t)) &\approx \langle \nabla_{\tilde{\mathbf{q}}(t)} L(\tilde{\mathbf{q}}),\ \tilde{\mathbf{q}}(t+1) - \tilde{\mathbf{q}}(t)\rangle && \text{(for small step size)} \\ &= -\eta\,\langle \mathbf{y}(t) - \mathbf{y}^*,\ \mathbf{z}(t)\rangle && \text{(for } \mathbf{z}(t) \in \partial g(\mathbf{y}(t)) \text{ from (22))} \\ &\leq \eta\,(g(\mathbf{y}^*) - g(\mathbf{y}(t)))\,. && (24) \end{aligned}$$

Here the last inequality follows from the convexity of $g(\cdot)$. As a result, the Lyapunov function strictly decreases (if steps are small enough) as long as the optimizer has not been reached.

In a typical use case of mirror descent, Φ is chosen such that $D_\Phi(\mathbf{y}_1, \mathbf{y}_2)$ is an appropriate notion of "proximity" between $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{M}_1$ that captures the geometry of $\mathcal{M}_1$. For example, in exponentiated gradient descent (Kivinen and Warmuth 1997) $\mathcal{M}_1$ is the probability simplex, and relative entropy function is used as Φ. Our approach will be similar, but with the twist that we will choose Φ^* to capture the geometry of the feasible region in the mirror space.

4.2 The Mirror Backpressure Policy

We will now generalize the idea leading to the VBP policy (Section 3.2) to allow a *function* of the queue lengths to be thought of as the dual variables in (14). The key observation we make is that such a policy *executes stochastic mirror descent on the partial dual problem (with flow constraints dualized)* (14), *with the aforementioned function being the inverse mirror map.* The queue lengths constitute the mirror point. This generalizes the previously known fact that VBP performs stochastic gradient descent on the partial dual problem (14). Our approach blending backpressure and mirror descent with a flexibly chosen mirror map is novel, to the best of our knowledge. We believe it can serve as a general framework for systematic design of near optimal backpressure-like control policies for queueing networks in settings with hairy practical constraints. We instantiate our approach to handle the constraint on the mirror space in our setting, namely, that the normalized queue lengths must lie in the m-dimensional probability simplex.

General Mirror Backpressure approach. We generalize the definition (15) (and equivalent definition (16)) of VBP by allowing the function $\nabla\Phi^*(\bar{\mathbf{q}})$ of normalized queue lengths that replaces the dual variables $\mathbf{y}$ to be flexibly chosen. Specifically, we define $\mathbf{y}(t) \triangleq \nabla\Phi^*(\bar{\mathbf{q}}(t))$ and

$$x_{ij'k}^{\text{MBP-G}}(t) \triangleq \begin{cases} 1 & \text{if } i = \operatorname{argmax}_{l \in \mathcal{N}(j')}\{w_{lj'k} + y_l(t) - y_k(t)\} \text{ and } w_{ij'k} + y_l(t) - y_k(t) \geq 0\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{25}$$

$$\Leftrightarrow \quad x_{ij'k}^{\text{MBP-G}}(t) \triangleq \arg\max_{\mathbf{x}\in\mathcal{X}} (\mathbf{w}^{\mathrm{T}} + \mathbf{y}(t)^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x})\,, \tag{26}$$

where MBP stands for Mirror Backpressure. Analogous to (17) we have

$$\mathbb{E}[\mathbf{R}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\text{MBP-G}}(t))] \in \partial g(\mathbf{y}(t))\,, \tag{27}$$

i.e., (using (5)) the expected change in queue lengths $\bar{\mathbf{q}}(t+1) - \bar{\mathbf{q}}(t) = -\mathbf{R}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\text{MBP-G}}(t))/K$ is proportional to the negative of a subgradient of $g(\cdot)$ at $y = \nabla\Phi^*(\bar{\mathbf{q}}(t))$. This is nothing but an iteration of mirror descent (22). In other words, *the Mirror Backpressure policy executes stochastic mirror descent on the partial dual problem* (14) with the mirror point $\bar{\mathbf{q}}(t)$ and the inverse mirror map $\nabla\Phi^*(\cdot)$. The key ingredients behind this simple but powerful approach are: (i) we dualize the *flow* constraints, (ii) the flow, by definition, is the negative rate of change of queue lengths (mirror point), (iii) the gradient of the partial dual objective is identical to the flow at the maximizing assignment, (iv) the MBP policy chooses the maximizing assignment with respect to the current dual variables. We remark that the MBP approach does not depend on the specific nature of the flow constraints (for example, it goes through in open networks as well). Crucially, we are free to choose the mirror map to handle relevant constraints.

Choosing our mirror map to handle underflow constraints. We modify the definition of the normalized queue lengths (the mirror point) to

$$\bar{\mathbf{q}} \triangleq \frac{1}{\tilde{K}}(\mathbf{q} + \delta\mathbf{1}) \quad \text{for } \tilde{K} \triangleq K + mB \text{ and } \delta \triangleq \frac{B}{\tilde{K}}\,. \tag{28}$$

This is the definition used henceforth in the paper. Note that $\bar{\mathbf{q}} \in \Delta^m$ since $\mathbf{1}^{\mathrm{T}}\mathbf{q} = K$. As a result Kullback-Leibler divergence emerges as an appropriate notion of "distance" in the mirror space, i.e., we would like

$$D_{\Phi^*}(\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2) = cD_{\mathrm{KL}}(\bar{\mathbf{q}}_1 \| \bar{\mathbf{q}}_2) \text{ for some } c \in (0,\infty)\,, \text{ where } D_{\mathrm{KL}}(\bar{\mathbf{q}}_1 \| \bar{\mathbf{q}}_2) \triangleq \sum_{i1\in V_S} q_{i1} \log\left(\frac{q_{i1}}{q_{i2}}\right)\,. \tag{29}$$

We achieve (29) by defining

$$\Phi^*(\bar{\mathbf{q}}) \triangleq -c(h(\bar{\mathbf{q}}) + \mathbf{1}^T\bar{\mathbf{q}})\,, \text{ where } h(\bar{\mathbf{q}}) \triangleq -\sum_{i\in V_S} \bar{q}_i \log \bar{q}_i \text{ is the entropy of } \bar{\mathbf{q}}\,.$$

(Note that $\mathbf{1}^T\bar{\mathbf{q}} = 1$ in for $\bar{\mathbf{q}} \in \Delta^m$; that term is included in the definition of $\Phi^*(\cdot)$ only to mildly simplify the inverse mirror map.) This choice of $\Phi^*(\cdot)$ corresponds to $\Phi(\mathbf{y}) = c\sum_{i\in V_S} e^{y_i/c}$ and the maps

$$\mathbf{y} = \nabla\Phi^*(\bar{\mathbf{q}}) = c\log(\bar{\mathbf{q}}) \quad \text{(Inverse Mirror map)}\,, \tag{30}$$

$$\bar{\mathbf{q}} = \nabla\Phi(\mathbf{y}) = \exp(\mathbf{y}/c) \quad \text{(Mirror map)}\,. \tag{31}$$

Here $\exp(\mathbf{y}) \triangleq (\exp(y_i))_{i \in V_S}$. We add a small positive value δ in the definition (28) to ensure that the $\bar{q}_i$s are not too small, thus avoiding the bad behavior of the logarithm function at 0.

We are now able to provide a formal description of our Mirror Backpressure policy. It is obtained by plugging (30) into (25), while carefully handling a boundary case to avoid underflow.

ALGORITHM 1: Mirror Backpressure (MBP) Policy

```
At the start of period $t$ compute $\mathbf{x}^{\mathrm{MBP}}(t)$ as:
for each origin-destination pair $(j', k)$ do
    $i \leftarrow \arg\max_{l \in \partial(j')} \{ w_{lj'k} + c(\log \bar{q}_l(t) - \log \bar{q}_k(t)) \}$
    if $w_{ij'k} + c(\log \bar{q}_i(t) - \log \bar{q}_k(t)) > 0$ and $\bar{q}_i(t) \geq 2\delta$ then
        $\mathbf{x}^{\mathrm{MBP}}_{j'k}(t) \leftarrow \mathbf{e}_i$, i.e., assign from $i$ for all type $(j', k)$ demand;
    else
        $\mathbf{x}^{\mathrm{MBP}}_{j'k}(t) \leftarrow \mathbf{0}$, i.e., drop all type $(j', k)$ demand;
    end
end
When demand in period $t$ is realized, dispatch as per $\mathbf{x}^{\mathrm{MBP}}(t)$.
```

Note that $\bar{q}_i(t) < 2\delta \Leftrightarrow q_i(t) < B$. Thus, MBP ensures that the no-underflow constraints are satisfied at all times by manually stopping service when the queue lengths at the assigned origin fall below threshold B. (By Condition 1, the total number of demand units with origin in the neighborhood $\mathcal{N}(i)$ of any supply location i in any period does not exceed B.) We define c as

$$c \triangleq c_0 w_{\max} \,. \tag{32}$$

We choose c proportional to $w_{\max}$ to maintain scale invariance. We will show later that the c_0 that minimizes the regret bound we obtain has order of magnitude $O(1)$ as K changes.

5 Main Result: Near Optimality of Mirror Backpressure

In this section we present our main result, namely, a performance guarantee for the Mirror Backpressure policy. We define the regret of any policy as the amount by which its expected per period payoff falls short of the upper bound in Proposition 1.

$$\mathrm{Regret}^{\pi}(T) \triangleq W^{\mathrm{OPT}} + \frac{(m-1)w_{\max}K}{T} - \bar{v}^{\pi}(T) \,.$$

In particular, $\mathrm{Regret}^{\pi}(T)$ is an upper bound on the gap between the per period payoff of π and that of the optimal policy for horizon T and any starting state.

Uniqueness of the optimal dual variables. Below is the main assumption we make on the model primitives.

Condition 2 (Unique dual optimum)**.** *Fix any* $i_1 \in V_S$*. There is a unique optimal solution* $\mathbf{y}_0^*$ *of the dual problem* (14) *satisfying* $(y_0^*)_{i_1} = 0$*. Then, as per* (19)*, the set of all optimal solutions is* $\{\mathbf{y}^*(\eta) \triangleq \mathbf{y}_0^* + \xi\mathbf{1}\,, \forall \xi \in \mathbb{R}\}$*, i.e., the set of all dual vectors obtained by translating each coordinate of* $\mathbf{y}_0^*$ *by the same (arbitrary) amount.*

The following condition (i.e., assumption) is stronger than Condition 2 (formalized in Proposition 2 below) but arguably provides a more insightful characterization of our assumption. For each demand type $(j' \in V_D, k \in V_S)$, define the *optimal pickup locations* as $\mathcal{N}_{\text{opt}}(j', k) \triangleq \{i \in \mathcal{N}(j') : x_{ij'k}^* > 0$ in some optimum $\mathbf{x}^*$ of (9)-(11)$\}$. Further, we say that (j', k) is a *partially satisfied demand type* if there exists an optimum $\mathbf{x}^*$ of (9)-(11) such that $0 < \sum_{i \in \mathcal{N}(j')} x_{ij'k}^* < 1$.

Condition 2′ (Every cut includes a partially satisfied demand type)**.** *For every subset* $I \subset V_S$*, there is some* $(i \in V_S, j' \in V_D, k \in V_S)$ *such that:*

- *Location* i *is an optimal pickup location for demand type* (j', k)*, i.e.,* $i \in \mathcal{N}_{\text{opt}}(j', k)$*.*
- *Exactly one of* i *and* k *is in* I*, i.e.,* $|\{i, k\} \cap I| = 1$*.*
- (j', k) *is a partially satisfied demand type.*

Proposition 2. *Condition 2′ implies Condition 2.*

Proposition 2 is proved in Appendix C.2.

Justification for Condition 2′. Observe that for Condition 2′ to hold, we need to have "connectedness" (if there is some subset I such that no vehicles enter or leave I in $\mathbf{x}^*$ then Condition 2′ does not hold). Connectedness is a natural requirement though, since if it is violated then one can always separately solve the subproblem for each connected component. However, connectedness does not imply Condition 2′; one can construct specific examples that are connected but violate Condition 2′. Nevertheless, we can show that such examples are atypical, in a restricted version of our model where there is no flexibility on dispatch location. (We make this restriction for theoretical tractability. Numerical results also suggest that Condition 2 holds for realistic demand matrices $\boldsymbol{\phi}$: it holds for all 5-minute, 1-hour, 2-hour, 5-hour and 10-hour average demand arrival rates in a 30-location tessellation of Manhattan from 6 a.m. to 4 p.m. Data source: decensored demand arrival rates estimated in Buchholz (2015).) In this restricted version of our model, assuming connectedness (suitably defined), we prove that Condition 2′ holds *generically.* Roughly, we show that it holds as long as there is not an exact equality between total demand to enter and leave some subset $I \subset V_S$.

We are now ready to state our main result.

Theorem 1. *Consider any $\epsilon > 0$ and any primitives $\mathbf{w}$, $\mathcal{F}$ and G that satisfy Condition 1 with constant $B = B(\mathcal{F}) < \infty$, as well as Condition 2. Let $\boldsymbol{\phi} = \boldsymbol{\phi}(\mathcal{F})$. Then there exists $K_1 = K_1(\mathbf{w}, \boldsymbol{\phi}, B, G, \epsilon) < \infty$ and $M = M(\mathbf{w}, \boldsymbol{\phi}, B, G, \epsilon) < \infty$ such that for any $K \geq K_1$, any horizon $T < \infty$, and any starting state $\mathbf{q}(0)$, the Mirror Backpressure (MBP) policy satisfies*

$$\text{Regret}^{\text{MBP}}(T) \leq M \left(\frac{K}{T} + \frac{1}{K^{1-\epsilon}} \right) .$$

There are several attractive features of the performance guarantee provided by Theorem 1 for the simple and practically attractive Mirror Backpressure policy.

The policy does not use knowledge of average demand arrival rates $\boldsymbol{\phi}$. At best the platform operator has access to an imperfect estimate of the average demand $\boldsymbol{\phi}$, so this is a very attractive feature of the policy. In contrast, the policy of Banerjee et al. (2016) requires a perfect estimate of the average demand, and will typically suffer a long run (steady state) per period regret of $\Omega(1)$ if the demand estimate is imperfect, even as $K \to \infty$.

Non-trivial regret bound for finite T. In contrast with Banerjee et al. (2016) which provides only a steady state bound for finite K, we are able to provide a performance guarantee for finite horizon and finite (large enough) K. In ride-hailing settings, it is natural that the number of vehicles K should increase proportionally to the demand arrival rates (known as the *large market* regime, see, e.g., Braverman et al. (2016)). Define T_K^{real} as the physical time for a system with K vehicles. As K increases in the large market regime, the primitives $\mathbf{w}, \boldsymbol{\phi}, B, G$ remains unchanged, while $T = O(KT_K^{\text{real}})$, i.e. the duration of a period decreases inversely proportional to K. We can rewrite our regret bound as

$$\text{Regret}^{\text{MBP}}(T) \leq M \left(\frac{1}{T_K^{\text{real}}} + \frac{1}{K^{1-\epsilon}} \right) ,$$

i.e., our transient regret is scale-invariant w.r.t. system size in the large market regime.

Scaling of the regret bound with number of cars K and horizon T. The scaling of the bound on regret with respect to the number of vehicles K and the horizon T may be almost tight for our policy: Note how stochastic gradient descent with a fixed step size τ on a conical objective (our objective is not conical overall, but generically is conical in the neighborhood of the optimum) takes $\Omega(1/\tau)$ steps before the value of the objective approaches the optimum, i.e., there is $\Omega(1)$ loss for $\Omega(1/\tau)$ periods, contributing an $\Omega(1/\tau)$ loss over $T > 1/\tau$ periods, and hence a per period regret contribution of $\Omega(1/(T\tau))$. Further, due to the fixed step size and locally conical objective, the per period loss remains $\Omega(\tau)$, thus leading to overall regret that is $\Omega(\max\{1/(T\tau), \tau\}) = \Omega(1/(T\tau) + \tau) = \Omega(K/T + 1/K)$ if we plug our step size $\tau = 1/K$. Our regret bound is only slightly larger. In fact, we *can* show an $O(K/T + 1/K)$ regret scaling at the cost of a larger minimum K; see Remark 1 at the end of the next section. It is worth noting that the consequent bound of $O(1/K)$ on steady state regret matches that provided by Banerjee et al. (2016) for a state independent policy, except that MBP crucially does not require knowledge of $b\phi$.

The next section sketches a proof of Theorem 1 via some key supporting lemmas.

6 Sketch of proof of near optimality of Mirror Backpressure

In this section we provide the main lemmas and ideas that lead to a proof of Theorem 1. Based on (23) and (29), we consider the following Lyapunov function in the analysis.

$$L(\bar{\mathbf{q}}) \triangleq D_{\mathrm{KL}}(\bar{\mathbf{q}}||\mathbf{q}^*)\,, \quad \text{where } \mathbf{q}^* \triangleq \nabla\Phi(\mathbf{y}^*)\,, \mathbf{y}^* \triangleq \mathbf{y}^*(\lambda_0)\,, \lambda_0 \triangleq -c\log\left(\mathbf{1}^{\mathrm{T}}\nabla\Phi(\mathbf{y}_0^*)\right)\,. \tag{33}$$

Recall that $\bar{\mathbf{q}}$ is defined in (28); $\mathbf{y}_0^*$ and $\mathbf{y}^*(\lambda)$ for $\lambda \in \mathbb{R}$ are defined in Condition 2. We choose $\mathbf{q}^*$ which not only is the mirror point of an optimal dual solution, but also rests in the m-dimensional probability simplex Δ^m. For notational simplicity, we refer to $\mathbf{y}^*(\lambda_0)$ as $\mathbf{y}^*$.

Under Condition 2, the gap between the dual function $g(\mathbf{y})$ and its minimizer $g(\mathbf{y}^*)$ can be lower bounded by a quantity proportional to $||\mathbf{y} - \mathbf{y}^*||_2$. This property is useful for establishing the negative drift of the Lyapunov function under the MBP policy. The property is formalized in the following lemma , proved in Appendix E.

Lemma 1. *If Condition 2 holds, then there exists $\beta > 0$ that depends on model primitives $(\mathbf{w}, \mathcal{F}, G)$ and c_0 defined in* (32) *such that for any $\mathbf{y} \in \left\{\nabla\Phi^*(\bar{\mathbf{q}}) : \bar{\mathbf{q}} \in \Delta^m, \bar{\mathbf{q}} > \mathbf{0}\right\}$, we have*

$$g(\mathbf{y}) - g(\mathbf{y}^*) \geq \beta||\mathbf{y} - \mathbf{y}^*||_2\,. \tag{34}$$

In order to show that the MBP policy has a non-trivial regret bound, we take the dual approach and first prove that $\nabla\Phi^*(\bar{\mathbf{q}}(t))$ converges to $\mathbf{y}^*$. To this end, we make the key observation that the MBP policy executes stochastic mirror descent on the dual function $g(\mathbf{y})$, except when certain entries of $\bar{\mathbf{q}}(t)$ are close to zero. Recall (26) which defines the nominal Mirror Backpressure policy (MBP-G) ignoring the no-underflow constraints. Henceforth, we rename MBP-G to the Nominal policy (Nom) to avoid confusion. Returning to the definition of MBP in Algorithm 1, we see that the MBP policy can be defined via the Nominal policy as

$$\mathbf{x}_{j'k}^{\mathrm{MBP}} = \begin{cases} \mathbf{x}_{j'k}^{\mathrm{Nom}} & \text{if } \bar{q}_{i(j'k)} \geq 2\delta\,, \\ 0 & \text{o.w.} \end{cases} \tag{35}$$

$$\text{where} \quad i(j'k) \triangleq \mathrm{argmax}_{i\in\partial(j')}\left\{w_{ij'k} + \nabla^*\Phi(\bar{q}_i(t)) - \nabla^*\Phi(\bar{q}_k(t))\right\}\,.$$

In words, MBP deviates from the Nominal policy only if at least one normalized queue length is smaller than 2δ.

We now have the building blocks to establish the negative drift of the Lyapunov function (33) under the MBP policy. For $\bar{\mathbf{q}}$ that lies in the "interior" of Δ^m, our analysis follows the standard proof of convergence for stochastic mirror descent algorithms (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003). When $\bar{\mathbf{q}}$ is close to the boundary of Δ^m, we utilize Lemma 1 and (35) to show, via a novel argument, that the forced demand drop does not offset the negative Lyapunov drift. The idea is that in when a forced demand drop occurs, the negative drift under the nominal policy (whose magnitude we bound from below using Lemma 1 along with a formal version of (24)) scales up a factor $\log K$ faster with K than the loss in drift resulting from the deviation (35) of MBP from the nominal policy, as K grows. As a result, the negative drift persists even near the boundary of Δ^m provided K is large enough.

Lemma 2. *Fix $\epsilon \in (0,1)$. Suppose Condition 1 and 2 hold, and let β be the positive constant in Lemma 1. Then there exists $K_0 = \exp\bigl(\text{poly}(m, B, 1/\beta, 1/\epsilon)\bigr) < \infty$ such that for all $K \geq K_0$, under the MBP policy we have*

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1)) - L(\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right] \leq -\frac{1}{2c\tilde{K}}\Bigl(g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\mathbf{q}^*\right)\Bigr) + \frac{m^2B^2}{2\tilde{K}^{2-\epsilon}}\,. \tag{36}$$

In Lemma 2 establishes that for sufficiently large K, the expected drift Lyapunov function is the sum of a negative term proportional to the current dual optimality gap, and a positive (variance) term which decays faster with K.

In the following lemma, we use Lemma 2 to bound the time average of the dual optimality gap by summing the expectation with respect to $\bar{\mathbf{q}}(t)$ of (36) over the first T periods, and then observing that the left-hand side forms a telescoping sum.

Lemma 3. *Suppose Condition 1 and 2 hold. Fix $\epsilon > 0$. For K_0 defined in Lemma 2, for all $K \geq K_0$, we have*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\Bigl[g\left(c\log\bar{\mathbf{q}}(t)\right)\Bigr] - g\left(c\log\mathbf{q}^*\right)\right) \leq \frac{2(c+2w_{\max})\tilde{K}m}{T} + \frac{cm^2B^2}{\tilde{K}^{1-\epsilon}}\,. \tag{37}$$

We establish Theorem 1 using the Lemmas 1 and 3 in Appendix G. The condition in Lemma 1 is similar to the "locally polyhedral" condition proposed by Huang and Neely (2009). The Lyapunov argument in Lemma 2 falls into the broad category of Lyapunov function based scheduling algorithms (see, e.g., Tassiulas and Ephremides 1992, Georgiadis et al. 2006, Neely 2006), with the key distinction that all aforementioned works used quadratic Lyapunov functions while we consider KL-divergence based Lyapunov function. A high level sketch of proof of Theorem 1 is as follows.

