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

Santiago R Balseiro, Omar Besbes, Dana Pizarro

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

Santiago R. Balseiro† Omar Besbes‡ Dana Pizarro§

first version: July 1, 2021, this version: December 15, 2022

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^2). 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, online matching, or order fulfillment, to name a few. Furthermore, we establish that the class of DRC^2 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^2 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, multi-secretary.

1 Introduction

Dynamic optimization problems with resource 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.

†Columbia University, Graduate School of Business. Email: srb2155@columbia.edu

‡Columbia University, Graduate School of Business. Email: ob2105@columbia.edu

§Universidad de O'Higgins, Institute of Engineering Sciences. Email: dana.pizarro@uoh.cl

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,[1] 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^2) problems, including problems with finite and continuum of actions and contexts. Notably, we show that this class admits as special cases a variety of problems studied separately in the literature. Broadly speaking, a DRC^2 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^2 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^2 problems encompass the following classical problems: *Network dynamic pricing* problems (see, e.g., Gallego and Van Ryzin 1997), *Multi-secretary* problems (see, e.g., Kleinberg 2005), *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^2 problem.

[1] We discuss the literature in detail when we discuss our model and present our main results and associated corollaries.

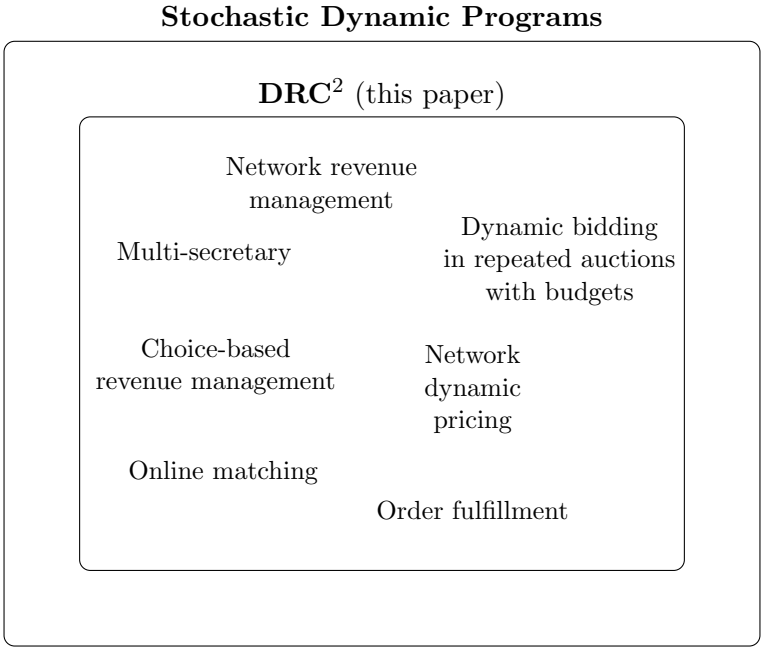


Figure 1: Definition of a new large class of dynamic decision problems (DRC^2) that encompasses many known problem classes.

A unified analysis of "fluid" certainty-equivalent control. Although from a theoretical perspective, DRC^2 problems can be formulated through a dynamic program, one natural question is whether the DRC^2 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^2 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^2 class of problems.

In more detail, solving even a special case of a DRC^2 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^2 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 loss between the optimal performance and that of the CE heuristic, and characterize the dependence of this loss on the "scale" of the system.

More specifically, we establish a hierarchy of sufficient conditions for the CE heuristic to lead to a "small" performance loss. The first layer of sufficient conditions identifies fundamental properties of the primal fluid problem that drive the CE performance and are stated in terms of the local strong-smoothness of the latter as a function of the resource vector. This condition highlights upfront a dichotomy between whether the primal problem is locally linear in the resource vector (i.e., weakly locally smooth) or locally quadratic (i.e., strictly locally smooth). Under weak local-smoothness the CE heuristic leads to losses that do not grow with T (where T is the length of the horizon), while under strict local-smoothness the CE heuristic leads at most logarithmic losses in T.

