



**HAL**  
open science

# Survey of Dynamic Resource Constrained Reward Collection Problems: Unified Model and Analysis

Santiago R Balseiro, Omar Besbes, Dana Pizarro

► **To cite this version:**

Santiago R Balseiro, Omar Besbes, Dana Pizarro. Survey of Dynamic Resource Constrained Reward Collection Problems: Unified Model and Analysis. 2021. hal-03363684v1

**HAL Id: hal-03363684**

**<https://hal.science/hal-03363684v1>**

Preprint submitted on 4 Oct 2021 (v1), last revised 18 Apr 2023 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Survey of Dynamic Resource Constrained Reward Collection Problems: Unified Model and Analysis\*

Santiago R. Balseiro<sup>†</sup>      Omar Besbes<sup>‡</sup>      Dana Pizarro<sup>§</sup>

October 4, 2021

## Abstract

Dynamic resource allocation problems arise under a variety of settings and have been studied across disciplines such as Operations Research and Computer Science. The present paper introduces a unifying model for a very large class of dynamic optimization problems, that we call *dynamic resource constrained reward collection* (DRC<sup>2</sup>). We show that this class encompasses a variety of disparate and classical dynamic optimization problems such as dynamic pricing with capacity constraints, dynamic bidding with budgets, network revenue management, on-line matching, or order fulfillment, to name a few. Furthermore, we establish that the class of DRC<sup>2</sup> problems, while highly general, is amenable to analysis. In particular, we characterize the performance of the fluid certainty equivalent control heuristic for this class. Notably, this very general result recovers as corollaries some existing specialized results, generalizes other existing results by weakening the assumptions required, but also yields new results in specialized settings for which no such characterization was available. As such, the DRC<sup>2</sup> class isolates some common features of a broad class of problems, and offers a new object of analysis.

**Keywords:** dynamic optimization, resource allocation, certainty equivalent, model predictive control, online matching, dynamic pricing, dynamic bidding, network revenue management.

## 1 Introduction

Dynamic optimization problems with resources constraints arise across a variety of disparate applications. For example, retailers dynamically price products with inventory constraints, airlines and hotels engage in dynamic allocation of limited seats or rooms, advertisers bid in real-time to fulfill campaigns with limited budget. Due to the importance and centrality of these problems, various classes of dynamic optimization problems have received significant attention in industry but also across academic communities in Operations Research, Computer Science, and Economics.

---

\*The work of Dana Pizarro has benefited from the AI Interdisciplinary Institute ANITI, which is funded by the French “Investing for the Future – PIA3” program under the Grant agreement ANR-19-P3IA-0004.

<sup>†</sup>Columbia University, Graduate School of Business. Email: [srb2155@columbia.edu](mailto:srb2155@columbia.edu)

<sup>‡</sup>Columbia University, Graduate School of Business. Email: [ob2105@columbia.edu](mailto:ob2105@columbia.edu)

<sup>§</sup>Université Toulouse 1 Capitole, Toulouse School of Economics. Email: [dana.pizarro@tse-fr.eu](mailto:dana.pizarro@tse-fr.eu)

A significant focus of the literature has been on the development of efficient algorithms to optimize performance subject to capacity constraints.

While the literature on these problems is rich and extensive,<sup>1</sup> studies have focused on specific applications, or classes of applications. As such, arguments are specialized for specific settings and do not directly apply to other settings, typically requiring to re-develop, from scratch, analyses and proofs when faced with a new type of dynamic optimization problem with resource constraints. While, from a practical perspective, problems such as those mentioned above can appear very different, these problems do admit some common mathematical structure. In the present work, we survey the literature, elucidate such common structure, and demonstrate that the latter can be captured by a general model we propose. In turn, we derive important theoretical implications of such commonalities.

**A unified model.** Our first main contribution is the introduction and definition of a general class of problems: *dynamic resource constrained reward collection* (DRC “squared” or for short DRC<sup>2</sup>) problems, including problems with finite and continuum of actions. Notably, we show that this class admits as special cases a variety of problems studied separately in the literature. Broadly speaking, a DRC<sup>2</sup> problem is defined as follows. A decision-maker endowed with some resources faces a finite (discrete) time horizon. At each period, the decision-maker is presented with a stochastic opportunity (independent of other periods), and must select an action; the action leads to some stochastic resource consumption and reward collection. The goal of the decision-maker is to select a sequence of actions to maximize her total expected rewards subject to the resource constraints. We assume that the decision-maker knows the distribution of the various stochastic components, and, as such, this problem can be formulated as a discrete and finite-time dynamic program, with the state given by the vector of resources available.

The DRC<sup>2</sup> class of problems generalizes and brings under the same umbrella a host of classical problems studied separately. Figure 1 provides a conceptual illustration of the class. In particular, we show in §3 how the proposed class of DRC<sup>2</sup> problems encompass the following classical problems: *Network dynamic pricing* problems (see, e.g., Gallego and Van Ryzin 1997), *Dynamic bidding in repeated auctions with budgets* (see, e.g., Balseiro et al. 2015), *Network revenue management* problems (see, e.g., Talluri and Van Ryzin 2006), *Choice-based revenue management* problems (see, e.g., Talluri and Van Ryzin 2004), *Order fulfillment* problems (see, e.g., Acimovic and Farias 2019), and *Online matching* problems (see, e.g., Aggarwal et al. 2011). For each of these problems, we explain how they map to a DRC<sup>2</sup> problem.

---

<sup>1</sup>We discuss the literature in detail when we discuss our model and present our main results and associated corollaries.

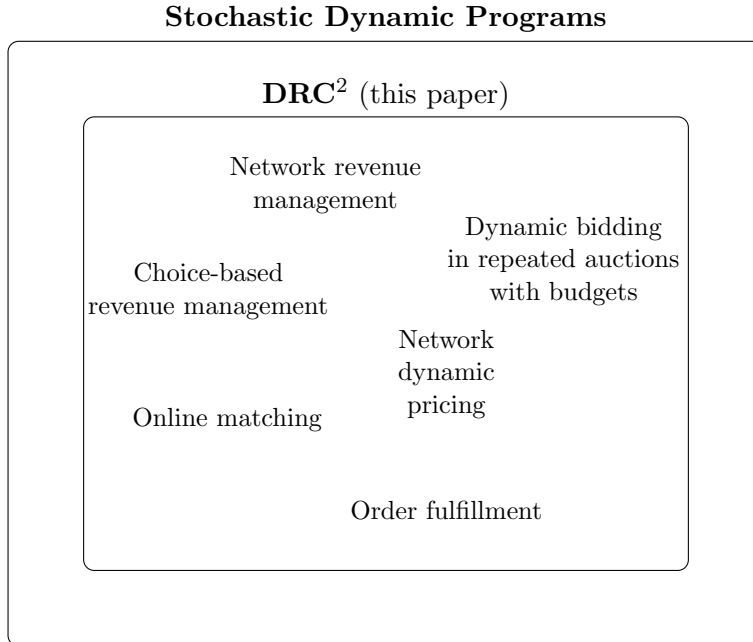


Figure 1: Definition of a new large class of dynamic decision problems (DRC<sup>2</sup>) that encompasses many known problem classes.

**A unified analysis of “fluid” certainty-equivalent control.** Although from a theoretical perspective, DRC<sup>2</sup> problems can be formulated through a dynamic program, one natural question is whether the DRC<sup>2</sup> formulation lends itself to analysis, beyond a generic analysis of a general dynamic program, that can be applied to all special cases, or whether problems should be specialized first to be able to derive properties of interest. We indeed demonstrate that the general DRC<sup>2</sup> formulation can lead to unified analysis through the study of a central heuristic in the stochastic dynamic optimization literature. Our second layer of contribution is in the analysis domain. In particular, we characterize the performance of a classical “fluid” certainty-equivalent control for the general DRC<sup>2</sup> class of problems.

In more detail, solving even a special case of a DRC<sup>2</sup> problem to optimality is typically impossible due to the curse of dimensionality; indeed the state space grows exponentially with the number of resources. This has brought forward the need for heuristics for such problems, and many such heuristics have been developed for subsets of the problems above. A notable heuristic for dynamic optimization problems is the so-called certainty-equivalent heuristic, which involves solving a deterministic problem in each period by using proxies for random quantities, implementing the prescribed decisions for that period, and repeating the process over time. Such certainty-equivalent heuristics have been shown to be near-optimal under some conditions in various special cases of DRC<sup>2</sup> problems. A notable example of a certainty-equivalent heuristic is the so-called “fluid” one,

in which the random quantities are replaced by their expectations. We will refer to this heuristic as CE. Such policies are sometimes also referred to as “re-solving” or “model predictive control” in the various related streams of literature. To analyze the performance of the heuristic, we measure the gap between the optimal performance and that of the CE heuristic, and characterize the dependence of this gap on the “scale” of the system.

More specifically, we establish two types of sufficient conditions for the CE heuristic to lead to a “small” performance gap. The first set of sufficient conditions are stated in terms of local strong-smoothness of the primal fluid problem as a function of the resource vector. These conditions, while simple to state, are in general harder to verify in many applications. The second, alternative, set of sufficient conditions are expressed in terms of the dual Lagrangian problem, which are easy to verify as they depend directly on the primitives of the problem. In particular, the analysis leads to a dichotomy between two fundamental cases: that when the set of actions is a continuum, and that when it is finite. For the former, the CE heuristic guarantees  $O(\log T)$  performance gap (where  $T$  is the length of the horizon), whereas for the latter, it guarantees  $O(1)$  gap, when the initial amount of resources are selected proportional to  $T$ .

In essence, the analysis establishes that the class of DRC<sup>2</sup> problems, under said sufficient conditions, is “easy” in that the CE heuristic is extremely effective. Intuitively, the CE heuristic enables the decision-maker to implement good decisions through the proxy problem while controlling very closely the path of the resource constraints.

In general, the primal problem solved in each step of CE heuristic needs to optimize over randomized controls. When the set of actions is a continuum, the resulting problems are over infinite-dimensional probability measures and, thus, challenging to solve. We provide simple conditions on the primitives under which the optimal controls are deterministic and strong duality holds (even when the underlying deterministic problem is non-convex). From a computational perspective, this leads to simpler implementations of the CE heuristic as, in many cases, the dual problem, which is always convex, can be alternatively solved. From a theoretical standpoint, we can leverage duality theory to provide simple conditions on the primitives that are simple to check and, at the same time, yield geometric insights into the structure of the problem.

We highlight here that the fact the CE heuristic is effective for specific DRC<sup>2</sup> instances is not new. The following are novel contributions of our work. First, the analysis as well as the types of ideas developed can be “lifted” and generalized at the DRC<sup>2</sup> level, and in turn, one may derive more general sufficient conditions for such performance, anchored in the raw primitives of the elements of a DRC<sup>2</sup> problem. Second, given such general sufficient conditions, the scope of problems for which such guarantees apply can be expanded without specialized arguments. In particular, after developing our theoretical results and the dichotomy between finite and continuum of actions, we

return to the classical problems in the literature and state the corollaries that one obtains from the general analysis of the CE heuristic. This allows to recover versions of various existing results but also to obtain such results under weaker conditions (see, e.g., the case of dynamic pricing (cf. §3.1,§6.1)), or to obtain altogether new results in the literature for the performance of the fluid CE heuristic (see, e.g., the case of dynamic bidding with budgets (cf. §3.2,§6.2) or the case of dynamic assortment optimization (cf. §3.4,§6.4)). Additionally, our results give performance bounds with better dependence on the number of resources and hold even when the underlying fluid problems in the CE heuristic are non-convex, as for example in the case of dynamic bidding with budgets. We discuss the related literature in detail when we discuss the various specialized problems.

Overall, this paper introduces a novel general formulation of dynamic optimization problems. We illustrate how this formulation lends itself to analysis through a unified analysis of the CE heuristic. As such, the DRC<sup>2</sup> class offers a “useful” and powerful intermediate class of problems between the specialized versions previously studied in the literature and a fully general dynamic program, and this work opens up the possibility of further generalizations of arguments developed for special cases of DRC<sup>2</sup> problems.

## 2 Model

We consider a dynamic decision-making problem with a finite time horizon  $T$ , over which a decision-maker collects rewards subject to resource constraints. We refer to this problem as the *Dynamic Resource Constrained Reward Collection* (DRC<sup>2</sup>) problem. There are  $L$  resources and the decision-maker is initially endowed with initial capacities  $C \in \mathbb{R}_+^L$  for the resources.

In each period  $t$ , an opportunity arises and each opportunity is characterized by a class  $\theta \in \Theta$ , where  $\Theta$  is a finite set of types.<sup>2</sup> Classes are drawn independently from a distribution  $p \in \Delta(\Theta)$ . Upon observing an opportunity, the decision-maker takes an action  $a \in \mathcal{A}$ , where  $\mathcal{A}$  is the set of feasible actions. Upon taking an action  $a$ , the decision-maker collects a reward that depends on the opportunity class  $\theta$ , the action  $a$ , and an idiosyncratic shock  $\epsilon$ . Shocks lie in a space  $\mathcal{E}$  and are drawn independently from a distribution  $f \in \Delta(\mathcal{E})$ .<sup>3</sup> Shocks are revealed to the decision-maker after an action is taken and are meant to capture exogenous factors that are idiosyncratic to the opportunity. We denote by  $r : \Theta \times \mathcal{A} \times \mathcal{E} \rightarrow \mathbb{R}$ , the reward function, where  $r(\theta, a, \epsilon)$  denotes the reward when the class is  $\theta$ , the action is  $a$ , and the shock is  $\epsilon$ . Taking an action consumes resources and we assume that the amount of resources consumed depends on the opportunity class  $\theta$  and the

---

<sup>2</sup>Here, we assume finiteness of the set of classes for expository purposes. We expect our results to extend to more general spaces.

<sup>3</sup>In the general model, to simplify notation, we make the probability distribution of  $\epsilon$  independent of the action and the class but any dependencies may be modeled by modifying reward and consumption functions, and in some applications, it might be convenient to consider some dependence.

value of the shock  $\epsilon$ . We denote by  $y : \Theta \times \mathcal{A} \times \mathcal{E} \rightarrow \mathbb{R}^L$ , the vector-valued resource consumption function. In particular  $y_l(\theta, a, \epsilon)$  represents the consumption of resource  $l$  if class  $\theta$  arrived, the decision maker chose an action  $a$ , and the shock was  $\epsilon$ .

To ensure that the problem is feasible, we assume there is a null action  $a_0$  in  $\mathcal{A}$  that consumes no resources and generates no reward. That is, for every opportunity class  $\theta$  and idiosyncratic shock  $\epsilon$ , we have  $r(\theta, a_0, \epsilon) = 0$  and  $y_l(\theta, a_0, \epsilon) = 0$  for every resource  $l$ .

We denote the history up to time  $t - 1$  as  $\mathcal{H}_{t-1} = \{\theta_s, a_s, \epsilon_s\}_{s=1}^{t-1}$ . We let  $\Pi$  denote the set of all non-anticipating policies, i.e., the set of policies such that the action at time  $t$ ,  $a_t$ , depends on the observed class of the opportunity in time  $t$  and the history up to (and including) time  $t - 1$ . That is, for a policy  $\pi$ ,  $a_t = a_t^\pi(\theta_t, \mathcal{H}_{t-1})$ .<sup>4</sup> The decision maker's objective is to choose a policy  $\pi \in \Pi$  that maximizes her expected rewards earned during the horizon. Taking into account that the consumption's constraints must hold almost surely, the stochastic optimization formulation of the decision-maker may be written as follows.

$$\begin{aligned}
 J^*(C, T) = & \sup_{\pi \in \Pi} \mathbb{E} \left( \sum_{t=1}^T r(\theta_t, a_t^\pi, \epsilon_t) \right) \\
 \text{s.t.} & \sum_{t=1}^T y_l(\theta_t, a_t^\pi, \epsilon_t) \leq C_l, \forall l \in [L] \text{ (a.s.)},
 \end{aligned} \tag{1}$$

where  $[L]$  denotes the set  $\{1, \dots, L\}$ .

Note that in general, this is a dynamic program with potentially a high number of dimensions and the curse of dimensionality precludes solving this problem to optimality. Given this, various heuristics can be considered and their performance can be assessed through the resulting optimality gap

$$J^*(C, T) - J^\pi(C, T),$$

where  $J^\pi(C, T)$  represents the expected reward obtained by the decision-maker if policy  $\pi$  is implemented. We refer to the expression above as the *reward loss* of the heuristic given by  $\pi$ .

In the following section, we survey part of the literature and review a set of problems that are particular cases of the proposed DRC<sup>2</sup> class of problems. For each problem, we show it can be modelled as a DRC<sup>2</sup> problem.

---

<sup>4</sup>To simplify notation we will write  $a_t^\pi$  to refer to the action taken at time  $t$  given the policy  $\pi$

### 3 Mapping Notable Problem Classes to a DRC<sup>2</sup> Problem

As it was mentioned above, many dynamic resource allocation problems can be modeled as a DRC<sup>2</sup> problem. In this section, we present some notable subclasses of problems and provide an explicit mapping to a DRC<sup>2</sup> problem. In Table 1, we present a high level overview of how various notable classes map to a DRC<sup>2</sup> problem, together with the subsections in which this mapping is established.

| Problem subclass                     | Elements of the DRC <sup>2</sup> problem |                                    |                                       |  |                                    |
|--------------------------------------|--|------------------------------------|---------------------------------------|--|------------------------------------|
|                                      | Arrival ( $\theta$ )                     | Action ( $a$ )                     | Idiosyncratic shock ( $\epsilon$ )    | Resource Constraint                      |                                    |
|                                      |  |                                    |                                       | LHS                                      | RHS                                |
| Network dynamic pricing<br>§3.1      | customer class                           | product prices                     | heterogeneity in customer values      | total resources consumed                 | initial inventory                  |
| Bidding in repeated auctions<br>§3.2 | impression's valuation                   | bid                                | exogenous auction uncertainty         | total payment                            | budget                             |
| Network revenue management<br>§3.3   | customer class                           | accept/reject                      | -                                     | total resources consumed                 | initial inventory                  |
| Choice-based NRM<br>§3.4             | customer class                           | offer set                          | heterogeneity in customer preferences | total resources consumed                 | initial inventory                  |
| Online matching<br>§3.5              | customer class                           | a resource                         | -                                     | total resources assigned                 | initial inventory                  |
| Order fulfillment<br>§3.6            | items requested                          | facilities used to fulfill request | -                                     | total items allocated from each facility | initial inventory in each facility |

Table 1: Comparison of some special subclasses of DRC<sup>2</sup> problems. The idiosyncratic shocks are meant to qualitatively capture the underlying exogenous factors associated with each opportunity as opposed to the distribution of the actual random variable  $\epsilon$ .

#### 3.1 Network Dynamic Pricing

In the Network Dynamic Pricing problem, a firm sells products to a sequence of customers over a finite horizon (see, e.g., Gallego and Van Ryzin [1997]). On the demand side, each buyer belongs to a class characterized by their valuation for the products, which influences her demand. On the firm side, at the beginning of each period, he observes the class of the customer who arrived and posts prices to maximize his expected revenue taking into account that each product consumes a subset of resources, whose inventories are finite and without replenishment.

The interest in dynamic pricing problems has grown during the last few decades. The design of near-optimal pricing policies that are easy to implement has been studied under several model



variants and heuristic policies are widely used in practice by firms. For example, [Kunnumkal and Topaloglu \[2010\]](#) and [Erdelyi and Topaloglu \[2011\]](#) consider a dynamic programming formulation. Both papers consider an airline network in which prices affect the probability of the arrival request. In the former, the authors propose a stochastic approximation algorithm for choosing prices dynamically and prove its convergence. In the latter, they develop two methods for making pricing decisions based on a decomposition of the original dynamic program. We refer the reader to the review papers and textbook of [Bitran and Caldentey \[2003\]](#), [Talluri and Van Ryzin \[2006\]](#), [Gallego et al. \[2019\]](#).

**Mapping to a DRC<sup>2</sup> problem.** In this setting, there is a set of  $N$  different products to sell during a finite, discrete horizon. An arrival class  $\theta \in \Theta$  represents a customer class or consumer segment, which will influence their demand. The set of actions  $\mathcal{A} = \mathbb{R}_+^N$  consists of the set of all feasible price vectors to post for the products.

For each customer class  $\theta$ , the idiosyncratic shock  $\epsilon$  captures heterogeneity in customer values. Then, given  $\theta$ , posted prices  $a \in \mathcal{A} = \mathbb{R}_+^N$  and a realization of the shock  $\epsilon$ , we denote by  $D(\theta, a, \epsilon) \in \mathbb{R}_+^N$  the induced vector of demand. The reward function captures the revenue of the firm and is given by  $r(\theta, a, \epsilon) = a^\top D(\theta, a, \epsilon)$ , and the consumption function is  $y(\theta, a, \epsilon) = A_\theta D(\theta, a, \epsilon)$ . In the latter expression,  $A_\theta \in \mathbb{R}^{L \times N}$  is a matrix where  $A_{l\theta}^n$  represents the units of resource  $l$  needed to serve a customer class  $\theta$  with a single unit of product  $n$ .

### 3.2 Dynamic Bidding in Repeated Auctions

A special case of a DRC<sup>2</sup> problem is the problem faced by a bidder participating in a sequence of repeated auctions to buy opportunities. The bidder has a budget constraint that limits his total expenditure over the horizon and aims to maximize his cumulative utility. This model is mainly motivated by internet advertising markets in which advertisers buy opportunities to display advertisements—an event referred to as an impression—via repeated auctions subject to budget constraints.

For instance, [Abhishek and Hosanagar \[2013\]](#) study this problem, where the goal is to compute optimal bids for multiple keywords in an advertiser’s portfolio. Motivated by ad exchanges, [Balseiro et al. \[2015\]](#) introduce a fluid mean-field equilibrium (FMFE) notion to study the strategic outcome of advertisers competing in repeated second-price auctions. [Fernandez-Tapia et al. \[2017\]](#) also study the problem of bidding in repeated auctions but they consider that the arrival of requests is a Poisson process and characterize the optimal bidding strategy via its Hamilton-Jacobi-Bellman equation.