Consider any period $t \leq T$. We have $W^{\mathrm{OPT}} \leq g(\bar{\mathbf{q}}(t)) = (\mathbf{w}^{\mathrm{T}} + c(\log \bar{\mathbf{q}}(t))^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{Nom}}(t))$ by weak duality and the definition of the Nominal policy. So, for given $\bar{\mathbf{q}}(t)$, we write

$$\begin{aligned}
&\mathbb{E}\left[W_{\mathrm{OPT}} - \mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\mathrm{MBP}}(t)) \,\middle|\, \bar{\mathbf{q}}(t)\right] \\
&= W_{\mathrm{OPT}} - \mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{MBP}}(t)) \\
&\leq \mathbf{w}^{\mathrm{T}}\left(\tilde{\boldsymbol{\phi}} \circ (\mathbf{x}^{\mathrm{Nom}}(t) - \mathbf{x}^{\mathrm{MBP}}(t))\right) + c(\log \bar{\mathbf{q}}(t))^{\mathrm{T}}\mathbf{R}\left(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{Nom}}(t)\right) \\
&= \underbrace{(\mathbf{w}^{\mathrm{T}} + c(\log \bar{\mathbf{q}}(t))^{\mathrm{T}}\mathbf{R})\left(\tilde{\boldsymbol{\phi}} \circ (\mathbf{x}^{\mathrm{Nom}}(t) - \mathbf{x}^{\mathrm{MBP}}(t))\right)}_{\text{(a)}} + \underbrace{c(\log \bar{\mathbf{q}}(t))^{\mathrm{T}}\mathbf{R}\left(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{MBP}}(t)\right)}_{\text{(b)}} . \qquad (38)
\end{aligned}$$

We show that the (a) term is small using that $\mathbf{x}^{\mathrm{MBP}}$ is close to $\mathbf{x}^{\mathrm{Nom}}$ as per (35), and in particular, when the two policies differ, the reduced cost $\mathbf{w} + c\mathbf{R}^{\mathrm{T}}(\log \bar{\mathbf{q}}(t))$ is small because the dispatched queue lengths are smaller than 2δ by definition of the policy (35). We show that the (b) term is also small by recalling that $\mathbf{R}\left(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{MBP}}(t)\right) = (\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1))/\tilde{K}$, and decomposing the other term $\log \bar{\mathbf{q}}(t)$ into a fixed part $\log \mathbf{q}^*$ and deviations from it $\log \bar{\mathbf{q}}(t) - \log \mathbf{q}^*$. The fixed part has a small contribution because $\sum_t(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1))$ is a telescoping sum, whereas the deviations term can again be controlled using Lemmas 1 and 3. Finally, we note that the c_0 that minimizes our regret bound has order of magnitude $O(1)$, hence we choose $c_0 = 1$.

Remark 1. *Following a very similar proof approach (including a minor change to Lemma 2, see Appendix F), we can get an ϵ-free regret bound of* $\mathrm{Regret}^{\mathrm{MBP}}(T) = O(K/T + 1/K)$*, but with a minimum required K which is* $\exp(\exp(\mathrm{poly}(m, B, 1/\beta, 1/\epsilon)))$*. We omit the details in the interest of space.*

7 Numerical Experiments

In this section, we simulate the MBP policy in a realistic environment using yellow cab data from NYC Taxi & Limousine Commission and Google Maps. Our theoretical model made the simplifying assumption that pickup and service of demand are *instantaneous*. We relax this assumption in our numerical experiments by adding realistic travel times. We consider the following two cases:

1. *Excess supply.* The number of cars in the system is slightly (5%) above the "fluid requirement" (see Section 7.1 for details on the "fluid requirement") to achieve the optimal payoff W^{OPT} in steady state. (Note that even with transit times it is still impossible to beat W^{OPT} in steady state since (10) and (11) must hold for the time average of $\mathbf{x}(t)$.)
2. *Scarce supply.* The number of cars fall short (by 25%) of the "fluid requirement", i.e., there aren't enough cars to realize the optimal solution of (9), hence W^{OPT} cannot be achieved.

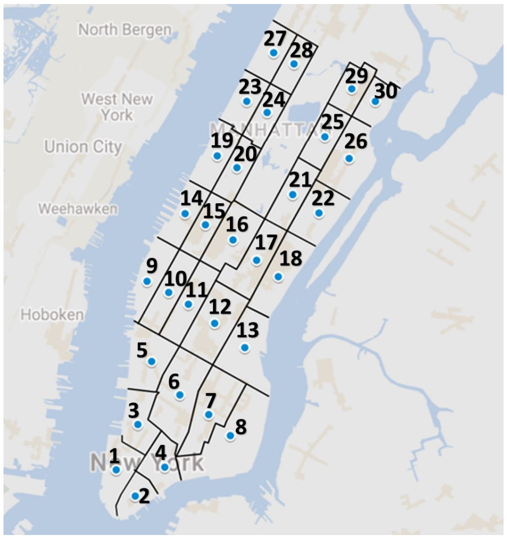

Figure 1: A 30 location model of Manhattan below 110-th street, excluding the Central Park. (Source: tessellation is based on Buchholz (2015), the figure is generated using Google Maps.)

Summary of findings. We find that in both cases, the MBP policy, which requires no knowledge of the demand arrival rates, significantly outperforms static (fluid-based) policy, even when the latter is provided with prior knowledge of exact demand arrival rates. The MBP policy also vastly outperforms the greedy non-idling policy, which demonstrates the importance of proactively dropping demand in certain situations in our setting. (We also simulated the MBP and greedy policy with time-varying demand arrival rates, where the demand arrival rate changes every 5 mins. Our MBP policy still significantly outperforms the greedy policy. Computing the static (fluid-based) policy and the time-varying version of W^{OPT} requires solving a continuous linear program, which is computationally very challenging, as is noted in Ozkan and Ward (2016).)

7.1 Simulation Setup and Benchmark Policies

Throughout the numerical experiments, we use the following model primitives.

- *Payoff structure.* Most major ride-hailing platforms take a commission proportional to the trip fare, which increases with trip distance/duration. Motivated by this, we present results for $w_{ij'k}$ set to be the travel time from j' to k. (However, the results were found to not be sensitive to the choice of $\mathbf{w}$. We also experimented with 100 randomly chosen payoff vectors $\mathbf{w}$, with each $w_{ij'k}$ drawn i.i.d. from Uniform(0,1), and the results were similar. Overall, performance is not heavily dependent on the choice of $\mathbf{w}$.)
- *Graph topology.* We consider a 30-location model of Manhattan below 110-th street excluding Central Park (see Figure 1), as defined by Buchholz (Buchholz 2015). We let pairs of regions

which share a non-trivial boundary be pickup compatible with each other. For example, regions 23 and 24 are compatible but regions 23 and 20 are not.

- *Demand arrival process, and pickup/service times.* We consider a stationary demand arrival process, whose rate is the average decensored demand from 8 a.m. to 12 p.m. estimated in Buchholz (2015) (see Appendix H for a full description). This period includes the morning rush hour and has significant imbalance of demand flow across geographical locations (for many customers the destination is in Midtown Manhattan). We estimate travel times between location pairs using Google Maps. (The average travel time across all demand is 13.1 minutes, and the average pickup time is about 4 minutes (it is policy dependent).)
- *Stationary upper bound, and number of cars.*
  - — *Excess supply.* We use as a baseline the fluid requirement K_{fl} on number of cars needed to achieve optimal payoff. A simple workload conservation argument (using Little's Law) gives the fluid requirement as follows. Applying Little's Law, if the optimal solution $\mathbf{x}^*$ of (9) is realized, the mean number of cars picking up customers is at least $\mathbf{d}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^*) = \sum_{i,j',k} d_{ij'k}\phi_{j'k}x^*_{ij'k}$, for $d_{ij'k} \triangleq \tilde{D}_{ij'} + D_{j'k}$, where $\tilde{D}_{ij}$ is the pickup time from i to j' and $D_{j'k}$ is the travel time from j to k. In our case, it turns out that $K_{\text{fl}} = 7,307$. We use $1.05 \times K_{\text{fl}}$ as the total number of cars in the system to study the excess supply case, i.e., there are 5% extra cars in the system assuming the optimal payoff W^{OPT} is achieved.
  - — *Scarce supply.* When the number of cars in the system is fewer than the fluid requirement, i.e. $K = \kappa K_{\text{fl}}$ with $\kappa < 1$, no policy can achieve a steady state performance of W^{OPT}. A tighter upper bound on the steady state performance is then the value of the static fluid problem (9)-(11) with an additional supply constraint:

$$\mathbf{d}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) \leq K\,. \tag{39}$$

  We denote the value of this problem for $K = \kappa K_{\text{fl}}$ by $W^{\mathrm{OPT}}(\kappa)$. We study the case $\kappa = 0.75$ as an example of scarce supply. For our simulation environment, it turns out that $W^{\mathrm{OPT}}(0.75) \approx 0.86 W^{\mathrm{OPT}}$, i.e., $0.86 W^{\mathrm{OPT}}$ is an upper bound on the per period payoff achievable in steady state.

We compare the performance of our MBP-based policy against the following two policies:

1. *Static (fluid-based) policy.* The fluid-based policy is a static randomization based on the solution to the fluid problem (9) (see, e.g., Banerjee et al. 2016, Ozkan and Ward 2016). See Appendix H for details.
2. *Greedy non-idling policy.* For each demand type (j', k), the greedy policy dispatches from

supply location i that has the highest payoff $w_{ij'k}$ among all compatible neighbors of j' which have at least one supply unit available. If there are ties, the policy prioritizes the supply location with shorter pickup time.

7.2 The Supply-Aware MBP Policy

We propose and study the following heuristic policy inspired by MBP, that additionally incorporates the supply constraint. We call it *supply-aware MBP*. The policy dispatches from:

$$\arg\max_{l\in\partial(j')} \left\{ w_{lj'k} + c(\log \bar{q}_l(t) - \log \bar{q}_k(t)) - v(t) d_{lj'k} \right\} \tag{40}$$

for demand type (j', k) if vehicles are available at that location (and drops the demand otherwise), where $v(t)$ is the current estimate of the shadow price for a "tightened" version of supply constraint (39). We define the tightened supply constraint as

$$\mathbf{d}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) \le 0.95K\,, \tag{41}$$

where the coefficient of K is the flexible "utilization" parameter, that we have set 0.95, meaning that we are aiming to keep 5% vehicles free on average, systemwide. (Keeping a small fraction of vehicles free is helpful in managing the stochasticity in the system. Note that the present paper does not study how to systematically choose the utilization parameter.) Then we update $v(t)$ as

$$v(t+1) = \left[v(t) + \frac{1}{K}(\mathbf{d}^{\mathrm{T}}(\tilde{\mathbf{a}} \circ \mathbf{x}(t)) - 0.95K) \right]^{+}\,,$$

which is equivalent to performing a projected stochastic subgradient step on the dual variable v in the dual problem of (9)-(11) along with (41) with constraint (41) dualized.

7.3 The Excess Supply Case

We simulate the (stationary) system from 8 a.m. to 12 p.m. with 100 randomly generated initial states (see Appendix H for details on the initial state generation). The simulation results on performance are shown in Figure 2. (W^{OPT} is still an upper bound on stationary performance when pickup and service times are included in our model. However, in this case a transient upper bound like (12) is difficult to derive. As a result, we use the ratio of average per period payoff to W^{OPT} as a performance measure, with the understanding that it may exceed 1 at early times.) The result confirms that the MBP policy significantly outperforms the static policy and greedy policy: the average payoff under MBP over 4 hours is about 103% of W^{OPT}, while the static policy and greedy policy only achieves 63% and 68% resp. of the upper bound. The static policy converges slowly to W^{OPT}, leading to poor transient performance. (For example, after running for 20 hours,

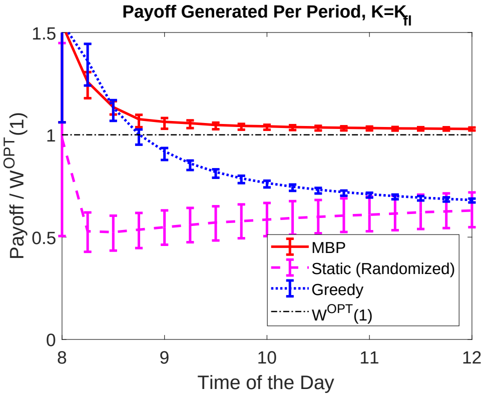

Figure 2: Payoff collected per period under the MBP policy, static fluid-based policy and greedy policy, relative to W^{OPT} defined in Proposition 1. We run 100 trials with random initial queue lengths; the error bars represent the performance between 95% and 5% quantiles, and the main line is the median.

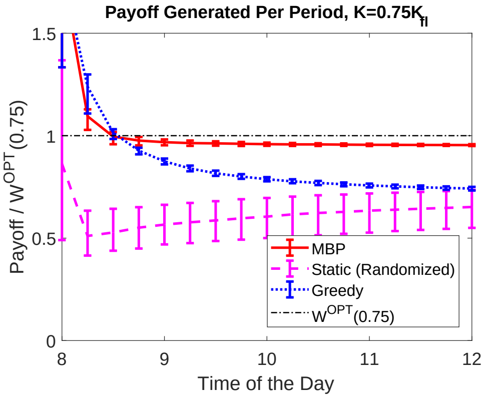

Figure 3: Payoff collected per period under the modified MBP policy, static fluid-based policy and greedy policy, relative to $W^{\mathrm{OPT}}(0.75)$, the value of the problem (9)-(11) along with constraint (39) for $K = 0.75K_{\mathrm{fl}}$. We run 100 trials with random initial queue lengths; the error bars represent the performance between 95% and 5% quantiles, and the main line is the median.

and the average payoff generated by static policy in the 20-th hour is $0.96W^{\mathrm{OPT}}$.) The performance of the greedy policy quickly deteriorates over time because it ignores the flow balance constraints and creates huge geographical imbalances in supply availability.

7.4 The Scarce Supply Case

In the scarce supply case, e.g., $K = 0.75K_{\mathrm{fl}}$, no policy can achieve a stationary performance of W^{OPT}; rather we have an steady state upper bound for $W^{\mathrm{OPT}}(0.75) \approx 0.86W^{\mathrm{OPT}}$. We use this as our benchmark.

Figure 3 shows that the MBP policy also vastly outperforms the static policy and greedy policy in the scarce supply case. MBP generates average per period payoff that is 95% of the benchmark $W^{\mathrm{OPT}}(0.75)$ over 4 hours, while the static policy and greedy policy only achieves 65% and 74% resp. of the benchmark over the same period. Reassuringly, the mean value of $v(t)$ in our simulations of Supply-aware MBP is within 10% of the optimal dual variable to the tightened supply constraint (41) in the problem (9)-(11) along with (41) (both values are close to 0.50). Again, we observe that the average performance of static policy improves as the time horizon gets longer, while the performance of greedy deteriorates.

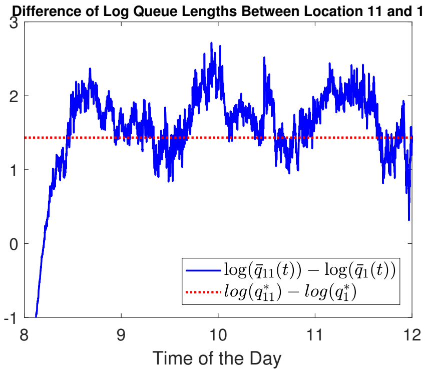

Figure 4: Behavior of the difference between logarithm of queue lengths at location 11 and 1 on a sample path (blue solid line) v.s. the difference between logarithm of optimal queue lengths obtained from optimal dual variable $\mathbf{y}^*$ (red dotted line).

7.5 Supply Queue Lengths and Dual Variables

Recall that MBP executes mirror descent on the dual problem (14), where the logarithm of normalized queue lengths (28) function as dual variables. Figure 4 shows the behavior of the difference between two log queue lengths (at location 11 and 1) on a sample path in the excess supply case. We observe that the difference of log queue lengths quickly converge to the difference of optimal log queue lengths (obtained from optimal dual variable $\mathbf{y}^*$), and then fluctuate around the optimum. This shows that our theoretical findings retain their power even with transit times.

8 MBP-based Joint-Pricing-Assignment Policies

In the following, we consider the joint-pricing-assignment (JPA) setting and design an MBP-based policy (we call it MBP-JPA). The platform's control problem is to set a price for each demand origin-destination pair, and decide an assignment plan at each period so as to maximize payoff over some time horizon. The proposed algorithm is a generalization of (vanilla) backpressure based joint-rate-scheduling control policies (see, e.g., Lin and Shroff 2004, Eryilmaz and Srikant 2006).

The JPA setting. The compatibility graph is the same as in Section 2. A JPA policy $\pi \in \mathcal{U}_{\text{JPA}}$ picks price $\mathbf{p}^\pi(t)$ and assignment $\mathbf{x}^\pi(t)$ at the beginning of every time slot t based on current (supply) queue lengths $\mathbf{q}(t)$. Similar to the JEA setting, we have $\mathbf{x}^\pi(t) \in \mathcal{X}$ where $\mathcal{X}$ is defined in (2). In response to price $p_{j'k}$, type (j', k)'s demand arrival $a_{j'k}(t)$ in the t-th period has distribution $\mathcal{F}_{j'k}(p_{j'k})$ with mean $\mu_{j'k}(p_{j'k})$. Assigning a supply unit from location $i \in \mathcal{N}(j')$ to serve a type (j', k) demand produces a payoff $w_{ij'} + p_{j'k}$, where $w_{ij'}$ models the cost of assignment.

The objective is to maximize payoff over some finite horizon:

$$\text{maximize}_{\pi\in\mathcal{U}_{\text{JPA}}}\ \bar{v}^{\pi}_{\text{JPA}}(T) \quad \text{for } \bar{v}^{\pi}_{\text{JPA}}(T) \triangleq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j'\in V_D, k\in V_S}\left(\sum_{i\in\partial(j')}(w_{ij'}+p_{j'k}(t))x_{ij'k}(t)\right)\right]. \tag{42}$$

We assume the following regularity conditions to hold for demand functions $(\mu_{j'k}(p_{j'k}))_{j',k}$.

Condition 3. *1. There exists $\bar{B} < \infty$ such that $\sum_{j'\in\mathcal{N}(i)}\sum_{k\in V_S}\mu_{j'k}(0) < \bar{B}$, $\forall i \in V_S$.*

2. For any $j' \in V_D$, $k \in V_S$, $\mu_{j'k}(p_{j'k})$ is differentiable and strictly decreasing, hence it has an inverse denoted as $p_{j'k}(\mu_{j'k})$ which is defined for $\mu_{j'k} \in (0, \mu_{j'k}(0))$.

3. The revenue functions $r_{j'k}(\mu_{j'k}) \triangleq \mu_{j'k}\cdot p_{j'k}(\mu_{j'k})$ are concave and twice continuously differentiable.

4. The customers have bounded willingness-to-pay, i.e. $\exists \bar{p} < \infty$ such that $\forall j, k$, $\mu_{j'k}(\bar{p}) = 0$.

These assumptions are quite standard in the revenue management literature, (see, e.g., Gallego and Van Ryzin 1994). Note that Condition 3 part 4 implies that $r'_{j'k}(0) \le \bar{p} < \infty$ and that $r_{j'k}(0) = 0$.

In the JPA setting, demand quantity plays a role in myopic revenues but also affects the distribution of supplies, and the chosen controls balance myopic revenues with maintaining a good spatial distribution of supply. We will show below that when pricing becomes an available lever, the platform will modulate the quantity of demand *always* through changing the prices (and serving all the resulting demand arrivals) rather than apply entry control (i.e. dropping some demand).

Parallel to the development of MBP in this paper, we first formulate the static JPA revenue maximization problem as an optimization problem, which we call `JPA-static`.

$$\texttt{JPA-static:}\qquad \text{maximize}_{\mathbf{x},\boldsymbol{\mu}} \sum_{j'\in V_D}\sum_{k\in V_S}\mu_{j'k}\sum_{i\in\mathcal{N}(j')}(w_{ij'}+p_{j'k}(\mu_{j'k}))x_{ij'k} \tag{43}$$

$$\text{s.t.} \sum_{j'\in\mathcal{N}(i)}\sum_{k\in V_S}\mu_{j'k}x_{ij'k} - \sum_{j'\in V_D}\sum_{l\in\mathcal{N}(j)}\mu_{j'i}x_{lj'i} = 0 \quad \forall i \in V_S\,, \tag{44}$$

$$\sum_{i\in\mathcal{N}(j')} x_{ij'k} \le 1 \quad \forall j' \in V_D,\ k \in V_S\,, \tag{45}$$

$$\mu_{j'k} \in [0, \mu_{j'k}(0)] \quad \forall j' \in V_D,\ k \in V_S\,, \tag{46}$$

$$x_{ij'k} \ge 0 \quad \forall i, k \in V_S,\ j' \in V_D\,. \tag{47}$$

Note that the objective (43) is not necessarily jointly concave in $\mathbf{x}, \boldsymbol{\mu}$. Luckily, under Condition

3, `JPA-static` admits an equivalent concave maximization. We call it `JPA-concave`.

$$\texttt{JPA-concave}: \qquad \text{maximize}_{\hat{\mathbf{x}}} \sum_{i\in V_S}\sum_{j'\in V_D}\sum_{k\in V_S} w_{ij'}\hat{x}_{ij'k} + \sum_{j'\in V_D}\sum_{k\in V_S} r_{j'k}\left(\sum_{i\in\partial(j')}\hat{x}_{ij'k}\right)$$

$$\text{s.t.} \sum_{j'\in\mathcal{N}(i)}\sum_{k\in V_S}\hat{x}_{ij'k} - \sum_{j'\in V_D}\sum_{l\in\mathcal{N}(j)}\hat{x}_{lj'i} = 0 \quad \forall i\in V_S \tag{48}$$

$$\sum_{i\in\partial(j')}\hat{x}_{ij'k} \le \mu_{j'k}(0) \quad \forall j'\in V_D,\ k\in V_S\,, \tag{49}$$

$$\hat{x}_{ij'k}\ge 0 \quad \forall i,k\in V_S,\ j'\in V_D\,.$$

Lemma 4. `JPA-concave` *is equivalent to* `JPA-static` *in the following sense: for any optimal solution* $\hat{\mathbf{x}}^*$ *of* `JPA-concave`*, define* $(\mathbf{x}^*,\boldsymbol{\mu}^*)$*:*

$$x^*_{ij'k} \triangleq \begin{cases} 0 & \text{if } \sum_{i\in\mathcal{N}(j')}\hat{x}^*_{ij'k} = 0\,, \\ \frac{\hat{x}^*_{ij'k}}{\sum_{i\in\mathcal{N}(j')}\hat{x}^*_{ij'k}} & \text{otherwise.}\end{cases} \qquad \mu^*_{j'k} \triangleq \sum_{i\in\mathcal{N}(j')}\hat{x}^*_{ij'k} \quad \forall j'\in V_D,\ i,k\in V_S\,, \tag{50}$$

then pair $(\mathbf{x}^*,\boldsymbol{\mu}^*)$ *is an optimal solution to* `JPA-static`*.*

Lemma 4 is proved in Appendix I. To obtain the MBP based policy for JPA policy, we first take the partial dual of `JPA-concave`. Let $\mathbf{y}$ be the dual variables of flow constraints (44) and only dualize these constraints. We have:

$$\text{minimize}_{\mathbf{y}}\ g_{\text{JPA}}(\mathbf{y}),$$

$$\text{for } g_{\text{JPA}}(\mathbf{y}) = \sum_{j'\in V_D}\sum_{k\in V_S}\max_{\{0\le\mu_{j'k}\le\mu_{j'k}(0)\}}\left(r_{j'k}\left(\mu_{j'k}\right) + \mu_{j'k}\max_{i\in\mathcal{N}(j')}\left(w_{ij'}+y_i-y_k\right)\right). \tag{51}$$

See Appendix I for the derivation.