The second layer of sufficient conditions explores the drivers of the local strong-smoothness of the fluid problem in terms of the primitives of a DRC^2 problem. In doing so, we highlight how the local strong-smoothness condition naturally emerges in a variety of settings. A central object for this set of sufficient conditions is the dual Lagrangian problem, and we elucidate how properties of the latter together with the nature of the problem (finite vs. continuum of actions, finite vs. continuum of contexts) impact the type of local smoothness of the primal fluid problem. In particular, the analysis leads to a dichotomy between two fundamental cases: that when the set of contexts and actions are finite (and the fluid problem is a finite linear program), and that when one of these is a continuum. When the set of contexts and actions are finite, the primal fluid problem is locally linear under some conditions and, in turn, we show that one can guarantee a reward loss that does not grow with T (where T is the length of the horizon). When either the set of contexts or the actions is a continuum, the primal problem is an infinite-dimensional program. In this case, we provide a variety of conditions that lead to its local strong-smoothness and, in turn, we can guarantee at most logarithmic losses in T for the CE heuristic. Our analysis also leads to some control over the constants that drive the losses.

In essence, the analysis establishes that the class of DRC^2 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 the 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 that the CE heuristic is effective for specific DRC^2 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^2 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^2 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 impact of the set of actions and contexts being finite or not, 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 us to recover versions of various existing results but also to obtain such results under weaker conditions (see, e.g., the case of dynamic pricing in §3.1 and §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 in §3.2 and §6.2 or the case of dynamic assortment optimization in §3.4 and §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, bringing under the same umbrella a variety of problems previously studied separately. We illustrate how this formulation lends itself to analysis through a unified analysis of the CE heuristic. As such, the DRC^2 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^2 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^2) 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 context $\theta \in \Theta$, where Θ is the set of contexts. The context includes auxiliary information available to the decision maker that allows to customize decisions. For example, it can capture a user's preferences, demographic or historical information about a consumer, information about an order of products that needs to be fulfilled, etc. Contexts are drawn independently from a distribution $p \in \Delta(\Theta)$, where we use $\Delta(\cdot)$ to represent the set of all probability distributions over a set. 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 context θ, the action a, and an idiosyncratic shock ϵ. Shocks lie in a space $\mathcal{E}$ and are drawn independently from a distribution $f \in \Delta(\mathcal{E})$. 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 $r : \Theta \times \mathcal{A} \times \mathcal{E} \to \mathbb{R}$, the reward function, where $r(\theta, a, \epsilon)$ denotes the reward when the context is θ, the action is a, and the shock is ϵ. Taking an action consumes resources and we assume that the amount of resources consumed depends on the context θ and the value of the shock ϵ. We denote by $y : \Theta \times \mathcal{A} \times \mathcal{E} \to \mathbb{R}^L$, the vector-valued resource consumption function. In particular $y_l(\theta, a, \epsilon)$ represents the consumption of resource l if context θ arrived, the decision maker chose an action a, and the shock was ϵ.

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 context θ and idiosyncratic shock ϵ, 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 Π 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 context of the opportunity in time t and the history up to (and including) time $t-1$. That is, for a policy π, $a_t = a_t^\pi(\theta_t, \mathcal{H}_{t-1})$.[2] 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:

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

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

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

[2] To simplify notation we will write a_t^π to refer to the action taken at time t given the policy π

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 π is implemented. We refer to the expression above as the *reward loss* of the heuristic given by π.

2.1 Discussion of the modeling assumptions

To simplify notation, we make the probability distribution of ϵ independent of the action and the context. In some applications, it might be convenient to allow for some dependence. Dependencies may be captured in our framework by modifying reward and consumption functions. We present the details of this reduction in Appendix A.1.