**Mapping to a DRC<sup>2</sup> problem.** In this setting, the decision maker is an advertiser. The advertiser is present in the market for  $T$  periods and one impression is auctioned per period. Upon the arrival of an impression at time  $t$ , the advertiser determines a real-valued valuation  $\theta_t \in \Theta \subset (0, \Theta_{\max}]$  for the impression, which is distributed according to  $p \in \Delta(\Theta)$ , and chooses an action  $a_t \in \mathcal{A} = [0, \Theta_{\max}]$  representing his bid in the auction. We denote by  $C$  the budget of the advertiser. The shock  $\epsilon$  captures all exogenous uncertainty in the auction, such as the bids of the competitors and any potential randomization of the auction. For simplicity, we assume that  $\epsilon$  is independent of the buyer’s valuation  $\theta$  but our model can be easily be modified to account for correlation. The auction is characterized by an allocation rule  $q : \mathcal{A} \times \mathcal{E} \rightarrow [0, 1]$  together with a payment rule  $m : \mathcal{A} \times \mathcal{E} \rightarrow \mathbb{R}$ , which determine the probability that the impression is allocated to the advertiser and his expected payment as a function of his bid and the exogenous shock, respectively. In §6.2 we explicitly obtain the allocation and payment rules when the advertisers bid in first and second-price auctions, which are two auctions commonly used in the literature. The reward earned by the advertiser and his budget consumption, given  $a, \epsilon$  and  $\theta$  can be expressed as  $r(\theta, a, \epsilon) = \theta q(a, \epsilon) - m(a, \epsilon)$  and  $y(\theta, a, \epsilon) = m(a, \epsilon)$ , respectively.

### 3.3 Network Revenue Management

Another notable special case of the DRC<sup>2</sup> class is a classical problem in the Revenue Management literature: the Network Revenue Management (NRM) problem. It was originally proposed in D’Sylva [1982], Glover et al. [1982] and Wang [1983] and analyzed in the seminal paper Talluri and Van Ryzin [1998]. Since then, it has been extensively studied in the literature and have also been the basis for various industry solutions. The books of Talluri and Van Ryzin [2006] and Gallego et al. [2019] provide extensive reviews.

In the NRM problem, the decision maker is a firm who is trying to dynamically allocate a limited amount of resources over a finite horizon. Resources are sold to heterogeneous consumers who arrive sequentially over time and belong to different classes depending on their consumption of resources and the fixed fare they pay. The distribution of customer classes is stationary. Upon a customer’s arrival, the firm has to decide whether to accept or reject the customer’s request. If the customer is accepted and there is enough remaining inventory to satisfy its request, she consumes the resources requested and pays the corresponding fare. Otherwise, no revenue is collected and no resource is used. The decision maker’s objective is to maximize the expected revenue earned during the selling horizon.

It is worth mentioning that some other problems, such as versions of dynamic knapsack problems (see, e.g., Papastavrou et al. 1996, Kleywegt and Papastavrou 1998, Arlotto and Xie 2018) and versions of the multisecretary problem (see, e.g., Karlin 1962, Sakaguchi and Saario 1995, Arlotto

and Gurvich 2019), can be seen as particular cases of the NRM problem and therefore they also belong to the DRC<sup>2</sup> class.

**Mapping to a DRC<sup>2</sup> problem.** In a NRM problem, a customer class can be captured by  $\theta \in \Theta$  and is characterized by their usage of resources and a fixed price they pay for the service. We let  $f_\theta$  denote the fare associated with class  $\theta$ . The decision maker’s feasible actions has two values,  $\mathcal{A} = \{0, 1\}$ , where we represent the action “accept” by 1 and “reject” by 0. In this problem the set of idiosyncratic shocks is empty; both reward and resource consumption are deterministic given the class. The reward if the customer belongs to class  $\theta$  and the decision maker chooses an action  $a$  is  $r(\theta, a) = f_\theta a$ . If we denote by  $A_\theta = (A_{l\theta})_l \in \mathbb{R}^L$  the consumption vector, where  $A_{l\theta}$  is the amount of resource  $l$  required to serve a class  $\theta$  customer, the consumption given that the decision maker chooses an action  $a$  and the customer class is  $\theta$  is given by  $y(\theta, a) = A_\theta a$ .

### 3.4 Choice-Based Network Revenue Management

This problem bears many similarities to the network revenue management problem. In this setting, a firm is trying to dynamically allocate a limited amount of products which are sold to heterogeneous consumers who arrive sequentially and belong to different classes characterized by their product preferences. The key difference with the NRM class is that, upon a customer’s arrival, the firm makes an offer and depending on the offer and on the customer’s preferences, the consumer selects a single product to buy.

This class of problems appears in the literature under different names depending on whether each product in the offer is a combination of one or more resources or whether there is a one-to-one mapping between products and resources. The first variant is the so-called choice-based problem whereas the second stream corresponds to a dynamic assortment optimization problem under capacity constraints.

In the first stream, the single-leg case was introduced by Talluri and Van Ryzin [2004] who provided an analysis of the optimal control policy under a general discrete choice model of demand. In a network setting, Gallego et al. [2004] was the first to study a choice-based NRM problem. They consider flexible products in a continuous time horizon and with arrivals following independent Poisson processes. A flexible product consists of a set of alternative products serving a customer class. That is, if a flexible product  $F$  is offered by the decision maker and accepted by the consumer, then the decision maker assigns him one of the products in  $F$ . Liu and Van Ryzin [2008] considered a choice-based network RM problem in which each consumer belongs to a market segment (customer class) characterized by a set of products (different for each segment) in which the consumer is interested and the decision maker has to decide a set of products to offer in each selling period.

Bront et al. [2009] consider the same problems as Liu and Van Ryzin [2008] but they allow customer classes to overlap.

In the second stream, Bernstein et al. [2015] consider a dynamic assortment problem in continuous time. In the problem they consider, all products have the same price and for each customer class they compute the probability that a customer belonging to that class chooses a product from the offer according to a MNL model. Golrezaei et al. [2014] also formulate a related dynamic assortment optimization problem. Their formulation is different in that it focuses on arbitrary, possibly adversarial, sequences of customer arrivals.

**Mapping to a DRC<sup>2</sup> problem.** In this setting we consider a set of  $N$  products, each of them consisting of a set of resources. Product  $n$  is priced at  $m_n$ . Given a customer class  $\theta$  and a product  $n$ , we denote by  $A_{l\theta}^n$  the amount of resource  $l$  needed to serve product  $n$  to customer  $\theta$ . The action set is given by  $\mathcal{A} = 2^{\{1, \dots, N\}}$ , where an action  $a \in \mathcal{A}$  represents a set of products to be offered to a consumer.

For each action  $a$  and customer class  $\theta$ , we define the shock random vector  $\epsilon_{\theta a} \in \mathcal{E}_a = \{\epsilon \in \{0, 1\}^N : \sum_{n \in [N]} \epsilon^n \leq 1, \epsilon^n = 0 \text{ for } n \notin a\}$ , where its  $n$ th component  $\epsilon_{\theta a}^n$  is 1 if and only if the customer selects product  $n$  from the offer  $a$ . Then,  $\epsilon_{\theta a} \sim \text{Multinomial}(1, g_{\theta a}^n)$  where  $g_{\theta a}^n$ <sup>5</sup> is the probability of the consumer choosing product  $n$  given that his class is  $\theta$  and the action taken is  $a$ . Here  $g_{\theta a}^n = 0$  if  $n \notin a$ , i.e., if the product does not belong to the offer set.

Given the consumer class  $\theta$ , the action  $a$ , and the shock realization  $\epsilon_{\theta a}$ , the reward function is given by  $r(\theta, a, \epsilon) = \sum_{n \in [N]} m_n \epsilon^n$  and the consumption function by  $y(\theta, a, \epsilon) = \sum_{n \in [N]} A_{\theta}^n \epsilon$ . Note that, conditional on the shock  $\epsilon$ , the resource consumption does not depend on the action. However, the action affects the distribution of  $\epsilon$ .

### 3.5 Online Matching

Another closely related class of problems is that of online matching. This problem is related to the NRM class, but now, instead of making accept/reject decisions and each opportunity consuming a subset of resources, each opportunity can be assigned to any resource and the decision maker needs to decide the resource to which to assign the opportunity.

The bipartite online matching problem was introduced by Karp et al. [1990] where they consider the case with arrivals in arbitrary order and with the goal of maximizing the total number of matches. Their results were extended by Aggarwal et al. [2011] to more general settings. It was also recently studied by Vera and Banerjee [2021]. Some special cases of the online matching

---

<sup>5</sup>Without loss of generality, we can apply suitable transformations to the reward and consumption functions to obtain a distribution of  $\epsilon$  that is independent of the class and the action.

problem were considered by [Feldman et al. \[2009\]](#), [Manshadi et al. \[2012\]](#) and [Devanur et al. \[2013\]](#).

**Mapping to a DRC<sup>2</sup> problem.** We have a bipartite graph with resources on one side and classes on the other side. An opportunity of class  $\theta$  arrives with probability  $p_\theta$  and the decision maker need to decide which resource to assign it to. Calling  $\mathcal{L}$  to the set of resources, each class  $\theta$  has a fare vector  $f_\theta \in \mathbb{R}_+^L$  and a resource consumption  $A_\theta \in \mathbb{R}_+^L$ . The action set is  $\mathcal{A} = \mathcal{L} \cup \{0\}$ , where the action 0 represents rejecting the request, i.e., the decision maker needs to decide the resource to which to assign the opportunity. Given an arrival  $\theta$  and an action  $a$ , the reward is given by  $r(\theta, a) = f_{\theta a} 1_{\{a \neq 0\}}$  and the consumption is  $y(\theta, a) = A_{\theta a} 1_{\{a \neq 0\}}$ . We assume that the bipartite graph is complete. Incomplete graphs can be modelled by setting  $f_{\theta j} = -\infty$  if assigning class  $\theta$  to resource  $j$  is not feasible.

### 3.6 Order Fulfillment

In this section we detail a class of problems faced by a retailer who needs to fulfill the orders they receive from different facilities. Specifically, in this problem orders arrive sequentially and a decision maker has to construct a fulfillment policy to decide from which facility each of the items in the arriving order should be fulfilled.

Many different variants of this DRC<sup>2</sup> problem have been studied in the literature (see, e.g., [Acimovic and Graves 2015](#), [Jasin and Sinha 2015](#), [Andrews et al. 2019](#)). For example, papers have considered different objectives to optimize, whether the model requires a demand forecast or not, multi or single-item approach, among others. We refer the reader [Acimovic and Farias \[2019\]](#) for a recent overview of order fulfillment problems.

On the other hand, some works in the existing literature consider additional constraints related, for instance, to the set of feasible facilities (or resources) from which is it possible to serve an order. [Asadpour et al. \[2019\]](#) consider an online allocation problem with equal numbers of types of resources and types of requests with the restriction that a request of type  $i$  can be served only by resources of type  $i$  and type  $i + 1$ . If both resources have zero inventory left, then the sale is lost. Their goal is to provide an upper bound on the difference between the performance with and without the above described restriction on fulfillment.

It is worth mentioning that some works consider the order fulfillment problem jointly with the pricing problem (see, e.g., [Harsha et al. 2019](#), [Lei et al. 2018a](#)) or jointly with both pricing and display problems (see, e.g., [Lei et al. 2018b](#)).

**Mapping to a DRC<sup>2</sup> problem.** We consider the setting where there are  $L$  different items that could be served from  $N$  different facilities. Each facility  $n$  is endowed with an inventory  $C_n \in \mathbb{R}_+^L$ ,

with  $C_{nl}$  representing the initial capacity of item  $l$  in facility  $n$ , and we consider that facility  $N$  is fictitious with infinite initial capacity of all items.

Arrival  $\theta$  occurs with probability  $p_\theta$  and corresponds to an order of products that needs to be fulfilled belonging to  $\Theta = 2^{\{1, \dots, L\}}$ . We assume one order includes at most one unit of each item. Then,  $\theta_l = 1$  if and only if item  $l$  is included on the offer  $\theta$  and 0, otherwise.

The decision maker has to construct a fulfillment policy to decide from which facility  $n \in [N]$  each of the items in  $\theta$  should be fulfilled in order to maximize his expected revenue. That is, the action set is given by  $\mathcal{A} = \{1, \dots, N\}^L$ , where given  $l \in [L]$ ,  $a_l = n$  means that item  $l$  is served from facility  $n$ . Furthermore, serving item  $l$  from facility  $n$  has an associated fare denoted by  $f_{ln}$ .

Given that the order is  $\theta$  and the decision maker chooses action  $a$ , the consumption of item  $l$  in the facility  $n$  is  $y_{ln}(\theta, a) = 1_{\{a_l=n\}}$ , and the reward is given by  $r(\theta, a) = \sum_{l \in \theta, n \in [N]} f_{ln} 1_{\{a_l=n\}}$ .

## 4 Performance Analysis of Certainty Equivalent Heuristic

As mentioned in §2, an optimal solution of the stochastic formulation of a DRC<sup>2</sup> problem is not easy to compute. A common and central heuristic in the theory of dynamic decision-making under uncertainty is based on the *certainty equivalent principle*: replace quantities by their expected values and take the best actions given the current history. Specifically, at each point of time  $t$ , we solve an optimization problem obtained by using the history up to  $t - 1$  and replacing the random quantities in problem (1) by their expectations.

That is, if we denote by  $\Phi$  the set of all class-dependent probability distributions  $\phi : \Theta \rightarrow \Delta(\mathcal{A})$ , and by  $\rho \in \mathbb{R}_+^L$  a non-negative parameter representing the vector of available inventory divided by the number of remaining periods, at time  $t$  we solve the following parametric programming problem for  $\rho = \rho_t$ , which we refer to as the *fluid problem*:

$$\begin{aligned} J(\rho) = \max_{\phi \in \Phi} \quad & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon (r(\theta, a, \epsilon)) d\phi_\theta(a) \\ \text{s.t.} \quad & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon (y_l(\theta, a, \epsilon)) d\phi_\theta(a) \leq \rho_l, \quad \forall l \in [L]. \end{aligned} \tag{2}$$

While we refer to (2) as the fluid problem as it uses deterministic quantities as inputs (random quantities are replaced by their expected values), we remark that the controls are, in general, randomized. Because the distributions of classes and shocks are independent and identically distributed (i.i.d.) and we allow for randomized actions, we can restrict attention without loss to static controls in the fluid problem. Fixing the parameter  $\rho$ , for each class  $\theta \in \Theta$  and a set of actions  $A \in \mathcal{A}$ , the decision variable  $\phi_\theta(A)$  gives the probability of choosing an action  $a \in A$  given

that the arrival belongs to class  $\theta$ .

In what follows, we assume that problem (2) admits an optimal solution. We denote by  $\phi_\rho^*$  an optimal solution when the parameter is  $\rho$ . We will provide sufficient conditions for existence of an optimal solution in §5.<sup>6</sup>

The Certainty Equivalence Principle leads to a natural heuristic for the decision-maker: at each point in time  $t$ , choose actions according to a solution  $\phi_{\rho_t}^*$  to (2) with  $\rho_t = c_t/(T - t + 1)$  and  $c_t$  the capacity remaining at beginning of time  $t$ . We call this heuristic the *Certainty Equivalent Heuristic* (CE) and denote the corresponding policy by  $\pi^{\text{CE}}$ . The heuristic is formally presented in Algorithm 1. The CE heuristic adjusts the parameter  $\rho$  dynamically according to the amount of resources remaining to avoid running out resources too early or avoid being overly constrained if resource consumption ends up being lower than expected.

The certainty equivalent heuristic has been extensively studied in the literature for specific applications. We return to these in §6. Our aim is to characterize its performance for the broader class of DRC<sup>2</sup> problems. Thus, as the family of DRC<sup>2</sup> problems encompasses a large number of applications that have been studied in the literature separately, by analyzing the performance of the CE heuristic, we shall recover some already known results and, in the process, obtain new results for other applications, while highlighting very general sufficient conditions to ensure “good” performance of the CE heuristic.

---

**Algorithm 1** Certainty Equivalent Heuristic (CE)

---

Initialize  $c_1 \leftarrow C$

**for**  $t = 1, \dots, T$  **do**

$\rho_t \leftarrow c_t/(T - t + 1)$

$\phi_{\rho_t}^* \leftarrow$  an optimal solution of Problem 2 with  $\rho = \rho_t$

    observe the opportunity class  $\theta_t$

    draw an action  $a_t$  with probability  $\phi_{\rho_t}^*(\theta_t, a_t)$

**if**  $y(\theta_t, a_t, \epsilon) \leq c_t \forall \epsilon \in \mathcal{E}$ , **then**

        choose the action  $a_t$

        observe the shock  $\epsilon_t$

$c_{t+1} \leftarrow c_t - y(\theta_t, a_t, \epsilon_t)$

**else**

        choose the null action  $a_0$

$c_{t+1} \leftarrow c_t$

**end**

**end**

---

<sup>6</sup>When an optimal solution does not exist, it suffices for our purpose to work with a  $1/T$  approximately optimal solution.

We note here for simplicity, we focus on a certainty equivalent control that adjusts decisions based on re-solving every period. We conjecture that the results developed in the present paper continue to hold when one re-solves less frequently. The latter has indeed been established for various special cases of DRC<sup>2</sup> problems (see, e.g., [Jasin and Kumar 2012](#)).

#### 4.1 Bound on the cumulative reward loss of the CE heuristic

In this section we study the performance of the CE heuristic for the general class of DRC<sup>2</sup> problems. To this end, we first introduce some definitions and conditions on the primitives.

We will assume that the reward and consumption functions are bounded. For a vector  $x \in \mathbb{R}^n$ , we denote by  $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$  its  $\ell^2$ -norm and denoted by  $\|x\|_\infty = \max_i |x_i|$  its  $\ell^\infty$ -norm.

**Assumption 1.** *The following hold:*

1. *There exists  $\bar{r}_\infty \in \mathbb{R}_{++}$  such that  $|r(\theta, a, \epsilon)| \leq \bar{r}_\infty$  for all  $\theta \in \Theta, a \in \mathcal{A}$ , and  $\epsilon \in \mathcal{E}$ .*
2. *There exist  $\bar{y}_2, \bar{y}_\infty \in \mathbb{R}_{++}$  such that  $\|y(\theta, a, \epsilon)\| \leq \bar{y}_2$  and  $\|y(\theta, a, \epsilon)\|_\infty \leq \bar{y}_\infty$  for all  $\theta \in \Theta, a \in \mathcal{A}$ , and  $\epsilon \in \mathcal{E}$ .*

Recall that  $\rho_1$  is the vector of initial inventory divided by the amount of periods to consider. We assume that, in the neighborhood of  $\rho_1$ , the optimal objective value of the fluid problem  $J(\rho)$  is locally smooth as well as that the consumption constraints are binding.

**Assumption 2.** *There exist  $\delta, K \in \mathbb{R}_{++}$  with  $\delta < \min_{j \in [L]}(\rho_1)_j$  such that for every  $\rho$  with  $\|\rho - \rho_1\| \leq \delta$ , it holds that:*

1. *The function  $J(\rho)$  satisfies  $J(\rho) \geq J(\rho_1) + \nabla J(\rho_1)(\rho - \rho_1) - \frac{K}{2}\|\rho - \rho_1\|^2$ .*
2. *The optimal solution  $\phi_\rho^*$  satisfies  $\mathbb{E}_{\theta \sim p, a \sim \phi_\rho^*, \epsilon \sim f}(y(\theta, a, \epsilon)) = \rho$ .*

We will refer to inequality given in Assumption 2.1 as  $J(\rho)$  admitting a  $K$ -lower downward quadratic ( $K$ -LDQ) envelope in  $\mathcal{N}(\rho_1, \delta)$ , where we denote by  $\mathcal{N}(\rho_1, \delta)$  the ball of radius  $\delta$  centered at  $\rho_1$ .<sup>7</sup> See Figure 2 (a) for an example of a function admitting a  $K$ -LDQ envelope and its envelope. This condition is a weaker and local notion of the  $K$ -strongly smooth condition for concave functions, which requires the inequality in Assumption 2.1 to hold for every pair of parameters  $\rho, \rho'$ . A sufficient condition for  $J(\rho)$  to admit a  $K$ -LDQ envelope is that its gradient is *locally  $K$ -Lipschitz continuous* at  $\rho_1$  for all  $\rho \in \mathcal{N}(\rho_1, \delta)$ , that is,

$$\|\nabla J(\rho_1) - \nabla J(\rho)\| \leq K\|\rho_1 - \rho\|.$$

---

<sup>7</sup>More specifically,  $\mathcal{N}(x, r)$  represents the set of all points at distance at most  $r$  from  $x$ , where the distance is the one induced by the  $\ell^2$ -norm:  $\mathcal{N}(x, r) = \{y \in \mathbb{R}^n : \|x - y\| \leq r\}$ .



See Lemma 3.4 in [Bubeck \[2014\]](#) for a proof of the previous fact.

We are now ready to state our performance bound of the CE heuristic under Assumptions 1 and 2. Specifically, in Theorem 1, we state that the reward loss of the heuristic given by CE is on the order of  $\log T$  or uniformly bounded, depending on the value of the constant  $K$ .

**Theorem 1.** *Let  $J^{\text{CE}}$  be the expected performance of Algorithm 1. Then, under Assumptions 1 and 2, the reward loss satisfies*

$$J^* - J^{\text{CE}} \leq \bar{y}_2^2 K \log T + \left[ \frac{2\bar{y}_\infty}{\underline{\rho}_1 - \delta} + \frac{14\bar{y}_\infty^2}{\delta^2} \right] J(\rho_1),$$

where  $\underline{\rho}_1$  the smallest component of vector  $\rho_1$ .

We note that the result above applies across all DRC<sup>2</sup> problems, and only requires Assumptions 1 and 2. Consider a regime in which  $C$  and  $T$  are scaled proportionally, i.e.,  $C = \rho_1 T$  for some  $\rho_1 \in \mathbb{R}_{++}^L$ . Theorem 1 implies that, in such a regime, the CE heuristic is asymptotically optimal in the sense that  $J^{\text{CE}}/J^* \rightarrow 1$  as  $T \rightarrow \infty$  because the reward collected by the CE heuristic grows as  $T \rightarrow \infty$ . Furthermore, the optimality gap is of  $O(\log T)$  if  $K > 0$  and of  $O(1)$  if  $K = 0$ . In other words, we see a clear distinction among DRC<sup>2</sup> problems driven by the value of  $K$  in Assumption 2.