Parallel to the derivation of Algorithm 1, we design the MBP-JPA policy by making it execute stochastic mirror descent on $g_{\text{JPA}}(\mathbf{y})$. The key ideas are as follows: The mean demand arrival rate under the policy will be the outer argmax in the definition (51) of $g_{\text{JPA}}(\mathbf{y})$, and the assignment decision achieves the inner argmax in the definition (51) of $g_{\text{JPA}}(\mathbf{y})$. The dual variable $\mathbf{y}$ is given by the inverse mirror map (30). When the supply queue length at i is shorter than $\bar{B}$, we protect it by dropping demand units arriving to i. The definition of normalized queue lengths $\bar{\mathbf{q}}$ and δ are the same as in (28), except that B is replaced by $\bar{B}$ in Condition 3.

Note that we can rewrite $\mu_{j'k}(p)$ as $\mu_{j'k}(0)\cdot(1-F_{j'k}(p))$, where $\mu_{j'k}(0)$ is the demand arrival rate when price is 0 (i.e. gross demand), and $F_{j'k}(\cdot)$ is the cumulative distribution function of the (random) willingness-to-buy of type (j',k) customers (i.e. price-response of demand). A key feature of MBP-JPA policy (see Algorithm 2 below) is that it *does not require prior knowledge of gross demand* $\boldsymbol{\mu}(\mathbf{0})$, in contrast to the fluid-based policies as in Banerjee et al. (2016). To see this, denote $\hat{\mu}_{j'k}(p) \triangleq 1 - F_{j'k}(p) = \mu_{j'k}(p)/\mu_{j'k}(0)$ $(\mu_{j'k}(0) \neq 0)$, we have:

$$g_{\text{JPA}}(\mathbf{y}) = \sum_{j' \in V_D} \sum_{k \in V_S} \mu_{j'k}(0) \max_{\{0 \le p_{j'k} \le \bar{p}\}} \left(p_{j'k} \cdot \hat{\mu}_{j'k}(p_{j'k}) + \hat{\mu}_{j'k}(p_{j'k}) \max_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k \right) \right), \tag{52}$$

i.e. the pricing under MBP-JPA (outer argmax of (52)) only depends on price response, not the gross demand. This is an attractive feature in many applications.

ALGORITHM 2: MBP-JPA Policy

```
At the start of period $t$ compute prices $\mathbf{p}^{\mathrm{MBP-JPA}}(t)$ and assignment decision $\mathbf{x}^{\mathrm{MBP-JPA}}(t)$ as:
for each origin-destination pair $(j', k)$ do
    $i \leftarrow \arg\max_{l \in \partial(j')} \{ w_{lj'} + c(\log \bar{q}_l(t) - \log \bar{q}_k(t)) \}$;
    if $\bar{q}_i(t) \ge 2\delta$ then
        $p_{j'k}^{\mathrm{MBP-JPA}}(t) \leftarrow \mathrm{argmax}_{0 \le p_{j'k} \le \bar{p}} \{ p_{j'k} \cdot \hat{\mu}_{j'k}(p_{j'k}) + \hat{\mu}_{j'k}(p_{j'k}) \cdot (w_{ij'} + c(\log \bar{q}_i(t) - \log \bar{q}_k(t))) \}$;
        $\mathbf{x}_{j'k}^{\mathrm{MBP-JPA}}(t) \leftarrow \mathbf{e}_i$, i.e., assign from $i$ for all type $(j', k)$ demand;
    else
        $p_{j'k}^{\mathrm{MBP-JPA}}(t) \leftarrow \bar{p}$ and $\mathbf{x}_{j'k}^{\mathrm{MBP-JPA}}(t) \leftarrow \mathbf{0}$ i.e., drop all type $(j', k)$ demand;
    end
end
When demand in period $t$ is realized as the result of $\mathbf{p}^{\mathrm{MBP-JPA}}(t)$, dispatch as per $\mathbf{x}^{\mathrm{MBP-JPA}}(t)$.
```

Analogous to Condition $2'$, we assume the following condition holds for the model primitives for the dynamic pricing case. For each demand type $(j' \in V_D, k \in V_S)$, define *optimal pickup locations (in JPA)* as $\mathcal{N}_{\text{opt}}(j', k) \triangleq \{ i \in \mathcal{N}(j') : \hat{x}^*_{ij'k} > 0$ in some optimum $\hat{\mathbf{x}}^*$ of `JPA-concave`$\}$. Further, we say that (j', k) is a *partially satisfied demand type (in JPA)* if there exists an optimum $\hat{\mathbf{x}}^*$ of `JPA-concave` such that $0 < \sum_{i \in \mathcal{N}(j')} \hat{x}^*_{ij'k} < \mu_{j'k}(0)$.

Condition 4 (Every cut includes a partially satisfied demand type (in JPA))**.** *For every subset $I \subset V_S$, there is some $(i \in V_S, j' \in V_D, k \in V_S)$ such that:*

- *Location i is an optimal pickup location (in JPA) for demand type (j', k), i.e., $i \in \mathcal{N}_{\text{opt}}(j', k)$.*
- *Exactly one of i and k is in I, i.e., $|\{i, k\} \cap I| = 1$.*
- *(j', k) is a partially satisfied demand type (in JPA).*

Similar to the discussion in Section 5, we can show that Condition 4 is slightly stronger than assuming the uniqueness (up to translating each coordinate by the same constant) of dual optimum $\mathbf{y}^*$. (See Appendix I for details)

Let $W_{\text{JPA}}^{\text{OPT}}$ be the optimal value of `JPA-concave`, $w_{\max} \triangleq \max_{i,j'} |w_{ij'}|$. Similar to Proposition 1, we have the following upper bound for $\bar{v}_{\text{JPA}}^{\pi}(T)$ where parameter c is chosen to be $w_{\max}$.

Proposition 3. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}(0)$, the expected payoff generated by any policy $\pi \in \mathcal{U}_{\text{JPA}}$ is upper bounded as:*

$$\bar{v}^{\pi}_{\text{JPA}}(T) \le W^{\text{OPT}}_{\text{JPA}} + \frac{(m-1)(w_{\max} + \bar{p})K}{T} \,. \tag{53}$$

Here $\bar{p}$ is the upper bound for willingness-to-pay defined in Condition 3. In particular, the long run payoff under any policy $\pi \in \mathcal{U}_{\text{JPA}}$ is bounded above as $\limsup_{T\to\infty} \bar{v}^{\pi}_{\text{JPA}}(T) \le W^{\text{OPT}}_{\text{JPA}}$.

The proof is in Appendix I. Define $\text{Regret}^{\pi}_{\text{JPA}}(T) \triangleq W^{\text{OPT}}_{\text{JPA}} + \frac{(m-1)(w_{\max}+\bar{p})K}{T} - \bar{v}^{\pi}_{\text{JPA}}(T)$. We have the following result for the regret of MBP-JPA policy.

Theorem 2. *Consider any $\epsilon > 0$ and any primitives $\mathbf{w}$, $(\mathcal{F}_{j'k})_{j',k}$, $(\mu_{j'k}(\cdot))_{j',k}$ and G that satisfy Condition 3 with constant $\bar{B} < \infty$, as well as Condition 4. Then there exists $K_1' = K_1'(\mathbf{w}, \boldsymbol{\mu}(\cdot), \bar{B}, G, \epsilon) < \infty$ and $M' = M'(\mathbf{w}, \boldsymbol{\mu}(\cdot), \bar{B}, G, \epsilon) < \infty$ such that for any $K \ge K_1'$, any horizon $T < \infty$, and any starting state $\mathbf{q}(0)$, the MBP-JPA policy satisfies*

$$\text{Regret}^{\text{MBP-JPA}}_{\text{JPA}}(T) \le M\left(\sqrt{\frac{K}{T} + \frac{1}{K^{1-\epsilon}}}\right) .$$

The proof is very similar to the proof of Theorem 1 with the following twist. The Lyapunov drift result (36) in Lemma 2 holds true for large enough K. We show that $g_{\text{JPA}}(\mathbf{y})$ is lower bounded by a quadratic function of $\mathbf{y} - \mathbf{y}^*$ rather than its norm (as in Lemma 1) near $\mathbf{y}^*$. As a result, the progress of mirror descent slows down when $\mathbf{y}$ approaches $\mathbf{y}^*$, and we will need to take the square root of the regret bound in Lemma 3. Using Jensen's inequality, we obtain the final regret bound which has the order as the square root of the regret bound in Theorem 1.

Similar to Remark 1, we can get an ϵ-free regret bound $\text{Regret}^{\text{MBP-JPA}}_{\text{JPA}}(T) = O(\sqrt{K/T + 1/K})$ with a minor change to the proof and a larger minimum required K. We omit the details in the interest of space.

9 Discussion

In this paper we consider the payoff maximizing dynamic control of a closed queueing network model of ride-hailing platforms. We propose a novel family of policies called Mirror Backpressure (MBP), which generalizes backpressure such that it executes mirror descent. The MBP policy overcomes the challenge stemming from the no-underflow constraint and supply externalities as the policy accounts for the geometry of the problem, and it does not require any knowledge of demand arrival rates. We prove that the MBP policy is able to achieve good transient performance, losing at most an $O(K/T + 1/K)$ fraction of the achievable payoff for joint-entry-assignment control, and

$O(\sqrt{K/T + 1/K})$ for joint-pricing-assignment control. Realistic numerical experiments corroborate our theoretical findings.

Before closing, we point out several interesting directions for future research, many of which we are actively pursuing.

1. *Travel times.* It would be of interest to incorporate travel times in the theory. We conjecture that an MBP policy which uses a carefully chosen function of current supply queue lengths, estimated time of arrival and destinations of in-transit cars will have good theoretical guarantees.
2. *Empty car repositioning.* Our model allows for the payoff to depend on dispatch location. As such, we may interpret our model in terms of empty car repositioning, especially if travel times can be incorporated.
3. *Time-varying demand arrivals.* Since our policy does not require any statistical knowledge of the demand arrival rates, it is promising for the situation where demand arrival rate is time-varying.
4. *Improved performance via "centering" MBP based on demand arrival rates.* Combining Lemmas 1 and 2, we expect that in steady state under MBP, $\bar{\mathbf{q}}(t)$ concentrates in a ball of radius $\sim m^2B^2/(q^*_{\min}\beta K^{1-\epsilon})$ around $\mathbf{q}^*$. If $\mathbf{y}^*$ is known (or learned by learning $\boldsymbol{\phi}$ via observing demand), we can modify the inverse mirror map to $y_i = y_i^* + c\log(mq_i)$ which leads to $\mathbf{q}^* = \mathbf{1}/m$. For the resulting "centered MBP" policy, based on the result of Huang and Neely (2009) and convergence of mirror descent, we are optimistic that the steady regret will be only $\exp(-\Omega(K^{(1-\epsilon)/2}))$ for a suitable choice of c.
5. *Eliminating the need for a unique dual optimum (Condition 2).* We believe we can replace this assumption with a mild connectivity assumption. We can leverage Corollary 2 (Appendix D, which only needs connectivity) to infer that for values of the normalized queue length vector $\bar{\mathbf{q}}$ which are close to the boundary of the simplex Δ^m, the inverse mirror point $\mathbf{y}$ is such that $g(\mathbf{y})$ is far from optimal for the partial dual problem (14) in any connected problem instance.
6. *Other applications of MBP.* MBP appears to be a powerful and general approach to obtain near optimal performance despite no-underflow constraints in control of queueing networks. It does not necessitate a heavy traffic assumption, and provides guarantees on both transient and steady state performance. For example, the matching queues problem studied by Gurvich and Ward (2014) is hard due to no-underflow constraints and the approach in Gurvich and Ward (2014) assumes stringent conditions on the network structure. MBP may be able to achieve provably near optimal performance for more general matching queue systems.

References

Adan, Ivo, Gideon Weiss. 2012. A loss system with skill-based servers under assign to longest idle server policy. *Probability in the Engineering and Informational Sciences* **26**(3) 307–321.

Adelman, Daniel. 2007. Price-directed control of a closed logistics queueing network. *Operations Research* **55**(6) 1022–1038.

Banerjee, Siddhartha, Daniel Freund, Thodoris Lykouris. 2016. Pricing and optimization in shared vehicle systems: An approximation framework. *CoRR abs/1608.06819* .

Banerjee, Siddhartha, Yash Kanoria, Pengyu Qian. 2018. State dependent control of closed queueing networks with application to ride-hailing .

Beck, Amir, Marc Teboulle. 2003. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters* **31**(3) 167–175.

Bertsimas, Dimitris, John N Tsitsiklis. 1997. *Introduction to linear optimization*, vol. 6. Athena Scientific Belmont, MA.

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2016. Spatial pricing in ride-sharing networks .

Braverman, Anton, Jim G Dai, Xin Liu, Lei Ying. 2016. Empty-car routing in ridesharing systems. *arXiv preprint arXiv:1609.07219* .

Buchholz, Nicholas. 2015. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. Tech. rep., Technical report, University of Texas at Austin.

Bumpensanti, Pornpawee, He Wang. 2018. A re-solving heuristic for dynamic resource allocation with uniformly bounded revenue loss. *arXiv preprint arXiv:1802.06192* .

Burke, James V, Michael C Ferris. 1993. Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization* **31**(5) 1340–1359.

Buỳić, Ana, Sean Meyn. 2015. Approximate optimality with bounded regret in dynamic matching models. *ACM SIGMETRICS Performance Evaluation Review* **43**(2) 75–77.

Cachon, Gerard P, Kaitlin M Daniels, Ruben Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* .

Caldentey, René, Edward H Kaplan, Gideon Weiss. 2009. Fcfs infinite bipartite matching of servers and customers. *Advances in Applied Probability* **41**(3) 695–730.

Chen, Gong, Marc Teboulle. 1993. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization* **3**(3) 538–543.

Dai, Jim G, Wuqin Lin. 2005. Maximum pressure policies in stochastic processing networks. *Operations Research* **53**(2) 197–218.

Dai, Jim G, Wuqin Lin. 2008. Asymptotic optimality of maximum pressure policies in stochastic processing networks. *The Annals of Applied Probability* **18**(6) 2239–2299.

Derman, Cyrus. 1970. Finite state markovian decision processes. Tech. rep.

Désir, Antoine, Vineet Goyal, Yehua Wei, Jiawei Zhang. 2016. Sparse process flexibility designs: is the long chain really optimal? *Operations Research* **64**(2) 416–431.

Eryilmaz, Atilla, R Srikant. 2006. Joint congestion control, routing, and mac for stability and fairness in wireless networks. *IEEE Journal on Selected Areas in Communications* **24**(8) 1514–1524.

Eryilmaz, Atilla, R Srikant. 2007. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *IEEE/ACM Transactions on Networking (TON)* **15**(6) 1333–1344.

Eryilmaz, Atilla, R Srikant. 2012. Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems* **72**(3-4) 311–359.

Gallego, Guillermo, Garrett Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science* **40**(8) 999–1020.

Georgiadis, Leonidas, Michael J Neely, Leandros Tassiulas, et al. 2006. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends® in Networking* **1**(1) 1–144.

Gurvich, Itai, Amy Ward. 2014. On the dynamic control of matching queues. *Stochastic Systems* **4**(2) 479–523.

Hall, Jonathan, Cory Kendrick, Chris Nosko. 2015. The effects of uber's surge pricing: A case study. *The University of Chicago Booth School of Business* .

Huang, Longbo, Michael J Neely. 2009. Delay reduction via lagrange multipliers in stochastic network optimization. *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on*. IEEE, 1–10.

Huang, Longbo, Michael J Neely. 2011. Utility optimal scheduling in processing networks. *Performance Evaluation* **68**(11) 1002–1021.

Jordan, William C, Stephen C Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Science* **41**(4) 577–594.

Kakade, S, Shai Shalev-Shwartz, Ambuj Tewari. 2009. Applications of strong convexity–strong smoothness duality to learning with matrices. *CoRR, abs/0910.0610* .

Kivinen, Jyrki, Manfred K Warmuth. 1997. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation* **132**(1) 1–63.

Lin, Xiaojun, Ness B Shroff. 2004. Joint rate control and scheduling in multihop wireless networks. *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 2. IEEE, 1484–1489.

Mairesse, Jean, Pascal Moyal. 2016. Stability of the stochastic matching model. *Journal of Applied Probability* **53**(4) 1064–1077.

Neely, Michael J. 2006. Energy optimal control for time-varying wireless networks. *IEEE transactions on Information Theory* **52**(7) 2915–2934.

Nemirovsky, Arkadii Semenovich, David Borisovich Yudin. 1983. Problem complexity and method efficiency in optimization. .

Ozkan, Erhun, Amy R Ward. 2016. Dynamic matching for real-time ridesharing .

Rockafellar, Ralph Tyrell. 2015. *Convex analysis*. Princeton university press.

Sason, Igal, Sergio Verdú. 2015. Upper bounds on the relative entropy and rényi divergence as a function of total variation distance for finite alphabets. *Information Theory Workshop-Fall (ITW), 2015 IEEE*. IEEE, 214–218.

Shi, Cong, Yehua Wei, Yuan Zhong. 2015. Process flexibility for multi-period production systems .

Srikant, Rayadurgam. 2012. *The mathematics of Internet congestion control*. Springer Science & Business Media.

Stolyar, Alexander L. 2003. Control of end-to-end delay tails in a multiclass network: Lwdf discipline optimality. *Annals of Applied Probability* 1151–1206.

Stolyar, Alexander L. 2004. Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *The Annals of Applied Probability* **14**(1) 1–53.

Stolyar, Alexander L. 2005. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems* **50**(4) 401–457.

Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

Tassiulas, Leandros. 1992. Dynamic link activation scheduling in multihop radio networks with fixed or changing connectivity .

Tassiulas, Leandros, Anthony Ephremides. 1992. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE transactions on automatic control* **37**(12) 1936–1948.

Waserhole, Ariel, Vincent Jost. 2016. Pricing in vehicle sharing systems: Optimization in queuing networks with product forms. *EURO Journal on Transportation and Logistics* **5**(3) 293–320.

Appendix

A Finite horizon payoff upper bound: Proof of Proposition 1

We establish two key lemmas to facilitate the proof of Proposition 1. The first lemma shows that the expected payoff cannot exceed the value of the finite horizon fluid problem.

Lemma 5. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}(0)$, the expected payoff generated by any policy π is upper bounded by the value of the finite horizon fluid problem defined by* (9)*,* (13) *and* (11)*.*

Proof. Let π be any feasible policy. It follows from Condition 1 that $\mathbf{a}(1)$ has finite support, which we denote as $\mathcal{A}$. Denote $p_{\mathbf{a}} \triangleq \mathbb{P}(\mathbf{a}(1) = \mathbf{a})$. For each $\mathbf{a} \in \mathcal{A}$, define

$$\mathbf{x}^{\mathbf{a}}(T) \triangleq \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathbf{x}(t)|\mathbf{a}(t) = \mathbf{a}]\,.$$

In words, $\mathbf{x}^{\mathbf{a}}(T)$ is the average over $1 \le t \le T$ of the average assignment in period t conditioned on the demand in period t being $\mathbf{a} \in \mathcal{A}$. (One might be tempted to think that $\mathbf{x}(t)$ is independent of $\mathbf{a}(t)$. However, because of the no-underflow constraint (6), $\mathbf{x}(t)$ is the *realized* dispatch decision hence it depends on $\mathbf{a}(t)$.) We decompose the time-average of payoff collected in the first T time slots as:

$$\bar{v}^{\pi}(T) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}(t) \circ \mathbf{x}(t))]$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\mathbf{a}\in\mathcal{A}}\mathbb{E}\left[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}(t))\mathbb{I}\left\{\mathbf{a}(t)=\mathbf{a}\right\}\right] \\
&= \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}(t))\,\middle|\mathbf{a}(t)=\mathbf{a}\right]\right) \\
&= \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a}}(T))\,,
\end{aligned}
$$

where we used the basic properties of conditional expectation. Similarly, for the time-average of the change of queue length we have:

$$
\frac{1}{T}\mathbb{E}[\mathbf{q}(T)-\mathbf{q}(0)] = \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\mathbf{R}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a}}(T))\,.
$$

Because there are only K supply units circulating in the system, the net outflow from any subset of supply locations $A\subset V_S$ should not exceed K. As a result, $\bar{v}^{\pi}(T)$ is upper bounded by the following optimization problem:

$$
\begin{aligned}
\max_{\mathbf{x}^{\mathbf{a}}}\ & \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a}}) && && (54)\\
\text{s.t.}\ & \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\mathbf{1}_S^{\mathrm{T}}\mathbf{R}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a}}) \le \frac{K}{T} && \forall S\subset V_S\,, && (55)\\
& \mathbf{x}^{\mathbf{a}}_{j'k}\in\mathrm{conv}(\mathcal{X}_{j'k}) && \forall j'\in V_D,\ k\in V_S,\ \mathbf{a}\in\mathcal{A}\,.
\end{aligned}
$$

Take the partial dual w.r.t. constraints (55), we have the the following Lagrange dual function:

$$
\begin{aligned}
\tilde{g}(\mathbf{y}) = & \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}} \sum_{j'\in V_D,k\in V_S}\ \max_{\mathbf{x}^{a}_{j'k}\in\mathrm{conv}(\mathcal{X}_{j'k})}\left(\mathbf{w}^{\mathrm{T}}_{j'k}(\tilde{\mathbf{a}}_{j'k}\circ\mathbf{x}^{\mathbf{a}}_{j'k}) + \sum_{S\subset V_S} y_S\left(\mathbf{1}_S^{\mathrm{T}}\mathbf{R}_{j'k}(\tilde{\mathbf{a}}_{j'k}\circ\mathbf{x}^{\mathbf{a}}_{j'k})\right)\right) \\
& - \sum_{S\subset V_S} y_S\frac{K}{T} \\
= & \sum_{j'\in V_D,k\in V_S}\ \sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}a_{j'k}\max_{i\in\mathcal{N}(j')}\left(w_{ij'k}\sum_{S\subset V_S} y_S\left(\mathbb{I}\{i\in S\}-\mathbb{I}\{k\in S\}\right)\right) - \sum_{S\subset V_S} y_S\frac{K}{T} \\
= & \sum_{j'\in V_D,k\in V_S} \phi_{j'k}\max_{i\in\mathcal{N}(j')}\left(w_{ij'k} + \sum_{S\subset V_S} y_S(\mathbb{I}\{i\in S\}-\mathbb{I}\{k\in S\})\right) - \sum_{S\subset V_S} y_S\frac{K}{T}\,.
\end{aligned}
$$

Here the last equality holds because $\boldsymbol{\phi}=\mathbb{E}[\mathbf{a}(1)]=\sum_{\mathbf{a}\in\mathcal{A}} p_{\mathbf{a}}\mathbf{a}$. Note that $\tilde{g}(\mathbf{y})$ equals to the partial dual function of the finite horizon fluid problem (9),(13) and (11), denoted by $g(\mathbf{y})$. Using the strong duality of linear programs, we have that the expected payoff generated by any policy π is upper bounded by the finite horizon fluid problem. □

In order to facilitate the second key lemma, we first prove a supporting lemma.

We call an assignment $\mathbf{x}$ a *directed acyclic assignment* if there is no sequence of node pairs $\mathcal{C}=((i_1,j_1'),(i_2,j_2'),\ldots,\,(i_\ell,j_\ell'))$ where $i_l\in V_S$ and $j_l'\in V_D$ for $l=1,2,\ldots,\ell$, such that

$$
\phi_{j_l'i_{l+1}}x_{i_lj_l'i_{l+1}} > 0 \qquad \forall\, l=1,2,\ldots,\ell\,, \tag{56}
$$

where $i_{\ell+1}\triangleq i_1, j_{\ell+1}\triangleq j_1$. In words, there is no cycle $\mathcal{C}$ such that there is a positive flow along $\mathcal{C}$.