In this paper, we assume that the decision maker knows all parameters and the probability distributions of all uncertainties. There exists another line of work studying problems when these quantities are unknown, and can be either stochastic or adversarial. When the context distribution is unknown but the shocks are either deterministic or from a known distribution, the problem is called the *online allocation problem* (see, e.g., Mahdian et al. 2012, Mirrokni et al. 2012, Agrawal et al. 2014, Ma and Simchi-Levi 2020, Li and Ye 2021, Balseiro et al. 2022). When the shock distribution is unknown and there is a single context, the problem is called the *multi-armed bandit with knapsack* [Badanidiyuru et al., 2013] and it includes important subclasses such as dynamic pricing with unknown demand [Besbes and Zeevi, 2009]. In the case when both the context and shock distributions are unknown, the problem is called the *contextual multi-armed bandit with knapsack* [Badanidiyuru et al., 2014].

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

3 Mapping Notable Problem Classes to a DRC^2 Problem

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

Problem subclass	Elements of the DRC^2 problem				
	Context (θ)	**Action** (a)	**Idiosyncratic shock** (ϵ)	**Resource Constraint**	
				LHS	RHS
Network dynamic pricing §3.1	customer class	product prices	heterogeneity in customer values	total resources consumed	initial inventory
Bidding in repeated auctions §3.2	impression's valuation	bid	exogenous auction uncertainty	total payment	budget
Network revenue management §3.3	customer class	accept/reject	- -	total resources consumed	initial inventory
Multi-secretary §3.3	candidate ability	hire/not hire	- -	total candidates hired	maximum hires
Choice-based NRM §3.4	customer class	offer set	heterogeneity in customer preferences	total resources consumed	initial inventory
Online matching §3.5	customer class	a resource	- -	total resources assigned	initial inventory
Order fulfillment §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^2 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 ϵ.

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, captured by the context, 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 polices 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 and Topaloglu [2019].

Mapping to a DRC^2 problem. In this setting, there is a set of N different products to sell during a finite, discrete horizon. The context $\theta \in \Theta$ 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 θ, the idiosyncratic shock ϵ captures heterogeneity in customer values. Then, given θ, posted prices $a \in \mathcal{A} = \mathbb{R}_+^N$ and a realization of the shock ϵ, 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 $y(\theta, a, \epsilon) = Q_\theta D(\theta, a, \epsilon)$. In the latter expression, $Q_\theta \in \mathbb{R}^{L \times N}$ is a matrix where $Q_{l\theta}^n$ represents the units of resource l needed to serve a customer in class θ with a single unit of product n.

3.2 Dynamic Bidding in Repeated Auctions

Consider 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 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^2 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 observes a real-valued valuation (the context) $\theta_t \in \Theta \subseteq [0, \Theta_{\max}]$ 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 ϵ 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 ϵ is independent of the buyer's valuation θ but our model can be easily be modified to account for

correlation using the reduction in Appendix A.1. The auction is characterized by an allocation rule $q : \mathcal{A}\times\mathcal{E} \to [0, 1]$ together with a payment rule $m : \mathcal{A}\times\mathcal{E} \to \mathbb{R}$; the former determines the probability that the impression is allocated to the advertiser and the latter his expected payment as a function of his bid and the exogenous shock. In Appendix D.2, we explicitly state the allocation and payment rules when the advertisers bid in first and second-price auctions, which are two auctions commonly used in practice and studied in the literature. The reward earned by the advertiser and his budget consumption, given a, ϵ and θ 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^2 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 has also been the basis for various industry solutions. The books of Talluri and Van Ryzin [2006] and Gallego and Topaloglu [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 contexts 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 2020) 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^2 class.

Mapping to a DRC^2 problem. In a NRM problem, a customer class can be captured by the context $\theta \in \Theta$ and is characterized by their usage of resources and a fixed price they pay for the service. We let r_θ denote the fare associated with class θ. 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 θ and the decision maker chooses an action a is $r(\theta, a) = r_\theta\, a$. If we denote by $Q_\theta = (Q_{l\theta})_l \in \mathbb{R}^L$ the consumption vector, where $Q_{l\theta}$ is the amount of resource l required to serve a customer of class θ, the consumption given that the decision maker chooses an action a and the customer class is θ is given by $y(\theta, a) = Q_\theta\, a$.