At a more detailed level, the dependency on the number of resources  $L$  enters our bound indirectly via the constants  $\bar{y}_2$ ,  $\bar{y}_\infty$ , and  $K$ . Interestingly, when resource consumption is uniformly bounded, i.e.,  $\bar{y}_\infty < \infty$ , the dependency on the number of resources is mostly driven by  $\bar{y}_2^2$ . While in the worst case we could have  $\bar{y}_2^2 = \Omega(L)$ , in many settings of interest, one will have  $\bar{y}_2^2 = O(1)$  and we obtain bounds that are independent of the number of resources. This could happen, for example, if every opportunity consumes only a finite subset of resources. Our dependence in the number of resources is better than some existing (specialized) results on the literature because we use concentration inequalities for multi-dimensional martingales instead of a union bound to control the stopping time associated to the first time a resource is close to being depleted, which would naturally lead to a linear dependence on  $L$ .

The proof of the theorem can be found in Appendix B. The proof leverages ideas pioneered by [Jasin and Kumar \[2012\]](#) and lifts them to more general settings than the one considered in their paper. In particular, we analyze the performance of the CE heuristic up to the stopping time  $\tau$ , where  $\tau$  is the first time that a resource is close to depletion or the ratio of capacity to time remaining  $\rho_t$  leaves the ball  $\mathcal{N}(\rho_1, \delta)$  defined in Assumption 2. Using that the fluid problem gives an upper bound on the optimal value of the stochastic problem, that is,  $J^* \leq TJ(C/T)$  (see, e.g.,

Gallego and Van Ryzin 1997 for a specialized argument), we can bound the reward loss as follows

$$J^* - J^{\text{CE}} \leq TJ(\rho_1) - J^{\text{CE}} \leq \mathbb{E} \left( \sum_{t=1}^{\tau} J(\rho_1) - \sum_{t=1}^{\tau} r(\theta_t, a_t^{\pi^{\text{CE}}}, \epsilon_t) \right) + \mathbb{E} \left( \sum_{t=\tau+1}^T J(\rho_1) \right), \quad (3)$$

where  $a_t^{\pi^{\text{CE}}}$  denotes the action taken by the CE heuristic and the second inequality follows because  $r(\theta_t, a_t^{\pi^{\text{CE}}}, \epsilon_t) \geq 0$  since the null action  $a_0$  is feasible. The second term of the right-hand side can be written as  $\mathbb{E}(T - \tau)J(C/T)$ , which is on the order of  $O(1)$ , as we establish in Lemma B-3. This follows because, under the CE heuristic, the ratio  $\rho_t$  behaves like a martingale by Assumption 2.2 and, as a result, the heuristic never runs out of resources nor  $\rho_t$  leaves the ball  $\mathcal{N}(\rho_1, \delta)$  too early. The first term is shown to be of order  $O(\log T)$ . To see this, note that up to time  $\tau$  actions are not constrained by resources and the expected reward at period  $t$  satisfies  $\mathbb{E} \left( r(\theta_t, a_t^{\pi^{\text{CE}}}, \epsilon_t) \mid \rho_t \right) = J(\rho_t)$  because the CE heuristic takes actions according to  $\phi_{\rho_t}^*$ . Therefore, using Assumption 2.1 we can upper bound the first term by  $\nabla J(\rho_1) \mathbb{E} \left( \sum_{t=1}^{\tau} (\rho_1 - \rho_t) \right) + K/2 \mathbb{E} \left( \sum_{t=1}^{\tau} \|\rho_1 - \rho_t\|^2 \right)$ . The first term is zero because  $\rho_t$  behaves like a martingale, while the second term can be bounded using that martingale differences are orthogonal. Putting everything together, we then conclude that  $J^* - J^{\text{CE}} = O(K \log T)$ .

We highlight here again that, although the general idea behind the proof of Theorem 1 is not novel, our contribution is to generalize this result to the class of DRC<sup>2</sup> problems and identify crisp sufficient conditions that allow to obtain a performance guarantees on the CE heuristic for every problem in this class without the need of exploiting specific structural features of each particular application. As we will see later, this allows to obtain in some cases weaker conditions or sharper bounds, but also to obtain new results altogether. More specifically, we see that it is not only possible to define a unified model for a very large of dynamic optimization problems as we established in §3, but it is also possible to identify general sufficient conditions that allow for a unified analysis of the CE heuristic and its performance.

Assumption 2 provides sufficient conditions to obtain good performance guarantees for the CE heuristic that involve, in a neighborhood of the “initial inventory per period”, the local smoothness of the optimal objective value of the fluid problem as well as that the consumption constraints are binding. These conditions, while simple to state, may not be easy to check in many applications. In the following section we provide simple assumptions on the primitives of the model, which are simple to check, that imply Assumption 2 and, therefore, also the bound in Theorem 1. Our conditions yield, in many cases, closed-form expressions for the values of  $\delta$  and  $K$  based on the primitives of the problem. In particular, we will show that for an important subclass of problems, Assumption 2 holds with  $K = 0$  and, therefore, we recover reward loss of  $O(1)$  instead of order

$O(\log T)$ .

## 5 Drivers of CE Heuristic Performance: General Sufficient Conditions through the Dual Problem

An important component to understand the drivers of the CE heuristic is to link the primitives of a DRC<sup>2</sup> problem to the performance that is achieved. In this section, we address this goal by giving conditions on the primitives of a DRC<sup>2</sup> problem that are sufficient for Assumption 2 to be satisfied. These conditions provide an unifying framework to understand how the structural features of different problems across the DRC<sup>2</sup> class impact the performance of the CE heuristic, rather than requiring a problem-specific analysis for each application.

We introduce a dual of Problem (2) in which we dualize the consumption constraints. To this end, let  $\mu \in \mathbb{R}_+^L$  be the vector of Lagrange multipliers associated with the consumption constraints of Problem (2). Let  $\bar{r} : \Theta \times \mathcal{A} \rightarrow \mathbb{R}_+$  denote the expected reward function, i.e.,  $\bar{r}(\theta, a) = \mathbb{E}_\epsilon(r(\theta, a, \epsilon))$ . In the same way, for each  $l \in [L]$ , let  $\bar{y} : \Theta \times \mathcal{A} \rightarrow \mathbb{R}_+^L$  denote the expected resource consumption function, i.e.,  $\bar{y}(\theta, a) = \mathbb{E}_\epsilon(y(\theta, a, \epsilon))$ . Then, the Lagrangian function is given by

$$\begin{aligned} \mathcal{L}(\phi, \mu) &= \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \bar{r}(\theta, a) d\phi_\theta(a) + \mu^\top \left( \rho - \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \bar{y}(\theta, a) d\phi_\theta(a) \right) \\ &= \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \left( \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \right) d\phi_\theta(a) + \mu^\top \rho. \end{aligned}$$

Define, for each  $\theta \in \Theta$ , the function  $g_\theta : \mathbb{R}^L \rightarrow \mathbb{R}$

$$g_\theta(\mu) = \sup_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \right\}. \quad (4)$$

The Lagrange dual function, for fixed  $\rho \geq 0$ , is given by

$$\Psi_\rho(\mu) = \sup_{\phi_a \in \Delta(\mathcal{A})} \mathcal{L}(\phi, \mu) = \mu^\top \rho + \sum_{\theta \in \Theta} p_\theta \sup_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \right\} = \mu^\top \rho + \sum_{\theta \in \Theta} p_\theta g_\theta(\mu),$$

where the second equality follows because the problem is separable over classes. The dual problem of Problem (2) is then given by

$$\inf_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu). \quad (5)$$

In what follows, we will assume that the dual solution for  $\rho = \rho_1$ , namely  $\mu^1$ , is interior; that is,  $\mu^1 > 0$ . We introduce this assumption to guarantee that all resource constraints are binding for

$\rho = \rho_1$ , but we believe our results still hold without this assumption.

## 5.1 Continuum of actions

We first assume that the set of actions  $\mathcal{A}$  is a continuum. Examples of applications with continuous actions are: network dynamic pricing problems and dynamic bidding in repeated auctions with budgets. In this case, we will give a closed-form expressions for values  $K > 0$  and  $\delta$  for which Assumption 2 holds and we obtain, under certain conditions that we explain below, a reward loss of the order of  $O(K \log T)$ .

We now present a condition under  $g_\theta$  that is sufficient to ensure that the fluid problem has zero duality gap for every positive parameter  $\rho$ . Furthermore, we prove that the problem admits an optimal, deterministic solution. The latter implies that when the set of actions is a continuum, under assumption SC 1, randomization in the fluid problem is not needed.

**SC 1.** *For each  $\theta \in \Theta$ ,  $g_\theta(\mu)$  is differentiable in  $\mu$ , and for every  $\mu \geq 0$ ,  $g_\theta(\mu)$  is achieved for an action  $a \in \mathcal{A}$ .*

**Proposition 1.** *Under Assumptions 1 and SC 1, strong duality holds, i.e.,  $J(\rho) = \inf_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu)$  for all  $\rho > 0$ . Furthermore,  $J(\rho)$  admits a deterministic optimal solution for all  $\rho$ .*

A proof is provided in Appendix C.1.1. As a result, we can write  $J(\rho)$  as the following non-linear program:

$$\begin{aligned} J(\rho) = \quad & \max_{a \in \mathcal{A}^{|\Theta|}} \sum_{\theta \in \Theta} p_\theta \mathbb{E}_\epsilon (r(\theta, a_\theta, \epsilon)) \\ & \text{s.t.} \quad \sum_{\theta \in \Theta} p_\theta \mathbb{E}_\epsilon (y_l(\theta, a_\theta, \epsilon)) \leq \rho_l, \quad \forall l \in [L]. \end{aligned}$$

The resulting non-linear program might be non-convex and, in general, challenging to solve. A straightforward corollary of Proposition 1 is that instead of the solving the primal problem, it is possible to construct deterministic controls for the CE heuristic by solving the dual problem at each step. In many cases the dual problem can be efficiently solved (e.g., using first-order methods) because it is always guaranteed to be convex and finite dimensional. For all  $\rho > 0$ , we show in the same result that the dual problem admits an optimal solution  $\mu^* \in \arg \max_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu)$ . A primal control can be constructed by solving, for each class  $\theta \in \Theta$ , for an action  $a_\theta^* \in \arg \max_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - \mu^{*\top} \bar{y}(\theta, a) \right\}$ . In other words, optimal actions maximize rewards minus the opportunity cost of consuming resources, where resources are priced according to the optimal dual variable  $\mu^*$ . We notice here that in some cases, such as the problem of dynamic bidding in repeated auctions (§3.2, §6.2), once dual variables are available, an optimal action can be computed in closed form.

In addition to condition SC 1, we need to make another regularity assumption over the function  $g_\theta$ , stated below, to ensure Assumption 2 holds for these particular cases of DRC<sup>2</sup>.

**SC 2.** *There exists positive real numbers  $\kappa$  and  $\nu \leq \underline{\mu}$  with  $\underline{\mu} = \min_{l \in [L]} \mu_l^1$  such that for all  $\theta \in \Theta$  and  $\mu \in \mathcal{N}(\mu^1, \nu)$ ,  $g_\theta(\mu)$  satisfies*

$$g_\theta(\mu) \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \frac{\kappa}{2} \|\mu - \mu^1\|^2. \quad (6)$$

In what follows we will refer to property (6) as  $g_\theta$  admitting a  $\kappa$ -lower upward quadratic ( $\kappa$ -LUQ) envelope in  $\mathcal{N}(\mu^1, \nu)$ . See Figure 2 (b) for an example of a function admitting a  $\kappa$ -LUQ and its envelope. Note that admitting a  $\kappa$ -LUQ envelope is a weaker and local notion of the  $\kappa$ -strongly convex condition, which requires (6) to hold for every pair of dual variables  $\mu, \mu'$ .

**Lemma 1.** *Suppose that Assumptions SC 1 and SC 2 hold. If  $\mu^1 > 0$ , Assumption 2 holds with  $K = 1/\kappa$  and  $\delta = (\nu\kappa)/2$ .*

A proof is provided in Appendix C.1.2. We obtain the following result as a corollary.

**Corollary 1.** *Suppose that Assumptions 1, SC 1, and SC 2 hold and that  $\mu^1 > 0$ . Then the reward loss of the certainty equivalent heuristic is of order  $O(\kappa^{-1} \log T)$  for DRC<sup>2</sup> problems with a continuum set of actions.*

**Geometric interpretation.** The deterministic function  $J(\rho)$  can be easily shown to be concave and non-decreasing. Assumption SC 2 states that the dual function admits a  $\kappa$ -LUQ envelope in a neighborhood of the Lagrange multiplier  $\mu^1$ . In Figure 2 (b) we represent the dual function for a one-resource problem. By duality, this allows us to prove the smoothness condition on the deterministic function stated in the first statement of Assumption 2, which is represented in Figure 2 (a) and consists of  $J(\rho)$  admitting a  $K$ -LDQ envelope.

### 5.1.1 Sufficient conditions on the primitives

The assumptions presented above are stated in terms of  $g_\theta(\mu)$ , which is a derived object, and, in general, might not be easy to verify. We now present sufficient conditions on the primitives of the problem for Assumptions SC 1 and SC 2 to hold, which, in turn, imply Assumption 2 and allows us to recover a reward loss of order  $O(\log T)$  for DRC<sup>2</sup> problems with a continuum set of actions.

**CA 1.** *For each  $\theta \in \Theta$ ,  $\bar{r}(\theta, a)$  is upper-semicontinuous in  $a$ .*

**CA 2.** *For each  $\theta \in \Theta$ ,  $\bar{y}(\theta, a)$  is continuous in  $a$ .*

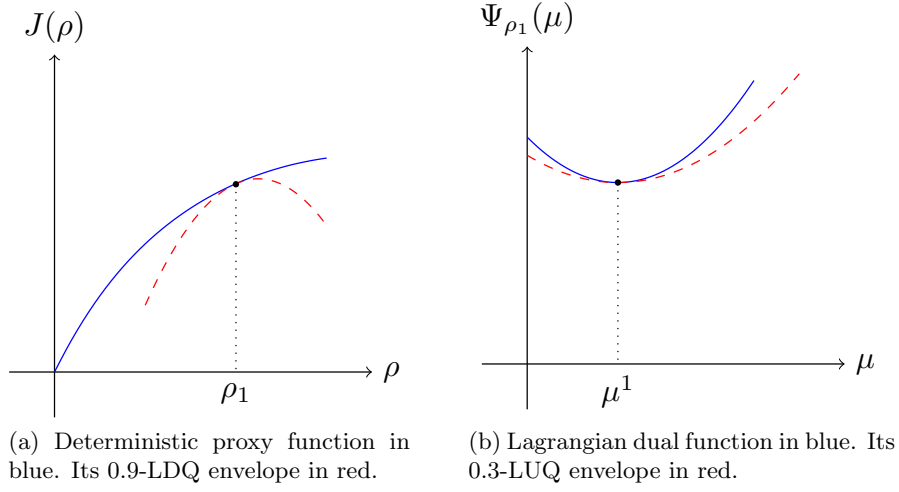


Figure 2: Deterministic proxy  $J$  and its Lagrangian dual function, and their envelopes.

**CA 3.** For each  $\mu \geq 0$  and  $\theta \in \Theta$ , the set  $A_\mu^* = \{a^* \in \mathcal{A} : a^* \in \arg \max_a \{\bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a)\}\}$  of maximizers is not empty. Furthermore, the set  $\{\bar{y}(\theta, a^*) : a^* \in A_\mu^*\}$  is a singleton.

**Proposition 2.** If the set of actions  $\mathcal{A}$  is compact and conditions CA 1, CA 2 and CA 3 hold, then condition SC 1 is fulfilled.

The proof of Proposition 2 follows directly from [Milgrom and Segal, 2002, Corollary 4] and is omitted. It follows from Proposition 2 that if  $\mathcal{A}$  is compact and CA 1, CA 2 and CA 3 hold, then there is a solution to the fluid problem (2).

For each  $\theta \in \Theta$ , let us denote by  $a_\theta^1$  a feasible action that maximizes  $\bar{r}(\theta, a) - (\mu^1)^\top \bar{y}(\theta, a)$ . We will assume the following extra conditions in order to bound the reward loss.

**CA 4.** For each  $\theta \in \Theta$ , the feasible action  $a_\theta^1$  is interior. That is, there exists a positive number  $\varphi$  such that  $\mathcal{N}(a_\theta^1, \varphi) \subseteq \mathcal{A}$  for all  $\theta$  in  $\Theta$ .

**CA 5.**  $\bar{r}(\theta, \cdot)$  admits a  $\kappa_r$ -LDQ envelope in  $\mathcal{N}(a_\theta^1, \varphi)$ . That is, for all  $\theta \in \Theta$ ,

$$\bar{r}(\theta, a) \geq \bar{r}(\theta, a_\theta^1) + \nabla \bar{r}(\theta, a_\theta^1)^\top (a - a_\theta^1) - \frac{\kappa_r}{2} \|a - a_\theta^1\|^2 \quad \forall a \in \mathcal{N}(a_\theta^1, \varphi).$$

**CA 6.** There exists a positive vector  $\kappa_y$  such that for all  $\theta \in \Theta$  and  $a \in \mathcal{N}(a_\theta^1, \varphi)$  the consumption function  $\bar{y}(\theta, \cdot)$  satisfies

$$\bar{y}(\theta, a) \leq \bar{y}(\theta, a_\theta^1) + \nabla \bar{y}(\theta, a_\theta^1)(a - a_\theta^1) + \frac{\kappa_y}{2} \|a - a_\theta^1\|^2, \quad (7)$$

where  $\nabla \bar{y}(\theta, \cdot)$  represents the Jacobian matrix.

We will refer to property (7) as the consumption function  $\bar{y}(\theta, \cdot)$  admitting a  $\kappa_y$ -upper upward quadratic ( $\kappa_y$ -UUQ) envelope in  $\mathcal{N}(a_\theta^1, \varphi)$ . Note that the  $\kappa_y$ -UUQ envelope condition is a weaker and local notion of the  $\kappa_y$ -strongly smooth condition for concave functions, which requires (7) to hold for every pair of action  $a, a'$ . Analogously to the relation made with the lower downward quadratic envelope notion, a sufficient condition for Assumption 6 to hold is that the gradient  $\nabla \bar{y}_j(\theta, \cdot)$  is locally Lipschitz continuous for every resource  $j \in [L]$ .

Given a real-valued matrix  $A$ , we denote by  $\|A\| = \sigma_{\max}(A)$ , where  $\sigma_{\max}(M)$  represents the largest singular value of matrix  $M$ . Recall that given a real-valued matrix  $A$ , its singular values are the square roots of the eigenvalues of matrix  $A^\top A$ . We are now ready to provide sufficient conditions for Assumption SC 2 to hold.

**Lemma 2.** *Suppose that Assumptions CA 1-CA 6 hold. Then, if  $\mathcal{A}$  is compact, SC 2 holds with  $\nu = \kappa\varphi/\sigma$  and  $\kappa = \kappa_r + (\nu + \|\mu^1\|)\|\kappa_y\|$  where  $\sigma = \min_{\theta \in \Theta} \sigma_\theta$ , with  $\sigma_\theta$  the minimum singular value of  $\nabla \bar{y}(\theta, a_\theta^1)$ .*

A proof is provided in Appendix C.1.3. We obtain the following result as a corollary

**Corollary 2.** *Suppose that Assumption 1 and Assumptions CA 1-CA 6 hold. If the set of feasible actions  $\mathcal{A}$  is compact and  $\mu^1 > 0$ , the reward loss of the certainty equivalent heuristic is on the order of  $O(\kappa^{-1} \log T)$  for the DRC<sup>2</sup> problems with a continuum set of actions.*

As we will see in the next section, conducting the analysis in the dual space allows to weaken the sufficient conditions on the primitives required to obtain performance guarantees for the CE heuristic. For instance, in the dynamic pricing case, if the reward function is not concave in the demand space, our results still hold if the conditions stated above are satisfied.

## 5.2 Finite set of actions

When the set of actions  $\mathcal{A}$  is finite, we will show that, under condition SC 3 below, the reward loss of the certainty equivalent heuristic is on the order of  $O(1)$ . Examples of problems with finite set of actions are: network revenue management problems, choice-based revenue management problems, order fulfillment problems, online matching problems, among others.

**SC 3.** *The dual problem (5) has a unique solution,  $\mu^1$ , for  $\rho = \rho_1$ .*

Note that Problem 2, in this special case, can be written as follows

$$\begin{aligned}
 J(\rho) = & \max_{\phi_\theta \in \Delta(\mathcal{A})} \sum_{\theta \in \Theta} p_\theta \bar{r}_\theta^\top \phi_\theta \\
 & \text{s.t.} \quad \sum_{\theta \in \Theta} p_\theta \bar{y}_\theta \phi_\theta \leq \rho,
 \end{aligned} \tag{8}$$

where  $\bar{r}_\theta = (\bar{r}(\theta, a))_{a \in \mathcal{A}} \in \mathbb{R}_+^{|\mathcal{A}|}$  is the vector of expected rewards for the different actions and  $\bar{y}_\theta = (\bar{y}(\theta, a))_{a \in \mathcal{A}} \in \mathbb{R}_+^{L \times |\mathcal{A}|}$  is the matrix of expected resource consumption. Problem (8) is a finite-dimensional linear programming problem. Furthermore, the feasible set is non empty and compact, and therefore there exists an optimal solution. In this case, (5) can be thought of as a “partial” dual problem in which we dualize the resource constraints but not the simplex constraint  $\sum_{a \in \mathcal{A}} \phi_\theta(a) \leq 1$ . Then, it follows that if the set of actions is finite, the duality gap is zero and  $J(\rho) = \inf_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu)$ .

Note that by Assumption SC 3, there exists non-degenerate optimal solution of (8) for  $\rho = \rho_1$ . Denote such solution by  $\phi_{\rho_1}^*$ . If the dual solution is interior, all resource constraints are binding by complementary slackness and this solution satisfies  $\sum_{\theta \in \Theta} p_\theta \bar{y}_\theta(\phi_{\rho_1}^*)_\theta = \rho_1$ .