Lemma 6. *Any feasible solution* $\mathbf{x}^F$ *of the finite horizon fluid problem satisfying* (13) *and* (11) *can be decomposed as*

$$\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}} \,, \tag{57}$$

where $\mathbf{x}^{\mathrm{S}}$ *is a feasible solution for the static fluid problem satisfying* (10) *and* (11), *and* $\mathbf{x}^{\mathrm{DAG}}$ *is a directed acyclic assignment satisfying* (13) *and* (11).

Proof. The existence of such a decomposition can be established using a standard network flow argument: Start with $\mathbf{x}^{\mathrm{S}} = \mathbf{0}$ and $\mathbf{x}^{\mathrm{DAG}} = \mathbf{x}^{\mathrm{F}}$. Then, iteratively, if $\mathbf{x}^{\mathrm{DAG}}$ includes a cycle $\mathcal{C}$ with a positive flow along $\mathcal{C}$ as above, move a flow of $u(\mathcal{C}) \triangleq \min_{1\le l\le \ell} \phi_{j'_l i_{l+1}} x_{i_l j'_l i_{l+1}}$ along $\mathcal{C}$ from $\mathbf{x}^{\mathrm{DAG}}$ to $\mathbf{x}^{\mathrm{S}}$, via the updates

$$\begin{aligned} x^{\mathrm{S}}_{i_l j'_l i_{l+1}} &\leftarrow x^{\mathrm{S}}_{i_l j'_l i_{l+1}} + u(\mathcal{C})/\phi_{j'_l i_{l+1}} \qquad \text{and} \\ x^{\mathrm{DAG}}_{i_l j'_l i_{l+1}} &\leftarrow x^{\mathrm{DAG}}_{i_l j'_l i_{l+1}} - u(\mathcal{C})/\phi_{j'_l i_{l+1}} \,, \end{aligned}$$

for all $l = 1, 2, \dots, \ell$. This iterative process maintains the following invariants which hold at the end of each iteration:

- $\mathbf{x}^{\mathrm{S}}$ remains feasible for the static fluid problem, in particular, it satisfies flow balance $\mathbf{R}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{S}}) = 0$.
- $\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}}$ remains true.
- $R(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) = R(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{F}})$ remains true, i.e., $\mathbf{x}^{\mathrm{DAG}}$ has the same net inflow/outflow from each supply node as $\mathbf{x}^{\mathrm{F}}$. In particular, $\mathbf{x}^{\mathrm{DAG}}$ satisfies approximate flow balance (13).

Moreover, the iterative process progresses monotonically: Observe that $\mathbf{x}^{\mathrm{S}}$ (weakly) increases monotonically, whereas $\mathbf{x}^{\mathrm{DAG}}$ (weakly) decreases monotonically (but preserves $\mathbf{x}^{\mathrm{DAG}} \ge \mathbf{0}$). Since we also know that $\mathbf{x}^{\mathrm{S}}$ is bounded, it follows that this iterative process converges. Moreover, when it converges, it must be that there is no remaining cycle with positive flow in $\mathbf{x}^{\mathrm{DAG}}$, else it contradicts the fact that the process has converged. Hence, $\mathbf{x}^{\mathrm{S}}, \mathbf{x}^{\mathrm{DAG}}$ at the end of the process provide the claimed decomposition. □

Using this supporting lemma, we now establish the second key lemma which shows that the value of the finite horizon fluid problem cannot be much larger than the value of the static fluid problem.

Lemma 7. *Let* W_T^{OPT} *be the value of the finite horizon fluid problem. This value is upper bounded in terms of the value* W^{OPT} *of the static fluid problem as*

$$W_T^{\mathrm{OPT}} \le W^{\mathrm{OPT}} + \frac{(m-1)K w_{\max}}{T} \,.$$

Proof. We appeal to the decomposition from Lemma 6 to decompose any feasible solution $\mathbf{x}^{\mathrm{F}}$ to the finite horizon fluid problem as

$$\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}} \,,$$

where $\mathbf{x}^{\mathrm{S}}$ is feasible for the static fluid problem and $\mathbf{x}^{\mathrm{DAG}}$ is a directed acyclic flow that is feasible for the finite horizon fluid problem, i.e., satisfying (13) and (11). Hence, the objective (9) of the finite horizon fluid problem can be written as the sum of two terms $\mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{F}}) = \mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{S}}) + \mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}})$, and each of the terms can be bounded from above. By definition of W^{OPT} we know that

$$\mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{S}}) \leq W^{\mathrm{OPT}} .$$

We will now show that $\mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) \leq \frac{(m-1)Kw_{\max}}{T}$. The lemma will follow, since this will imply an upper bound of $W^{\mathrm{OPT}} + \frac{(m-1)Kw_{\max}}{T}$ on the objective for any $\mathbf{x}^{\mathrm{F}}$ satisfying (13) and (11).

Consider $\mathbf{x}^{\mathrm{DAG}}$. Since it is a directed acyclic assignment, there is an ordering $(k_1, k_2, \dots, k_m)$ of the nodes in V_S such that all assignments move supply from an earlier nodes to a later node in this ordering. More precisely, it holds that

$$x^{\mathrm{DAG}}_{k_l,j',k_r} = 0 \qquad \forall\, m \geq l > r \geq 1\,, j' \in \mathcal{N}(k_l)\,. \tag{58}$$

Now consider the subsets $A_\ell \triangleq \{k_1, k_2, \dots, k_\ell\} \subset V_S$ for $\ell = 1, 2, \dots, m-1$. Note that from (58), x^{DAG} does not move any supply from $V_S \backslash A_\ell$ to A_ℓ. Hence we have

$$\mathbf{1}^{\mathrm{T}}_{A_\ell}\mathbf{R}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) \;=\; \sum_{i \in A_\ell, k \in V_S \backslash A_\ell, j' \in \mathcal{N}(i)} \phi_{j'k} x^{\mathrm{DAG}}_{ij'k} \;\leq\; K/T \quad \forall\, l = 1, 2, \dots, m-1\,, \tag{59}$$

where we made use of (13) to obtain the upper bound. Further, note that for each $x^{\mathrm{DAG}}_{k_l,j',k_r}$ with $l < r$, the term $\phi_{j'k_r} x^{\mathrm{DAG}}_{k_l j' k_r}$ is part of the above sum for $\ell = l$. It follows that

$$\begin{aligned}
\mathbf{1}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) \;&=\; \sum_{1 \leq l < r \leq m, j' \in \mathcal{N}(k_l)} \phi_{j'k_r} x^{\mathrm{DAG}}_{k_l j' k_r} \\
&\leq \sum_{1 \leq \ell < m} \sum_{i \in A_\ell, k \in V_S \backslash A_\ell, j' \in \mathcal{N}(i)} \phi_{j'k} x^{\mathrm{DAG}}_{ij'k} \\
&\leq (m-1)K/T\,,
\end{aligned} \tag{60}$$

using (59) in the second step. Now, recalling the definition (1) of $w_{\max}$ and using (60), we deduce that

$$\mathbf{w}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) \leq w_{\max}\mathbf{1}^{T}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\mathrm{DAG}}) \leq w_{\max}K(m-1)/T\,.$$

This completes the proof. □

Proof of Proposition 1. The proposition follows immediately from Lemmas 5 and 7. □

B An example where the dual optimum lies outside the positive orthant

Example 1. *Consider a setting with two locations, whose clones are the supply and demand locations $V_S = \{1, 2\}$ and $V_D = \{1', 2'\}$. Suppose that each supply location is only pickup compatible with itself $E = \{(1, 1'), (2, 2')\}$, and that there is no demand from any location to itself $\phi_{1',1} = \phi_{2',2} = 0$,* *whereas* $\phi_{1',2} > \phi_{2',1} > 0$ *and* $w_{1,1',2} > 0$, $w_{2,2',1} > 0$. *Since there is more demand from location* 1 *to* 2 *than vice versa, some of the former demand must be dropped to satisfy the flow constraint* (10). *The unique optimum of the primal* (9)-(11) *is* $x^*_{1,1'} = \phi_{2',1}/\phi_{1',2}$ *and* $x^*_{2,2'} = 1$, *and the optima of the dual* (14) *form a one dimensional family* $\mathbf{y}^*(\eta) = (y^*_1(\eta), y^*_2(\eta)) = (\eta, w_{1,1',2} + \eta)$ *for all* $\eta \in \mathbb{R}$. *(Note that since demand* $(1', 2)$ *is* partially *satisfied in the primal optimum, it follows from strong duality that* $\max_{j\in\mathcal{N}(1')} w_{j,1',2} + y^*_j - y^*_2 = 0 \Leftrightarrow w_{1,1',2} + y^*_1 - y^*_2 = 0 \Leftrightarrow y^*_2 = y^*_1 + w_{1,1',2}$.*) This family intersects* $y_1 + y_2 = 1$ *at* $\mathbf{y}^* = \big((1 - w_{1,1',2})/2, (1 + w_{1,1',2})/2\big)$. *This point* $\mathbf{y}^*$ *lies outside the positive orthant for any instance where* $w_{1,1',2} > 1$.

C Justification of Condition 2 and Condition 2′

C.1 Condition 2′ is generic

We show that Condition 2′ holds generically within the following restriction of our model: for every supply node $i \in V_S$ there is a "clone" demand node $i' \in V_D$ and vice versa; in effect V_S and V_D are the same set. Moreover, each location is only pickup compatible with itself, i.e., $E = \{(i, i') : i \in V_S\}$. So the control problem boils down to choosing whether or not to serve each demand unit. (The near optimal MBP policy is an idling policy that sometimes drops a demand if the destination already has many vehicles relative to the origin. This prevents all the supply from piling up in a subset of locations while other locations are starved.) Henceforth in this section we consider the restricted model. Note that this restricted setting is similar to the baseline setting in Banerjee et al. (2016).

Condition 5 (Connectivity). *For every subset* $I \subset V_S$, *there is pair of nodes* $i \in I$, $k \in V_S \backslash I$ *such that* $\phi_{i'k} > 0$, $\phi_{k'i} > 0$, $w_{ii'k} > 0$ *and* $w_{kk'i} > 0$.

Lemma 8. *Under Condition 5, in any optimum* $\mathbf{x}^*$ *of static fluid problem* (9)-(11), *there is a positive flow of vehicles across every cut, i.e., for every subset* $I \subset V_S$ *we have*

$$\sum_{i\in I} \sum_{k\in V_S\backslash I} \phi_{i'k} x_{ii'k} > 0\,. \tag{61}$$

Proof of Lemma 8. Consider any feasible $\mathbf{x}$ which satisfies (10) and (11) but violates (61) for some subset I. We will show that $\mathbf{x}$ is not optimal for the problem (9)-(11). We construct a modified feasible assignment $\hat{\mathbf{x}}$ which is identical to $\mathbf{x}$ except that, for $i \in I$, $k \in V_S \backslash I$ given by Condition 5, we define

$$\hat{x}_{ii'k} = \frac{\min(\phi_{i'k}, \phi_{k'i})}{\phi_{i'k}}\,, \quad \hat{x}_{kk'i} = \frac{\min(\phi_{i'k}, \phi_{k'i})}{\phi_{k'i}}\,,$$

in contrast to $x_{ii'k} = x_{kk'i} = 0$. By definition $\hat{\mathbf{x}}$ satisfies constraints (10) and (11). Also, it achieves a value of the objective (9) which is $(w_{ii'k} + w_{kk'i}) \min(\phi_{i'k}, \phi_{k'i}) > 0$ larger than the objective value achieved by $\mathbf{x}$. This completes the proof. □

Assuming connectivity (Condition 5), we show that Condition 2′ follows from the following Generalized Imbalance condition in the restricted model.

Condition 6 (Generalized Imbalance)**.** *Let $\mathcal{D} = \{(i', k) : \phi_{i'k} > 0\}$ be the demand types that exist. Then there is no pair of disjoint subsets of demand types $\mathcal{D}_1, \mathcal{D}_2 \subset \mathcal{D}, \mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$ with $\mathcal{D}_1$ non-empty, and an associated subset of supply locations $I \subset V_S$ such that:*

- *All demand types $(i', k) \in \mathcal{D}_1$ have origin $i' \in I$ and destination $k \in V_S \backslash I$. All demand types $(i', k) \in \mathcal{D}_2$ have origin $i' \in V_S \backslash I$ and destination $k \in I$.*
- *The total demand of types in $\mathcal{D}_1$ is identical to the total demand of types in $\mathcal{D}_2$, i.e.,*

$$\mathbf{1}_{\mathcal{D}_1}^{\mathrm{T}} \boldsymbol{\phi} = \mathbf{1}_{\mathcal{D}_2}^{\mathrm{T}} \boldsymbol{\phi} \,. \tag{62}$$

In words, for every subset I, it holds that the total demand wanting to go from I to $V_S \backslash I$ is not equal to the total demand wanting to go from $V_S \backslash I$ to I.

The following fact is easy to establish (we skip the proof), since the exact balance of demands (62) holds only on a knife edge.

Fact 1 (Generalized Imbalance is generic)**.** *The Generalized Imbalance condition (Condition 6) is generic in the following sense. Fix the set of demand types $\mathcal{D}$. Then the set of demand arrival rate vectors $(\phi_{i',k})_{(i',k)\in\mathcal{D}}$ for which Condition 6 holds is open and dense in $\mathbb{R}_{++}^{|\mathcal{D}|}$.*

Now we show that Condition 6 implies Condition 2′. Note that in the restricted model under consideration, Condition 2′ simplifies to: *For every subset $I \subset V_S$, there is some $(i \in V_S, k \in V_S)$ such that:*

- *Exactly one of i and k is in I, i.e., $|\{i, k\} \cap I| = 1$.*
- *(i', k) is a partially satisfied demand type.*

Proposition 4. *Any instance of the restricted model which satisfies connectivity (Condition 5) and Generalized Imbalance (Condition 6) also satisfies Condition 2′.*

Proof. Consider any subset I and suppose Condition 5 holds. We will establish the result by showing that if Condition 2′ is violated for I then Condition 6 must be violated for I. Let $\mathbf{x}^*$ be an optimal solution to the problem (9)-(11). Suppose Condition 2′ is violated for I. Then all "cut crossing" demand types, i.e., demand types with origin in I and destination in $V_S \backslash I$ or vice versa, are either fully unsatisfied or fully satisfied under $\mathbf{x}^*$. Let $\mathcal{D}_1$ be the set of demand types with origin in I and destination in $V_S \backslash I$ which are fully satisfied under $\mathbf{x}^*$. Similarly, let $\mathcal{D}_2$ be the set of demand types with origin in $V_S \backslash I$ and destination in I which are fully satisfied under $\mathbf{x}^*$. Since Condition 5 holds, by Lemma 8 we have (61) for I, so we deduce that $\mathcal{D}_1$ is non-empty. Since $\mathbf{x}^*$ satisfies the flow constraint (10), it must be that $\mathbf{1}_{\mathcal{D}_1}^{\mathrm{T}} \boldsymbol{\phi} = \mathbf{1}_{\mathcal{D}_2}^{\mathrm{T}} \boldsymbol{\phi}$, i.e., Condition 6 is violated for I. □

Combining Proposition 4 with Fact 1, we get that Condition 2′ is generic in the restricted model.

Corollary 1. *Condition 2′ holds generically in the restricted model in the following sense. Fix the set of demand types $\mathcal{D}$. Then the set of demand arrival rate vectors $(\phi_{i',k})_{(i',k)\in\mathcal{D}}$ for which Condition 2′ holds is open and dense in $\mathbb{R}_{++}^{|\mathcal{D}|}$.*

C.2 Condition 2′ implies Condition 2: Proof of Proposition 2

Proof of Proposition 2. By definition, for all $\mathbf{y} \in \mathbb{R}^m$ and any $\xi \in \mathbb{R}$, we have $g(y + \xi\mathbf{1}) = g(y)$. It follows that for any optimal solution $\mathbf{y}^*$ to (14), the set $\{\mathbf{y}^*(\xi) \triangleq \mathbf{y}^* + \xi\mathbf{1}\}_{\xi\in\mathbb{R}}$ are all optima of (14). Hence, to establish Condition 2, it suffices to show that there is a unique optimal solution to (14) satisfying $y^*_{i_1} = 0$. We prove that this last statement holds under Condition 2′.

Consider any optimal solution $\mathbf{y}^*$ to (14) with $y^*_{i_1} = 0$. We will show inductively that all coordinates of y^* are uniquely determined. We proceed by writing down the full dual to (9)-(11). The full dual is

$$\text{minimize}_{\mathbf{y}\in\mathbb{R}^m,\mathbf{z}\in\mathbb{R}^{n\times m}} \sum_{j'\in V_D,k\in V_S} \phi_{j'k} z_{j'k} \tag{63}$$

$$\text{s.t.} \quad z_{j'k} \geq w_{ij'k} + y_i - y_k \quad \forall\, j' \in V_D\,, k \in V_S\,, i \in \mathcal{N}(j')\,, \tag{64}$$

$$z_{j'k} \geq 0 \quad \forall\, j' \in V_D\,, k \in V_S\,. \tag{65}$$

where $\mathbf{z}$ are the dual variables to the demand constraint (11). Since $\mathbf{y}^*$ is an optimal solution to the partial dual (14), there is some $\mathbf{z}^*$ such that $(\mathbf{y}^*, \mathbf{z}^*)$ is an optimal solution to the full dual (63)-(65). Let I be the set of nodes at which the value of $\mathbf{y}^*$ has been proved to be uniquely determined. Our induction base is $I = \{i_1\}$ since $y^*_{i_1} = 0$. As long as $I \subset V_S$, we can add a node to I as follows. Invoke Condition 2′ with subset I. Then there exists some $i \in I$, $j' \in V_D$ and $k \in V_S \backslash I$ such that (i) location i is an optimal pickup location for partially satisfied demand type (j', k), or (ii) location k is an optimal pickup location for partially satisfied demand type (j', i). Consider case (i). Since i is an optimal pickup location for demand type (j', k), it follows from complementary slackness that constraint (64) is tight for j', k, i, i.e., $z^*_{j'k} = w_{ij'k} + y^*_i - y^*_k$. Furthermore, since (j', k) is a partially satisfied demand type, it follows from complementary slackness that $z^*_{j'k} = 0$. Combining we conclude that $0 = w_{ij'k} + y^*_i - y^*_k \Rightarrow y^*_k = w_{ij'k} + y^*_i$. Since $i \in I$ we already knew that y^*_i is uniquely determined, and we can now deduce that y^*_k is also uniquely determined. As a result, we can add node k to I. Case (ii) can be handled very similarly via the deduction $0 = w_{kj'i} + y^*_k - y^*_i$, and again we are able to add node k to I. Induction then completes the proof that all coordinates of y^* are uniquely determined. □

D Supporting fact: $\mathbf{q}^*$ has no coordinate that is too small

Let $q^*_{\min} \triangleq \min_{i\in V_S} q^*_i$. One can easily prove a lower bound on $q^*_{\min}$ via the following fact.

Fact 2. *Suppose Condition 2 holds. Then, we have that* $\max_{i\in V_S} \mathbf{y}_i^* - \min_{k\in V_S} \mathbf{y}_k^* \le 2(m-1)w_{\max}$*, where* $\mathbf{y}^*$ *is any optimum of problem* (14)*, for instance the one defined in* (33)*.*

Proof of Fact 2. Let $\mathbf{x}^*$ be a primal optimum. Then we know that

$$\begin{aligned} W^{\mathrm{OPT}} &= g(y^*) = \max_{\mathbf{x}\in\mathcal{X}} (\mathbf{w}^{\mathrm{T}} + (\mathbf{y}^*)^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}) \\ &= \sum_{j'\in V_D, k\in V_S} \phi_{j'k} \left[\max_{i\in\mathcal{N}(j')} \ (w_{ij'k} + y_i^* - y_k^*) \right]^+ \\ &= (\mathbf{w}^{\mathrm{T}} + \mathbf{y}^{\mathrm{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^*) \end{aligned} \tag{66}$$

Now, if the mean demand $\boldsymbol{\phi}$ is not strongly connected, then the dual optimum will have an additional degree of freedom for every strongly connected component which contradicts Condition 2, hence $\boldsymbol{\phi}$ is strongly connected (in the sense that for every ordered pair of locations (i, k), there is a sequence of demand types with positive arrival rate and associated dispatch locations that would take a vehicle from i eventually to k). Order the nodes in V_S in decreasing order of y_i^* as $y_{i_1}^* \ge y_{i_2}^* \ge \ldots \ge y_{i_m}^*$. If the difference between consecutive values is at most $2w_{\max}$ then we are done. Suppose not (we will obtain a contradiction). Then there is some ℓ such that $y_{i_\ell}^* - y_{i_{\ell+1}}^* > 2w_{\max}$. Let $A \triangleq \{y_{i_l}\}_{1\le l\le \ell}$. Since $\boldsymbol{\phi}$ is strongly connected, there is some demand type (j', k) that can (depending on the dispatch decision) take a vehicle from A to $V_S\backslash A$, i.e., (j', k) is such that $A \cap \mathcal{N}(j') \ne \emptyset$ and $k \in V_S\backslash A$. Let $i \in A \cap \mathcal{N}(j')$. Then $w_{ij'k} + y_i^* - y_k^* \ge -w_{\max} + y_i^* - y_k^* > w_{\max} > 0$ and also $w_{ij'k} + y_i^* - y_k^* > w_{\hat{i}j'k} + y_{\hat{i}}^* - y_k^*$ for any $\hat{i} \in V_S\backslash A$ since $y_i^* - y_{\hat{i}}^* > 2w_{\max} \ge w_{\hat{i}j'k} - w_{ij'k}$. It follows from (66) that x^* *does* dispatch from A to serve all such demand (j', k). On the other hand, it never dispatches from $V_S\backslash A$ to serve demand with destination in A since $w_{\max} + y_k^* - y_i^* < -w_{\max} < 0$ for any $k \in V_S\backslash A$ and $i \in A$. But then we can deduce that $\mathbf{x}^*$ takes vehicles from A to $V_S\backslash A$ but not vice versa, i.e., it violates the flow constraint. Thus we have obtained a contradiction. □

Corollary 2. *Under Condition 2, we have*

$$q_{\min}^* \ge e^{-(2m/c_0) - \log m}, \qquad \text{where } q_{\min}^* \triangleq \min_{i\in V_S} q_i^* . \tag{67}$$

Proof of Corollary 2. Since $\mathbf{1}^{\mathrm{T}}\mathbf{q}^* = 1$ by definition, it follows that $q_{\max}^* \triangleq \max_{i\in V_S} q_i^* \ge 1/m$. Using Fact 2, it follows that

$$\begin{aligned} & c\log(q_{\min}^*/q_{\max}^*) = \min_{k\in V_S} \mathbf{y}_k^* - \max_{i\in V_S} \mathbf{y}_i^* \ge -2(m-1)w_{\max} \\ \Rightarrow \quad & q_{\min}^* \ge \exp(-2(m-1)w_{\max}/c)/m = e^{-2(m-1)/c_0}/m . \end{aligned}$$

Taking natural log on both sides, we further obtain that $\log q_{\min}^* \ge -2(m-1)/c_0 - \log m$. □

E Lower Bound for the Partial-Dual Function: Proof of Lemma 1

Proof. Let $\mathbf{y} = \nabla\Phi^*(\bar{\mathbf{q}})$ for some $\bar{\mathbf{q}} \in \Delta^m \cap \mathbb{R}^m_{++}$. Let $\mathcal{Y}^*$ be the set of optimal solutions of dual problem (14). Using Theorem 3.5 of Burke and Ferris (1993), we have that $\mathcal{Y}^*$ is a set of weak sharp minima, i.e. there exists $\alpha > 0$ such that

$$g(\mathbf{y}) - g(\mathbf{y}^*) \geq \alpha||\mathbf{y} - \mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2\,, \tag{68}$$

where $\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})$ is the orthogonal projection of $\mathbf{y}$ on $\mathcal{Y}^*$. It can be easily verified that

$$\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y}) = \mathbf{y}^* + \frac{1}{m}\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle \mathbf{1}\,. \tag{69}$$

Claim 1: If it is true that

$$|\cos\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle| = \frac{|\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle|}{\sqrt{m}||\mathbf{y} - \mathbf{y}^*||_2} \leq \frac{1}{\gamma} \qquad \text{for some } \gamma > 1\,, \tag{70}$$

where $\cos\langle \mathbf{a}, \mathbf{b}\rangle$ is the cosine of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, then (34) holds for some $\beta > 0$.