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 in the form of a set of options 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. Jasin and Kumar [2012] considers a model with customer choice that is slightly more general that the choice-based NRM as it allows for more general reward functions and resource consumption distributions. Still, their model fits in the DRC^2 class with appropriately defined reward and resource consumption functions.

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 Multinomial Logit 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^2 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 f_n. Contexts represent customer classes. Given a customer class θ and a product n, we denote by $A^n_{l\theta}$ the amount of resource l needed to serve product n to customer θ. The action set is given by $\mathcal{A} \subseteq 2^{\{1,\dots,N\}}$, 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 θ, we define the shock random vector $\epsilon_{\theta a} \in \mathcal{E}_a = \left\{\epsilon \in \{0,1\}^N : \sum_{n\in[N]} \epsilon^n \leq 1,\ \epsilon^n = 0 \text{ for } n \notin a\right\}$, where its nth component $\epsilon^n_{\theta a}$ is 1 if and only if the customer selects product n from the offer a. Then, $\epsilon_{\theta a} \sim \text{Multinomial}(1, g_{\theta a})$ where $g^n_{\theta a}$is the probability of the consumer choosing product n given that his class is θ and the action taken is a. Here $g^n_{\theta a} = 0$ if $n \notin a$, i.e., if the product does not belong to the offer set.

Given the consumer class θ, the action a, and the shock realization $\epsilon_{\theta a}$, the reward function is given by $r(\theta, a, \epsilon) = \sum_{n\in[N]} f_n \epsilon^n$ and the consumption function by $y(\theta, a, \epsilon) = \sum_{n\in[N]} Q^n_\theta \epsilon^n$. Note that, conditional on the shock ϵ, the resource consumption does not depend on the action. However, the action affects the distribution of ϵ. In Appendix A.1, we show how one may apply suitable transformations to the reward and consumption functions to obtain an equivalent problem in which the random shock is independent of the class and the action.

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 one 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 problem were considered by Feldman et al. [2009], Manshadi et al. [2012] and Devanur et al. [2013].

Mapping to a DRC^2 problem. We have a bipartite graph with resources on one side and contexts on the other side. An opportunity with context θ arrives with probability p_θ and the decision maker needs to decide which resource to assign it to. Calling $\mathcal{L}$ to the set of resources, each context θ has a reward vector $f_\theta \in \mathbb{R}^L_+$ and a resource consumption $Q_\theta \in \mathbb{R}^L_+$. 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 with context θ and an action a, the reward is given by $r(\theta, a) = f_{\theta a} 1_{\{a \neq 0\}}$ and the consumption of resource $l \in \mathcal{L}$ is given by $y_l(\theta, a) = Q^a_\theta 1_{\{a=l\}}$. We assume that the bipartite graph is complete. Incomplete graphs can be modelled by setting $f_{\theta j} = -\infty$ if assigning context θ 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^2 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 approaches, 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^2 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 θ occurs with probability p_θ 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, $l \in \theta$ if and only if item l is included in the order θ.

The decision maker has to construct a fulfillment policy to decide from which facility $n \in [N]$ each of the items in θ 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 reward denoted by f_{ln}.

Given that the order is θ 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^2 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 ($\mathcal{P}$) by their expectations.