Considering the standard form of Problem (8) (see Appendix A.1 for details), let  $B \in \mathbb{R}^{(L+|\Theta|) \times (L+|\Theta|)}$  be the corresponding optimal basis matrix and denote by  $B_{\rho_1}^{-1}$  the submatrix of  $B^{-1}$  associated to the resource constraints. In the following lemma, we show that Assumption 2 holds under Assumption SC 3 whenever  $\mu^1 > 0$ .

**Lemma 3.** *Suppose that Assumption SC 3 holds and  $\mu^1 > 0$ . Then Assumption 2 holds with  $K = 0$  and  $\delta = \phi_{\min}^* / \|B_{\rho_1}^{-1}\|$ , where  $\phi_{\min}^* = \min_{\theta \in \Theta, a \in \mathcal{A}} \{\phi_\theta^*(a) : \phi_\theta^*(a) > 0\}$ .*

The proof is provided in Appendix C.1.4. A direct result from Lemma 3 is that the reward loss if the action space is finite is on the order of  $O(1)$ .

**Corollary 3.** *Suppose that Assumptions 1 and SC 3 hold and that the dual solution is interior. Then, the reward loss of the certainty equivalent heuristic is bounded by a constant for DRC<sup>2</sup> problems with finite set of actions.*

**Geometric interpretation.** Let us now provide a geometric interpretation for assumption SC 3 as well as for the deterministic and dual functions for the case of finite actions.

First, note that problem (8) is a LP and therefore the deterministic function  $J(\rho)$  is a concave piece-wise linear function (see Bertsimas and Tsitsiklis 1997 for more details). Moreover, due to the nature of the problem, it will be non-decreasing. In Figure 3 (a), the function  $J(\rho)$  is plotted for a problem with one resource and two classes. Every optimal dual variable  $\mu$  for  $\Psi_\rho(\mu)$  gives a super-gradient to  $J(\rho)$ . Therefore, the slope of each straight-line segment is equal to the Lagrange multiplier associated to the consumption constraint, and the corresponding interval gives the values of the right-hand side range for the consumption constraint  $\rho$  for which the same dual variable is optimal.

Figure 3 (b) and (c) plot the dual function  $\Psi_\rho(\mu)$  as a function of  $\mu$  for two different possible values of the parameter  $\rho$ . In Figure 3 (b), we take  $\rho = \rho_1^1$ , a value where  $J(\rho)$  has a kink. In this



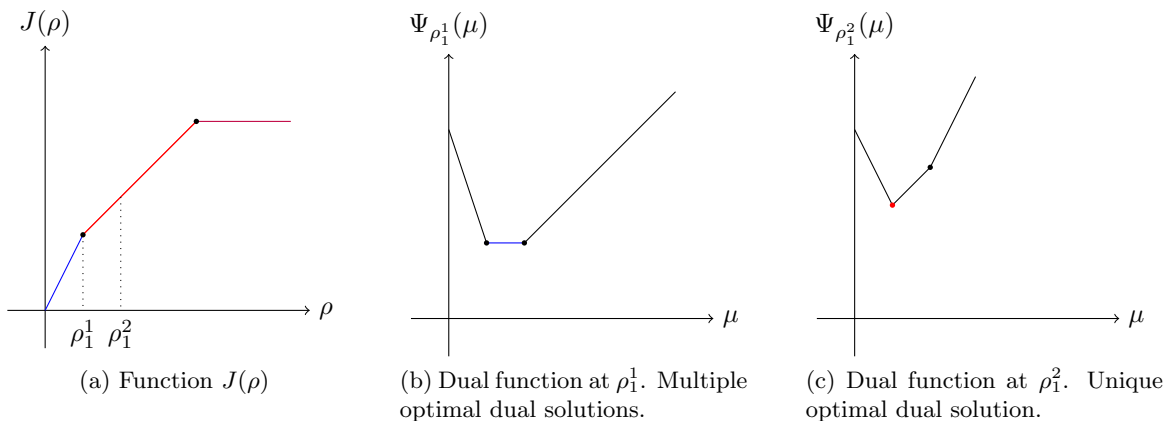


Figure 3: Function  $J$  and dual function  $\Psi_\rho$  for two different parameters.

case, the dual problem has an infinite solutions (blue segment in the figure) and every dual solution is a super-gradient of  $J(\rho_1)$ . Figure 3 (c), we plot the dual function at  $\rho = \rho_1^2$ , a value belonging to an interval where  $J(\rho)$  is smooth. There, the set of super-gradients is a singleton and, as a result, the dual optimal solution is unique (red dot in the figure). Thus, Assumption SC 3 is equivalently asking that the parameters  $\rho_1$  lies in the interior of an interval where the deterministic function  $J(\rho)$  is smooth.

Note that in the case of a finite set of actions it is not necessary to assume that the dual function admits a  $\kappa$ -LUQ envelope (condition SC 2). As we can see in Figure 3 (b), a lower upward quadratic envelope is obtained for free when the optimal dual solution is unique and the action set is finite, because the dual problem is piece-wise linear.

In the following section, we revisit the applications presented in §3. For each problem, we provide problem-specific sufficient conditions for our assumptions to hold, and we establish the implications of Theorem 1. As we will see, our results allow to recover some existing results in the literature as special cases, sometimes under weaker assumptions, and also uncover new results for other classes of problems studied in the literature.

## 6 CE Heuristic Performance: Corollaries Across Subclasses of Problems

In this section we revisit the applications discussed in §3. For the dynamic pricing and dynamic bidding problems, which are the classes with continuum of actions, we present sufficient conditions for Assumption 2 together with the performance characterization of the CE heuristic, and we then comment on the specific connections to the existing literature in the study of the CE heuristic. For

the other classes, with finite set of actions, we give a closed form for the value of  $\delta$  in Assumption 2 and, then, we comment on the connection with earlier analyses of the CE heuristic in the literature.

## 6.1 Network Dynamic Pricing

This class of problems was presented in §3.1. As highlighted there, this is a central class of problems widely studied. Next we present conditions on the primitives that ensure logarithmic loss, and then discuss connections to the literature on the CE heuristic in this context.

Let  $\bar{D}(\theta, a) = \mathbb{E}_\epsilon[D(\theta, a, \epsilon)]$  denote the expected demand and  $\bar{r}(\theta, a) = a^\top \bar{D}(\theta, a)$  the corresponding expected reward function. The fluid problem can be expressed as follows:

$$\begin{aligned} J(\rho) = \max_{\phi \in \Phi} & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \bar{r}(\theta, a) d\phi_\theta(a) \\ \text{s.t.} & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} A_\theta \bar{D}(\theta, a) d\phi_\theta(a) \leq \rho. \end{aligned} \tag{9}$$

In this problem, there is a continuum of actions. We map conditions CA 1-CA 6 to sufficient conditions for this particular problem. The following conditions together with  $\mu^1 > 0$  and compactness of the set of actions  $\mathcal{A}$  are sufficient for Lemma 2 to hold.

- The expected demand function  $\bar{D}(\theta, a)$  is continuous in  $a$ .
- The expected resource consumption  $A_\theta \bar{D}(\theta, a)$  at a maximizer of  $(a - A_\theta \mu)^\top \bar{D}(\theta, a)$  is unique.
- For each  $\theta \in \Theta$ , the price vector maximizing  $(a - A_\theta^\top \mu)^\top \bar{D}(\theta, a)$ , namely  $a_\theta^1$ , is interior. That is, there exists a positive number  $\varphi$  such that  $\mathcal{N}(a_\theta^1, \varphi) \subseteq \mathcal{A}$  for all  $\theta$  in  $\Theta$ .
- The expected revenue function  $\bar{r}(\theta, \cdot)$  admits a  $\kappa_r$ -LDQ envelope in  $\mathcal{N}(a_\theta^1, \varphi)$ .
- There exists a positive vector  $\kappa_y$  such that the expected demand function  $\bar{D}(\theta, \cdot)$  admits a  $\kappa_y$ -UUQ envelope in  $\mathcal{N}(a_\theta^1, \varphi)$ .

In particular, under the conditions above, from Lemma 1, Assumption 2 holds with  $K = 1/\kappa$  and  $\delta = (\nu\kappa)/2$ , where  $\kappa = \kappa_r + (\nu + \|\mu^1\|)\|\kappa_y\|$ ,  $\nu = \kappa\varphi/\sigma$ , and  $\sigma$  is a lower bound on the minimum singular value of  $A_\theta \nabla \bar{D}(\theta, a_\theta^1)$ . Therefore, by Corollary 1, the revenue loss of the certainty equivalent heuristic is on the order of  $O(\log T)$  for the network dynamic pricing problem with a continuum set of feasible prices. Another implication of our result is that the optimal pricing policy associated with the deterministic proxy is deterministic and the decision maker does not need to randomize over posted prices.

Finally, note that, if we consider a finite set of feasible prices, the fluid problem reduces to a linear program and therefore it is enough to assume the dual solution for  $\rho = \rho_1$  is unique and interior to obtain a constant revenue loss for the CE heuristic.

**Connection with earlier analysis of the CE heuristic.** The CE heuristic for this problem was previously analyzed in [Maglaras and Meissner \[2006\]](#) and [Jasin \[2014\]](#). The former established that a CE heuristic for the pricing problem will always yield an asymptotic weak decrease in the revenue loss compared to a static control. [Jasin \[2014\]](#) considers a single customer class and present a certainty equivalent heuristic akin to the CE one. Our general result recovers the bound in [Jasin \[2014\]](#) on the logarithmic revenue loss but our sufficient conditions are weaker than his. It is worth mentioning that this problem is typically analyzed in the demand space in the literature, i.e., for each class  $\theta$ , the decision variables are the expected demands  $\lambda = \bar{D}(\theta, a)$  instead of the prices  $a$  (see, e.g., [Jasin 2014](#)). In many cases, this leads to a more tractable problem because constraints become linear and, under additional conditions, the objective becomes concave. However, further assumptions are needed for this reformulation of the problem to go through. For example, it is typically assumed that the reward function is concave in the demand space and the demand function is invertible. Our result yield similar performance guarantees for the CE heuristic and only requires local smoothness properties of the revenue function, which typically leads to weaker assumptions. This is an important departure from previous work, even when specializing the analysis. We see here, how by lifting the formulation to a DRC<sup>2</sup> problem, we are not only able to recover existing results through a generalized argument, but also to weaken the assumptions needed for such results to hold.

## 6.2 Dynamic Bidding in Repeated Auctions

This class of problems was presented in §3.2. Next we present conditions on the primitives that ensure logarithmic loss, and then discuss connections to the literature on the CE heuristic in this context.

To simplify some of the notation in what follows, given an action  $a$ , we introduce the interim allocation and interim payment variables defined as follows:  $\bar{q}(a) = \mathbb{E}_\epsilon[q(a, \epsilon)]$ ,  $\bar{m}(a) = \mathbb{E}_\epsilon[m(a, \epsilon)]$ . For the particular setting described in §3.2, the fluid problem (2) is equivalent to the following problem:

$$\begin{aligned}
 J(\rho) = \max_{\phi \in \Phi} & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} (\theta \bar{q}(a) - \bar{m}(a)) d\phi_\theta(a) \\
 \text{s.t.} & \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \bar{m}(a) d\phi_\theta(a) \leq \rho.
 \end{aligned} \tag{10}$$

For each  $\theta \in \Theta$ , let  $g_\theta(\mu) = \max_{a \in \mathcal{A}} \{\theta \bar{q}(a) - (\mu + 1) \bar{m}(a)\}$ . Assumption SC 1 requires that  $g_\theta(\mu)$  is differentiable and that  $g_\theta(\mu)$  is achieved by an action. Under these conditions, Proposition 1 implies that strong duality holds and the Problem (10) admits a deterministic optimal solution. In this application, we can characterize an optimal bidding strategy in terms of an optimal bidding function for the static auction without budget constraints, which we denote by  $\beta : \Theta \rightarrow \mathcal{A}$ . That is, given an advertiser with valuation  $\theta$ , the optimal bidding strategy for the static auction (ignoring budget constraints) satisfies

$$\beta(\theta) \in \arg \max_{a \in \mathcal{A}} \{\theta \bar{q}(a) - \bar{m}(a)\}.$$

We have the following result. A proof is provided in Appendix C.2.1.

**Proposition 3.** *Under Assumption SC 1, an optimal solution of (10) is to bid  $\beta(\theta/(1+\mu^*))$  when the value is  $\theta$ , where  $\mu^*$  is the optimal solution of the dual problem of (10).*

If in addition Assumption SC 2 holds, from Lemma 1 we obtain that if  $\mu^1 > 0$ , Assumption 2 holds with  $K = 1/\kappa$  and  $\delta = (\nu\kappa)/2$ . Below, we study the particular cases of second-price auction and first-price auction. Specifically, we provide sufficient conditions on the primitives of the problem for conditions SC 1 and SC 2 to be satisfied.

**Second-price auctions.** In a second-price auction, the bidder with the highest bid wins the auction and pays the second-highest bid. In this case, we reduce the definition of  $\epsilon$  to a random variable capturing the maximum bid of the competitors and take  $\mathcal{E} = \mathbb{R}_+$ . Again, we assume  $\epsilon$  is distributed according to  $f$ , with density function  $f'$ . We assume that ties are broken in favor of the decision maker. The allocation and payment functions are given by  $q(a, \epsilon) = 1_{\{a \geq \epsilon\}}$  and  $m(a, \epsilon) = \epsilon 1_{\{a \geq \epsilon\}}$ , respectively.

Suppose that the following conditions hold:

- The distribution of the maximum competing bid  $f$  is absolutely continuous and strictly increasing.
- The density  $f'$  is locally  $\xi$ -Lipschitz continuous with respect to  $a_\theta^1$  in  $\mathcal{N}(a_\theta^1, \varphi)$ .

If  $\mu^1 > 0$ , it is possible to show—see Lemma A-1 in Appendix A.2—that these conditions are sufficient to apply Proposition 2 and Lemma 2 and, in turn, Corollary 2 holds. This leads to a revenue loss of  $O(\log T)$ .

**First-price auctions.** In a first-price auction, the winner is the highest bidder but pays his bid. Again, we reduce the definition of  $\epsilon$  to a random variable capturing the maximum bid of the

competitors. We assume  $\epsilon$  is distributed according to  $f$ , with density function  $f'$ . The allocation and payment functions are given by  $q(a, \epsilon) = 1_{\{a \geq \epsilon\}}$  and  $m(a, \epsilon) = a 1_{\{a \geq \epsilon\}}$ , respectively.

Suppose the following conditions hold:

- The distribution of the maximum competing bid  $f$  is absolutely continuous.
- The function  $M(a) = a + f(a)/f'(a)$  is strictly increasing.
- The bid  $a_\theta^1$  maximizing  $\theta \bar{q}(a) - (1 + \mu^1) \bar{m}(a)$  is interior. That is, there exists a positive number  $\varphi$  such that  $\mathcal{N}(a_\theta^1, \varphi) \subset \mathcal{A}$  for all  $\theta \in \Theta$ .
- The density  $f'$  is locally  $\xi$ -Lipschitz continuous with respect to  $a_\theta^1$  in  $\mathcal{N}(a_\theta^1, \varphi)$ .

Moreover, if we have  $\mu^1 > 0$ , then it can be proved—see Lemma A-2 in Appendix A.2—that assumptions CA 1-CA 6 hold and, therefore, by Corollary 2 we obtain an  $O(\log T)$  revenue loss.

**Connection with earlier analysis of the CE heuristic.** While the problem of bidding in repeated auctions with budgets has been studied in the past (see, e.g., [Abhishek and Hosanagar 2013](#), [Balseiro et al. 2015](#)), to the best of our knowledge, this is the first result that characterizes the reward loss of a certainty equivalent heuristic with resolving for the advertiser’s decision problem. We note here that this problem is an important illustration of the value of the unified DRC<sup>2</sup> model we propose, and the associated highly general sufficient conditions for the logarithmic loss of the CE heuristic. Even more, we note that our analysis does not require convexity assumptions to hold.

### 6.3 Network Revenue Management

This class of problems was presented in §3.3. Note that here, in contrast to the two problems exposed before, the set of actions is finite. In what follows, we first obtain the value of  $\delta$  that guarantees a constant reward loss for this class of problems and then we comment on the connection of our result with the existing literature on the analysis of the CE heuristic.

Notice that in this case, Problem (8) can be written as follows

$$\begin{aligned}
 J(\rho) = & \max_{\phi_\theta(1) \in [0,1]} \sum_{\theta \in \Theta} p_\theta f_\theta \phi_\theta(1) \\
 \text{s.t.} & \sum_{\theta \in \Theta} p_\theta A_\theta \phi_\theta(1) \leq \rho.
 \end{aligned} \tag{11}$$

Because the set of action consists of  $\mathcal{A} = \{0, 1\}$ , it is enough to consider decision variables  $\phi_\theta(1)$  for all  $\theta \in \Theta$  because  $\phi_\theta(0) = 1 - \phi_\theta(1)$ . Then, for the definition of  $\phi_{\min}^*$  involved in  $\delta$ , we now need to take into account both the non-negative variables and those that are strictly

smaller than one. That is,  $\phi_{\min}^* = \min_{\theta \in \Theta} \{\phi_{\theta}^*(1) : \phi_{\theta}^*(1) > 0\} \wedge \min_{\theta \in \Theta} \{1 - \phi_{\theta}^*(1) : \phi_{\theta}^*(1) < 1\}$ , where  $x \wedge y$  denotes the minimum between  $x$  and  $y$ . Furthermore, if we denote by  $A$  the matrix whose  $\theta^{th}$  column consists of the vector  $p_{\theta}A_{\theta}$ ,  $B_{\rho_1}$  represents the submatrix of  $A$  in which we only consider the columns associated to variables  $0 < \phi_{\theta}^*(1) < 1$ , and we set  $A_{\rho_1} = B_{\rho_1}$ . Then, under Assumption SC 3, if  $\mu^1 > 0$ , Assumption 2 holds with  $K = 0$  and  $\delta = \phi_{\min}^* / \|A_{\rho_1}^{-1}\|$ , where  $\phi_{\min}^* = \min_{\theta \in \Theta} \{\phi_{\theta}^*(1) : \phi_{\theta}^*(1) > 0\} \wedge \min_{\theta \in \Theta} \{1 - \phi_{\theta}^*(1) : \phi_{\theta}^*(1) < 1\}$ .

**Connection with earlier analysis of the CE heuristic.** The question of approximating optimal performance through simple policies has also received significant attention around the network revenue management problem.

Jasin and Kumar [2012] consider a more general version of the NRM problem than ours (they consider a NRM problem with customer choice), and provide a constant revenue loss guarantee for the CE heuristic under the nondegeneracy assumption. In that sense, we recover their result for the setting where there is no customer choice. Wu et al. [2015] study a variant with one resource in which the decision-maker can take one of many actions (not just accept/reject) and show that the CE heuristic attains constant revenue loss under non-degeneracy while  $O(\sqrt{T})$  revenue loss under degeneracy. Bumpensanti and Wang [2018] assume that arrivals follow a Poisson process and show that the CE heuristic could have a  $\Theta(\sqrt{T})$  revenue loss when the fluid problem at  $\rho = C/T$  is degenerate, which highlights the necessity of the non-degeneracy assumption for the loss considered here (we further comment on alternative losses in §7).

## 6.4 Choice-Based Network Revenue Management

This class of problems was presented in §3.4. In this case, the set of actions is finite and the fluid problem can be expressed as

$$\begin{aligned}
 J(\rho) = & \max_{\phi \in \Phi} \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} \phi_{\theta}(a) p_{\theta} \sum_{n \in [N]} m_n g_{\theta a}^n \\
 \text{s.t.} & \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} \phi_{\theta}(a) p_{\theta} \sum_{n \in [N]} A_{\theta}^n g_{\theta a}^n \leq \rho.
 \end{aligned} \tag{12}$$

We can define the vector of variables and the associated matrix involved in the constraints of problem (12) by setting each column to  $p_{\theta} \sum_{n \in [N]} A_{\theta}^n g_{\theta a}^n$ . Then, under assumption SC 3, if the optimal dual solution of (12) for  $\rho = \rho_1$  is interior, we can apply Lemma 3 and Assumption 2 holds for  $K = 0$  and  $\delta$  defined as in the statement of the lemma. Therefore, we obtain a constant bound on the revenue loss of the CE heuristic for the choice-based network revenue management problem.

**Connection with earlier analysis of the CE heuristic.** Our formulation shows that a certainty equivalent heuristic admits strong performance guarantees under Assumption SC 3 and also assuming  $\mu^1 > 0$ . Corollary 3 leads, to the best of our knowledge, to the first such result for this subclass of dynamic assortment optimization problems.

## 6.5 Online Matching

Here we revisit the problem presented in §3.5 and perform a similar analysis to the one we conducted for the NRM problem. That is, we interpret the value of  $\delta$  for this class of problems. First, note that the fluid problem is given by

$$\begin{aligned}
 J(\rho) = \max_{\phi \in \Phi} & \sum_{\theta \in \Theta} \sum_{a \in L} \phi_{\theta}(a) p_{\theta} f_{\theta a} \\
 \text{s.t.} & \sum_{\theta \in \Theta} p_{\theta} \text{diag}(A_{\theta}) \phi_{\theta} \leq \rho,
 \end{aligned} \tag{13}$$

where  $\phi_{\theta}^{\top} = (\phi_{\theta}(1), \dots, \phi_{\theta}(L))$  and where for a vector  $x \in \mathbb{R}^L$ ,  $\text{diag}(x) \in \mathbb{R}^{L \times L}$  is a diagonal matrix with diagonal entry  $i$  given by  $x_i$ .

Here, as in the network revenue management problem, we can define the matrix  $A \in \mathbb{R}^{L \times (L|\Theta)}$  where  $A = (\text{diag}(p_{\theta} A_{\theta}))_{\theta \in \Theta}$  is obtained by horizontally stacking the diagonal matrices  $\text{diag}(p_{\theta} A_{\theta}) \in \mathbb{R}^{L \times L}$ . Let  $A_{\rho_1}$  be the submatrix of  $A$  corresponding to the columns associated to the basic variables. Then, if the dual optimal solution of problem (13) for  $\rho = \rho_1$  is unique and interior, Assumption 2 holds with  $K = 0$  and  $\delta = \phi_{\min}^* / \|A_{\rho_1}^{-1}\|$ , where  $\phi_{\min}^* = \min_{\theta \in \Theta, a \in [L]} \{\phi_{\theta}^*(a) : \phi_{\theta}^*(a) > 0\}$ , and the constant bound for the revenue loss is obtained.