Proof of Claim 1: To see this, note that by Pythagorean equality and (69), we have

$$||\mathbf{y} - \mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2^2 = ||\mathbf{y} - \mathbf{y}^*||_2^2 - ||\mathbf{y}^* - \mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2^2 = ||\mathbf{y} - \mathbf{y}^*||_2^2 - \frac{1}{m}|\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle|^2\,. \tag{71}$$

Hence if (70) holds, we have

$$\begin{aligned} ||\mathbf{y} - \mathbf{y}^*||_2 &\geq \frac{\gamma}{\sqrt{m}}|\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle| \quad \text{for some } \gamma > 1 \\ &= \frac{\alpha}{\sqrt{m(\alpha^2 - \tilde{\beta}^2)}}|\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle| \quad \text{for some } \tilde{\beta} \in (0, \alpha)\,. \end{aligned} \tag{72}$$

Therefore

$$\begin{aligned} ||\mathbf{y} - \mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2^2 &= ||\mathbf{y} - \mathbf{y}^*||_2^2 - \frac{1}{m}|\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle|^2 && \text{(using (71))} \\ &\geq \frac{\tilde{\beta}^2}{m\alpha^2}|||\mathbf{y} - \mathbf{y}^*||_2^2 \quad \text{for some } \tilde{\beta} \in (0, \alpha). && \text{(using (72))} \end{aligned}$$

Plug the above inequality into (68), we have

$$g(\mathbf{y}) - g(\mathbf{y}^*) \geq \alpha||\mathbf{y} - \mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2 \geq \frac{\tilde{\beta}}{\sqrt{m}}||\mathbf{y} - \mathbf{y}^*||_2 \quad \text{for some } \tilde{\beta} \in (0, \alpha)\,,$$

and the lemma holds with $\beta = \tilde{\beta}/\sqrt{m}$. The claim is proved.

In the following, we prove (70). There are two cases:

1. $\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle \geq 0$. Since it is always true that $||\mathbf{y} - \mathbf{y}^*||_2 \geq \frac{1}{\sqrt{m}}||\mathbf{y} - \mathbf{y}^*||_1$, if suffices to prove that

$$\mathcal{T} \triangleq ||\mathbf{y} - \mathbf{y}^*||_1 - \gamma\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^*\rangle \geq 0 \quad \text{for some } \gamma > 1\,. \tag{73}$$

We have

$$\begin{aligned} \mathcal{T} &= c\sum_{i \in V_S} \log\left(\frac{\max\{\bar{q}_i, q_i^*\}}{\min\{\bar{q}_i, q_i^*\}}\right) - c\gamma\sum_{i \in V_S}\log\left(\frac{\bar{q}_i}{q_i^*}\right) \\ &= c\sum_{\{i \in V_S : \bar{q}_i \geq q_i^*\}} \log\left(\frac{\bar{q}_i}{q_i^*}\right) + c\sum_{\{i \in V_S : \bar{q}_i < q_i^*\}} \log\left(\frac{q_i^*}{\bar{q}_i}\right) - c\gamma\sum_{i \in V_S}\log\left(\frac{\bar{q}_i}{q_i^*}\right) \end{aligned}$$

$$= c(\gamma+1) \sum_{\{i\in V_S : \bar{q}_i < q_i^*\}} \log\left(\frac{q_i^*}{\bar{q}_i}\right) - c(\gamma-1) \sum_{\{i\in V_S : \bar{q}_i \geq q_i^*\}} \log\left(\frac{\bar{q}_i}{q_i^*}\right) \quad . \tag{74}$$

Using the concavity of logarithmic function, for $0 < x_1 \leq x_2$ we have $\frac{1}{x_2}(x_2 - x_1) \leq \log(x_2) - \log(x_1) = \log\left(\frac{x_2}{x_1}\right) \leq \frac{1}{x_1}(x_2 - x_1)$. Hence

$$\mathcal{T} \geq \frac{c(\gamma+1)}{\max_{\{i\in V_S:\bar{q}_i<q_i^*\}} q_i^*} \sum_{\{i\in V_S:\bar{q}_i<q_i^*\}} \left(q_i^* - \bar{q}_i\right) - \frac{c(\gamma-1)}{\min_{\{i\in V_S:\bar{q}_i\geq q_i^*\}} q_i^*} \sum_{\{i\in V_S:\bar{q}_i\geq q_i^*\}} \left(\bar{q}_i - q_i^*\right) . \tag{75}$$

Since $\bar{\mathbf{q}}, \mathbf{q}^* \in \Delta^m$, we have

$$\sum_{\{i\in V_S:\bar{q}_i<q_i^*\}} \left(\bar{q}_i - q_i^*\right) + \sum_{\{i\in V_S:\bar{q}_i\geq q_i^*\}} \left(\bar{q}_i - q_i^*\right) = 0 \, . \tag{76}$$

Plugging (76) into (75) and noting that $\max_i q_i^* < 1$, we have

$$\begin{aligned} \mathcal{T} &\geq c \sum_{\{i\in V_S:\bar{q}_i<q_i^*\}} \left(q_i^* - \bar{q}_i\right) \cdot \left(\frac{\gamma+1}{\max_{\{i\in V_S:\bar{q}_i<q_i^*\}} q_i^*} - \frac{\gamma-1}{\min_{\{i\in V_S:\bar{q}_i\geq q_i^*\}} q_i^*}\right) \\ &= c \sum_{\{i\in V_S:\bar{q}_i<q_i^*\}} \left(q_i^* - \bar{q}_i\right) \left(\frac{\gamma+1}{q^*_{\max}} - \frac{\gamma-1}{q^*_{\min}}\right) . \end{aligned}$$

Note that $\mathcal{T} \geq 0$ for $\gamma \in \left(1, \frac{1+q^*_{\min}}{1-q^*_{\min}}\right)$, hence (73) holds.

2. $\langle \mathbf{1}, \mathbf{y} - \mathbf{y}^* \rangle = -\rho < 0$. We first reformulate (70) as the following claim: there exists $\gamma > 1$ such that for all $\rho > 0$ and

$$V(\rho) \triangleq \min_{\mathbf{y}} ||\mathbf{y} - \mathbf{y}^*||_2 \qquad \text{s.t. } \mathbf{1}^{\mathrm{T}}\mathbf{y} = \mathbf{1}^{\mathrm{T}}\mathbf{y}^* - \rho, \quad \sum_{i\in V_S} e^{y_i/c} = 1 \, , \tag{77}$$

we have

$$V(\rho) \geq \frac{\gamma}{\sqrt{m}}\rho \, . \tag{78}$$

We have the following claim:

Claim 2: (78) holds for $\gamma = \sqrt{1 + \frac{1}{4m}}$.

Proof of Claim 2: We consider a relaxed problem of (77), where the last equality is relaxed by inequality $\sum_{i\in V_S} e^{y_i/c} \geq 1$. Note that the optimal value of relaxed problem is a lower bound for $V(\rho)$. Suppose the claim is not true, then there exists $\mathbf{y}$ that satisfies the following simultaneously:

$$||\mathbf{y} - \mathbf{y}^*||_2^2 \leq \left(1 + \frac{1}{4m}\right)\frac{\rho^2}{m} \, , \quad \mathbf{1}^{\mathrm{T}}\mathbf{y} = \mathbf{1}^{\mathrm{T}}\mathbf{y}^* - \rho \, , \quad \sum_{i\in V_S} e^{y_i/c} \geq 1 \, . \tag{79}$$

For such $\mathbf{y}$, it is easy to verify that

$$\left\langle \mathbf{y} - \left(\mathbf{y}^* - \frac{\rho}{m}\mathbf{1}\right), \frac{\rho}{m}\mathbf{1} \right\rangle = 0 \, .$$

Therefore by Pythagorean theorem we have

$$\left\|\mathbf{y}-\left(\mathbf{y}^*-\frac{\rho}{m}\right)\right\|_\infty^2 \le \left\|\mathbf{y}-\left(\mathbf{y}^*-\frac{\rho}{m}\mathbf{1}\right)\right\|_2^2 = ||\mathbf{y}-\mathbf{y}^*||_2^2 - \left\|\frac{\rho}{m}\mathbf{1}\right\|_2^2 \le \frac{\rho^2}{4m^2}\,.$$

As a result,

$$\begin{aligned}\sum_{i\in V_S} e^{y_i/c} = \sum_{i\in V_S} e^{y_i^*/c}e^{(y_i-y_i^*)/c} &\le\ e^{-\frac{\rho}{cm}+\frac{||\mathbf{y}-\mathbf{y}^*+\frac{\rho}{m}\mathbf{1}||_\infty}{c}}\sum_{i\in V_S} e^{y_i^*/c}\\ &\le e^{-\frac{\rho}{cm}+\frac{\rho}{2cm}} \le e^{-\frac{\rho}{2cm}} < 1\,,\end{aligned}$$

contradicting the last inequality in (79). Hence Claim 2 is true, and (70) holds with $\sqrt{1+\frac{1}{4m}}$.

To sum up, (70) holds with

$$\gamma = \min\left\{\frac{1+q_{\min}^*}{1-q_{\min}^*}, \sqrt{1+\frac{1}{4m}}\right\} > 1\,.$$

This concludes the proof. □

F Negative Lyapunov Drift: Proof of Lemma 2

We need the following technical facts of KL divergence in the following analysis. For $\mathbf{p},\mathbf{q},\mathbf{r} \in \Delta_m \cap \mathbb{R}_{++}$,

1. (Theorem 1, Sason and Verdú 2015)

$$D_{\mathrm{KL}}(\mathbf{p},\mathbf{q}) \le \log\left(1+\frac{1}{8\min_i q_i}||\mathbf{p}-\mathbf{q}||_1^2\right)\,. \tag{80}$$

2. (Three-point lemma of Bregman divergence, see e.g. Chen and Teboulle 1993)

$$D_{\mathrm{KL}}(\mathbf{r},\mathbf{p}) + D_{\mathrm{KL}}(\mathbf{p},\mathbf{q}) - D_{\mathrm{KL}}(\mathbf{r},\mathbf{q}) = \langle \log\mathbf{q} - \log\mathbf{p}, \mathbf{r}-\mathbf{p}\rangle\,. \tag{81}$$

Proof of Lemma 2. Our proof consists of three steps:

- *Step 1. Relate the Lyapunov drift with current optimality gap.* We have:

$$\begin{aligned}&\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1)) - L(\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right]\\ =\ &\mathbb{E}\left[D_{\mathrm{KL}}(\bar{\mathbf{q}}(t+1)||\mathbf{q}^*) - D_{\mathrm{KL}}(\bar{\mathbf{q}}(t)||\mathbf{q}^*)|\bar{\mathbf{q}}(t)\right]\\ =\ &\underbrace{\mathbb{E}\left[\langle\log\bar{\mathbf{q}}(t) - \log\mathbf{q}^*, \bar{\mathbf{q}}(t+1)-\bar{\mathbf{q}}(t)\rangle|\bar{\mathbf{q}}(t)\right]}_{\mathcal{T}_0}\\ &+\underbrace{\mathbb{E}\left[D_{\mathrm{KL}}(\bar{\mathbf{q}}(t+1)||\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right]}_{\mathcal{T}_1}\,. \qquad\qquad \text{(Three-point Lemma (81))}\end{aligned}$$

We decompose term $\mathcal{T}_0$ into sum of the Lyapunov drift under the Nominal policy (denoted as $\mathcal{T}_2$), and MBP's deviation from the nominal policy (due to forced demand drop when the dispatch location is almost empty, denoted as $\mathcal{T}_3$):

$$\mathcal{T}_0 = \frac{1}{\tilde{K}}\mathbb{E}\left[\langle\log\bar{\mathbf{q}}(t) - \log\mathbf{q}^*, -\mathbf{R}(\tilde{\mathbf{a}}(t)\circ\mathbf{x}(t))\rangle\,\middle|\bar{\mathbf{q}}(t)\right]$$

$$= \underbrace{\frac{1}{\tilde{K}}\mathbb{E}\left[\langle \log \bar{\mathbf{q}}(t) - \log \mathbf{q}^*, -\mathbf{R}(\tilde{\mathbf{a}}(t) \circ \mathbf{x}^{\mathrm{Nom}}(t))\rangle \,\middle|\, \bar{\mathbf{q}}(t)\right]}_{\mathcal{T}_2}$$

$$+ \underbrace{\frac{1}{\tilde{K}}\mathbb{E}\left[\sum_{j',k} a_{j'k}(t)\langle \log \bar{\mathbf{q}}(t) - \log \mathbf{q}^*, \mathbf{R}_{j'k}\mathbf{x}^{\mathrm{Nom}}_{j'k}(t)\rangle \mathbb{1}\left\{\bar{q}_{i(j'k)} \le 2\delta\right\} \,\middle|\, \bar{\mathbf{q}}(t)\right]}_{\mathcal{T}_3} .$$

It's easy to verify that

$$\mathbb{E}\left[\mathbf{R}(\tilde{\mathbf{a}}(t) \circ \mathbf{x}^{\mathrm{Nom}}(t)) \,\middle|\, \bar{\mathbf{q}}(t)\right] \in \partial g\left(c \log \bar{\mathbf{q}}(t)\right) .$$

Using the convexity of $g(\cdot)$, we have

$$\mathcal{T}_2 \le \frac{1}{c\tilde{K}}\Big(g\left(c\log \mathbf{q}^*\right) - g\left(c \log \bar{\mathbf{q}}(t)\right)\Big) . \tag{82}$$

Expand $\mathcal{T}_3$, we have

$$\begin{aligned}
\mathcal{T}_3 =& \frac{1}{\tilde{K}}\mathbb{E}\left[\sum_{j',k} a_{j'k}(t)\left(\log \bar{q}_{i(j'k)}(t) - \log \bar{q}_k(t) - \log q^*_{i(j'k)} + \log q^*_k\right) \mathbb{1}\left\{\bar{q}_{i(j'k)} \le 2\delta\right\} \,\middle|\, \bar{\mathbf{q}}(t)\right] \\
\le& \frac{\mathbf{1}^{\mathrm{T}}\boldsymbol{\phi}\mathbf{1}}{\tilde{K}} \max_{ij'k}\left(\log\left(\frac{2\delta}{\bar{q}_k}\right) + \log\left(\frac{\bar{q}^*_k}{\bar{q}^*_{i(j'k)}}\right)\right) \\
\le& \frac{mB}{\tilde{K}}\log\left(\frac{2}{q^*_{\min}}\right) .
\end{aligned} \tag{83}$$

Here the last inequality holds because $\bar{q}_k \ge \delta$, $q^*_k < 1$ for any $k \in V_S$, and $\mathbf{1}^{\mathrm{T}}\boldsymbol{\phi}\mathbf{1} \le mB$.

For the term $\mathcal{T}_1$, using (80) we have

$$\mathcal{T}_1 \le \mathbb{E}\left[\log\left(1 + \frac{||\mathbf{R}(\tilde{\mathbf{a}}(t)\circ \mathbf{x}(t))||_1^2}{8\tilde{K}^2 \min_i \bar{q}_i(t)}\right) \,\middle|\, \bar{\mathbf{q}}(t)\right] .$$

Note that

$$\begin{aligned}
||\mathbf{R}(\tilde{\mathbf{a}}(t)\circ\mathbf{x}(t))||_1 =& \sum_{i\in V_S} |\text{net inflow of demand into } i| \\
\le& \sum_{i \in V_S}|\text{inflow of demand into } i| + \sum_{i\in V_S}|\text{outflow of demand out of } i| \\
\le& \, 2(\text{total number of period } t \text{ demands}) \\
\le& \, 2mB ,
\end{aligned}$$

where the last inequality holds because at most B demands arise in the neighborhood of each $i \in V_S$ by Condition 1. Combined with fact that $\log(1+x) \le x$ for $x \ge 0$, we have

$$\mathcal{T}_1 \le \frac{m^2B^2}{2\tilde{K}^2 \min_i \bar{q}_i(t)} . \tag{84}$$

To sum up, we have

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1)) - L(\bar{\mathbf{q}}(t)) | \bar{\mathbf{q}}(t)\right] \le \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 . \tag{85}$$

We will derive upper bound for $\mathcal{T}_1$ and $\mathcal{T}_3$ to show that they do not offset the negative Lyapunov drift $\mathcal{T}_2$.

- *Step 2. Upper bounding $\mathcal{T}_1$.* Fix $\epsilon \in (0,1)$. Define

$$K_0 = 4B \exp\left(\frac{2m/c_0 + \log m}{\epsilon} + \frac{2m^2 B}{\beta\epsilon} + 1\right) . \tag{86}$$

For $K \geq K_0$, we claim that for $\min_{i\in V_S} \bar{q}_i(t) \leq \frac{1}{\tilde{K}^\epsilon}$

$$\mathcal{T}_1 \leq \frac{1}{4c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) , \tag{87}$$

for $\min_{i\in V_S} \bar{q}_i(t) \geq \frac{1}{\tilde{K}^\epsilon}$,

$$\mathcal{T}_1 \leq \frac{m^2 B^2}{2\tilde{K}^{2-\epsilon}} . \tag{88}$$

Hence for any $\bar{\mathbf{q}}(t)$, we have

$$\mathcal{T}_1 \leq \frac{1}{4c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) + \frac{m^2 B^2}{2\tilde{K}^{2-\epsilon}} . \tag{89}$$

Note that (88) can be derived by plugging $\min_{i\in V_S} \bar{q}_i(t) \geq \frac{1}{\tilde{K}^\epsilon}$ into (84).

Now we prove (87). First note that for $K \geq K_0$, we have

$$\frac{1}{\tilde{K}^\epsilon} \leq \exp\left(-\frac{2m/c_0 + \log m}{\epsilon}\epsilon\right) \leq q^*_{\min} ,$$

where the last inequality follows from Corollary 2. Using Lemma 1, we have

$$\frac{1}{c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) \geq \frac{\beta}{c\tilde{K}} ||c \log \bar{\mathbf{q}}(t) - c \log \mathbf{q}^*||_2 \geq \frac{\beta}{\tilde{K}} \left(\log q^*_{\min} - \log \frac{1}{\tilde{K}^\epsilon} \right) . \tag{90}$$

Plug Corollary 2 into (90), we have

$$\frac{1}{c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) \geq \frac{\beta}{\tilde{K}} \left(-\log m - 2m/c_0 + \epsilon \log \tilde{K} \right) ,$$

Plug in $K \geq K_0$, we have

$$\begin{aligned} \frac{1}{c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) &\geq \frac{\beta}{\tilde{K}} \left(-\log m - 2m/c_0 + \epsilon \left(\frac{\log m + 2m/c_0}{\epsilon} + \frac{2m^2 B}{\beta\epsilon} \right) \right) \\ &= \frac{2m^2 B}{\tilde{K}} \geq 4\mathcal{T}_1 , \end{aligned} \tag{91}$$

which is equivalent to (87).

- *Step 3. Upper bounding $\mathcal{T}_3$.* First note that $\mathcal{T}_3 \neq 0$ only when there exists $i \in V_S$ such that $\bar{q}_i \leq 2\delta$. Define

$$K_1 = 2B \exp\left(\frac{4mB}{\beta}(\log 2 + 2m/c_0 + \log m) + 2m/c_0 + \log m\right) .$$

For $K \geq K_1$, we have

$$\frac{1}{c\tilde{K}} \Big(g\left(c \log \bar{\mathbf{q}}(t)\right) - g\left(c \log \mathbf{q}^*\right) \Big) \geq \frac{\beta}{c\tilde{K}} ||c \log \bar{\mathbf{q}}(t) - c \log \mathbf{q}^*||_2 \geq \frac{\beta}{\tilde{K}} \left(\log q^*_{\min} - \log \frac{2B}{\tilde{K}} \right) . \tag{92}$$

Using Corollary 2 and $K \geq K_1$, we have

$$\begin{aligned}\frac{1}{c\tilde{K}}\Big(g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\mathbf{q}^*\right)\Big) &\geq \frac{\beta}{\tilde{K}}\left(-\log m-2m/c_0-\log 2B+\log K_1\right)\\ &=\frac{\beta}{\tilde{K}}\frac{4mB}{\beta}(\log 2+2m/c_0+\log m)\\ &\geq \frac{4mB}{\tilde{K}}\log\left(\frac{2}{q^*_{\min}}\right)\\ &\geq 4\mathcal{T}_3\,,\end{aligned}$$

where the last inequality follows from (83). Hence

$$\mathcal{T}_3 \leq \frac{1}{4c\tilde{K}}\Big(g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\mathbf{q}^*\right)\Big)\,.$$

All combined, for $K \geq \max\{K_0, K_1\} = \Omega\Big(\exp\left(\mathrm{poly}(m, B, 1/\beta, 1/\epsilon, 1/c_0)\right)\Big)$, we have

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1))-L(\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right] \leq -\frac{1}{2c\tilde{K}}\Big(g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\mathbf{q}^*\right)\Big)+\frac{m^2B^2}{2\tilde{K}^{2-\epsilon}}\,.$$

□

We remark that we can get an ϵ-free bound by replacing $\tilde{K}^\epsilon$ with $\log\tilde{K}$ in cases (2) and (3) and performing the same analysis. For $\tilde{K} = \Omega\left(\exp\Big(\exp\left(\mathrm{poly}(m, B, 1/\beta, 1/c_0)\right)\Big)\right)$, we get

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1))-L(\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right] \leq -\frac{1}{2c\tilde{K}}\Big(g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\mathbf{q}^*\right)\Big)+\frac{m^2B^2}{2\tilde{K}^2}\,.$$

G Proofs of Lemma 3 and Theorem 1

Proof of Lemma 3. Fix $\epsilon > 0$. For $K \geq K_0$, taking expectations w.r.t. $\bar{\mathbf{q}}(t)$ on both sides of (36) in Lemma 2, we have

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1))\right]-\mathbb{E}\left[L(\bar{\mathbf{q}}(t))\right] \leq -\frac{1}{2c\tilde{K}}\left(\mathbb{E}\Big[g\left(c\log\bar{\mathbf{q}}(t)\right)\Big]-g\left(c\log\mathbf{q}^*\right)\right)+\frac{m^2B^2}{2\tilde{K}^{2-\epsilon}}\,.$$

Take the telescopic sum from 1 to T and rearrange the terms. We get

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\Big[g\left(c\log\bar{\mathbf{q}}(t)\right)\Big]-g\left(c\log\mathbf{q}^*\right)\right)\\ &\leq \frac{2c\tilde{K}}{T}\left(\mathbb{E}\left[D_{\mathrm{KL}}(\bar{\mathbf{q}}(1)||\bar{\mathbf{q}}^*)\right]-\mathbb{E}\left[D_{\mathrm{KL}}(\bar{\mathbf{q}}(T+1)||\bar{\mathbf{q}}^*)\right]+\frac{m^2B^2}{2\tilde{K}^{2-\epsilon}}T\right)\\ &\leq \frac{2c\tilde{K}}{T}\cdot\max_{\mathbf{p}\in\Delta^m}\log\left(1+\frac{||\mathbf{p}-\mathbf{q}^*||_1^2}{8q^*_{\min}}\right)+\frac{cm^2B^2}{\tilde{K}^{1-\epsilon}}(\text{using (80)})\\ &\leq \frac{2c\tilde{K}}{T}\log\left(1+\frac{1}{2q^*_{\min}}\right)+\frac{cm^2B^2}{\tilde{K}^{1-\epsilon}} \qquad\qquad (\text{since } \|\mathbf{p}_1-\mathbf{p}_2\|_1\leq 2\ \forall\ \mathbf{p}_1,\mathbf{p}_2\in\Delta^m)\,.\end{aligned}$$

Using Corollary 2 we have

$$\log\left(1+\frac{1}{2q^*_{\min}}\right) \le \log\left(1+me^{2m/c_0}/2\right) \le \log(3e^{2m/c_0}) = 2m/c_0 + \log 3\,,$$

where the last inequality uses $m \ge 1$. Substituting in the above bound and recalling $c = c_0 w_{\max}$ yields

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right)\right] - g\left(c\log\mathbf{q}^*\right)\right) \le \frac{2(2m+c_0\log 3)w_{\max}\tilde{K}}{T} + \frac{c_0 w_{\max} m^2 B^2}{\tilde{K}^{1-\epsilon}}\,.$$

as needed. □

Proof of Theorem 1. We first bound the amount by which the average payoff collected per period in T periods can fall short of W_{OPT}; the bound on regret will follow immediately.