That is, if we denote by Φ the set of all context-dependent probability distributions $\phi : \Theta \to \Delta(\mathcal{A})$, 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*:

$$\begin{aligned} \bar{J}(\rho) = \quad & \sup_{\phi \in \Phi} \ \mathbb{E}_{\theta \sim p, a \sim \phi(\theta), \epsilon \sim f} \left[r(\theta, a, \epsilon) \right] \\ & \text{s.t} \ \ \mathbb{E}_{\theta \sim p, a \sim \phi(\theta), \epsilon \sim f} \left[y_l(\theta, a, \epsilon) \right] \leq \rho_l, \ \ \forall l \in [L]. \end{aligned} \tag{$\mathcal{P}_{\text{FLUID}}$}$$

While we refer to ($\mathcal{P}_{\text{FLUID}}$) 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 contexts 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. For each context $\theta \in \Theta$, the decision variable $\phi(\theta)$ gives a probability distribution over actions $a \in A$ conditional on the arrival belonging to context θ.

In what follows, we assume that problem ($\mathcal{P}_{\text{FLUID}}$) admits an optimal solution. We denote by ϕ^*_ρ an optimal solution when the parameter is ρ. We will provide sufficient conditions for existence of an optimal solution in §5. When an optimal solution does not exist or it is computationally intractable, the analysis we develop can also be applied when an approximately optimal solution is computable. In particular, we can control the loss stemming from using an approximately optimal solution. In Appendix A.2, we provide a detailed analysis on the additional cumulative losses one incurs.

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 ($\mathcal{P}_{\text{FLUID}}$) 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 π^{CE}. The heuristic is formally presented in Algorithm 1. The CE heuristic adjusts the parameter ρ dynamically according to the amount of resources remaining to avoid running out of 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^2 problems. Thus, as the family of DRC^2 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 $t = 1, \ldots, T$ **do**
  $\rho_t \leftarrow c_t/(T - t + 1)$
  $\phi^*_{\rho_t} \leftarrow$ an optimal solution of Problem ($\mathcal{P}_{\mathrm{FLUID}}$) with $\rho = \rho_t$
  observe the context θ_t
  draw an action a_t from the distribution$\phi^*_{\rho_t}(\theta_t)$
  if $y(\theta_t, a_t, \epsilon) \leq c_t \ \forall \epsilon \in \mathcal{E}$, **then**
    choose the action a_t
    observe the shock ϵ_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

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^2 problems (see, e.g., Jasin and Kumar 2012).

Before proceeding to the analysis of the performance of the CE heuristic, we show that the fluid problem ($\mathcal{P}_{\mathrm{FLUID}}$) gives an upper bound on the optimal value of the stochastic problem ($\mathcal{P}$), a result that we will use to obtain the bound for the reward loss of the CE heuristic in the next section. Although versions of this result have been proven many times for some special cases and under

stronger assumptions (see, e.g., Gallego and Van Ryzin 1997), here we present a generic result in the context of the set of DRC^2 problems. We prove the result in the primal space, where the key observation is that the fluid problem allows for randomized policies. Therefore, for every dynamic policy, we can construct a feasible policy for the fluid problem that attains the same objective, by taking time averages and expectations over histories. The proof can be found in Appendix B.1.

Proposition 1. *The optimal value of the stochastic problem ($\mathcal{P}$) is upper bounded by T times the value of the fluid problem ($\mathcal{P}_{\text{FLUID}}$) for $\rho = C/T$. That is,*

$$J^*(C,T) \leq T\bar{J}(C/T).$$

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^2 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. This assumption is well motivated in practice as opportunities typically consumer a small number of resources relative to the total initial capacities. For a vector $x \in \mathbb{R}^n$, we denote by $\|x\| = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$ its ℓ^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 ρ_1 is the vector of initial inventory divided by the number of periods to consider. Given $\phi^*_{\rho_1}$, an optimal solution of ($\mathcal{P}_{\text{FLUID}}$) for $\rho = \rho_1$, we partition the resources $[L]$ into the set of resources $\mathcal{C}$ for which the corresponding resource constraint of problem ($\mathcal{P}_{\text{FLUID}}$) is binding at $\rho = \rho_1$ and $\mathcal{U}$ the set of resources for which the constraint does not bind. That is,

$$\mathcal{C} := \left\{ l \in [L] : \mathbb{E}_{\theta\sim p, a\sim\phi^*_{\rho_1}(\theta),\epsilon\sim f}[y(\theta,a,\epsilon)] = \rho_l \right\} \text{ and}$$
$$\mathcal{U} := \left\{ l \in [L] : \mathbb{E}_{\theta\sim p, a\sim\phi^*_{\rho_1}(\theta),\epsilon\sim f}[y(\theta,a,\epsilon)] < \rho_l \right\}.$$