**Connection with earlier analysis of the CE heuristic.** Although the online matching problem has been extensively studied under different setting variants (see, e.g., Karp et al. 1990, Aggarwal et al. 2011, Vera and Banerjee 2021), to the best of our knowledge, our paper is the first studying the performance of the CE heuristic for the online matching problem.

## 6.6 Order Fulfillment

The order fulfillment problem was presented in §3.6 and in this case the fluid problem can be expressed as follows

$$\begin{aligned}
 J(\rho) = \max_{\phi \in \Phi} & \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} p_{\theta} \phi_{\theta}(a) \sum_{l \in \theta} \sum_{n \in N} f_{ln} 1_{\{a_l = n\}} \\
 \text{s.t.} & \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} p_{\theta} \phi_{\theta}(a) 1_{\{a_l = n\}} \leq \rho_{ln}.
 \end{aligned} \tag{14}$$

In this case, as in the previous ones, we can write the constraints of problem (14) in matrix form and, thus, obtain an expression for  $\delta$  involved in the assumption. We do not give an explicit formula for  $\delta$  to avoid introducing more notation. Moreover, if the dual optimal solution of problem (14) for  $\rho = \rho_1$  is unique and interior, Assumption 2 holds with  $K = 0$  and we recover a constant revenue loss bound for the order fulfillment problem.

**Connection with earlier analysis of the CE heuristic.** While there are related literature studying heuristic performances (see, e.g., Acimovic and Graves 2015, Jasin and Sinha 2015, Andrews et al. 2019), to the best of our knowledge our result is the first that provides guarantees for the CE heuristic.

## 7 Concluding Remarks

In this paper, we show that several classical dynamic optimization problems, which are usually studied separately in the literature, possess structural similarities that allow them to be studied under a common framework. More explicitly, we introduce a large class of problems, that we called DRC<sup>2</sup>, which encompasses many notable problems that have been studied individually in the literature.

In addition to presenting this novel, unified model, we exploit the common features of problems in the DRC<sup>2</sup> class to study the performance of a fluid certainty equivalent control heuristic. More specifically, we establish some sufficient conditions to obtain good performance guarantees (see §4 and §5), which depends on whether the set of actions is a continuum or finite. For the former case, we prove that the CE heuristic guarantees  $O(\log T)$  revenue loss, whereas for the latter it guarantees a constant revenue loss. This result leads us to recover certain existing results in the literature for some of the classical problems in DRC<sup>2</sup>, sometimes under weaker conditions, but also to obtain new ones for others (see §6).

The present work opens many avenues for future research. A first one involves the use of different benchmarks to analyze the performance of the CE heuristic or variations of it. While in this paper we use the fluid problem as a benchmark, some authors consider tighter benchmarks for particular problems in the DRC<sup>2</sup> class. One of them is known as *hindsight optimum* (see, e.g. Bumpensanti and Wang 2018, Vera and Banerjee 2021), which is the problem obtained when all uncertainties are revealed in advance. This benchmark is usually considered in the situations without idiosyncratic shocks such as the network revenue management problem or the online matching problem as knowledge of the idiosyncratic shocks confers too much power to the decision maker. (Stronger bounds can be obtained in the presence of idiosyncratic shocks by only granting the



decision maker advance knowledge of the classes.) Another natural benchmark is the optimal value of the dynamic program, which was recently considered by Wang and Wang [2020] to evaluate the performance of the CE heuristic in a dynamic pricing problem. We also refer the reader Vera et al. [2021] for a novel framework that allows using various information augmented benchmarks, enabling to obtain constant regret across a broad family of problems. An interesting direction would be to see if comparisons to tighter benchmarks can be applied or generalized to the  $\text{DRC}^2$  class.

A second important and related avenue is to study alternative heuristics at the  $\text{DRC}^2$  level and weaken the assumptions required for strong performance. For instance, for the dynamic pricing problem Jasin [2014] shows it is possible to attain revenue losses of similar order using a heuristic that solves a single optimization problem at the beginning of the selling horizon and then adjusts controls linearly. There has also been an important stream of recent papers that consider the hindsight optimum benchmark while also relaxing the non-degeneracy assumption in the NRM problem. Reiman and Wang [2008] propose a heuristic that resolves the deterministic problem once at a judiciously chosen time that obtains a revenue loss of order  $o(\sqrt{T})$ . In a related setting, Bumpensanti and Wang [2018] propose a heuristic that has a  $O(1)$  revenue loss. The idea is to only re-solve the deterministic problem only a few selected times, using the approach of Reiman and Wang [2008] recursively, while applying thresholds to the controls. Vera and Banerjee [2021] propose a meta-algorithm based on statistical predictions of the hindsight benchmark that leads to a constant upper bound on the revenue loss for NRM and online matching problems. Thus, an interesting question is whether existing arguments and analyses for these particular problems and heuristics can be lifted to derive similar results for the broader class of  $\text{DRC}^2$  problems that accommodates, e.g., a continuum of actions.

A third important direction pertains to further expanding the class of  $\text{DRC}^2$  problems for which a unified analysis is possible. One direction could be to include additional flexibility in inventory evolution and decisions. For instance, Vera et al. [2020], in the context of a NRM problem, consider possible replenishment of resources and/or the possibility of delays in serving requests, and proves a constant regret of a policy that involves re-solving the fluid problem at each time period.

Finally, another promising direction would be to relax informational assumptions and understand if algorithms that jointly learn and optimize can be analyzed in a unified manner in the  $\text{DRC}^2$  class.

## References

Vibhanshu Abhishek and Kartik Hosanagar. Optimal bidding in multi-item multislot sponsored search auctions. *Operations Research*, 61(4):855–873, 2013.

- Jason Acimovic and Vivek F Farias. The fulfillment-optimization problem. In *Operations Research & Management Science in the Age of Analytics*, pages 218–237. INFORMS, 2019.
- Jason Acimovic and Stephen C Graves. Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management*, 17(1):34–51, 2015.
- Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1253–1264. SIAM, 2011.
- John M Andrews, Vivek F Farias, Aryan I Khojandi, and Chad M Yan. Primal-dual algorithms for order fulfillment at urban outfitters, inc. *Interfaces*, 49(5):355–370, 2019.
- Alessandro Arlotto and Itai Gurvich. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 9(3):231–260, Sep 2019. ISSN 1946-5238. doi: 10.1287/stsy.2018.0028.
- Alessandro Arlotto and Xinchang Xie. Logarithmic regret in the dynamic and stochastic knapsack problem. *arXiv preprint arXiv:1809.02016*, 2018.
- Arash Asadpour, Xuan Wang, and Jiawei Zhang. Online resource allocation with limited flexibility. *Management Science*, 2019.
- Santiago R Balseiro, Omar Besbes, and Gabriel Y Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.
- Fernando Bernstein, A Gürhan Kök, and Lei Xie. Dynamic assortment customization with limited inventories. *Manufacturing & Service Operations Management*, 17(4):538–553, 2015.
- Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.
- Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.
- Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.
- Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. A column generation algorithm for choice-based network revenue management. *Operations research*, 57(3):769–784, 2009.
- Sébastien Bubeck. Convex optimization: Algorithms and complexity. *arXiv preprint arXiv:1405.4980*, 2014.
- Pornpawee Bumpensanti and He Wang. A re-solving heuristic with uniformly bounded loss for network revenue management. *arXiv preprint arXiv:1802.06192*, 2018.
- Nikhil R Devanur, Kamal Jain, and Robert D Kleinberg. Randomized primal-dual analysis of ranking for online bipartite matching. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 101–107. SIAM, 2013.

- Ewalds D'Sylva. O and d seat assignment to maximize expected revenue. *Unpublished internal report, Boeing Commercial Airplane Company, Seattle, WA*, 1982.
- Alexander Erdelyi and Huseyin Topaloglu. Using decomposition methods to solve pricing problems in network revenue management. *Journal of Revenue and Pricing Management*, 10(4):325–343, 2011.
- Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and Shan Muthukrishnan. Online stochastic matching: Beating  $1-1/e$ . In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 117–126. IEEE, 2009.
- Joaquin Fernandez-Tapia, Olivier Guéant, and Jean-Michel Lasry. Optimal real-time bidding strategies. *Applied Mathematics Research eXpress*, 2017(1):142–183, 2017.
- Guillermo Gallego and Garrett Van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations research*, 45(1):24–41, 1997.
- Guillermo Gallego, Garud Iyengar, Robert Phillips, and Abhay Dubey. Managing flexible products on a network. *Available at SSRN 3567371*, 2004.
- Guillermo Gallego, Huseyin Topaloglu, et al. *Revenue management and pricing analytics*, volume 209. Springer, 2019.
- Fred Glover, Randy Glover, Joe Lorenzo, and Claude McMillan. The passenger-mix problem in the scheduled airlines. *Interfaces*, 12(3):73–80, 1982.
- Negin Golrezaei, Hamid Nazerzadeh, and Paat Rusmevichientong. Real-time optimization of personalized assortments. *Management Science*, 60(6):1532–1551, 2014.
- Joseph F Grcar. A matrix lower bound. *Linear algebra and its applications*, 433(1):203–220, 2010.
- Allan Gut. *Probability: a graduate course*, volume 75. Springer Science & Business Media, 2013.
- Pavithra Harsha, Shivaram Subramanian, and Joline Uichanco. Dynamic pricing of omnichannel inventories: Honorable mention—2017 m&som practice-based research competition. *Manufacturing & Service Operations Management*, 21(1):47–65, 2019.
- Stefanus Jasin. Reoptimization and self-adjusting price control for network revenue management. *Operations Research*, 62(5):1168–1178, 2014.
- Stefanus Jasin and Sunil Kumar. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research*, 37(2):313–345, 2012.
- Stefanus Jasin and Amitabh Sinha. An lp-based correlated rounding scheme for multi-item e-commerce order fulfillment. *Operations Research*, 63(6):1336–1351, 2015.
- Samuel Karlin. Stochastic models and optimal policy for selling an asset. *Studies in applied probability and management science*, 1962.

- Richard M Karp, Umesh V Vazirani, and Vijay V Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 352–358, 1990.
- Anton J Kleywegt and Jason D Papastavrou. The dynamic and stochastic knapsack problem. *Operations research*, 46(1):17–35, 1998.
- Sumit Kunnumkal and Huseyin Topaloglu. A stochastic approximation algorithm for making pricing decisions in network revenue management problems. *Journal of Revenue and Pricing Management*, 9(5):419–442, 2010.
- Yanzhe Lei, Stefanus Jasin, and Amitabh Sinha. Joint dynamic pricing and order fulfillment for e-commerce retailers. *Manufacturing & Service Operations Management*, 20(2):269–284, 2018a.
- Yanzhe Murray Lei, Stefanus Jasin, Joline Uichanco, and Andrew Vakhutinsky. Randomized product display (framing), pricing, and order fulfillment for e-commerce retailers. *Stefanus and Uichanco, Joline and Vakhutinsky, Andrew, Randomized Product Display (Framing), Pricing, and Order Fulfillment for E-commerce Retailers (November 9, 2018)*, 2018b.
- Qian Liu and Garrett Van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310, 2008.
- Constantinos Maglaras and Joern Meissner. Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing & Service Operations Management*, 8(2):136–148, 2006.
- Vahideh H Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.
- Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.
- Jason D Papastavrou, Srikanth Rajagopalan, and Anton J Kleywegt. The dynamic and stochastic knapsack problem with deadlines. *Management Science*, 42(12):1706–1718, 1996.
- Iosif Pinelis. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pages 1679–1706, 1994.
- Martin I Reiman and Qiong Wang. An asymptotically optimal policy for a quantity-based network revenue management problem. *Mathematics of Operations Research*, 33(2):257–282, 2008.
- R Tyrrell Rockafellar. *Convex analysis*. Princeton university press, 1970.
- Sheldon M Ross, John J Kelly, Roger J Sullivan, William James Perry, Donald Mercer, Ruth M Davis, Thomas Dell Washburn, Earl V Sager, Joseph B Boyce, and Vincent L Bristow. *Stochastic processes*, volume 2. Wiley New York, 1996.
- Minoru Sakaguchi and Vesa Saario. A class of best-choice problems with full information. *Mathematica japonicae*, 41(2):389–398, 1995.
- Kalyan Talluri and Garrett Van Ryzin. An analysis of bid-price controls for network revenue management. *Management science*, 44(11-part-1):1577–1593, 1998.

- Kalyan Talluri and Garrett Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.
- Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.
- Alberto Vera and Siddhartha Banerjee. The bayesian prophet: A low-regret framework for online decision making. *Management Science*, 67(3):1368–1391, 2021.
- Alberto Vera, Alessandro Arlotto, Itai Gurvich, and Eli Levin. Dynamic resource allocation: The geometry and robustness of constant regret. *Working paper*, 2020.
- Alberto Vera, Siddhartha Banerjee, and Itai Gurvich. Online allocation and pricing: Constant regret via bellman inequalities. *Operations Research*, 2021.
- KW Wang. Optimum seat allocation for multi-leg flights with multiple fare types. In *AGIFORS PROCEEDINGS*-, 1983.
- Yining Wang and He Wang. Constant regret re-solving heuristics for price-based revenue management. *arXiv preprint arXiv:2009.02861*, 2020.
- Huasen Wu, Rayadurgam Srikant, Xin Liu, and Chong Jiang. Algorithms with logarithmic or sublinear regret for constrained contextual bandits. *arXiv preprint arXiv:1504.06937*, 2015.

# Electronic Companion: Appendix for Survey of Dynamic Resource Constrained Reward Collection of Problems: Unified Model and Analysis

Santiago R. Balseiro<sup>1</sup>, Omar Besbes<sup>2</sup>, and Dana Pizarro<sup>3</sup>

## A Additional material

### A.1 Finite set of actions

Problem 8 is a linear program. In particular, introducing the set of slack variables  $\{x_1 \dots x_L\}$ , the standard form is given by

$$\begin{aligned}
 J(\rho) = \max \quad & \sum_{\theta=1}^{\Theta} p_{\theta} \bar{r}_{\theta} \phi_{\theta} \\
 \text{s.t.} \quad & \sum_{\theta \in \Theta} p_{\theta} \bar{y}_{l\theta} \phi_{\theta} + x_l = \rho_l \quad \forall l \in [L] \\
 & \sum_{a \in \mathcal{A}} \phi_{\theta}(a) = 1 \quad \forall \theta \in \Theta \\
 & \phi_{\theta}(a) \geq 0 \quad \forall \theta \in \Theta, \forall a \in \mathcal{A} \\
 & x_i \geq 0 \quad \forall i \in [L].
 \end{aligned} \tag{A-1}$$

### A.2 Dynamic bidding in repeated auctions

#### A.2.1 Second-price auctions

**Lemma A-1.** *If  $f$  absolutely continuous and strictly increasing and  $f'$  is locally  $\xi$ -Lipschitz continuous in  $\mathcal{N}(a_{\theta}^1, \varphi)$ , then conditions CA 1-CA 6 hold.*

*Proof.* Let us see that conditions CA 1- CA 6 hold.

- **Conditions CA 1 and CA 2:** By hypothesis  $f$  is absolutely continuous and therefore both  $\bar{q}$  and  $\bar{m}$  are continuous.
- **Condition CA 3:** For each  $\theta \in \Theta$ , let us define the function  $G_{\theta} : \mathcal{A} \rightarrow \mathbb{R}$  by  $G_{\theta}(a) = \theta f(a) - \int_0^a x df(x)$ . Note that  $G'_{\theta}(a) = (\theta - a)f'(a)$  and  $\theta > 0$ , then  $\lim_{a \searrow 0} G'_{\theta}(a) > 0$  and

<sup>1</sup>Columbia University, Graduate School of Business. Email: [srb2155@columbia.edu](mailto:srb2155@columbia.edu).

<sup>2</sup>Columbia University, Graduate School of Business. Email: [ob2105@columbia.edu](mailto:ob2105@columbia.edu).

<sup>3</sup>Université Toulouse 1 Capitole, Toulouse School of Economics. Email: [dana.pizarro@tse-fr.eu](mailto:dana.pizarro@tse-fr.eu).

$\beta(\theta) \neq 0$ . If  $\theta \neq \Theta_{\max}$ , we also have  $\lim_{a \rightarrow 0} G'_\theta(a) < 0$ , and therefore  $\beta(\theta) \in \arg \max_a G_\theta(a)$  is interior. Assume then first that  $\theta \neq \Theta_{\max}$ . In this case, we can compute the first order condition, obtaining that  $\beta(\theta)$  satisfies the equation

$$(\theta - \beta(\theta))f'(\beta(\theta)) = 0.$$

Therefore, as the cumulative distribution function  $f$  is strictly increasing, the unique optimum is to bid truthfully, as it is known in the literature.

Otherwise, if  $\theta = \Theta_{\max}$ ,  $G_\theta$  is strictly increasing for all  $a \in \mathcal{A}$  and therefore  $\beta(\theta) = \Theta_{\max}$ .

We then conclude that CA 3 holds.

- **Condition CA 4:** Due to the truthfulness property of the second price auction, from Proposition 3, it follows that  $a_\theta^1 = \theta/(1 + \mu^1)$  which belongs to  $(0, \Theta_{\max})$  due to the bid is positive, and thus condition CA 4 holds.
- **Conditions CA 5 and CA 6:** Note first that  $\bar{r}(\theta, a) = \theta f(a) - \bar{m}(\theta, a)$  and  $\bar{y}(\theta, a) = \bar{m}(a)$ , because  $f$  absolutely continuous. Then, it is enough to show the conditions hold for  $h(\theta, \cdot) = \theta f(\cdot)$  and  $\bar{m}(a) = \int_0^a x df(x)$ . Specifically, we will show that if the density function  $f'$  is locally  $\xi$ -Lipschitz continuous in  $\mathcal{N}(a_\theta^1, \varphi)$ , then the gradient of  $h(\theta, \cdot) = \theta f(\cdot)$  is locally  $(\xi\Theta_{\max})$ -Lipschitz continuous in  $\mathcal{N}(a_\theta^1, \varphi)$  and  $\bar{m}'(a) = a f'(a)$  is locally  $((\varphi + \Theta_{\max}/(\mu^1 + 1))\xi + \eta)$ -Lipschitz continuous in  $\mathcal{N}(a_\theta^1, \varphi)$ .

To see the former note that

$$\|\nabla_a h(\theta, a) - \nabla_a h(\theta, a_\theta^1)\| = \theta |f'(a) - f'(a_\theta^1)| \leq \Theta_{\max} \xi |a - a_\theta^1|,$$

where the equality follows from the gradient of  $h$  and the inequality holds due to the locally  $\xi$ -Lipschitz continuity of  $f'$  and because  $\theta \leq \Theta_{\max}$ .

For the latter, we first show that  $f'(a_\theta^1) \leq \eta$  with  $\eta = 1/\varphi + \xi\varphi$ . Because the density  $f'$  is locally  $\xi$ -Lipschitz continuous in  $\mathcal{N}(a_\theta^1, \varphi)$ , we have that  $f'(a_\theta^1) \leq f'(x) + \xi\varphi$  for all  $x \in [a_\theta^1, a_\theta^1 + \varphi]$ . Integrating over  $x \in [a_\theta^1, a_\theta^1 + \varphi]$  we obtain that  $f'(a_\theta^1)\varphi \leq 1 + \xi\varphi^2$  because  $f'$  integrates to at most one. The result follows by dividing by  $\varphi$ . We now show that  $\bar{m}'(a)$

is locally Lipschitz continuous:

$$\begin{aligned}
|\bar{m}'(a) - \bar{m}'(a_\theta^1)| &= |af(a) - a_\theta^1 f(a_\theta^1)| \\
&= |a [f'(a) - f'(a_\theta^1)] + f'(a_\theta^1)(a - a_\theta^1)| \\
&\leq a|f'(a) - f'(a_\theta^1)| + f'(a_\theta^1)|a - a_\theta^1| \\
&\leq \left( \left( \varphi + \frac{\Theta_{\max}}{\mu^1 + 1} \right) \xi + \eta \right) |a - a_\theta^1|,
\end{aligned}$$

where the first inequality holds applying triangle inequality and the last follows from the bound of  $f'(a_\theta^1)$ , together with the locally  $\xi$ -Lipschitz continuity of  $f'$ , the equality  $a_\theta^1 = \theta/(1 + \mu^1)$  and the bound  $a_\theta^1 \leq \Theta_{\max}$ . The proof is completed.  $\square$

### A.3 First-price auctions

**Lemma A-2.** *If  $f$  absolutely continuous,  $M(a) = a + f(a)/f'(a)$  strictly increasing, the bid  $a_\theta^1$  maximizing  $\theta\bar{q}(a) - (\mu^1 + 1)\bar{m}(a)$  is interior, and the density function  $f'$  is locally  $\xi$ -Lipschitz in  $\mathcal{N}(a_\theta^1, \varphi)$ , and  $f'(a_\theta^1)$  is upper bounded by  $\eta$ , conditions CA 1- CA 6 hold.*

*Proof.* As in the lemma for second-price auctions, conditions CA 1 and CA 2 holds because  $f$  is absolutely continuous. On the other hand, the bidder's problem in the static first price auction without budget constraints is to find a bid function  $\beta(\theta)$  maximizing  $(\theta - a)f(a)$ . Note that  $\arg \max \theta f(a) - af(a) = \arg \max \theta' f(a) - (\mu^1 + 1)af(a)$ , where  $\theta' = \theta/(\mu^1 + 1)$ , is interior by hypothesis and then, computing the first order condition, we obtain that  $\beta(\theta)$  should satisfy

$$f'(\beta(\theta))\theta - f(\beta(\theta)) - \beta(\theta)f'(\beta(\theta)) = 0. \quad (\text{A-2})$$

Then, we have  $\theta = \beta(\theta) + f(\beta(\theta))/f'(\beta(\theta)) = M(\beta(\theta))$  and by hypothesis we can compute the inverse of  $M$  and therefore  $\beta(\theta) = M^{-1}(\theta)$ . Thus, payments at the optimal solution are unique and assumption CA 3 holds.

Note that condition CA 4 is directly assumed in the statement of the lemma, and therefore it holds.