Consider the t-th period for $1 \le t \le T$. We know from weak duality that $W^{\text{OPT}} \le g(\mathbf{y})\ \forall\, \mathbf{y} \in \mathbb{R}^m$. Plugging in $\mathbf{y} = \bar{\mathbf{q}}(t)$ and appealing to the definition of $\mathbf{x}^{\text{Nom}}(t)$, we deduce that

$$W^{\text{OPT}} \le g(\bar{\mathbf{q}}(t)) = (\mathbf{w}^{\text{T}} + c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R})(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\text{Nom}}(t))\,.$$

It follows that, for a given $\bar{\mathbf{q}}(t)$ we have

$$\begin{aligned}
&\mathbb{E}\left[W_{\text{OPT}} - \mathbf{w}^{\text{T}}(\tilde{\mathbf{a}} \circ \mathbf{x}^{\text{MBP}}(t)) \,\middle|\, \bar{\mathbf{q}}(t)\right] && (93)\\
&= W_{\text{OPT}} - \mathbf{w}^{\text{T}}(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\text{MBP}}(t))\\
&\le \mathbf{w}^{\text{T}}\left(\tilde{\boldsymbol{\phi}} \circ (\mathbf{x}^{\text{Nom}}(t) - \mathbf{x}^{\text{MBP}}(t))\right) + c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R}\left(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\text{Nom}}(t)\right)\\
&= \underbrace{(\mathbf{w}^{\text{T}} + c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R})\left(\tilde{\boldsymbol{\phi}} \circ (\mathbf{x}^{\text{Nom}}(t) - \mathbf{x}^{\text{MBP}}(t))\right)}_{\text{(a)}} + \underbrace{c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R}\left(\tilde{\boldsymbol{\phi}} \circ \mathbf{x}^{\text{MBP}}(t)\right)}_{\text{(b)}}\,. && (94)
\end{aligned}$$

We bound terms (a) and (b) separately. We start with term (a). Decomposing term (a) into individual demand types (j', k) and using (35), we obtain

$$\text{Term (a)} = \sum_{j',k}\phi_{j'k}\left(\mathbf{w}^{\text{T}}_{j'k} + c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R}_{j'k}\right)\mathbf{x}^{\text{Nom}}_{j'k}(t)\mathbb{1}\left\{\bar{q}_{i(j'k)}(t) \le 2\delta\right\} \qquad (95)$$

where $i(j'k)$ was defined in (35). Now, $\mathbf{x}^{\text{Nom}}_{j'k} = \mathbf{e}_{i(j'k)}$ by definition, hence

$$\begin{aligned}
\left(\mathbf{w}^{\text{T}}_{j'k} + c(\log\bar{\mathbf{q}}(t))^{\text{T}}\mathbf{R}_{j'k}\right)\mathbf{x}^{\text{Nom}}_{j'k}(t) =& w_{i(j'k)j'k} + c\left(\log\bar{q}_{i(j'k)}(t) - \log\bar{q}_k(t)\right)\\
\le& w_{\max} + c\log 2 && (96)
\end{aligned}$$

for $\bar{q}_{i(j'k)}(t) \le 2\delta$, since $\bar{q}_k(t) \ge \delta$ by definition (28) of $\bar{\mathbf{q}}$, and $w_{i(j'k)j'k} \le w_{\max}$. To control the indicator in (95), note that

$$\begin{aligned}
&\bar{q}_{i(j'k)}(t) \le 2\delta\\
\Rightarrow\quad& \|\log\mathbf{q}(t) - \log\mathbf{q}^*\|_2 \ge \log(q^*_{\min}) - \log(2\delta)\\
\Rightarrow\quad& g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right) \ge \beta c\left(\log q^*_{\min} - \log(2\delta)\right) \ge \beta c(-2m - \log m - \log(2\delta))\,,
\end{aligned}$$

using Lemma 1 and then Corollary 2. For $K \ge 4Be^{2m/c_0 + \log m + 1}$ we have $\delta = B/\tilde{K} < B/K$ leading to $-2m/c_0 - \log m - \log(2\delta) \ge \log(2e)$, and so

$$\bar{q}_{i(j'k)}(t) \le 2\delta$$

$$\Rightarrow g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right) \geq \beta c\log(2e) \text{ for } K \geq 4mBe^{2m/c_0+1}\,. \tag{97}$$

Note how we exploited here the fact that when at least one queue is very small the gap $g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)$ grows with K like $\log(1/\delta) \sim \log K$.

Deploying (96) and (97) in (95), we obtain

$$\text{Term (a)} \leq (\mathbf{1}^T\boldsymbol{\phi}\mathbf{1})\, w_{\max}(1+c_0\log 2)\mathbb{1}\left\{g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right) \geq \beta c\log 3\right\}.$$

Taking expectation with respect to $\bar{\mathbf{q}}(t)$ under MBP and using Markov's inequality, we obtain

$$\begin{aligned}\mathbb{E}_{\bar{\mathbf{q}}(t)}[\text{Term (a)}] &\leq \frac{(\mathbf{1}^T\boldsymbol{\phi}\mathbf{1})\, w_{\max}(1+c_0\log 2)}{\beta c\log 3}\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right]\\ &\leq \frac{mB(1+c_0\log 2)}{\beta c_0\log 3}\,\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right], \qquad \text{for } K \geq 4mBe^{2m/c_0+1}\,,\end{aligned} \tag{98}$$

since $\mathbf{1}^T\boldsymbol{\phi}\mathbf{1} \leq mB$ by Condition 1.

We now turn our attention to term (b) in (94). Note that given the definition (28) of $\bar{\mathbf{q}}$, we have $\bar{\mathbf{q}}(t+1) - \bar{\mathbf{q}}(t) = -\mathbf{R}(\tilde{\mathbf{a}}\circ\mathbf{x})/\tilde{K}$. Hence, for given $\bar{\mathbf{q}}(t)$, under MBP, we have

$$\mathbf{R}(\tilde{\boldsymbol{\phi}}\circ\mathbf{x}^{\mathrm{MBP}}(t)) = \tilde{K}\,\mathbb{E}[\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)|\bar{\mathbf{q}}(t)]\,.$$

Substituting in term (b) in (94), we then get

$$\begin{aligned}\text{Term(b)} =& \tilde{K}c\,(\log\bar{\mathbf{q}}(t))^{\mathrm{T}}\,\mathbb{E}[\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)|\bar{\mathbf{q}}(t)]\\ =& c\tilde{K}\,(\log\mathbf{q}^*)^{\mathrm{T}}\,\mathbb{E}[(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1))|\bar{\mathbf{q}}(t)]\\ &+ \tilde{K}\,\mathbb{E}[\left(c\log\bar{\mathbf{q}}(t) - c\log\mathbf{q}^*\right)^{\mathrm{T}}\left(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\right)|\bar{\mathbf{q}}(t)]\end{aligned} \tag{99}$$

Towards controlling the second term, note that $\|\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\|_2 \leq \|\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\|_1 \leq 2mB/\tilde{K}$ where the second inequality holds since at most B vehicles are dispatched from each of the m locations in V_S in period t given Condition 1, and each dispatch decrements one coordinate of $\mathbf{q}$ and increments another. Now we use Cauchy-Schwarz inequality to control the second term in (99),

$$\begin{aligned}\left(c\log\bar{\mathbf{q}}(t) - c\log\mathbf{q}^*\right)^{\mathrm{T}}\left(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\right) \leq& \|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2\cdot\|\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\|_2\\ \leq& \|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2\cdot\frac{2mB}{\tilde{K}}\\ \leq& \frac{2mB}{\tilde{K}\beta}\left(g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right)\,,\end{aligned} \tag{100}$$

where we used Lemma 1 in the last step. Taking the expectation with respect to $\bar{\mathbf{q}}(t)$ in (99) and using (100), we get

$$\mathbb{E}_{\bar{\mathbf{q}}(t)}[\text{Term(b)}] = c\tilde{K}(\log\mathbf{q}^*)^{\mathrm{T}}\mathbb{E}[(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1))] + \frac{2mB}{\beta}\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right]\,. \tag{101}$$

We now return to (94). Taking expectation of (94) with respect to $\bar{\mathbf{q}}(t)$ under MBP, using the expressions for each of the terms from (98) and (101) and taking the average over $1 \leq t \leq T$, for $K \geq 4mBe^{2m/c_0+1}$ we get

$W^{\mathrm{OPT}} - \bar{v}^{\mathrm{MBP}}(T)$

$$\le \frac{c\tilde{K}}{T}(\log \mathbf{q}^*)^{\mathrm{T}}\mathbb{E}[(\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1))] + \left(\frac{1+c_0\log 2}{c_0\log 3}+2\right)\frac{mB}{\beta T}\sum_{t=1}^{T}\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\bar{\mathbf{q}}^*\right)\right], \tag{102}$$

where the first term is the result of a telescoping sum. Using the Cauchy-Schwarz inequality, we have

$$(\log \mathbf{q}^*)^{\mathrm{T}}(\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1)) \le \|\log \mathbf{q}^*\|_2 \, \|\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1)\|_2 \le \|\log \mathbf{q}^*\|_1 \, \|\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1)\|_1 .$$

Note that $\|\log \mathbf{q}^*\|_1 \le m|\log q^*_{\min}| \le m(2m/c_0+\log m)$ using Corollary 2, and $\|\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1)\|_1 \le 2$ since $\bar{\mathbf{q}}(1), \bar{\mathbf{q}}(T+1) \in \Delta^m$. Hence,

$$(\log \mathbf{q}^*)^{\mathrm{T}}\mathbb{E}[(\bar{\mathbf{q}}(1)-\bar{\mathbf{q}}(T+1))] \le 2m(2m/c_0+\log m) . \tag{103}$$

Using (103) in (102), and bounding the second term using Lemma 3, we have, for $K \ge \max(K_0, 4mBe^{2m/c_0+1})$ that

$$\begin{aligned} W^{\mathrm{OPT}} - \bar{v}^{\mathrm{MBP}}(T) &\le (4m^2 + 2c_0 m\log m)\frac{w_{\max}\tilde{K}}{T} + \left(\frac{1}{c_0}+3\right)\frac{mB}{\beta} \\ &\quad \cdot \left(\frac{2(2m+c_0\log 3)w_{\max}\tilde{K}}{T} + \frac{c_0 w_{\max} m^2 B^2}{\tilde{K}^{1-\epsilon}}\right) \\ &\le 6w_{\max}m^2\left(c_0+2+\frac{2B}{\beta}\left(\frac{1}{c_0}+4+\frac{3c_0}{m}\right)\right)\frac{K}{T} + \frac{(1+3c_0)w_{\max}m^3B^3/\beta}{K^{1-\epsilon}}, \end{aligned} \tag{104}$$

using $c = c_0 w_{\max}$ and $K \ge 4Bm > 2Bm \Rightarrow \tilde{K} = K + mB \in (K, 3K/2)$. From the proof of Lemma 1 we know that $\beta \ge O(\sqrt{q^*_{\min}/m}) = O(\frac{1}{m}e^{-m/c_0})$. Minimize RHS w.r.t. c_0, it is not hard to see that the minimizer should have order of magnitude $O(1)$. Hence we choose $c_0 = 1$. Finally, plugging in the definition of Regret we lose another additive amount $w_{\max}(m-1)K/T$, leading to

$$\mathrm{Regret}^{\mathrm{MBP}}(T) \le 6w_{\max}m^2(3+16B/\beta)\,\frac{K}{T} + \frac{4w_{\max}m^3B^3/\beta}{K^{1-\epsilon}} \qquad \text{for } K \ge \max(K_0, 4mBe^{2m+1}) . \tag{105}$$

This establishes the theorem with

$$\begin{aligned} K_1 &\triangleq \max(K_0, 4mBe^{2m+1}) = \exp(\mathrm{poly}(m, B, 1/\beta, 1/\epsilon)) \\ M &\triangleq \max(6w_{\max}m^2(3+16B/\beta), 4w_{\max}m^3B^3/\beta) = \mathrm{poly}(w_{\max}, m, B, 1/\beta) . \end{aligned}$$

Note here β is a function of the primitives $(\mathbf{w}, \boldsymbol{\phi}, G)$ and $\beta > 0$ under Condition 2 by Lemma 1.

□

H Simulation Settings

Model Primitives.

- *Demand arrival process.* Using the estimation in Buchholz (2015), which is based on Manhattan's taxi trip data during August and September in 2012, we obtain the (average) demand arrival rates for each origin-destination pair during the day (7 a.m. to 4 p.m.).

- *Pickup/service times.* We extract the pairwise travel time between region centroids (marked by the dots in Figure 1) using Google Maps, denoted by D_{ij}'s ($i, j = 1, \cdots, 30$). We use D_{jk} as service time for customers traveling from j to k. For each customer at j who is picked up by a supply from i we add a pickup time [1] of $\tilde{D}_{ij} = \max\{D_{ij}, 2 \text{ minutes}\}$.

Benchmark policy: static fluid-based policy. We consider the fluid-based randomized policy (Banerjee et al. 2016, Ozkan and Ward 2016) as a benchmark. Let $\mathbf{x}^*$ be a solution of (9). When a type (j, k) demand arrives at location j, the randomized fluid-based policy dispatches from location $i \in \mathcal{N}(j)$ with probability x^*_{ijk}.

Initial state generation. We first uniformly sample 100 points from the simplex $\{\mathbf{q} : \sum_{i\in V_S} q_i = K\}$, which are used as the system's initial states at 6 a.m. (note that all the cars are free). Then we "warm-up" the system by employing the static policy from 6 a.m. to 8 a.m., assuming the demand arrival process during this period to be stationary (with the average demand arrival rate during this period as mean). Finally, we use the system's states at 8 a.m. as the initial states for the simulations in Section 7.

I The Joint-Pricing-Assignment Case: Proof of Theorem 2

In this section, we prove all the results in Section 8.

I.1 Equivalent Concave Formulation of the JPA Static Problem: Proof of Lemma 4

Proof of Lemma 4. The proof consists of two steps.

Step 1. We first show that there always exists an optimal solution of `JPA-static` where constraints (45) are binding for all $j' \in V_D$, $k \in V_S$. To see this, let $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ be any optimal solution of `JPA-static`. Define

$$\tilde{x}_{ij'k} \triangleq \begin{cases} 0 & \text{if } \sum_{i\in\mathcal{N}(j')} x^*_{ij'k} = 0\,, \\ \frac{x^*_{ij'k}}{\sum_{i\in\mathcal{N}(j')} x^*_{ij'k}} & \text{otherwise.} \end{cases} \qquad \tilde{\mu}_{j'k} \triangleq \mu^*_{j'k} \sum_{i\in\mathcal{N}(j')} x^*_{ij'k} \quad \forall j' \in V_D,\, i, k \in V_S\,.$$

We claim that $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}})$ is also an optimal solution of `JPA-static`. First note that $\mu^*_{j'k} x^*_{ij'k} = \tilde{\mu}_{j'k} \tilde{x}_{ij'k}$ holds for all i, j', k, hence the flow balance constraints (44) hold for $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}})$. We have $\sum_{i\in\mathcal{N}(j')} \tilde{x}_{ij'k} = 1$, hence demand constraints (45) hold. Non-negativity constraints (47) trivially hold. From the definition of $\tilde{\mu}_{j'k}$, we have $\tilde{\mu}_{j'k} \le \mu^*_{j'k} \le \mu_{j'k}(0)$, so constraints (46) also hold.

Finally, we examine the objective (43):

$$(43) = \sum_{j'\in V_D} \sum_{k\in V_S} \mu^*_{j'k} \sum_{i\in\mathcal{N}(j')} (w_{ij'} + p_{j'k}(\mu^*_{j'k})) x^*_{ij'k}$$

[1] We use the inflated D_{ij}'s as pickup times to account for delays in finding or waiting for the customer.

$$
\begin{aligned}
&= \sum_{j'\in V_D}\sum_{k\in V_S}\tilde{\mu}_{j'k}\sum_{i\in\mathcal{N}(j')}(w_{ij'}+p_{j'k}(\mu^*_{j'k}))\tilde{x}_{ij'k} && (\text{since } \mu^*_{j'k}x^*_{ij'k}=\tilde{\mu}_{j'k}\tilde{x}_{ij'k})\\
&\le \sum_{j'\in V_D}\sum_{k\in V_S}\tilde{\mu}_{j'k}\sum_{i\in\mathcal{N}(j')}(w_{ij'}+p_{j'k}(\tilde{\mu}_{j'k}))\tilde{x}_{ij'k} && (\text{since } \tilde{\mu}_{j'k}\le\mu^*_{j'k} \text{ and } p_{j'k}(\cdot) \text{ is decreasing})\,.
\end{aligned}
$$

This shows that $(\tilde{\mathbf{x}},\tilde{\boldsymbol{\mu}})$ is an optimal solution of `JPA-static`, and the demand constraints (45) are binding for $(\tilde{\mathbf{x}},\tilde{\boldsymbol{\mu}})$.

Step 2. We now show that each optimal solution of `JPA-concave`, $\hat{\mathbf{x}}^*$ can be converted to an optimal solution of `JPA-static`. Plug in the result obtained in step 1 and perform a change of variable on `JPA-static`:

$$
\hat{x}_{ij'k}=x_{ij'k}\mu_{j'k}\,,
$$

then we reduce `JPA-static` to `JPA-concave`. Therefore for any optimal solution $\hat{\mathbf{x}}^*$, the pair $(\mathbf{x}^*,\boldsymbol{\mu}^*)$ defined via (50) is an optimal solution to is `JPA-static`. This concludes the proof. □

I.2 Partial Dual of `JPA-concave` and Condition 4

The Lagrange dual function $g_{\mathrm{JPA}}(\mathbf{y})$ of `JPA-concave` can be derived as follows.

$$
\begin{aligned}
g_{\mathrm{JPA}}(\mathbf{y}) = &\max_{\{\mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}_{j'k}\le\mu_{j'k}(0),\hat{\mathbf{x}}_{j'k}\ge\mathbf{0},\forall j',k\}} \sum_{i\in V_S}\sum_{j'\in V_D}\sum_{k\in V_S} w_{ij'}\hat{x}_{ij'k} + \sum_{j'\in V_D}\sum_{k\in V_S} r_{j'k}\left(\sum_{i\in\partial(j')}\hat{x}_{ij'k}\right)\\
&+\sum_{i\in V_S} y_i\left(\sum_{j'\in\mathcal{N}(i)}\sum_{k\in V_S}\hat{x}_{ij'k}-\sum_{j'\in V_D}\sum_{l\in\mathcal{N}(j')}\hat{x}_{lj'i}\right)\\
= &\sum_{j'\in V_D}\sum_{k\in V_S}\max_{\{\mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}_{j'k}\le\mu_{j'k}(0),\hat{\mathbf{x}}_{j'k}\ge\mathbf{0}\}}\left(\sum_{i\in\mathcal{N}(j')}\big(w_{ij'}+y_i-y_k\big)\,\hat{x}_{ij'k}+r_{j'k}\left(\sum_{i\in\partial(j')}\hat{x}_{ij'k}\right)\right)\\
= &\sum_{j'\in V_D}\sum_{k\in V_S}\max_{\{0\le\mu_{j'k}\le\mu_{j'k}(0)\}}\max_{\{\mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}_{j'k}=\mu_{j'k},\hat{\mathbf{x}}_{j'k}\ge\mathbf{0}\}}\left(\sum_{i\in\mathcal{N}(j')}\big(w_{ij'}+y_i-y_k\big)\,\hat{x}_{ij'k}+r_{j'k}\left(\mu_{j'k}\right)\right)\\
= &\sum_{j'\in V_D}\sum_{k\in V_S}\max_{\{0\le\mu_{j'k}\le\mu_{j'k}(0)\}}\left(r_{j'k}\left(\mu_{j'k}\right)+\mu_{j'k}\max_{i\in\mathcal{N}(j')}\big(w_{ij'}+y_i-y_k\big)\right)\,.
\end{aligned}
$$

Note that g_{JPA} is translation invariant:

$$
g_{\mathrm{JPA}}(\mathbf{y}+\xi\mathbf{1})=g_{\mathrm{JPA}}(\mathbf{y})\quad\forall y\in\mathbb{R}^m,\xi\in\mathbb{R}\,. \tag{106}
$$

In the following, we show that Condition 4 implies the uniqueness (up to translating each coordinate by the same constant) of optimum of partial dual $g_{\mathrm{JPA}}(\mathbf{y})$.

Proposition 5. *Suppose Condition 4 holds, then the following is true. Fix any $i_1\in V$. There is a unique optimal solution $(\mathbf{y}')^*$ of the dual problem (51) satisfying $(\mathbf{y}')^*_{i_1}=0$. Then, as per (106), the set of all optimal solutions is $\{(\mathbf{y}')^*(\eta)\triangleq(\mathbf{y}')^*+\xi\mathbf{1}\,,\forall\xi\in\mathbb{R}\}$, i.e., the set of all dual vectors obtained by translating each coordinate of $(\mathbf{y}')^*$ by the same (arbitrary) amount.*

Proof of Proposition 5. Proposition 5 can be proved inductively in the same fashion as Proposition 2. We only prove the induction step below.