Note that $\mathcal{C}$ and $\mathcal{U}$ are complement sets, i.e., $\mathcal{C} \cup \mathcal{U} = [L]$. Given a vector v, we will denote by $v|_{\mathcal{C}}$ and $v|_{\mathcal{U}}$ the restriction of v to the components of the set $\mathcal{C}$ and $\mathcal{U}$, respectively.

We will assume that, in the "neighborhood" of ρ_1, the optimal objective value of the fluid problem $\bar{J}(\rho)$ is locally smooth as well as that the consumption constraints corresponding to the

set $\mathcal{C}$ stay binding. We require the assumption to hold in the set of

$$\mathcal{N}(\rho_1, \delta, \mathcal{C}) = \left\{\rho \in \mathbb{R}_+^L : \|\rho|_{\mathcal{C}} - \rho_1|_{\mathcal{C}}\| \leq \delta \text{ and } \rho|_{\mathcal{U}} - \rho_1|_{\mathcal{U}} \geq -\delta \mathbf{1}\right\}, \tag{1}$$

which are the vectors that are δ close to ρ_1 for the binding resources and at least $-\delta$ larger for those resources that are not binding. Intuitively, resources which are not binding can increase without changing the optimal solution so these are only restricted from below. When all resources are binding, i.e., $\mathcal{C} = [L]$, we write $\mathcal{N}(\rho_1, \delta)$ to represent the set of all points at distance at most δ from ρ_1.

Assumption 2. *There exist $\delta, K \in \mathbb{R}_{++}$ with $\delta < \min_{l\in[L]} \rho_{1,l}$ such that for every $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$ it holds that:*

1. *The function $\bar{J}(\rho)$ satisfies $\bar{J}(\rho) \geq \bar{J}(\rho_1) + \nabla\bar{J}(\rho_1)(\rho - \rho_1) - \frac{K}{2}\|\rho|_{\mathcal{C}} - \rho_1|_{\mathcal{C}}\|^2$.*
2. *There exists an optimal solution ϕ_ρ^* satisfying $\mathbb{E}_{\theta\sim p, a\sim\phi_\rho^*(\theta), \epsilon\sim f}\left[y(\theta, a, \epsilon)|_{\mathcal{C}}\right] = \rho|_{\mathcal{C}}$.*

We will refer to the inequality given in Assumption 2.1 as $\bar{J}(\rho)$ admitting a *K-lower downward quadratic* (K-LDQ) *envelope* in $\mathcal{N}(\rho_1, \delta, \mathcal{C})$. See Figure 4 (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 ρ, ρ'. When all resources are binding, a sufficient condition for $\bar{J}(\rho)$ to admit a K-LDQ envelope is that its gradient is *locally K-Lipschitz continuous* at ρ_1 for all $\rho \in \mathcal{N}(\rho_1, \delta)$, that is,

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

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

The second part of Assumption 2 requires that all resources that are binding at ρ_1 remain binding at optimal solutions in a neighborhood of ρ_1. Because the lower quadratic envelope in the first part is independent of the unconstrained resources, we are also implicitly precluding unconstrained resources from becoming constrained and impacting performance.

We are now ready to state our performance bound of the CE heuristic under Assumptions 1 and 2. Specifically, in Theorem 1, we bound the reward loss of the heuristic given by CE as a function of the parameters in Assumptions 1 and 2.

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

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

where $\underline{\rho}_1$ *is the smallest component of the vector* ρ_1.