It remains to see smoothness of both the expected reward  $\bar{r}(\theta, a) = (\theta - a)f(a)$  and expected payment  $\bar{y}(a) = af(a)$  functions, but it is enough to show that the gradient of  $\bar{y}(\theta, a)$  is locally Lipschitz continuous. To this end, note that

$$|\nabla_a \bar{y}(\theta, a)| = |f(a) + af(a) - f(a_\theta^1) - a_\theta^1 f'(a_\theta^1)| \leq |f(a) - f(a_\theta^1)| + |af(a) - a_\theta^1 f'(a_\theta^1)|,$$

where the last expression in the inequality can be bound by using the local Lipschitz continuity



of  $f'$  together with the upper bound for  $f'$  (as in the case of the second-price action, we have that  $f'(a_\theta^1)$  is bounded) by using the mean value theorem. The remaining algebra is similar to the second-price case and the proof is completed.  $\square$

## B Proof Theorem 1

The goal of this section is to proof Theorem 1, which is our main result regarding the performance of the heuristic CE for the set of DRC<sup>2</sup> problems. To do that, we assume that Assumption 1 and Assumption 2 hold and we first introduce some processes and random variables, as well as technical results, that will be useful to obtain the desired result.

In what follows we will denote by  $y_t$  the resource consumption at time  $t$  if the decision maker follows the policy  $\pi^{\text{CE}}$ . That is,  $y_t = y(\theta_t, a_t^{\pi^{\text{CE}}}, \epsilon_t)$ . Let us consider the process  $\{M_t\}_{t \geq 1}$  up to time  $T$  consisting in, at each time period, the accumulated difference between the resource vector consumption and its expectation, divided by the remaining horizon. More specifically, for each  $t \in [T]$ ,

$$M_t = \sum_{s=1}^t \frac{\mathbb{E}(y_s | \rho_s) - y_s}{T - s}.$$

Let us define the stopping time  $\tau$ . To this end, we need to introduce two random variables. On one hand, we define  $\tau_\delta$  to be the first time  $t$  such that  $M_t$  has  $\ell^2$ -norm greater or equal than  $\delta$ , where  $\delta$  is defined in Assumption 2. That is,

$$\tau_\delta = \min_{t \in [T]} \{t : \|M_t\| \geq \delta\}.$$

If  $\|M_t\|$  is at most  $\delta$  for all  $t \in [T]$ , we set  $\tau_\delta = \infty$ . On the other hand, we define  $\tau_-$  as the first time at which there exists a resource such that its consumption under the policy  $\phi_{\rho_t}^*$  is *close to* over capacity. That is,

$$\tau_- = \min_{t \in [T]} \left\{ t : \exists l \in [L] \text{ s.t. } c_{t,l} - y_l(\theta_t, a^{\phi_{\rho_t}^*}, \epsilon_t) < \bar{y}_\infty \right\}.$$

As above, if  $c_{t,l} - y_l(\theta_t, a^{\phi_{\rho_t}^*}, \epsilon_t)$  is greater or equal to  $\bar{y}_\infty$  for all  $t \in [T]$  and  $l \in [L]$ , we set  $\tau_- = \infty$ . Then, we define the random variable  $\tau$  as the minimum between  $\tau_\delta$  and  $\tau_-$ , and the number of periods  $T$ , i.e.,

$$\tau = \min \{\tau_\delta, \tau_-, T\}.$$

Because  $y(\theta_t, a^{\phi_{\rho_t}^*}, \epsilon_t) \leq \bar{y}_\infty$  by Assumption 1, we have that the actions of the policy up to time  $\tau_-$  are not constrained by resources and they are taken according to an optimal solution of the CE

problem. To see this, note that if  $t = \tau_-$ , then  $c_{t,l} = c_{t-1,l} - y_l(\theta_{t-1}, a^{\phi^*_{\rho_{t-1}}}, \epsilon_{t-1}) \geq \bar{y}_\infty$  and, thus,  $c_{t,l} - y_l(\theta_t, a^{\phi^*_{\rho_t}}, \epsilon_t) \geq 0$ .

Note that both  $\tau_\delta$  and  $\tau_-$  are stopping times with respect to the filtration  $\{\mathcal{F}_t\}_{t \geq 1}$ , with  $\mathcal{F}_t = \sigma(\theta_1, \dots, \theta_t, a_1, \dots, a_t, \epsilon_1, \dots, \epsilon_t)$ , the history up to the end of period  $t$ , and thus we obtain that  $\tau$  is also a stopping time with respect to the same filtration  $\{\mathcal{F}_t\}_{t \geq 1}$ .

Furthermore, the process  $\{M_t\}_{t \geq 1}$  is a martingale with respect to the filtration  $\{\mathcal{F}_t\}_{t \geq 1}$ . In fact, for each  $t$ , from Assumption 1.2 follows that  $\mathbb{E}(y_t | \rho_t) - y_t \leq \bar{y}_\infty < \infty$ , and therefore  $\mathbb{E}(\|M_t\|) < \infty$  for all  $t$ . On the other hand, for each  $t$ , it holds that

$$M_{t+1} - M_t = \frac{\mathbb{E}(y_{t+1} | \rho_{t+1}) - y_{t+1}}{T - t - 1}$$

and  $\mathbb{E}(\mathbb{E}(y_{t+1} | \rho_{t+1}) - y_{t+1} | \mathcal{F}_t) = 0$ , concluding that

$$\mathbb{E}(M_{t+1} | \mathcal{F}_t) = M_t \quad \forall t \geq 1.$$

Since  $\{M_t\}_{t \geq 1}$  is a martingale and  $\tau$  an stopping time, it turns out that the stopped process  $\{M_{t \wedge \tau}\}_{t \geq 1}$  is also a martingale with respect to the filtration  $\{\mathcal{F}_t\}_{t \geq 1}$ .

We are now ready to present some properties for the process and the stopping time defined above, which will be needed to prove the bound for the reward loss. The first result states that up to time  $\tau_\delta$ , the random variable  $M_t$  can be expressed as the difference between  $\rho_{t+1}$  and  $\rho_1$  and that  $\rho_t$  belongs to the ball centered at  $\rho_1$  with radius  $\delta$ .

**Lemma B-1.** *Under Assumption 2, if  $t$  is at most  $\tau_\delta$ , it holds that:*

1.  $\rho_{t+1} - \rho_1 = M_t$
2.  $\|\rho_t - \rho_1\| < \delta$ .

*Proof of Lemma B-1.* We will proceed by induction on  $t$ , dividing the proof into two steps, the first corresponds to prove the base case and the other the induction step.

**Step 1.** Note that for  $t = 1$ , statement 2 of the lemma follows trivially and we are then under the hypothesis of Assumption 2, obtaining  $\mathbb{E}(y_1 | \rho_1) = \rho_1$ . Therefore, we can express  $M_1$  as follows:

$$M_1 = \frac{\mathbb{E}(y_1 | \rho_1) - y_1}{T - 1} = \frac{\rho_1 - y_1}{T - 1}. \quad (\text{B-1})$$

From the definition of  $\rho_1$  and  $\rho_2$ , we have that  $y_1 = \rho_1 T - \rho_2(T - 1)$  and replacing in (B-1)

follows that

$$M_1 = \frac{\rho_1 - \rho_1 T + \rho_2(T-1)}{T-1} = \rho_2 - \rho_1,$$

obtaining the first statement of the lemma and completes the base case.

**Step 2.** Now, assume that Lemma B-1 holds for all  $s$  smaller or equal than a fixed  $t < \tau_\delta$  and let us prove that both statements also hold for  $t+1$ .

As in the base case, we will first prove statement 2 and we then use it to prove statement 1. That is, let us show that  $\|\rho_{t+1} - \rho_1\| < \delta$ . Applying the induction hypothesis to  $t$ , it holds that  $\|\rho_{t+1} - \rho_1\| = \|M_t\|$ . On the other hand,  $t < \tau_\delta$  and thus  $\|M_t\| < \delta$ , concluding that  $\|\rho_{t+1} - \rho_1\| < \delta$ , and the second statement follows.

In the remainder of the proof, we show that  $\rho_{t+2} - \rho_1 = M_{t+1}$ . Note that

$$\rho_{t+2} - \rho_1 = \sum_{s=1}^{t+1} \rho_{s+1} - \rho_s = \sum_{s=1}^{t+1} \frac{\rho_s(T-s+1) - y_s}{T-s} - y_s = \sum_{s=1}^{t+1} \frac{\rho_s - y_s}{T-s},$$

where the first equality is obtained by using a telescoping sum and the second holds because  $\rho_{s+1} = c_{s+1}/(T-s)$ ,  $c_s = \rho_s(T-s+1)$  and  $c_{s+1} = c_s - y_s$ . By the induction hypothesis, together with the statement 2 we already proved for  $s = t+1$ , it holds that  $\|\rho_s - \rho_1\| < \delta$  for all  $s \leq t+1$ . Therefore we can apply Assumption 2 to the expression above obtaining that

$$\rho_{t+2} - \rho_1 = \sum_{s=1}^{t+1} \frac{\mathbb{E}(y_s|\rho_s) - y_s}{T-s} = M_{t+1},$$

and the lemma follows.  $\square$

Since  $\{M_{t \wedge \tau}\}_{t \geq 1}$  is a zero mean martingale with respect to the filtration  $\{\mathcal{F}_t\}_{t \geq 1}$ , a direct consequence of Lemma B-1 is that also the stopped process  $\{\rho_{t \wedge \tau}\}_{t \geq 1}$  is a martingale with respect to the same filtration.

The following lemma is a technical result we need to prove Lemma B-3, which, in turn, states that the expected number of remaining periods after the stopping time  $\tau$  is bounded by a constant. Specifically, the following result gives sufficient conditions on  $t$  to be a lower bound for the stopping time  $\tau_-$ .

**Lemma B-2.** *Assume that Assumption 1 and 2 hold, and define  $T^- = T + 1 - 2\frac{\bar{y}_\infty}{\rho_1 - \delta}$ , where  $\rho_1$  is the smallest component of vector  $\rho_1$ . If  $t \leq T^-$  and  $t < \tau_\delta$ , then  $t < \tau_-$ .*

*Proof of Lemma B-2.* Due to the definition of the stopping time  $\tau_-$ , we have to show that for all  $s \leq$

$t$ , the consumption is at most the available capacity (minus the maximum possible consumption), i.e.,  $y(\theta_s, a^{\phi_{\rho_s^*}}, \epsilon_s) \leq c_s - \mathbf{1}\bar{y}_\infty$ . Take  $s \leq t$ . By hypothesis,  $s < \tau_\delta$  and by Lemma B-1 it holds that  $\|\rho_s - \rho_1\| < \delta$ . In particular,  $|(\rho_s - \rho_1)_l| < \delta \forall l \in [L]$ , obtaining

$$\rho_s > \rho_1 - \mathbf{1}\delta, \quad (\text{B-2})$$

where  $\mathbf{1}$  denotes the vector of ones of size  $L$ . On the other hand, note that

$$c_s > (T - s + 1)(\rho_1 - \mathbf{1}\delta) \geq \frac{2\bar{y}_\infty}{\underline{\rho}_1 - \delta}(\rho_1 - \mathbf{1}\delta) \geq 2 \cdot \mathbf{1}\bar{y}_\infty \geq \mathbf{1}\bar{y}_\infty + y(\theta_s, a^{\phi_{\rho_s^*}}, \epsilon_s),$$

where the strict inequality follows from the definition of  $\rho_s$ , together with inequality (B-2); the second inequality holds because  $t \leq T^-$  and  $\underline{\rho}_1 > \delta$ ; the third due to the definition of  $\underline{\rho}_1$ ; and the last because  $y(\theta_s, a^{\phi_{\rho_s^*}}, \epsilon_s) \leq \mathbf{1}\bar{y}_\infty$  from Assumption 1.2. We then conclude that  $\tau_-$  is greater than  $t$  and the proof is completed.  $\square$

We next prove that the expected number of remaining periods after the stopping time  $\tau$  is upper bounded by a constant that does not depend on  $T$ , which is a key result to obtain the main theorem.

**Lemma B-3.** *If Assumptions 1 and 2 hold, then  $\mathbb{E}(T - \tau) = O(1)$ . More specifically,*

$$\mathbb{E}(T - \tau) < \frac{2\bar{y}_\infty}{\underline{\rho}_1 - \delta} + 14 \frac{\bar{y}_\infty^2}{\delta^2}.$$

*Proof of Lemma B-3.* We will prove the result by bounding the expected value of  $\tau$ , which is equivalent to the expression  $\sum_{t=1}^{\infty} \mathbb{P}(\tau \geq t)$  because  $\tau$  is a non-negative random variable. From the definition of  $\tau$ , the probability of  $\tau$  being greater than  $T$  is zero, and then,

$$\mathbb{E}(\tau) = \sum_{t=1}^T \mathbb{P}(\tau_\delta \wedge \tau_- \geq t) \geq \sum_{t=1}^{T^- - 1} \mathbb{P}(\tau_\delta \wedge \tau_- \geq t) \quad (\text{B-3})$$

where the last inequality follows just splitting the horizon and because probabilities are non-negative.

On the other hand,

$$\begin{aligned}
\sum_{t=1}^{T^- - 1} \mathbb{P}(\tau_\delta \wedge \tau_- \geq t) &= \sum_{t=1}^{T^- - 1} \mathbb{P}\left(\min_{s \in [T]} \{s : \|M_s\| \geq \delta\} \geq t\right) \\
&= \sum_{t=1}^{T^- - 1} \mathbb{P}(\|M_s\| < \delta \forall s \in [t]) \\
&= T^- - 1 - \sum_{t=1}^{T^- - 1} \mathbb{P}\left(\max_{s \in [t]} \|M_s\| \geq \delta\right),
\end{aligned}$$

where the first equality is obtained by Lemma B-2 (since  $t < T^-$ ,  $\tau_\delta \wedge \tau_- = \tau_\delta$ ) and the last one because  $\mathbb{P}(\|M_s\| < \delta \forall s \in [t]) = 1 - \mathbb{P}(\max_{s \in [t]} \|M_s\| \geq \delta)$ .

Then, using the equality above in (B-3) it holds that

$$\mathbb{E}(\tau) \geq T^- - 1 - \sum_{t=1}^{T^- - 1} \mathbb{P}\left(\max_{s \in [t]} \|M_s\| \geq \delta\right).$$

In the remainder of the proof we will upper bound  $\sum_{t=1}^{T^- - 1} \mathbb{P}(\max_{s \in [t]} \|M_s\| \geq \delta)$ , and we proceed by applying Theorem 3.5 in Pinelis [1994]. To this end, note first that  $(\mathbb{R}^L, \|\cdot\|)$  is a separable Banach space, and since  $\|x+y\| + \|x-y\| \leq 2\|x\|^2 + 2\|y\|^2$  holds for all  $x, y \in \mathbb{R}_+$ , it is  $(2, 1)$ -smooth. Define, for each  $t$ , the martingale  $\{M_{t \wedge s}\}_{s \geq 1}$ . From the definition of  $M_s$  it follows that

$$M_s - M_{s-1} = \frac{\mathbb{E}(y_s | \rho_s) - y_s}{T - s},$$

and therefore

$$\sum_{s=1}^t \left\| \frac{\mathbb{E}(y_s | \rho_s) - y_s}{T - s} \right\|_\infty^2 \leq \frac{(2\bar{y}_\infty)^2}{T - t},$$

where the inequality follows from Assumption 1.2, and using that  $\sum_{s=1}^t 1/(T-s)^2 \leq \int_0^t 1/(T-s)^2 < 1/(T-t)$ .

Then, we are under the hypothesis of the theorem mentioned above, and applying it together with the inequality  $\mathbb{P}(\max_{s \in [t]} \|M_s\| \geq \delta) \leq 1$ , we obtain

$$\mathbb{P}\left(\max_{s \in [t]} \|M_s\| \geq \delta\right) \leq 1 \wedge 2 \exp\left(-\frac{\delta^2(T-t)}{8\bar{y}_\infty^2}\right).$$

Summing over  $t$  and using the bound obtained above, we have

$$\begin{aligned} \sum_{t=1}^{T^- - 1} \mathbb{P} \left( \max_{s \in [t]} \|M_s\| \geq \delta \right) &\leq \sum_{t=1}^T \left( 2 \exp \left( -\frac{\delta^2 (T-t)}{8\bar{y}_\infty^2} \right) \wedge 1 \right) \\ &\leq \int_0^T \left( 2 \exp \left( -\frac{\delta^2 (T-t)}{8\bar{y}_\infty^2} \right) \wedge 1 \right) dt \\ &\leq \frac{8\bar{y}_\infty^2}{\delta^2} (\log 2 + 1), \end{aligned}$$

where the second inequality follows from bounding the summation by the integral and the last inequality from Lemma D-1.

Putting all together and using that  $8(\log 2 + 1) \leq 14$  we conclude

$$\mathbb{E}(T - \tau) < T - T^- + 1 + 14 \frac{\bar{y}_\infty^2}{\delta^2} = \frac{2\bar{y}_\infty}{\rho_1 - \delta} + 14 \frac{\bar{y}_\infty^2}{\delta^2},$$

and the desired result is obtained.  $\square$

We are now ready to prove Theorem 1 by combining the technical results already presented.

*Proof of Theorem 1.* We have to bound  $J^* - J^{\text{CE}}$ , which is upper bounded by  $TJ(\rho_1) - J^{\text{CE}}$  because  $J^* \leq TJ(C/T)$  (see, e.g., Gallego and Van Ryzin 1997). Thus, it is enough to bound  $TJ(\rho_1) - J^{\text{CE}}$ . By dividing the horizon from 1 to  $\tau$  and from  $\tau$  to  $T$ , we obtain

$$TJ(\rho_1) - J^{\text{CE}} \leq \underbrace{\mathbb{E} \left( \sum_{t=1}^{\tau} J(\rho_1) - \sum_{t=1}^{\tau} r(\theta_t, a_t^{\text{CE}}, \epsilon_t) \right)}_{(A)} + \underbrace{\mathbb{E} \left( \sum_{t=\tau+1}^T J(\rho_1) \right)}_{(B)}, \quad (\text{B-4})$$

and we then have to bound (A) and (B), which will be done in Part 1 and Part 2, respectively.

**Part 1.** We bound (A) by dividing the proof into three steps. First, we show that the expected reward earned up to time  $\tau$  considering the policy given by the CE heuristic equals the expected reward until time  $\tau$  of the fluid problem for  $\rho = \rho_t$  at time  $t$ . Let us prove that

$$\mathbb{E} \left( \sum_{t=1}^{\tau} r(\theta_t, a_t^{\text{CE}}, \epsilon_t) \right) = \mathbb{E} \left( \sum_{t=1}^{\tau} J(\rho_t) \right). \quad (\text{B-5})$$

To this end, consider the sequence of zero mean, i.i.d. random variables  $\{X_t\}_{t \geq 1}$  given by

$$X_t = r(\theta_t, a_t^{\text{CE}}, \epsilon_t) - \mathbb{E}_{\theta, \epsilon} (r(\theta_t, a_t^{\text{CE}}, \epsilon_t) | \rho_t).$$

Then, letting  $N_s = \sum_{t=1}^s X_t$  it holds that  $\{N_s\}_{s \geq 1}$  is a martingale relative to the filtration  $\{\mathcal{F}_t\}_{t \geq 1}$  previously defined (see e.g. [Ross et al. 1996](#) page 296). Therefore, due to  $\tau$  being a stopping time with respect to the same filtration, we can apply the Martingale Stopping Theorem ([Ross et al. 1996](#), Theorem 6.6.2), which in turns implies that  $\mathbb{E}(N_\tau) = \mathbb{E}(N_1) = 0$ . On the other hand, by the definition of the fluid problem, we have that  $\mathbb{E}_{\theta, \epsilon} (r(\theta_t, a_t^{\text{CE}}, \epsilon_t) | \rho_t) = J(\rho_t)$  for all  $t \leq \tau$ . Then, (B-5) follows from combining these equations.

Second, using (B-5) and applying Assumption 2 (the hypothesis is fulfilled because  $t \leq \tau$  and then by Lemma B-1 we have  $\|\rho_t - \rho_1\| < \delta$ ) we obtain

$$(A) = \mathbb{E} \left( \sum_{t=1}^{\tau} (J(\rho_1) - J(\rho_t)) \right) \leq \underbrace{\mathbb{E} \left( \sum_{t=1}^{\tau} -\nabla J(\rho_1)(\rho_t - \rho_1) \right)}_{(A_1)} + \underbrace{\mathbb{E} \left( \sum_{t=1}^{\tau} \frac{K}{2} \|\rho_t - \rho_1\|^2 \right)}_{(A_2)}.$$

From the linearity of the expectation, for the first term we obtain that

$$(A_1) = -\nabla J(\rho_1) \mathbb{E} \left( \sum_{t=1}^{\tau} \rho_t - \rho_1 \right) = 0,$$

where we used that  $\{\rho_{t \wedge \tau}\}_{t \geq 1}$  is a martingale, and thus the term  $(A_1)$  vanishes in the bound.

In the third and last step, we bound  $(A_2)$ . Using Lemma B-1 have that

$$(A_2) = \frac{K}{2} \mathbb{E} \left( \sum_{t=2}^{\tau} \|M_{t-1}\|^2 \right) \leq \frac{K}{2} \mathbb{E} \left( \sum_{t=2}^T \|M_{t-1}\|^2 \right) = \frac{K}{2} \sum_{t=2}^T \mathbb{E} (\|M_{t-1}\|^2),$$

where the inequality follows because  $t \leq T$  and using that the summands are positive, and the last from the linearity of expectations. Therefore, it is enough to bound  $\mathbb{E} (\|M_t\|^2)$ . Because martingale increments are orthogonal (see e.g. [Gut 2013](#), Chapter 10 Lemma 4.1), we have that

$$\mathbb{E} (\|M_t\|^2) = \mathbb{E} \left\| \sum_{s=1}^t \frac{\mathbb{E}(y_s | \rho_s) - y_s}{T-s} \right\|^2 = \sum_{s=1}^t \frac{1}{(T-s)^2} \mathbb{E} \|\mathbb{E}(y_s | \rho_s) - y_s\|^2. \quad (\text{B-6})$$

Furthermore, by definition of  $\ell^2$ -norm, it holds that

$$\begin{aligned}\mathbb{E} \|\mathbb{E}(y_s|\rho_s) - y_s\|^2 &= \sum_{l \in [L]} \mathbb{E} \left( (\mathbb{E}(y_{s,l}|\rho_s) - y_{s,l})^2 \right) = \sum_{l \in [L]} \mathbb{E}(\text{Var}(y_{s,l}|\rho_s)) \\ &\leq \sum_{l \in [L]} \mathbb{E}(\mathbb{E}(y_{s,l}^2|\rho_s)) = \mathbb{E} \|y_s\|^2 \leq \bar{y}_2^2,\end{aligned}$$

where the inequality follows because  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$  and the last inequality follows from Assumption 1.2.