We first obtain the complementary slackness condition of `JPA-concave`. Let $\mathbf{y} \in \mathbb{R}^m$ be the dual variable of constraints (48), $\mathbf{z} \in \mathbb{R}^{m^2}$ be the dual variables of constrains (49), and $g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z})$ be the (full) Lagrangian dual function. Let $r_{j'k}(\mu_{j'k}) \triangleq -\infty$ for $\mu_{j'k} > \mu_{j'k}(0)$, $\forall j', k$. We simplify $g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z})$ as follows.

$$\begin{aligned}
&g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z}) \\
&= \max_{\hat{\mathbf{x}} \geq \mathbf{0}} \sum_{i \in V_S} \sum_{j' \in V_D} \sum_{k \in V_S} w_{ij'} \hat{x}_{ij'k} + \sum_{j' \in V_D} \sum_{k \in V_S} r_{j'k} \left(\sum_{i \in \partial(j')} \hat{x}_{ij'k} \right) \\
&\qquad + \sum_i y_i \left(\sum_{j' \in \mathcal{N}(i)} \sum_{k \in V_S} \hat{x}_{ij'k} - \sum_{j' \in V_D} \sum_{l \in \mathcal{N}(j')} \hat{x}_{lj'i} \right) + \sum_{j' \in V_D} \sum_{k \in V_S} z_{j'k} \left(\mu_{j'k}(0) - \sum_{i \in \partial(j')} \hat{x}_{ij'k} \right) \\
&= \sum_{j' \in V_D} \sum_{k \in V_S} \max_{\hat{\mathbf{x}}_{j'k} \geq \mathbf{0}} \left(\sum_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k - z_{j'k} \right) \hat{x}_{ij'k} + r_{j'k} \left(\sum_{i \in \partial(j')} \hat{x}_{ij'k} \right) \right) \\
&\qquad + \sum_{j' \in V_D} \sum_{k \in V_S} z_{j'k} \mu_{j'k}(0) \\
&= \sum_{j' \in V_D} \sum_{k \in V_S} \sup_{\mu_{j'k} \geq 0} \max_{\{\mathbf{1}^{\mathrm{T}} \hat{\mathbf{x}}_{j'k} = \mu_{j'k}, \hat{\mathbf{x}}_{j'k} \geq \mathbf{0}\}} \left(\sum_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k - z_{j'k} \right) \hat{x}_{ij'k} + r_{j'k} \left(\mu_{j'k} \right) \right) \\
&\qquad + \sum_{j' \in V_D} \sum_{k \in V_S} z_{j'k} \mu_{j'k}(0) \\
&= \sum_{j' \in V_D} \sum_{k \in V_S} \left(\max_{0 \leq \mu_{j'k} \leq \mu_{j'k}(0)} \left(r_{j'k} \left(\mu_{j'k} \right) + \mu_{j'k} \max_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k - z_{j'k} \right) \right) + \mu_{j'k}(0) z_{j'k} \right) .
\end{aligned}$$

Let $\mu^*_{j'k}$ be the value that achieves the outer maximum at $(\mathbf{y}, \mathbf{z})$. For partially satisfied demand (j', k), we have $\mu_{j'k} \in (0, \mu_{j'k}(0))$. The first order condition implies

$$r'_{j'k}(\mu^*_{j'k}) + \max_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k \right) - z_{j'k} = 0 \,. \tag{107}$$

Furthermore, suppose $(\mathbf{y}^*, \mathbf{z}^*)$ is a minimizer of $g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z})$ subject to $\mathbf{z} \geq 0$. Since

$$\frac{\partial g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z})}{z_{j'k}} = -\mu^*_{j'k} + \mu_{j'k}(0) > 0 \,,$$

optimality condition implies that $z^*_{j'k} = 0$.

Now we focus back on the induction step. Since $\mathbf{y}^*$ is an optimal solution to the partial dual (51), there is some $\mathbf{z}^*$ such that $(\mathbf{y}^*, \mathbf{z}^*)$ is an optimal solution to the full dual problem $\mathrm{minimize}_{\mathbf{y}, \mathbf{z} \geq \mathbf{0}}\, g_{\mathrm{JPA}}(\mathbf{y}, \mathbf{z})$. Let I be the set of nodes at which the value of $\mathbf{y}^*$ has been proved to be uniquely determined. Our induction base is $I = \{i_1\}$ since $y^*_{i_1} = 0$. As long as $I \subset V_S$, we can add a node to I as follows. Invoke Condition 4 with subset I. Then there exists some $i \in I$, $j' \in V_D$ and $k \in V_S \backslash I$ such that (i) location i is an optimal pickup location for partially satisfied demand type (j', k), or (ii) location k is an optimal pickup location for partially satisfied demand type (j', i). Consider case (i). Since demand type (j', k) is partially satisfied and i is an optimal pickup location for this demand type, using the first order condition (107) we have

$$w_{ij'} + y_i^* - y_k^* - z_{j'k}^* + r'_{j'k}(\mu_{j'k}^*) = 0\,.$$

Furthermore, it follows from complementary slackness derived above that $z_{j'k}^* = 0$. Combining we conclude that

$$y_k^* = w_{ij'} + y_i^* + r'_{j'k}(\mu_{j'k}^*)\,.$$

Since $i \in I$ we already knew that y_i^* is uniquely determined, and we can now deduce that y_k^* is also uniquely determined. As a result, we can add node k to I. Case (ii) can be handled very similarly, and again we are able to add node k to I. Induction then completes the proof that all coordinates of y^* are uniquely determined. □

I.3 Finite Performance Upper Bound for any JPA Policy: Proof of Proposition 3

Similar to the proof of Proposition 3, we establish the result by first proving two key lemmas. The first lemma shows that the expected payoff cannot exceed the value of the finite horizon fluid problem.

Lemma 9. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}(0)$, the expected payoff generated by any JPA policy π is upper bounded by the value of the finite horizon fluid problem:*

$$\text{maximize}_{\hat{\mathbf{x}}}\ (\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}} + \sum_{j'\in V_D, k\in V_S} r_{j'k}\left(\mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}_{j'k}\right) \tag{108}$$

$$\text{s.t. } |\mathbf{1}_S^{\mathrm{T}}\mathbf{R}\hat{\mathbf{x}}| \le \frac{K}{T} \qquad \forall\, S \subset V_S\,, \tag{109}$$

$$0 \le \mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}_{j'k} \le \mu_{j'k}(0),\ \hat{\mathbf{x}}_{j'k} \ge 0 \qquad \forall j' \in V_D\,, k \in V_S\,. \tag{110}$$

Here $\mathbf{w}_{ij'k} \triangleq w_{ij'}$. The above problem can be obtained from `JPA-concave` *by replacing flow balance constraint* (48) *with "approximate" flow balance constraint* (109).

Proof. Let π be any feasible JPA policy. Fix initial state $\mathbf{q}(0)$. It follows from Condition 3 that the demand arrival $\mathbf{a}$ given any chosen demand arrival rate $\boldsymbol{\mu}$ (or, equivalently, price $\mathbf{p}$) has finite support. We denote the support as $\mathcal{A}$. Denote $\rho_{\mathbf{a},\boldsymbol{\mu}} \triangleq \mathbb{P}(\mathbf{a}(1) = \mathbf{a}|\boldsymbol{\mu}(1) = \boldsymbol{\mu})$. Denote the probability distribution of the demand arrival rate (as a result of pricing decision) $\boldsymbol{\mu}$ under policy π during the t-th period as $\sigma_t^\pi(\cdot)$. (Note that we fixed the initial state, hence $\sigma_t^\pi(\cdot)$ only depend on π and t.) Denote $\mathcal{V} \triangleq \Pi_{j'\in V_D, k\in V_S}[0, \mu_{j'k}(0)]$. We decompose the time-average of payoff collected in the first T time slots as:

$$\begin{aligned}&\bar{v}_{\mathrm{JPA}}^{\pi}(T)\\ =\ &\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}(t)\circ\mathbf{x}(t)) + \sum_{j',k} p_{j'k}(\mu_{j'k}(t))\left(a_{j'k}(t)\sum_{i\in\mathcal{N}(j')} x_{ij'k}(t)\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathcal{V}}\mathbb{E}\left[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}(t)\circ\mathbf{x}(t))+\sum_{j',k}p_{j'k}(\mu_{j'k})\left(a_{j'k}(t)\sum_{i\in\mathcal{N}(j')}x_{ij'k}(t)\right)\middle|\boldsymbol{\mu}(t)=\boldsymbol{\mu}\right]d\sigma_t^{\pi}(\boldsymbol{\mu})\\
&\le \frac{1}{T}\sum_{t=1}^{T}\int_{\mathcal{V}}\mathbb{E}\left[\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}(t)\circ\mathbf{x}(t))+\sum_{j',k}p_{j'k}(\mu_{j'k})\cdot a_{j'k}(t)\middle|\boldsymbol{\mu}(t)=\boldsymbol{\mu}\right]d\sigma_t^{\pi}(\boldsymbol{\mu})
\end{aligned}
$$

where we used the basic properties of conditional expectation and the fact that $\sum_{i\in\mathcal{N}(j')}x_{ij'k}\le 1$. Let $\sigma^{\pi}(\cdot)\triangleq\frac{1}{T}\sum_{t=1}^{T}\sigma_t^{\pi}(\cdot)$. Denote $\mathbf{x}^{\mathbf{a},\boldsymbol{\mu}}(t)\triangleq\mathbb{E}[\mathbf{x}(t)|\mathbf{a}(t)=\mathbf{a},\boldsymbol{\mu}(t)=\boldsymbol{\mu}]$. Rearranging the terms, we have

$$
\begin{aligned}
&\bar{v}^{\pi}_{\mathrm{JPA}}(T)\\
&=\frac{1}{T}\sum_{t=1}^{T}\int_{\mathcal{V}}\left(\sum_{\mathbf{a}\in\mathcal{A}}\rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbb{E}[\mathbf{x}(t)|\mathbf{a}(t)=\mathbf{a},\boldsymbol{\mu}(t)=\boldsymbol{\mu}])+\sum_{j',k}r_{j'k}(\mu_{j'k})\right)d\sigma_t^{\pi}(\boldsymbol{\mu})\\
&\le\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\mathbf{a}\in\mathcal{A}}\int_{\mathcal{V}}\rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a},\boldsymbol{\mu}}(t))d\sigma_t^{\pi}(\boldsymbol{\mu})\right)+\sum_{j',k}r_{j'k}\left(\int_{\mathcal{V}}\mathbf{e}_{j'k}^{\mathrm{T}}\boldsymbol{\mu}d\sigma^{\pi}(\boldsymbol{\mu})\right),
\end{aligned}
$$

where the last inequality uses Jensen's inequality and the concavity of revenue function $r_{j'k}(\cdot)$'s. Similarly, for the time-average of the change of queue length we have:

$$
\frac{1}{T}\mathbb{E}[\mathbf{q}(T)-\mathbf{q}(0)]=\frac{1}{T}\sum_{t=1}^{T}\int_{\mathcal{V}}\sum_{\mathbf{a}\in\mathcal{A}_{\boldsymbol{\mu}}}\rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{R}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a},\boldsymbol{\mu}}(t))d\sigma_t^{\pi}(\boldsymbol{\mu})\,.
$$

Because there are only K supply units circulating in the system, the net outflow from any subset of supply locations $A\subset V_S$ should not exceed K. As a result, $\bar{v}^{\pi}(T)$ is upper bounded by the following optimization problem:

$$
\max_{\mathbf{x}^{\mathbf{a},\mathbf{p}}(t),\sigma_t^{\pi}(\cdot)}\ \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\mathbf{a}\in\mathcal{A}}\int_{\mathcal{V}}\rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{w}^{\mathrm{T}}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a},\boldsymbol{\mu}}(t))d\sigma_t^{\pi}(\boldsymbol{\mu})\right)+\sum_{j',k}r_{j'k}\left(\int_{\mathcal{V}}\mathbf{e}_{j'k}^{\mathrm{T}}\boldsymbol{\mu}d\sigma^{\pi}(\boldsymbol{\mu})\right) \tag{111}
$$

$$
\text{s.t. }\frac{1}{T}\sum_{t=1}^{T}\int_{\mathcal{V}}\sum_{\mathbf{a}\in\mathcal{A}_{\boldsymbol{\mu}}}\rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{1}_S^{\mathrm{T}}\mathbf{R}(\tilde{\mathbf{a}}\circ\mathbf{x}^{\mathbf{a},\boldsymbol{\mu}}(t))d\sigma_t^{\pi}(\boldsymbol{\mu})\le\frac{K}{T}\qquad\forall S\subset V_S\,, \tag{112}
$$

$$
\mathbf{x}_{j'k}^{\mathbf{a},\boldsymbol{\mu}}(t)\in\mathrm{conv}(\mathcal{X}_{j'k})\qquad\forall j'\in V_D,\ k\in V_S,\ \mathbf{a},\ t\,.
$$

Denote $\bar{\mu}_{j'k}\triangleq\int_{\mathcal{V}}\mathbf{e}_{j'k}^{\mathrm{T}}\boldsymbol{\mu}d\sigma^{\pi}(\boldsymbol{\mu})$. Take the partial dual w.r.t. constraints (112), we have the the following Lagrange dual function:

$$
\begin{aligned}
\tilde{g}_{\mathrm{JPA}}(\mathbf{y})&=\frac{1}{T}\sum_{t=1}^{T}\sup_{\sigma_t^{\pi}(\cdot)}\int_{\mathcal{V}}\sum_{\mathbf{a}\in\mathcal{A}}\rho_{\mathbf{a},\boldsymbol{\mu}}\sum_{j',k}\max_{\mathbf{x}_{j'k}^{\mathbf{a},\boldsymbol{\mu}}\in\mathrm{conv}(\mathcal{X}_{j'k})}\Big(a_{j'k}\mathbf{w}_{j'k}^{\mathrm{T}}\mathbf{x}_{j'k}^{\mathbf{a},\boldsymbol{\mu}}(t)+r_{j'k}\left(\bar{\mu}_{j'k}\right)\\
&\qquad+\sum_{S\subset V_S}y_S\left(a_{j'k}\mathbf{1}_S^{\mathrm{T}}\mathbf{R}_{j'k}\mathbf{x}_{j'k}^{\mathbf{a},\boldsymbol{\mu}}(t)\right)\Big)\,d\sigma_t^{\pi}(\boldsymbol{\mu})-\sum_{S\subset V_S}y_S\frac{K}{T}\\
&=\frac{1}{T}\sum_{t=1}^{T}\sup_{\sigma_t^{\pi}(\cdot)}\left(\int_{\mathcal{V}}\sum_{\mathbf{a}\in\mathcal{A}}\rho_{\mathbf{a},\boldsymbol{\mu}}\sum_{j',k}a_{j'k}\max_{i\in\mathcal{N}(j')}\left(w_{ij'}+\sum_{S\subset V_S}y_S(\mathbb{I}\{i\in S\}-\mathbb{I}\{k\in S\})\right)d\sigma_t^{\pi}(\boldsymbol{\mu})\right.
\end{aligned}
$$

$$+\sum_{j',k} r_{j'k}(\bar{\mu}_{j'k})\Bigg) - \sum_{S\subset V_S} y_S \frac{K}{T}$$

$$= \frac{1}{T}\sum_{t=1}^{T} \sup_{\sigma_t^\pi(\cdot)} \Bigg(\int_{\mathcal{V}} \sum_{j',k} \mu_{j'k} \max_{i\in\mathcal{N}(j')} \Bigg(w_{ij'} + \sum_{S\subset V_S} y_S(\mathbb{I}\{i\in S\} - \mathbb{I}\{k\in S\}) \Bigg) d\sigma_t^\pi(\boldsymbol{\mu})$$

$$+\sum_{j',k} r_{j'k}(\bar{\mu}_{j'k})\Bigg) - \sum_{S\subset V_S} y_S \frac{K}{T}$$

$$= \sum_{j',k} \max_{0\le\bar{\mu}_{j'k}\le\mu_{j'k}(0)} \Bigg(\bar{\mu}_{j'k} \max_{i\in\mathcal{N}(j')} \Bigg(w_{ij'} + \sum_{S\subset V_S} y_S(\mathbb{I}\{i\in S\} - \mathbb{I}\{k\in S\}) \Bigg) + \sum_{j',k} r_{j'k}(\bar{\mu}_{j'k}) \Bigg)$$

$$- \sum_{S\subset V_S} y_S \frac{K}{T}\,.$$

Here the second last equality holds because $\boldsymbol{\mu} = \sum_{\mathbf{a}\in\mathcal{A}} \rho_{\mathbf{a},\boldsymbol{\mu}}\mathbf{a}$. Note that $\tilde{g}_{\text{JPA}}(\mathbf{y})$ equals to the partial dual function of the finite horizon fluid problem defined in the Lemma, denoted by $g_{\text{JPA}}(\mathbf{y})$. Using the strong duality of linear programs and infinite linear programs (see, e.g., Rockafellar 2015), we have that the expected payoff generated by any policy π is upper bounded by the finite horizon fluid problem. □

Using the exact same proof, we can show that Lemma 6 still hold for the finite horizon fluid problem (108). We state the result below and skip the proof.

Lemma 10. *Any feasible solution $\mathbf{x}^F$ of the finite horizon fluid problem satisfying* (109) *and* (110) *can be decomposed as*

$$\hat{\mathbf{x}}^{\text{F}} = \hat{\mathbf{x}}^{\text{S}} + \hat{\mathbf{x}}^{\text{DAG}}\,,$$

where $\hat{\mathbf{x}}^{\text{S}}$ is a feasible solution for the static fluid problem satisfying (48) *and* (49)*, and $\hat{\mathbf{x}}^{\text{DAG}}$ is a directed acyclic assignment satisfying* (109) *and* (49)*.*

Using this supporting lemma, we now establish the second key lemma which shows that the value of the finite horizon (JPA) fluid problem cannot be much larger than the value of the static (JPA) fluid problem.

Lemma 11. *Let $(W_{\text{JPA}}^{\text{OPT}})_T$ be the value of the finite horizon (JPA) fluid problem. This value is upper bounded in terms of the value $W_{\text{JPA}}^{\text{OPT}}$ of the static fluid problem as*

$$(W_{\text{JPA}}^{\text{OPT}})_T \le W_{\text{JPA}}^{\text{OPT}} + \frac{(m-1)(w_{\max}+\bar{p})K}{T}\,. \tag{113}$$

Here $\bar{p}$ is the upper bound for the willingness-to-pay defined in Condition 3.

Proof. We appeal to the decomposition from Lemma 10 to decompose any feasible solution $\hat{\mathbf{x}}^{\text{F}}$ to the finite horizon fluid problem as

$$\hat{\mathbf{x}}^{\text{F}} = \hat{\mathbf{x}}^{\text{S}} + \hat{\mathbf{x}}^{\text{DAG}}\,,$$

where $\hat{\mathbf{x}}^{\mathrm{F}}$ is feasible for the static (JPA) fluid problem and $\hat{\mathbf{x}}^{\mathrm{DAG}}$ is a directed acyclic flow that is feasible for the finite horizon (JPA) fluid problem, i.e., satisfying (109) and (49). Hence, the objective (108) of the finite horizon (JPA) fluid problem can be written as

$$\begin{aligned}
&(\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{F}} + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{F}}_{ij'k}\right) \\
&= (\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{S}} + (\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{DAG}} + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \left(\hat{x}^{\mathrm{S}}_{ij'k} + \hat{x}^{\mathrm{DAG}}_{ij'k}\right)\right) \\
&\le (\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{S}} + (\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{DAG}} + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{S}}_{ij'k}\right) + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{DAG}}_{ij'k}\right) .
\end{aligned}$$

Here the inequality follows from the subadditivity of non-negative concave functions. By definition of $W^{\mathrm{OPT}}_{\mathrm{JPA}}$ we know that

$$(\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{S}} + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{S}}_{ij'k}\right) \le W^{\mathrm{OPT}}_{\mathrm{JPA}} .$$

We will now show that $(\mathbf{w})^{\mathrm{T}}\mathbf{x}^{\mathrm{DAG}} + \sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{DAG}}_{ij'k}\right) \le \frac{(m-1)(w_{\max}+\bar{p})K}{T}$. The lemma will follow, since this will imply an upper bound of $W^{\mathrm{OPT}}_{\mathrm{JPA}} + \frac{(m-1)(w_{\max}+\bar{p})K}{T}$ on the objective for any $\hat{\mathbf{x}}^{\mathrm{F}}$ satisfying (109) and (49).

Consider $\hat{\mathbf{x}}^{\mathrm{DAG}}$. Since it is a directed acyclic assignment, there is an ordering $(k_1, k_2, \ldots, k_m)$ of the nodes in V_S such that all assignments move supply from an earlier nodes to a later node in this ordering. More precisely, it holds that

$$x^{\mathrm{DAG}}_{k_l,j',k_r} = 0 \qquad \forall\, m \ge l > r \ge 1 \,, j' \in \mathcal{N}(k_l) \,. \tag{114}$$

Now consider the subsets $A_\ell \triangleq \{k_1, k_2, \ldots, k_\ell\} \subset V_S$ for $\ell = 1, 2, \ldots, m-1$. Note that from (114), $\hat{\mathbf{x}}^{\mathrm{DAG}}$ does not move any supply from $V_S \backslash A_\ell$ to A_ℓ. Hence we have

$$\mathbf{1}^{\mathrm{T}}_{A_\ell}\mathbf{R}\hat{\mathbf{x}}^{\mathrm{DAG}} \;=\; \sum_{i\in A_\ell, k\in V_S\backslash A_\ell, j'\in\mathcal{N}(i)} \hat{x}^{\mathrm{DAG}}_{ij'k} \;\le\; K/T \quad \forall\, l = 1, 2, \ldots, m-1 \,, \tag{115}$$

where we made use of (109) to obtain the upper bound. Further, note that for each $\hat{x}^{\mathrm{DAG}}_{k_l,j',k_r}$ with $l < r$, the term $\hat{x}^{\mathrm{DAG}}_{k_l j' k_r}$ is part of the above sum for $\ell = l$. It follows that

$$\mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{DAG}} \;=\; \sum_{1\le l<r\le m, j'\in\mathcal{N}(k_l)} \hat{x}^{\mathrm{DAG}}_{k_l j' k_r} \;\le\; \sum_{1\le\ell<m}\; \sum_{i\in A_\ell, k\in V_S\backslash A_\ell, j'\in\mathcal{N}(i)} \hat{x}^{\mathrm{DAG}}_{ij'k} \le (m-1)K/T \,, \tag{116}$$

using (115) in the second step. Now we deduce that

$$\begin{aligned}
&(\mathbf{w})^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{DAG}} \le \left(\max w_{ij'}\right) \mathbf{1}^{\mathrm{T}}\hat{\mathbf{x}}^{\mathrm{DAG}} \le \left(\max w_{ij'}\right) K(m-1)/T \,, \\
&\sum_{j',k} r_{j'k}\left(\sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{DAG}}_{ij'k}\right) \le \sum_{j',k} \bar{p} \sum_{i\in\mathcal{N}(j')} \hat{x}^{\mathrm{DAG}}_{ij'k} \le \bar{p}K(m-1)/T \,.
\end{aligned}$$

This completes the proof. □

Proof of Proposition 3. The proposition follows immediately from Lemmas 9 and 11. □

I.4 Lower Bound for the Partial Dual Function

Recall the definition of $\mathbf{y}^*$ in (33). We show in the following Lemma that if Condition 4 holds, we can lower bound the optimality gap $g(\mathbf{y}) - g(\mathbf{y}^*)$ by a multiple of quadratic term $||\mathbf{y}-\mathbf{y}^*||_2^2$ when $\mathbf{y}$ is close to $\mathbf{y}^*$, and by a multiple of norm $||\mathbf{y}-\mathbf{y}^*||_2$ when $\mathbf{y}$ is far from $\mathbf{y}^*$.

Lemma 12. *If Condition 4 holds, then there exists $\beta', \tau > 0$ that depends on model primitives $(\mathbf{w}, \boldsymbol{\mu}, G)$ such that for or any $\mathbf{y} \in \left\{\nabla\Phi^*(\bar{\mathbf{q}}) : \bar{\mathbf{q}} \in \Delta^m, \bar{\mathbf{q}} > \mathbf{0}\right\}$, we have*

$$g(\mathbf{y}) - g(\mathbf{y}^*) \geq \begin{cases} \beta'||\mathbf{y}-\mathbf{y}^*||_2^2 & \text{for } \mathbf{y} : ||\mathbf{y}-\mathbf{y}^*||_2 < \tau\,, \\ \beta'||\mathbf{y}-\mathbf{y}^*||_2 & \text{for } \mathbf{y} : ||\mathbf{y}-\mathbf{y}^*||_2 \geq \tau\,. \end{cases} \tag{117}$$

We prove Lemma 12 in this subsection.