Using expression (B-6), we finally obtain

$$(A_2) \leq \frac{K}{2} \bar{y}_2^2 \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{(T-s)^2} = \frac{K}{2} \bar{y}_2^2 \sum_{s=1}^{T-1} \sum_{t=s+1}^T \frac{1}{(T-s)^2} = \frac{K}{2} \bar{y}_2^2 \sum_{s=1}^{T-1} \frac{1}{T-s} \leq \frac{K}{2} \bar{y}_2^2 (\gamma + \log(T)),$$

where the second equation follows from exchanging the order of summations and the last inequality because  $\sum_{s=1}^{T-1} (T-s)^{-1} = \sum_{s=1}^{T-1} s^{-1} \leq \gamma + \log(T)$ , where  $\gamma$  is the Euler–Mascheroni constant.

Putting all together we get

$$\mathbb{E} \left( \sum_{t=1}^{\tau} (J(\rho_1) - J(\rho_t)) \right) \leq \frac{K}{2} \bar{y}_2^2 (\gamma + \log T) \leq K \bar{y}_2^2 \log(T), \quad (\text{B-7})$$

because  $\gamma + \log T \leq \log(T)$  for  $T \geq 2$ , and the proof of Part 1 is complete.

**Part 2.** It only remains to bound the second term in (B-4). Note that

$$\mathbb{E} \left( \sum_{t=\tau+1}^T J(\rho_1) \right) = \mathbb{E}(T - \tau) J(C/T) \leq \left[ \frac{2\bar{y}_\infty}{\rho_1 - \delta} + \frac{14\bar{y}_\infty^2}{\delta^2} \right] J(C/T), \quad (\text{B-8})$$

where the inequality follows from applying Lemma B-3.

**Putting everything together.** Using (B-7) together with (B-8) in (B-4) we get

$$J^* - J^{\text{CE}} \leq \bar{y}_2^2 K \log T + \left[ \frac{2\bar{y}_\infty}{\rho_1 - \delta} + \frac{14\bar{y}_\infty^2}{\delta^2} \right] J(C/T),$$

and the result follows. □



## C Missing proofs

### C.1 Proofs of Section 5

#### C.1.1 Proof of Proposition 1

*Proof of Proposition 1.* We divide the proof into two steps. First, we prove that an optimal solution to the dual problem exists, namely  $\mu^*$ . Then, we define  $\phi^*$  properly and we apply Proposition 5.1.5 in Bertsekas [1997] to prove that  $\phi^*$  is primal solution and  $\mu^*$  is in fact a Lagrangian multiplier and that therefore there is no duality gap, obtaining the desire result.

**Step 1.** Note that for each  $\theta \in \Theta$ ,  $g_\theta$  is convex because it is defined as the supremum of a family of linear functions and therefore the dual problem is a convex problem.

To prove the existence of optimal dual solution  $\mu^*$ , we first prove  $\Psi_\rho$  is differentiable (and thus continuous) and then we argue that the domain of the dual problem can be restricted to a compact set, achieving the result applying the extreme value theorem. For the former, it is enough to note that  $g_\theta(\mu)$  is differentiable by Assumption SC 1, and in particular we have that  $\Psi_\rho$  is continuous. On the other hand, we can prove that we can restrict the domain of the dual problem to the hypercube  $[0, \bar{\mu}_\rho]^L$ , for  $\bar{\mu}_\rho = \bar{r}_\infty / \bar{\rho}$ , where  $\bar{\rho} = \min_{l \in [L]} \rho_l$  and  $\bar{r}_\infty$  is the positive real number provided by Assumption 1.1. We have that  $\bar{\mu}_\rho < \infty$  because  $\rho > 0$ . Let us check that every  $\mu \notin [0, \bar{\mu}_\rho]^L$  is suboptimal. Take  $\mu \notin [0, \bar{\mu}_\rho]^L$ , and define  $L_1 = \{l \in [L] : \mu_l > \bar{\mu}_\rho\}$  the components of  $\mu$  greater than  $\bar{\mu}_\rho$ . Then, we have

$$\Psi_\rho(\mu) \geq \rho^\top \mu \geq \sum_{l \in L_1} \rho_l \bar{\mu}_l = \sum_{l \in L_1} \bar{r}_\infty \frac{\rho_l}{\rho} \geq \bar{r}_\infty \geq \Psi_\rho(0),$$

where the first inequality holds because  $\bar{r}(\theta, a_0) = \bar{y}(\theta, a_0) = 0$ , for all  $\theta \in \Theta$  and therefore  $g_\theta(\mu) \geq 0$ ; the second follows from the non-negativity of vectors  $\mu$  and  $\rho$ , the third inequality holds because  $L_1$  contains at least one element and  $\rho_l \geq \rho$ , and the last one follows because  $\Psi_\rho(0) = \sum_{\theta \in \Theta} p_\theta \max_{a \in \mathcal{A}} \bar{r}(\theta, a)$  and  $\bar{r}_\infty \geq \bar{r}(\theta, a)$  for all  $\theta \in \Theta$ , and  $a \in \mathcal{A}$ . Then, we have  $\Psi_\rho(0) \leq \Psi_\rho(\mu)$  and together with the extreme value theorem we conclude that for each  $\rho > 0$  there exist  $\mu^*$  optimal dual solution satisfying  $\mu^* \in [0, \bar{\mu}_\rho]^L$ .

**Step 2.** Given  $\rho > 0$ , take  $\mu^*$  an optimal dual solution and, for each  $\theta \in \Theta$ , define  $\phi_\theta^*$  a distribution that assigns probability one to an action  $a_\theta^* \in \arg \max_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - \mu^{*\top} \bar{y}(\theta, a) \right\}$ . Such actions are guaranteed to exist by Assumption SC 1. Let us now show that  $(\phi^*, \mu^*)$  is an optimal solution-Lagrange multiplier pair. We will proceed by using Proposition 5.1.5 in Bertsekas [1997]. That is,

we need to check primal and dual feasibility, Lagrangian optimality and complementary slackness.

1. *Primal and dual feasibility.* Dual feasibility follows because  $\mu^* \geq 0$ . For primal feasibility, note that from the envelope theorem applied to  $g_\theta$  (see, e.g., Theorem 1 in [Milgrom and Segal 2002](#)), the gradient of  $\Psi_\rho$  evaluated at  $\mu^*$  is given by

$$\nabla \Psi_\rho(\mu^*) = \rho + \sum_{\theta \in \Theta} p_\theta \nabla g_\theta(\mu^*) = \rho - \sum_{\theta \in \Theta} p_\theta \bar{y}(\theta, a_\theta^*), \quad (\text{C-1})$$

where we used that by [SC 1](#) the value function  $g_\theta(\mu)$  is differentiable and achieved for an action  $a_\theta^*$ , and that the gradient of  $\bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a)$  with respect to  $\mu$  exists and is given by  $\bar{y}(\theta, a)$ . Because  $\mu^*$  is an optimal dual solution and the constraint set is convex, by [Proposition 2.1.2](#) in [Bertsekas \[1997\]](#), the first-order conditions are given by

$$\nabla \Psi_\rho(\mu^*)^\top (\mu - \mu^*) \geq 0, \quad \forall \mu \in \mathbb{R}_+^L. \quad (\text{C-2})$$

Letting  $\mu_l \rightarrow \infty$ , we obtain that  $\nabla \Psi_\rho(\mu^*) \geq 0$ , which, in turn, implies that  $\sum_{\theta \in \Theta} p_\theta \bar{y}(\theta, a_\theta^*) \leq \rho$  by [\(C-1\)](#). Primal feasibility follows.

2. *Complementary slackness.* If  $\mu_l^* = 0$ , we trivially have  $(\nabla \Psi_\rho(\mu^*))_l \mu_l^* = 0$  and complementary slackness follows. If  $\mu_l^* > 0$ , we can take  $\nu > 0$  with  $\mu_l^* + \nu$  and  $\mu_l^* - \nu$  belonging to  $\mathbb{R}_+$ . Using [\(C-2\)](#), we obtain that  $(\nabla \Psi_\rho(\mu^*))_l \nu \geq 0$  and  $(\nabla \Psi_\rho(\mu^*))_l \nu \leq 0$ . Thus, it holds that  $(\nabla \Psi_\rho(\mu^*))_l = 0$  and complementary slackness follows.

3. *Lagrangian optimality.* Note that

$$\begin{aligned} \arg \max_{\phi \in \Phi} \mathcal{L}(\phi, \mu^*) &= \arg \max_{\phi \in \Phi} \left\{ \mu^* \rho + \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \left( \bar{r}(\theta, a) - \mu^{*\top} \bar{y}(\theta, a) \right) d\phi_\theta(a) \right\} \\ &= \left\{ \arg \max_{\phi_\theta \in \Delta(\mathcal{A})} \int_{\mathcal{A}} \left( \bar{r}(\theta, a) - \mu^{*\top} \bar{y}(\theta, a) \right) d\phi_\theta(a) \right\}_{\theta \in \Theta} \\ &= \left\{ \arg \max_{a \in \mathcal{A}} g_\theta(\mu^*) \right\}_{\theta \in \Theta}, \end{aligned}$$

where the second equality holds because we can separate the problem for each  $\theta$ . But note that  $g_\theta(\mu^*)$  is maximized at  $a_\theta^*$  and thus we have Lagrangian optimality.

Therefore, the four conditions hold and the proof is complete.  $\square$

### C.1.2 Proof of Lemma 1

*Proof of Lemma 1.* We have to show that if  $\rho \in \mathcal{N}(\rho_1, \delta)$ , with  $\delta = (\nu\kappa)/2$  then it holds that

1.  $J(\rho) \geq J(\rho_1) + \nabla J(\rho_1)(\rho - \rho_1) - \frac{1}{2\kappa}\|\rho - \rho_1\|^2$ ,
2.  $\sum_{\theta \in \Theta} p_\theta \int_a \bar{y}(\theta, a) d\phi_\theta^*(a) = \rho$ .

**Part 1.** We first extend the strong convexity lower bound of  $g_\theta$  to the entire domain. Given  $\theta \in \Theta$ , by SC 2,  $g_\theta$  admits a  $\kappa$ -LUQ envelope in  $I^\nu = \mathcal{N}(\mu^1, \nu)$ . Then, for all  $\mu \in I^\nu$ .

$$g_\theta(\mu) \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \frac{\kappa}{2}\|\mu - \mu^1\|^2. \quad (\text{C-3})$$

We next extend the lower bound to every feasible dual variable. Consider  $\mu \geq 0$  with  $\mu \notin I^\nu$ . Take  $\alpha$  such that  $\alpha\mu + (1 - \alpha)\mu^1 = \hat{\mu}$  where  $\hat{\mu}$  is in the boundary of the ball  $I^\nu$ , i.e.,  $\hat{\mu} \in \partial I^\nu$ . The latter is possible because  $\mu \notin I^\nu$ . Note that  $\hat{\mu} - \mu^1 = \alpha(\mu - \mu^1)$ . Taking  $\ell^2$ -norm in both sides, we get that  $\alpha = \|\hat{\mu} - \mu^1\|/\|\mu - \mu^1\|$ . Moreover,  $\alpha \in (0, 1)$  because  $\mu^1$  is interior since  $\nu > 0$  and  $\mu \notin I^\nu$ . Because  $g_\theta$  is convex, we have

$$\alpha g_\theta(\mu) + (1 - \alpha)g_\theta(\mu^1) \geq g_\theta(\alpha\mu + (1 - \alpha)\mu^1) = g_\theta(\hat{\mu}),$$

which can be reordered to give

$$\begin{aligned} g_\theta(\mu) &\geq \frac{1}{\alpha}g_\theta(\hat{\mu}) - \frac{1 - \alpha}{\alpha}g_\theta(\mu^1) \\ &\geq \frac{1}{\alpha}g_\theta(\mu^1) + \frac{1}{\alpha}\nabla g_\theta(\mu^1)^\top (\hat{\mu} - \mu^1) + \frac{\kappa}{2\alpha}\|\hat{\mu} - \mu^1\|^2 - \frac{1 - \alpha}{\alpha}g_\theta(\mu^1) \\ &= g_\theta(\mu^1) + \frac{1}{\alpha}\nabla g_\theta(\mu^1)^\top (\hat{\mu} - \mu^1) + \frac{\kappa}{2\alpha}\|\hat{\mu} - \mu^1\|^2, \\ &= g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \frac{\kappa\nu}{2}\|\mu - \mu^1\|, \end{aligned}$$

where the second inequality follows by (C-3) with  $\mu = \hat{\mu}$  and the second equality from  $\hat{\mu} - \mu^1 = \alpha(\mu - \mu^1)$  together with  $\|\hat{\mu} - \mu^1\|^2 = \alpha\|\mu - \mu^1\|\|\hat{\mu} - \mu^1\| = \alpha\nu\|\mu - \mu^1\|$  since  $\|\hat{\mu} - \mu^1\| = \nu$  because  $\hat{\mu}$  lies at the boundary of the ball  $I^\nu$ . Combining both cases we obtain that

$$g_\theta(\mu) \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \kappa\ell(\mu - \mu^1), \quad (\text{C-4})$$

where

$$\ell(z) = \begin{cases} \frac{1}{2}\|z\|^2 & \text{if } \|z\| \leq \nu \\ \frac{\nu}{2}\|z\| & \text{otherwise.} \end{cases}$$

The function  $\ell : \mathbb{R}^L \rightarrow \mathbb{R}$  is, unfortunately, not convex. We restore convexity while preserving the lower bound by shrinking the radius of ball in half and shifting down the cone outside the ball. In particular, consider the function  $f^* : \mathbb{R}^L \rightarrow \mathbb{R}$  given by

$$f^*(z) = \begin{cases} \frac{1}{2}\|z\|^2 & \text{if } \|z\| \leq \frac{\nu}{2} \\ \frac{\nu}{2}\|z\| - \frac{1}{8}\nu^2 & \text{otherwise.} \end{cases}$$

The function is easily shown to be convex and satisfies  $\ell(z) \geq f^*(z)$  for all  $z \in \mathbb{R}^L$ . (Actually,  $f^*(z)$  is the largest convex function satisfying  $\ell(z) \geq f^*(z)$ .) Putting everything together we obtain that

$$g_\theta(\mu) \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1). \quad (\text{C-5})$$

**Part 2.** Let us prove the first statement. Using this lower bound on  $g_\theta$  to bound  $J(\rho)$  we have

$$\begin{aligned} J(\rho) &\geq \min_{\mu \geq 0} \left\{ \rho^\top \mu + \sum_{\theta \in \Theta} p_\theta \left( g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1) \right) \right\} \\ &= \rho_1^\top \mu^1 + \sum_{\theta \in \Theta} p_\theta g_\theta(\mu^1) + (\mu^1)^\top (\rho - \rho_1) + \min_{\mu \geq 0} \left\{ (\rho - \rho_1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1) \right\} \\ &= J(\rho_1) + \nabla J(\rho_1)^\top (\rho - \rho_1) + \underbrace{\min_{z \geq \mu^1} \left\{ (\rho - \rho_1)^\top z + \kappa f^*(z) \right\}}_{(E)}, \end{aligned}$$

where the first equality holds because the dual solution is interior, i.e.,  $\mu^1 > 0$ , and then the first order conditions for the dual problem imply that  $\rho_1 + \sum_{\theta \in \Theta} p_\theta \nabla g_\theta(\mu^1) = 0$  and the last equality follows from performing the change of variables  $\mu - \mu^1 = z$  and because  $J(\rho_1) = \rho_1^\top \mu^1 + \sum_{\theta \in \Theta} p_\theta g_\theta(\mu^1)$  together with  $\nabla J(\rho_1) = \mu^1$  from the envelope theorem. Note that envelope theorem applies to  $J$  because both  $g_\theta$  and  $J$ —in a neighborhood of  $\rho_1$ —are continuously differentiable (it follows from Assumption SC 1, and from the concavity of  $J$  and Theorem 25.5 in Rockafellar 1970, respectively).

In the remainder of the proof we lower bound the error term  $(E)$ . We have that

$$(E) \geq \min_{z \in \mathbb{R}^L} \left\{ (\rho - \rho_1)^\top z + \kappa f^*(z) \right\} = -\kappa \max_{z \in \mathbb{R}^L} \left\{ \left( \frac{\rho_1 - \rho}{\kappa} \right)^\top z - f^*(z) \right\} = -\kappa f^{**} \left( \frac{\rho_1 - \rho}{\kappa} \right),$$

where the first inequality follows from relaxing the constraint that  $z \geq -\mu^1$ , the first equality from factoring  $\kappa > 0$  and changing the direction of the optimization, and the last one by denoting  $f^{**}(x) = \max_{z \in \mathbb{R}^L} \{x^\top z - f^*(z)\}$  to be the convex conjugate of  $f^*(z)$ . Invoking Lemma D-2 with  $\varphi = \nu/2$ , we obtain that  $f^{**}(x) = f(x)$  with  $f(x) = \frac{1}{2}\|x\|^2$  if  $\|x\| \leq \nu/2$  and  $f(x) = \infty$  otherwise

because the function  $f(x)$  is proper (because  $\nu > 0$ ), closed, and convex (because every squared norm is convex). Therefore, if  $\|\rho - \rho_1\| \leq \nu\kappa/2$ , we have

$$(E) \geq -\kappa f\left(\frac{\rho_1 - \rho}{\kappa}\right) = -\frac{1}{2\kappa}\|\rho - \rho_1\|^2.$$

Putting it all together, we conclude that for  $\rho$  such that  $\|\rho - \rho_1\| \leq (\nu\kappa)/2$ ,

$$J(\rho) \geq J(\rho_1) + \nabla J(\rho_1)^\top(\rho - \rho_1) - \frac{1}{2\kappa}\|\rho - \rho_1\|^2.$$

**Part 3.** Let us now see the second statement. More specifically, we will show that if  $\|\rho_1 - \rho\| \leq \delta$  then

$$\sum_{\theta \in \Theta} p_\theta \int_{a \in \mathcal{A}} \bar{y}(\theta, a) d\phi_\theta^*(a) = \rho,$$

where  $\phi^*$  is an optimal solution of the fluid problem when the resource vector is  $\rho$ .

Note that by complementary slackness we know that for all  $i \in [L]$

$$\mu_i \left( \rho - \sum_{\theta \in \Theta} p_\theta \int_{a \in \mathcal{A}} \bar{y}(\theta, a) d\phi_\theta^*(a) \right)_i = 0,$$

where  $\mu$  is the optimal solution of (5). That is,

$$\mu \in \arg \min_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu),$$

with  $\Psi_\rho(\mu) = \rho^\top \mu + \sum_{\theta \in \Theta} p_\theta g_\theta(\mu)$  the Lagrange dual function. We prove the result by showing that  $\mu_i > 0 \forall i \in [L]$  for every optimal solution when the resource vector  $\rho$  satisfies  $\|\rho - \rho_1\| < \delta$ .

For all  $\mu \geq 0$ , we have from (C-4) that

$$\begin{aligned} \Psi_\rho(\mu) &\geq \rho^\top \mu + \sum_{\theta \in \Theta} p_\theta \left( g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top(\mu - \mu^1) + \kappa \ell(\mu - \mu^1) \right) \\ &= \rho^\top \mu + \sum_{\theta \in \Theta} p_\theta g_\theta(\mu^1) + (\mu^1 - \mu)^\top \rho_1 + \kappa \ell(\mu - \mu^1) \\ &= \Psi_\rho(\mu^1) + \kappa \ell(\mu - \mu^1) + (\mu^1 - \mu)^\top(\rho_1 - \rho). \end{aligned}$$

Let us define  $U_0 = \{\mu \in \mathbb{R}_+^L : \mu_j = 0 \text{ for some } j \in [L]\}$  the set of dual feasible solutions with some zero component. We will prove that if  $\rho$  is suitably chosen, then  $\min_{\mu \in U_0} \Psi_\rho(\mu) > \Psi_\rho(\mu^1)$  and therefore the optimal solution of the dual problem satisfies  $\mu_i > 0$  for all  $i \in [L]$  since all dual

solutions in the boundary  $U_0$  are strictly dominated by  $\mu^1$ . To this end, it is sufficient to show that

$$(I) = \min_{\mu \in U_0} \left\{ \kappa \ell(\mu - \mu^1) + (\mu^1 - \mu)^\top (\rho_1 - \rho) \right\} > 0.$$

Suppose that  $\|\mu - \mu^1\| \leq \nu$ . In this case,  $\ell(z) = \frac{1}{2}\|z\|^2$ . Before proceeding we note that  $\min_{\mu \in U_0} \|\mu - \mu^1\| \geq \underline{\mu}$ . In fact, if  $\mu \in U_0$  there exist  $j \in [L]$  such that  $\mu_j = 0$  and therefore  $(\mu - \mu^1)_j = -\mu_j^1$ , obtaining that  $\|\mu - \mu^1\| \geq \mu_j^1 \geq \underline{\mu}$ . Then, using Cauchy-Schwartz we obtain that

$$\begin{aligned} (I) &\geq \min_{\mu \in U_0} \left\{ \frac{\kappa}{2} \|\mu - \mu^1\|^2 - \|\rho_1 - \rho\| \|\mu - \mu^1\| \right\} = \min_{\mu \in U_0} \|\mu - \mu^1\| \left( \frac{\kappa}{2} \|\mu^1 - \mu\| - \|\rho_1 - \rho\| \right) \\ &\geq \min_{\mu \in U_0} \|\mu - \mu^1\| \left( \frac{\kappa \underline{\mu}}{2} - \|\rho_1 - \rho\| \right) \geq \underline{\mu} \left( \frac{\kappa \underline{\mu}}{2} - \|\rho_1 - \rho\| \right) > 0, \end{aligned}$$

where the second inequality follows because  $\min_{\mu \in U_0} \|\mu - \mu^1\| \geq \underline{\mu}$ , the third and fourth inequalities because  $\|\rho_1 - \rho\| < (\nu \kappa)/2 \leq (\underline{\mu} \kappa)/2$  because  $\nu \leq \underline{\mu}$ .

Suppose that  $\|\mu - \mu^1\| > \nu$ . In this case,  $\ell(z) = \nu \|z\|/2$  and using again Cauchy-Schwartz we obtain

$$(I) \geq \min_{\mu \in U_0} \left( \frac{\kappa \nu}{2} - \|\rho_1 - \rho\| \right) \|\mu - \mu^1\| \geq \left( \frac{\kappa \nu}{2} - \|\rho_1 - \rho\| \right) \nu > 0,$$

where the third inequality follows because  $\|\rho - \rho_1\| < (\kappa \nu)/2$ . The result follows.  $\square$

### C.1.3 Proof of Lemma 2

*Proof of Lemma 2.* We will show that under assumptions CA 1-CA 6,  $g_\theta(\mu)$  admits a  $\kappa$ -LUQ envelope in  $\mathcal{N}(\mu^1, \nu)$  for  $\nu = \kappa \varphi / \sigma$  and  $\kappa = \kappa_r + (\nu + \|\mu^1\|) \|\kappa_y\|$ .

From CA 5,  $\bar{r}(\theta, \cdot)$  admits a  $\kappa_r$ -LDQ envelope in  $\mathcal{N}(a_\theta^1, \varphi)$ . That is, for all  $\theta \in \Theta$ ,

$$\bar{r}(\theta, a) \geq \bar{r}(\theta, a_\theta^1) + \nabla \bar{r}(\theta, a_\theta^1)^\top (a - a_\theta^1) - \frac{\kappa_r}{2} \|a - a_\theta^1\|^2 \quad \forall a \in \mathcal{N}(a_\theta^1, \varphi).$$

On the other hand, from CA 6,  $\bar{y}(\theta, \cdot)$  admits a  $\kappa_y$ -UUQ in  $\mathcal{N}(a_\theta^1, \varphi)$ . That is, for all  $\theta \in \Theta$ ,

$$\bar{y}(\theta, a) \leq \bar{y}(\theta, a_\theta^1) + \nabla \bar{y}(\theta, a_\theta^1) (a - a_\theta^1) + \frac{\kappa_y}{2} \|a - a_\theta^1\|^2 \quad \forall a \in \mathcal{N}(a_\theta^1, \varphi).$$

Combining these two inequalities we obtain that, for  $a \in \mathcal{N}(a_\theta^1, \varphi)$ , we have

$$\begin{aligned}
& \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \\
& \geq \bar{r}(\theta, a_\theta^1) - \mu^\top \bar{y}(\theta, a_\theta^1) + (\nabla \bar{r}(\theta, a_\theta^1) - \nabla \bar{y}(\theta, a_\theta^1)^\top \mu)^\top (a - a_\theta^1) - \frac{\kappa_r + \mu^\top \kappa_y}{2} \|a - a_\theta^1\|^2 \\
& \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + (\nabla \bar{y}(\theta, a_\theta^1)^\top (\mu^1 - \mu))^\top (a - a_\theta^1) - \frac{\kappa_r + \mu^\top \kappa_y}{2} \|a - a_\theta^1\|^2, \quad (\text{C-6})
\end{aligned}$$

where the equality follows because  $g_\theta(\mu^1) = \bar{r}(\theta, a_\theta^1) - (\mu^1)^\top \bar{y}(\theta, a_\theta^1)$ , because  $\nabla \bar{r}(\theta, a_\theta^1) = \nabla \bar{y}(\theta, a_\theta^1)^\top \mu^1$  from the first order condition of  $g_\theta$  (by assumption CA 4,  $a_\theta^1$  is interior), and because  $\nabla g_\theta(\mu^1) = -\bar{y}(\theta, a_\theta^1)$  from the envelope theorem applied to  $g_\theta$  (using compactness of  $\mathcal{A}$  and Assumptions CA 1-CA 3 we can apply Corollary 4 in Milgrom and Segal 2002).

We now proceed to bound  $g_\theta(\mu)$ . Fix  $\mu \in \mathcal{N}(\mu^1, \nu)$ . We have

$$\begin{aligned}
g_\theta(\mu) &= \max_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \right\} \geq \max_{\{a: \|a - a_\theta^1\| \leq \varphi\}} \left\{ \bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \right\} \\
&\geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \max_{\{x: \|x\| \leq \varphi\}} \left\{ \left( \nabla \bar{y}(\theta, a_\theta^1)^\top (\mu^1 - \mu) \right)^\top x - \frac{\kappa}{2} \|x\|^2 \right\} \\
&= g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + h \left( \nabla \bar{y}(\theta, a_\theta^1)^\top (\mu^1 - \mu) \right),
\end{aligned}$$

where the first inequality follows from restricting the optimization to  $a \in \mathcal{A}$  such that  $\|a - a_\theta^1\| \leq \varphi$ ; the second from (C-6), using Cauchy-Schwartz and the triangle inequality to bound  $\mu^\top \kappa_y \leq \|\mu\| \|\kappa_y\| \leq (\|\mu - \mu^1\| + \|\mu^1\|) \|\kappa_y\| \leq (\nu + \|\mu^1\|) \|\kappa_y\|$ , setting  $\kappa = \kappa_r + (\nu + \|\mu^1\|) \|\kappa_y\|$ , and making the change of variables  $x = a - a_\theta^1$ ; the second equality follows from setting  $h(z) = \max_{\{x: \|x\| \leq \varphi\}} \left\{ z^\top x - \frac{\kappa}{2} \|x\|^2 \right\}$ .

Note that  $h(z)$  is the convex conjugate of  $\kappa f(x)$  with  $f(x)$  defined in the statement of Lemma D-2. Using that the convex conjugate of the scaled function  $\kappa f(x)$  is given by  $\kappa f^*(z/\kappa)$  (see, e.g., Boyd and Vandenberghe 2009, Section 3.3.2) together with Lemma D-2 and that the dual norm to the Euclidean norm is the Euclidean norm, we obtain that

$$h(z) = \begin{cases} \frac{1}{2\kappa} \|z\|^2 & \text{if } \|z\| \leq \kappa\varphi \\ \varphi \|z\| - \frac{1}{2}\kappa\varphi^2 & \text{otherwise.} \end{cases}$$

Given that  $\sigma$  is a lower bound on the smallest singular value of  $\nabla \bar{y}(\theta, a_\theta^1)$ , it holds that  $\|\nabla \bar{y}(\theta, a_\theta^1)^\top z\| \geq \sigma \|z\|$  for all  $z$  (see Grcar 2010, Lemma 3.3). We can equivalently write  $h(z) = \frac{1}{\kappa} \min(\kappa\varphi, \|z\|) \cdot \|z\| - \frac{1}{2\kappa} \min(\kappa\varphi, \|z\|)^2$ , which implies that  $h(\tilde{z}) \geq h(z)$  whenever  $\|\tilde{z}\| \geq \|z\|$  since  $h$  is increasing in  $\|z\|$ . This yields that  $h(\nabla \bar{y}(\theta, a_\theta^1)^\top z) \geq h(\sigma z)$  for all  $z \in \mathbb{R}^L$ .

Therefore, if  $\mu \in \mathcal{N}(\mu^1, \nu)$ , we have that  $\|\mu - \mu^1\| \leq \nu = \kappa\varphi/\sigma$ , in which case  $h(\sigma z) = (\sigma^2/2\kappa)\|z\|^2$ , which yields

$$g_\theta(\mu) \geq g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + \frac{\sigma^2}{2\kappa} \|\mu - \mu^1\|^2,$$

and the result follows.  $\square$

### C.1.4 Proof of Lemma 3

*Proof of Lemma 3.* Let us prove that the first statement of Assumption 2 holds by showing that the function  $J(\cdot)$  is linear over the set  $\mathcal{N}(\rho_1, \delta) = \{\rho : \|\rho - \rho_1\| \leq \delta\}$  with  $\delta$  given in the statement of this result. Take  $\rho \in \mathcal{N}(\rho_1, \delta)$ . That is,  $\rho = \rho_1 + \epsilon v$  for some  $v$  unitary vector and  $\epsilon$  a positive real number smaller than  $\delta$ . Let  $\xi^\top = (\rho_1, \mathbf{1})$  be the corresponding right hand side of problem (A-1) for  $\rho = \rho_1$  and  $u^\top = (v, \mathbf{0})$ , where  $\mathbf{1}$  and  $\mathbf{0}$  denote a vector of ones and zeros, with a proper size, respectively. Note that in this case,  $B^{-1}u = B_{\rho_1}^{-1}v$  because the last  $|\Theta|$  components of  $u$  are zero and thus  $\|B^{-1}u\| = \|B_{\rho_1}^{-1}v\| \leq \|B_{\rho_1}^{-1}\|$  where the last equality hold because  $v$  is an unitary vector. Then, by Lemma D-3 we have that if  $0 \leq \epsilon \leq \frac{\phi_{\min}^*}{\|B_{\rho_1}^{-1}\|}$ ,  $B$  is an optimal basis for the standard problem with right hand side  $\xi + \epsilon u$  and therefore the optimal basic variable vector, namely  $\phi_B$ , can be computed as  $B^{-1}(\xi + \epsilon u)$ . Let us define  $c$  the objective function coefficient vector of problem (A-1). That is,  $c_i = p_i \bar{r}_i$  if  $i \in \Theta$  and 0 otherwise. Thus, calling  $c_B$  to the coefficient vector associated to the basic variables, it holds that

$$\begin{aligned} J(\rho) &= c_B^\top B^{-1}(\xi + \epsilon u) \\ &= c_B^\top B^{-1}\xi + \epsilon c_B^\top B^{-1}u \\ &= c_B^\top B^{-1}\xi + c_B^\top B_{\rho_1}^{-1}(\epsilon v) \\ &= J(\rho_1) + \nabla J(\rho_1)(\rho - \rho_1), \end{aligned}$$

where the last equality follows because  $B$  is optimal basis of problem (A-1) and  $\epsilon v = \rho - \rho_1$ . We then have that  $J(\cdot)$  is linear over  $\mathcal{N}(\rho_1, \delta)$  and the first statement holds with  $K = 0$ .

Note that the second statement follows directly because, by hypothesis, the constraints are binding for  $\rho_1$  and taking  $\rho \in \mathcal{N}(\rho_1, \delta)$ , by Lemma D-3- applied to  $\xi + \epsilon u$  with  $\epsilon \leq \delta$ ,  $u^\top = (v, \mathbf{0})$  and  $\|v\| = 1$ - holds that the optimal basis does not change.

Therefore, we conclude that Assumption 2 holds for  $\delta = \frac{\phi_{\min}^*}{\|B_{\rho_1}^{-1}\|}$ , and  $K = 0$ .  $\square$



## C.2 Proofs of Section 6

### C.2.1 Proof of Proposition 3

*Proof of Proposition 3.* Recall that there exist  $\mu^*$  optimal dual solution satisfying  $\mu^* \in [0, \bar{\mu}]$  (see Step 1 in the proof of Proposition 1). Thus, it is enough to show that  $\left(\beta \left(\frac{\theta}{1+\mu^*}\right), \mu^*\right)$  is an optimal solution- Lagrange multiplier pair. We will proceed by using Proposition 5.1.5 in Bertsekas [1997]. That is, we need to check primal and dual feasibility, Lagrangian optimality and complementary slackness.

1. *Dual feasibility.* It follows directly because we take  $\mu^*$  optimal dual solution.
2. *Primal feasibility and complementary slackness.* To check primal feasibility and complementary slackness we will apply Proposition 2.1.2 in Bertsekas [1997], which gives us that, as  $\mu^*$  is optimal dual solution, we have that

$$\Psi'_\rho(\mu^*)(\mu - \mu^*) \geq 0, \quad \forall \mu \in [0, \bar{\mu}],$$

where the derivative of  $\Psi_\rho$  is given by

$$\Psi'_\rho(\mu) = \rho + \sum_{\theta \in \Theta} p_\theta g'_\theta(\mu) = \rho - \sum_{\theta \in \Theta} p_\theta \bar{m} \left( \beta \left( \frac{\theta}{1+\mu} \right) \right). \quad (\text{C-7})$$

If  $\mu^* = 0$ ,  $\Psi'_\rho(0)\mu \geq 0$  and therefore  $\Psi'_\rho(0) \geq 0$ . Note that we also have  $\Psi_\rho(\mu^*)\mu^* = 0$ , and then primal feasibility and complementary slackness follows by (C-7) because  $\sum_{\theta \in \Theta} p_\theta \bar{m} \left( \beta \left( \frac{\theta}{1+\mu^*} \right) \right)$  is the expected payment under the optimal bidding strategy.

If  $\mu^* > 0$ , there exists  $\nu > 0$  such that  $\mu^* + \nu$  and  $\mu^* - \nu$  belongs to  $[0, \bar{\mu}]$ . Therefore both  $\Psi'_\rho(\mu^*)\nu$  and  $\Psi'_\rho(\mu^*)(-\nu)$  are non-negative, obtaining  $\Psi'_\rho(\mu^*) = 0$ , and primal feasibility and complementary slackness hold.

3. *Lagrangian optimality.* Note that

$$\begin{aligned} \arg \max_{\phi \in \Phi} \mathcal{L}(\phi, \mu^*) &= \arg \max_{\phi \in \Phi} \left\{ \mu^* \rho + \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} (\theta \bar{q}(a) - (1 + \mu^*) \bar{m}(a)) d\phi_\theta(a) \right\} \\ &= \left\{ \arg \max_{\phi_\theta \in \Delta(\mathcal{A})} \int_{\mathcal{A}} (\theta \bar{q}(a) - (1 + \mu^*) \bar{m}(a)) d\phi_\theta(a) \right\}_{\theta \in \Theta} \\ &= \left\{ \arg \max_{a \in \mathcal{A}} g_\theta(a, \mu^*) \right\}_{\theta \in \Theta}, \end{aligned}$$

where the second equality holds because we can separate the problem for each  $\theta$ . But note that  $g_\theta(a, \mu^*)$  is maximized at  $a = \beta(\theta/(\mu^* + 1))$  and thus we have Lagrangian optimality.

Therefore, the four conditions holds and the proof is completed.  $\square$

## D Auxiliary Results

The following lemma is a technical result we need to prove Lemma B-3.

**Lemma D-1.** *For every  $a > 1, b > 0$  we have*

$$\int_0^T a \exp(-b(T-t)) \wedge 1 dt \leq \frac{1}{b}(\log(a) + 1).$$

*Proof.* We assume that  $a \exp(-bT) < 1$ . Otherwise, the integrand is always one and the bound trivially holds because  $T < \log(a)/b$  if  $a \exp(-bT) > 1$ . Let  $\tilde{T} \in [0, T]$  be such that  $a \exp(-b(T - \tilde{T})) = 1$ , which is always guaranteed to exist because the function  $t \mapsto a \exp(-b(T - t))$  is continuous, increasing, and evaluates to  $a \exp(-bT) < 1$  at  $t = 0$  and  $a > 1$  at  $t = T$ . Then, by partitioning the integral at  $\tilde{T}$  we obtain

$$\begin{aligned} \int_0^T a \exp(-b(T-t)) \wedge 1 dt &= \int_0^{\tilde{T}} a \exp(-b(T-t)) dt + T - \tilde{T} \\ &= \frac{a}{b} \exp(-b(T-t)) \Big|_0^{\tilde{T}} + T - \tilde{T} \\ &= \frac{a}{b} \exp(-b(T-\tilde{T})) - \frac{a}{b} \exp(-bT) + T - \tilde{T} \\ &\leq \frac{1}{b} + T - \tilde{T}, \end{aligned}$$

where the last inequality follows from our choice of  $\tilde{T}$  and discarding the second term. On the other hand, as  $a \exp(-b(T - \tilde{T})) = 1$  we have that  $T - \tilde{T} = \frac{\log a}{b}$ , and therefore we conclude that

$$\int_0^T a \exp(-b(T-t)) \wedge 1 dt \leq \frac{1}{b}(\log a + 1). \quad \square$$

**Lemma D-2.** *Let  $\|x\|$  be a norm in the Euclidean space and let  $\|z\|_* = \max_{\|x\| \leq 1} \{z^\top x\}$  be its dual norm. Let  $f(x) = \frac{1}{2}\|x\|^2$  if  $\|x\| \leq \varphi$  and  $f(x) = \infty$  otherwise. Then, its convex conjugate*

$f^*(z) = \max_x \{z^\top x - f(x)\} = \max_{x: \|x\| \leq \varphi} \{z^\top x - \frac{1}{2}\|x\|^2\}$  is given by

$$f^*(z) = \begin{cases} \frac{1}{2}\|z\|_*^2 & \text{if } \|z\|_* \leq \varphi \\ \varphi\|z\|_* - \frac{1}{2}\varphi^2 & \text{otherwise.} \end{cases}$$

*Proof.* Note that the convex conjugate can be more compactly written as  $\min(\varphi, \|z\|_*) \cdot \|z\|_* - \frac{1}{2} \min(\varphi, \|z\|_*)^2$ . We first show that the latter expression provides an upper bound and then show that the upper can be attained by choosing a suitable feasible solution.

For the upper bound, use Cauchy-Schwartz inequality to obtain that

$$f^*(z) \leq \max_{x: \|x\| \leq \varphi} \left\{ \|z\|_* \|x\| - \frac{1}{2}\|x\|^2 \right\} = \max_{\ell \in \mathbb{R}: 0 \leq \ell \leq \varphi} \left\{ \|z\|_* \ell - \frac{1}{2}\ell^2 \right\},$$

where the equality follows because we can equivalently optimize over the attainable norm values in  $[0, \varphi]$ . The objective value of the latter problem is a downward parabola with maximum at  $\ell = \|z\|_*$ . The claim follows because the optimal solution is  $\ell = \min(\varphi, \|z\|_*)$ .

For the lower bound, fix  $z$  and let  $\tilde{x} = \arg \max_{\|x\| \leq 1} \{z^\top x\}$ , i.e., a vector satisfying  $\|z\|_* = z^\top \tilde{x}$ . Such a vector exists because the dual norm always admits an optimal solution by Weierstrass theorem (the objective is continuous and the feasible set is compact). Consider the solution  $x = \min(\varphi, \|z\|_*) \tilde{x}$ . This solution is feasible because  $\|x\| = \min(\varphi, \|z\|_*) \|\tilde{x}\| \leq \varphi$  since  $\|\tilde{x}\| \leq 1$ . Therefore,

$$\begin{aligned} f^*(z) &\geq z^\top x - \frac{1}{2}\|x\|^2 = z^\top \tilde{x} \cdot \min(\varphi, \|z\|_*) - \frac{1}{2}\|\tilde{x}\|^2 \cdot \min(\varphi, \|z\|_*)^2 \\ &\geq \min(\varphi, \|z\|_*) \cdot \|z\|_* - \frac{1}{2} \min(\varphi, \|z\|_*)^2, \end{aligned}$$

where the last inequality follows because  $\|z\|_* = z^\top \tilde{x}$  and  $\tilde{x} \leq 1$ . The result follows.  $\square$

**Lemma D-3.** Consider the general linear program problem

$$\begin{aligned} &\max_x c^\top x \\ &\text{s.t } Ax = \xi + \epsilon u, \\ &x \geq 0, \end{aligned} \tag{D-1}$$

where  $c, x, \mathbf{0}, \mathbf{1} \in \mathbb{R}^n$ ,  $\xi, u \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{m \times n}$  matrix of rank  $m$  and  $\epsilon$  a real parameter. Define  $x_{\min}^* = \min\{x_i^* : x_i^* > 0\}$ , where  $x^*$  denotes a non-degenerate optimal solution of problem (D-1) for  $\epsilon = 0$ , and denote by  $A_B \in \mathbb{R}^{m \times m}$  its associated basis matrix. If  $\delta = x_{\min}^* / \|A_B^{-1} u\|$ , then  $A_B$  remains optimal for problem (D-1) for all  $0 \leq \epsilon \leq \delta$ .

*Proof.* By permuting its columns, matrix  $A$  can be written as  $A = (A_B|A_N)$ , where  $A_B \in \mathbb{R}^{m \times m}$  is the submatrix containing the columns associated to the basic variables of  $x^*$  and  $A_N \in \mathbb{R}^{m \times (n-m)}$  is the submatrix corresponding to the non-basic variables of  $x^*$ . Furthermore, we can write  $x^* = (x_B^*, \mathbf{0})$ , where  $x_B^* = A_B^{-1}\xi \in \mathbb{R}^m$  is the subvector of basic variables and  $\mathbf{0} \in \mathbb{R}^{n-m}$ . Note that non-degeneracy of  $x^*$  implies that  $x_B^* > 0$ .

Note that  $\delta > 0$  is well defined because  $\{x_i^* : x_i^* > 0\}$  is not empty due to the non-degeneracy condition on  $x^*$ . Take  $\epsilon \leq \delta$ . We will prove that  $A_B$  is an optimal basis for (D-1), that is,  $x = (x_B, \mathbf{0})$  with  $x_B = A_B^{-1}(\xi + \epsilon u)$  is an optimal solution for Problem (D-1). Changing the right-hand side of the equality constraints does not change the reduced cost vector and, therefore, it is enough to show that  $x_B$  is non-negative.

To this end, take  $j \in \{1, \dots, m\}$  such that  $(A_B^{-1}u)_j < 0$ . Note that if does not exist such  $j$ , the desired inequality follows trivially because  $x_j^* = (A_B^{-1}\xi)_j > 0$  since  $x^*$  is non-degenerate. Otherwise, we have that

$$(x_B)_j = (A_B^{-1}(\xi + \epsilon u))_j = (x_B^*)_j + \epsilon (A_B^{-1}u)_j \geq x_{\min}^* - \epsilon \|A_B^{-1}u\| \geq x_{\min}^* - \delta \|A_B^{-1}u\| = 0,$$

where the first equation follows from the definition of  $x_B$ , the second because  $(x_B^*)_j$  is a basic variable, the first inequality from the definition of  $x_{\min}^*$  together with  $|x_j| \leq \|x\|$  for every  $x_j \in \mathbb{R}^m$ , the second inequality because  $\epsilon \leq \delta$ , and the last from the definition of  $\delta$ . The proof is completed.  $\square$