We first prove the following *locally quadratic* condition for partial dual function $g_{\mathrm{JPA}}(\mathbf{y})$.

Lemma 13. *Suppose Condition 3 and 4 holds and denote the corresponding maximizing $\boldsymbol{\mu}$ as $\boldsymbol{\mu}^*$, the unique minimizer with $y_{i_1} = 0$ as $\mathbf{y}^*$. Then there exists constant $\tau(\boldsymbol{\mu}^*) > 0$ that only depends on $\boldsymbol{\mu}^*$, $L(\boldsymbol{\mu}(\cdot)) < \infty$ that only depends on demand functions $\boldsymbol{\mu}(\cdot)$ such that*

$$g_{\mathrm{JPA}}(\mathbf{y}^* + \mathbf{v}) - g_{\mathrm{JPA}}(\mathbf{y}^*) \geq \frac{1}{H(\boldsymbol{\mu}(\cdot))}||\mathbf{v}||_2^2\,, \quad \forall \mathbf{v} \in \left\{\mathbf{v} : \mathbf{1}^{\mathrm{T}}\mathbf{v} = 0,\, ||\mathbf{v}||_2 \leq \tau(\boldsymbol{\mu}^*)\right\}\,. \tag{118}$$

Proof. To facilitate the proof, we introduce the following auxiliary notations ($\forall j' \in V_D, k \in V_S$):

$$z_{j'k}(\mathbf{y}) \triangleq \max_{i \in \mathcal{N}(j')} \left(w_{ij'} + y_i - y_k\right)\,, \qquad z^*_{j'k} \triangleq \max_{i \in \mathcal{N}(j')} \left(w_{ij'} + y^*_i - y^*_k\right)\,,$$

$$\tilde{r}_{j'k}(\mu) \triangleq \begin{cases} r_{j'k}(\mu) & \text{for } \mu \in [0, \mu_{j'k}(0)] \\ -\infty & \text{otherwise} \end{cases}\,.$$

Now we can rewrite the partial dual function as

$$g_{\mathrm{JPA}}(\mathbf{y}) = \sum_{j' \in V_D} \sum_{k \in V_S} (-\tilde{r}_{j'k})^*(z_{j'k}(\mathbf{y}))\,,$$

where $(-\tilde{r}_{j'k})^*(\cdot)$ is the Fenchel conjugate of function $-\tilde{r}_{j'k}(\cdot)$.

We first establish the local strong convexity property of functions $(-\tilde{r}_{j'k})^*(\cdot)$. Using Condition 3, we know that $r''_{j'k}(\cdot)$ is continuous on bounded interval $[0, \mu_{j'k}(0)]$, hence its range is also bounded. Mathematically, there exists $H < \infty$ such that for any j', k, $r''_{j'k}(\mu) \geq -H$. As a result, for any j', k $r_{j'k}$ is strongly smooth on $[0, \mu_{j'k}(0)]$, i.e. $\forall \mu_0, \mu_1 \in [0, \mu_{j'k}(0)]$,

$$r_{j'k}(\mu_1) - r_{j'k}(\mu_0) \geq r'_{j'k}(\mu_0)(\mu_1 - \mu_0) - \frac{H}{2}(\mu_1 - \mu_0)^2\,.$$

Using the duality between strong smoothness and strong convexity (see, e.g. Kakade et al. 2009), we have that $(-\tilde{r}_{j'k})^*(\cdot)$ is $\frac{1}{H}$-strongly convex on $[-r'_{j'k}(0), -r'_{j'k}(\mu_{j'k}(0))]$, i.e. $\forall z_0, z_1 \in [-r'_{j'k}(0), -r'_{j'k}(\mu_{j'k}(0))]$,

$$(-r_{j'k})^*(z_1) - (-r_{j'k})^*(z_0) \geq ((-r_{j'k})^*)'(z_0)(z_1 - z_0) + \frac{1}{2H}(z_1 - z_0)^2\,. \tag{119}$$

Now we use the above inequality to lower bound $g(\mathbf{y}^* + \mathbf{v})$. Since $\mathbf{y}^*$ is the (unconstrained) minimizer of function g_{JPA}, by first order condition we have:

$$\mathbf{0} \in \partial g_{\text{JPA}}(\mathbf{y}^*) \,. \tag{120}$$

Note that functions $(-\tilde{r}_{j'k})^*(\cdot)$ are differentiable everywhere. Specifically, $((-\tilde{r}_{j'k})^*)'(z^*_{j'k}) = \mu^*_{j'k}$. Hence it follows from (120) that there exists $\mathbf{h}_{j'k} \in \partial z_{j'k}(\mathbf{y}^*)$ for each j', k such that

$$\mathbf{0} = \sum_{j' \in V_D} \sum_{k \in V_S} \mu^*_{j'k} \mathbf{h}_{j'k} \in \partial g_{\text{JPA}}(\mathbf{y}^*) \,. \tag{121}$$

Define:

$$\tau \triangleq \min_{j',k:\mu^*_{j'k} \in (0,\mu_{j'k}(0))} \left\{ \min \left\{ r'_{j'k}(0) - r'_{j'k}(\mu^*_{j'k}),\, r'_{j'k}(\mu^*_{j'k}) - r'_{j'k}(\mu_{j'k}(0)) \right\} \right\} \,.$$

Note that each argument is strictly positive due to strict concavity of $r_{j'k}$ and there are finite number of them, we have $\tau > 0$. Note that for j', k such that $\mu^*_{j'k} \in (0, \mu_{j'k}(0))$, for $z_{j'k}$ where $|z_{j'k} - z^*_{j'k}| \le \tau(\boldsymbol{\mu}^*)$, we have $z_{j'k}, z^*_{j'k} \in [-r'_{j'k}(0), -r'_{j'k}(\mu_{j'k}(0))]$. Combining (119) and (121), and use the convexity of $g_{\text{JPA}}(\cdot)$, we have the following inequality for $\mathbf{v}$'s such that $||\mathbf{z}(\mathbf{y}^*+\mathbf{v}) - \mathbf{z}^*||_\infty \le \tau$:

$$\begin{aligned} g_{\text{JPA}}(\mathbf{y}^* + \mathbf{v}) - g_{\text{JPA}}(\mathbf{y}^*) = & \sum_{j' \in V_D} \sum_{k \in V_S} \left((-\tilde{r}_{j'k})^*(z_{j'k}(\mathbf{y}^* + \mathbf{v})) - (-\tilde{r}_{j'k})^*(z_{j'k}(\mathbf{y}^*)) \right) \\ \ge & \sum_{j' \in V_D} \sum_{k \in V_S} \mu^*_{j'k} \mathbf{h}_{j'k} + \frac{1}{2H} \sum_{j',k:\mu^*_{j'k} \in (0,\mu_{j'k}(0))} (z_{j'k}(\mathbf{y}^* + \mathbf{v}) - z^*_{j'k})^2 \\ = & \frac{1}{2H} \sum_{j',k:\mu^*_{j'k} \in (0,\mu_{j'k}(0))} (z_{j'k}(\mathbf{y}^* + \mathbf{v}) - z^*_{j'k})^2 \,. \end{aligned} \tag{122}$$

As a final step, we analyze the sensitivity of $\mathbf{z}$ when $\mathbf{y}$ deviates from $\mathbf{y}^*$ and complete the proof. We first claim that for any $\mathbf{v}$ such that $\mathbf{1}^{\mathrm{T}}\mathbf{v} = 0$ and $||\mathbf{v}||_2 = \gamma > 0$, we can always divide V_S into two non-overlapping, non-empty subsets I and $V_S \backslash I$ such that

$$\forall i \in I,\, k \in V_S \backslash I\,,\, v_i - v_k > \frac{\gamma}{m^{3/2}} \,.$$

This claim can be proved by contradiction: if it doesn't hold, then we can show that $\max_{i \in V_S} v_i - \min_{i \in V_S} v_k \le (m-1)\frac{\gamma}{m^{3/2}}$. Since it is always true that $\min_{i \in V_S} v_k \le \frac{1}{m}\mathbf{1}^{\mathrm{T}}\mathbf{v} = 0$, we have

$$||\mathbf{v}||_2 \le \sqrt{m} \max_{i \in V_S} |v_i| \le \sqrt{m}(m-1)\frac{\gamma}{m^{3/2}} < \gamma \,.$$

Now consider such a subset I. Using Condition 4, one of the two following scenarios will happen:

1. There exists $i_a \in I, k_a \in V_S \backslash I$ and $j'_a \in \mathcal{N}(i_a)$ such that $\mu^*_{j'_a k_a} \in (0, \mu_{j'_a k_a}(0))$, and i_a is the optimal dispatch location. In this case

$$\begin{aligned} z_{j'_a k_a}(\mathbf{y}^* + \mathbf{v}) = & \max_{i \in \mathcal{N}(j'_a)} \left(w_{ij'_a} + y^*_i - y^*_{k_a} + v_i - v_{k_a} \right) \\ & \ge w_{i_a j'_a} + y^*_{i_a} - y^*_{k_a} + v_{i_a} - v_{k_a} \\ & = z_{j'_a k_a}(\mathbf{y}^*) + \frac{\gamma}{m^{3/2}} \,, \end{aligned} \tag{123}$$

2. There exists $i_a \in I, k_a \in V_S \backslash I$ and $j'_a \in \mathcal{N}(k_a)$ such that $\mu^*_{j'_a i_a} \in (0, \mu_{j'_a i_a}(0))$, and k_a is the

optimal dispatch location. In this case

$$\begin{aligned} z_{j'_a i_a}(\mathbf{y}^*+\mathbf{v}) &= \max_{k\in\mathcal{N}(j'_a)}\left(w_{kj'_a}+y^*_k-y^*_{i_a}+v_k-v_{i_a}\right) \\ &\le \max_{k\in\mathcal{N}(j'_a)}\left(w_{kj'_a}+y^*_k-y^*_{i_a}-\frac{\gamma}{m^{3/2}}\right) \\ &= z_{j'_a i_a}(\mathbf{y}^*)-\frac{\gamma}{m^{3/2}}\,. \end{aligned} \tag{124}$$

In both cases, for any j', k we have

$$|z_{j'k}(\mathbf{y}^*+\mathbf{v})-z_{j'k}(\mathbf{y}^*)|\le 2||\mathbf{v}||_\infty \le 2||\mathbf{v}||_2\,, \tag{125}$$

hence for $||\mathbf{v}||_2 \le \frac{1}{2}\tau$ we have $||\mathbf{z}(\mathbf{y}^*+\mathbf{v})-\mathbf{z}^*||_\infty \le \tau$. For such $\mathbf{v}$, plugging (123) and (124) into the last term of inequality (122), we have

$$g_{\mathrm{JPA}}(\mathbf{y}^*+\mathbf{v})-g_{\mathrm{JPA}}(\mathbf{y}^*)\ge \frac{1}{2m^3H}||\mathbf{v}||^2\,.$$

Let $\tau(\boldsymbol{\mu}^*)\triangleq\frac{1}{2}\tau$, $H(\boldsymbol{\mu}(\cdot))\triangleq 2m^3H$, we obtain the desired result. □

Using the above result and the fact that $g_{\mathrm{JPA}}(\mathbf{y})$ is convex, we can establish the following lemma. It provides a lower bound for $g_{\mathrm{JPA}}(\mathbf{y}+\mathbf{v})$ using $||\mathbf{v}||$ when $||\mathbf{v}||$ is large enough, where $\mathbf{1}^{\mathrm{T}}\mathbf{v}=0$.

Lemma 14. *Suppose the same conditions as in Lemma 13 hold. For any* $\mathbf{v}$ *that satisfies* $\mathbf{1}^{\mathrm{T}}\mathbf{v}=0$ *and* $||\mathbf{v}||_2 > s\tau(\boldsymbol{\mu}^*)$ *where* $s\in(0,1)$*, we have*

$$g_{\mathrm{JPA}}(\mathbf{y}^*+\mathbf{v})-g_{\mathrm{JPA}}(\mathbf{y}^*)\ge \frac{s\tau(\boldsymbol{\mu}^*)}{2H(\boldsymbol{\mu}(\cdot))}||\mathbf{v}||_2\,.$$

Proof. For notation simplicity, we denote $\tau_0\triangleq\tau(\boldsymbol{\mu}^*)$ and $H_0\triangleq H(\boldsymbol{\mu}(\cdot))$. Suppose $\mathbf{v}$ satisfy $\mathbf{1}^{\mathrm{T}}\mathbf{v}=0$ and $||\mathbf{v}||_2 > s\tau_0$. Note that $\left\|\frac{s\tau_0}{2||\mathbf{v}||_2}\mathbf{v}\right\|_2=\frac{s\tau_0}{2}<s\tau_0$. Using the convexity of $g_{\mathrm{JPA}}(\cdot)$, we have

$$g_{\mathrm{JPA}}\left(\mathbf{y}^*+\frac{s\tau_0}{2||\mathbf{v}||_2}\mathbf{v}\right)=g_{\mathrm{JPA}}\left(\left(1-\frac{s\tau_0}{2||\mathbf{v}||_2}\right)\mathbf{y}^*+\frac{s\tau_0}{2||\mathbf{v}||_2}(\mathbf{y}^*+\mathbf{v})\right) \tag{126}$$

$$\le\left(1-\frac{s\tau_0}{2||\mathbf{v}||_2}\right)g_{\mathrm{JPA}}\left(\mathbf{y}^*\right)+\frac{s\tau_0}{2||\mathbf{v}||_2}g_{\mathrm{JPA}}(\mathbf{y}^*+\mathbf{v})\,. \tag{127}$$

On the other hand, using Lemma 13, we have

$$g_{\mathrm{JPA}}\left(\mathbf{y}^*+\frac{s\tau_0}{2||\mathbf{v}||_2}\mathbf{v}\right)\ge g_{\mathrm{JPA}}(\mathbf{y}^*)+\frac{1}{H_0}\frac{s^2\tau_0^2}{4||\mathbf{v}||_2^2}||\mathbf{v}||_2^2= g_{\mathrm{JPA}}(\mathbf{y}^*)+\frac{s^2\tau_0^2}{4H_0}\,. \tag{128}$$

Combining (126) and (128) and rearrange the terms, we have

$$g_{\mathrm{JPA}}(\mathbf{y}^*+\mathbf{v})-g_{\mathrm{JPA}}(\mathbf{y}^*)\ge\frac{s\tau_0}{2H}||\mathbf{v}||_2\,.$$

This concludes the proof. □

Now we are ready to prove Lemma 12:

Proof of Lemma 12. We proved in Lemma 1 that (70) holds, i.e.

$$|\cos\langle\mathbf{1},\mathbf{y}-\mathbf{y}^*\rangle|=\frac{|\langle\mathbf{1},\mathbf{y}-\mathbf{y}^*\rangle|}{\sqrt{m}||\mathbf{y}-\mathbf{y}^*||_2}\le\frac{1}{\gamma}\qquad\text{for some }\gamma>1\,.$$

Using (71) and the first inequality of (72), we have

$$\left(1-\frac{1}{\gamma^2}\right)||\mathbf{y}-\mathbf{y}^*||_2^2\leq ||\mathbf{y}-\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2^2\leq ||\mathbf{y}-\mathbf{y}^*||_2^2\,.$$

For $\mathbf{y}$ such that $||\mathbf{y}-\mathbf{y}^*||\leq\tau$, since we also have $||\mathbf{y}-\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2\leq\tau$, we can apply Lemma 13:

$$g(\mathbf{y})-g(\mathbf{y}^*)\geq\frac{1}{H}||\mathbf{y}-\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2^2\geq\frac{1}{H}\left(1-\frac{1}{\gamma^2}\right)||\mathbf{y}-\mathbf{y}^*||_2^2\,.$$

For $\mathbf{y}$ such that $||\mathbf{y}-\mathbf{y}^*||\geq\tau$, since we have $||\mathbf{y}-\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2\geq\sqrt{1-\frac{1}{\gamma^2}}\tau$, we can apply Lemma 14:

$$g(\mathbf{y})-g(\mathbf{y}^*)\geq\sqrt{1-\frac{1}{\gamma^2}}\frac{\tau}{2H}||\mathbf{y}-\mathcal{P}_{\mathcal{Y}^*}(\mathbf{y})||_2\geq\left(1-\frac{1}{\gamma^2}\right)\frac{\tau}{2H}||\mathbf{y}-\mathbf{y}^*||_2\,.$$

Let $\beta'\triangleq\frac{1}{H}\left(1-\frac{1}{\gamma^2}\right)\cdot\min\left\{1,\frac{\tau}{2}\right\}$. This concludes the proof. □

I.5 Proof of Theorem 2

The proof of Theorem 2 is very similar to the proof of Theorem 1. We outline the proof below and only emphasize on the parts that are different

Lemma 15. *Fix $\epsilon\in(0,1)$. Suppose Condition 3 and 4 hold, and let β' be the positive constant defined in Lemma 12. Then there exists $K_0'=\Omega\left(\exp\left(\text{poly}(m,\bar{B},1/\beta',1/\epsilon,\tau)\right)\right)<\infty$ such that for all $K\geq K_0'$, under the MBP-JPA policy we have*

$$\mathbb{E}\left[L(\bar{\mathbf{q}}(t+1))-L(\bar{\mathbf{q}}(t))|\bar{\mathbf{q}}(t)\right]\leq-\frac{1}{2c\tilde{K}}\left(g\left(c\log\bar{\mathbf{q}}(t)\right)-g\left(c\log\mathbf{q}^*\right)\right)+\frac{m^2\bar{B}^2}{2\tilde{K}^{2-\epsilon}}\,. \tag{129}$$

Proof Sketch. The proof is almost exactly the same as the proof of Lemma 2, with the following minor changes. The upper bound for total number of demands requesting service from any supply location was B in Lemma 2, it is now replaced by $\bar{B}$ defined in Condition 3. The "slope" parameter β in Lemma 2 is replaced by β' defined in Lemma 12. To ensure the lower bound in Lemma 14 kicks in, we need to let

$$K\geq K_2\triangleq\exp\left(\frac{\tau}{w_{\max}\epsilon}+\frac{3m}{\epsilon}\right)\,,$$

so as to satisfy $||c\log(\tilde{K}^{-\epsilon})-c\log\mathbf{q}^*||_2>\tau$. All combined, for $K=\Omega\left(\exp\left(\text{poly}(m,\bar{B},1/\beta',1/\epsilon,\tau)\right)\right)$, the desired result (129) holds.

Lemma 16. *Suppose Condition 3 and 4 hold. Fix $\epsilon>0$. For K_0' defined in Lemma 15, for all $K\geq K_0'$, we have*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right)\right]-g\left(c\log\mathbf{q}^*\right)\right)\leq\frac{6w_{\max}m\tilde{K}}{T}+\frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}\,. \tag{130}$$

Proof Sketch. Proceed exactly the same as in Lemma 3, sum the both sides of (129), we obtain the desired result.

Proof sketch of Theorem 2. The key difference in the proof of Theorem 2 is using the dual optimality gap to bound LHS of (100): using Lemma 12, (100) is replaced by

$$\begin{aligned}
&\left(c\log\bar{\mathbf{q}}(t) - c\log\mathbf{q}^*\right)^{\mathrm{T}}\left(\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\right)\\
\leq&\|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2\cdot\|\bar{\mathbf{q}}(t) - \bar{\mathbf{q}}(t+1)\|_2\\
\leq&\|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2\cdot\frac{2m\bar{B}}{\tilde{K}}\\
\leq&\begin{cases}\frac{2m\bar{B}}{\tilde{K}}\frac{1}{\sqrt{\beta'}}\sqrt{g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)} & \text{for } \|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2\leq\tau\,,\\ \frac{2m\bar{B}}{\tilde{K}}\frac{1}{\beta'}\left(g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right) & \text{for } \|c\log\bar{\mathbf{q}}(t) - c\log\bar{\mathbf{q}}^*\|_2 > \tau\,.\end{cases}\\
\leq&\frac{2m\bar{B}}{\tilde{K}}\left(\frac{1}{\sqrt{\beta'}}\sqrt{g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)} + \frac{1}{\beta'}\left(g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right)\right)\,.
\end{aligned}$$

Hence (104) is replaced by

$$\begin{aligned}
&W_{\mathrm{JPA}}^{\mathrm{OPT}} - \bar{v}^{\text{MBP-JPA}}(T)\\
\leq&\frac{2c\tilde{K}m^2}{T} + \frac{3m\bar{B}}{\beta'}\left(\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}\right)\\
&+\frac{2m\bar{B}}{T}\frac{1}{\sqrt{\beta'}}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)}\right]\\
\leq&\frac{2c\tilde{K}m^2}{T} + \frac{3m\bar{B}}{\beta'}\left(\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}\right)\\
&+\frac{2m\bar{B}}{T}\frac{1}{\sqrt{\beta'}}\sum_{t=1}^{T}\sqrt{\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right]} \qquad \text{(Jensen's inequality)}\\
\leq&\frac{2c\tilde{K}m^2}{T} + \frac{3m\bar{B}}{\beta'}\left(\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}\right)\\
&+\frac{2m\bar{B}}{\sqrt{\beta'}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[g\left(c\log\bar{\mathbf{q}}(t)\right) - g\left(c\log\bar{\mathbf{q}}^*\right)\right]} \qquad \text{(Jensen's inequality)}\,.
\end{aligned} \tag{131}$$

Plug in Lemma 16, for $K \geq \max(K_0', 4Be^{3m+1})$ we have

$$\begin{aligned}
&W_{\mathrm{JPA}}^{\mathrm{OPT}} - \bar{v}^{\text{MBP-JPA}}(T)\\
\leq&\frac{2c\tilde{K}m^2}{T} + \frac{3m\bar{B}}{\beta'}\left(\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}\right) + \frac{2m\bar{B}}{\sqrt{\beta'}}\sqrt{\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}}\\
\leq&\,3w_{\max}m^2(1+9\bar{B}/\beta)\,\frac{K}{T} + \frac{(3w_{\max}m^3\bar{B}^3/\beta)}{K^{1-\epsilon}} + \frac{2m\bar{B}}{\sqrt{\beta'}}\sqrt{\frac{6w_{\max}m\tilde{K}}{T} + \frac{w_{\max}m^2\bar{B}^2}{\tilde{K}^{1-\epsilon}}}\,,
\end{aligned}$$

using $c = w_{\max}$ and $K \geq 4Be^{3m+1} > 2Bm \Rightarrow \tilde{K} = K + mB \in (K, 3K/2)$. Finally, plugging in the definition of Regret (which is based on Proposition 3) we lose another additive amount $(w_{\max} + \bar{p})(m-1)K/T$.

This establishes the theorem with

$K_1' \triangleq \max(K_0', 4Be^{3m+1}) = \exp\left(\mathrm{poly}(m, \bar{B}, 1/\beta', 1/\epsilon, \tau)\right)$
$M' \triangleq \max(4(w_{\max}+\bar{p})m^2(1+9\bar{B}/\beta'), 3w_{\max}m^3\bar{B}^3/\beta, 2\sqrt{6}\sqrt{w_{\max}}m^{3/2}\bar{B}/\sqrt{\beta'}, 2\sqrt{w_{\max}}m^2\bar{B}^2/\sqrt{\beta'})$

$$= \mathrm{poly}\left(w_{\max}, m, \bar{B}, \bar{p}, 1/\beta'\right) \; .$$