We note that the result above applies across all DRC^2 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 $J^{\mathrm{CE}}(C,T)/J^*(C,T) \to 1$ as $T \to \infty$ because the reward collected by the CE heuristic grows as $T \to \infty$. Furthermore, the optimality gap is of order $O(\log T)$ if $K > 0$ and of order $O(1)$ if $K = 0$. In other words, we see a clear distinction among DRC^2 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 (e.g., those in Jasin and Kumar 2012) 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 resources is close to being depleted, which would naturally lead to a linear dependence on L. Finally, as it is common in the operations research literature, the bound provided in Theorem 1 is instance dependent, i.e., it depends on the parameters of the particular problem at hand. (This is in contrast to, e.g., bounds in the multi-armed bandit [Bubeck and Cesa-Bianchi, 2012] or online convex optimization [Hazan et al., 2016] literatures which are worst-case over a large class of instances and depend on few key parameters such as the size of the action space and the length of the horizon.)

The proof of the theorem can be found in Appendix B.2. 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 τ, where τ is the first time that a resource is close to depletion or the ratio of capacity to time remaining ρ_t leaves the ball $\mathcal{N}(\rho_1, \delta, \mathcal{C})$ defined in Assumption 2. Using that the fluid problem gives an upper bound on the optimal value of the stochastic problem, that is, $J^*(C,T) \leq T\bar{J}(C/T)$ (see Proposition 1), we can bound the reward loss as follows

$$\begin{aligned} J^*(C,T) - J^{\mathrm{CE}}(C,T) &\leq T\bar{J}(\rho_1) - J^{\mathrm{CE}}(C,T) \\ &\leq \mathbb{E}\left[\sum_{t=1}^{\tau} \bar{J}(\rho_1) - \sum_{t=1}^{\tau} r(\theta_t, a_t^{\pi^{\mathrm{CE}}}, \epsilon_t)\right] + \mathbb{E}\left[\sum_{t=\tau+1}^{T} \bar{J}(\rho_1)\right], \end{aligned} \tag{2}$$

where $a_t^{\pi^{\mathrm{CE}}}$ 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 $\mathbb{E}[T - \tau] \cdot \bar{J}(C/T)$, which is of order $O(1)$, as we establish in Lemma B-3. This follows because, under the CE heuristic, the ratio ρ_t behaves like a martingale for the binding resources (and a submartingale for the ones that are not binding) by Assumption 2.2 and, as a result, the heuristic never runs out of resources nor ρ_t leaves the set $\mathcal{N}(\rho_1, \delta, \mathcal{C})$ too early. The first term is shown to be of order $O(\log T)$. To see this, note that up to time τ actions are not constrained by resources and the expected reward at period t satisfies $\mathbb{E}\left[r(\theta_t, a_t^{\pi^{\text{CE}}}, \epsilon_t) \mid \rho_t\right] = \bar{J}(\rho_t)$ because the CE heuristic takes actions according to $\phi^*_{\rho_t}$. Therefore, using Assumption 2.1 together with the fact that $\frac{\partial \bar{J}}{\partial \rho_l}(\rho_1) = 0$ for every resource $l \in \mathcal{U}$ that is not binding, we can upper bound the first term by

$$\mathbb{E}\left[\sum_{t=1}^{\tau} \sum_{l \in \mathcal{C}} \frac{\partial \bar{J}}{\partial \rho_l}(\rho_1)(\rho_{1,l} - \rho_{t,l})\right] + (K/2)\mathbb{E}\left[\sum_{t=1}^{\tau} \|\rho_1|_{\mathcal{C}} - \rho_t|_{\mathcal{C}}\|^2\right].$$

The first term is zero because, for the binding resources, ρ_t behaves like a martingale, while the second term can be bounded using the fact that martingale differences are orthogonal. Putting everything together, we then conclude with the bound in the Theorem.

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^2 problems and identify crisp sufficient conditions that allow us 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 us 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 class 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," local smoothness of the optimal objective value of the fluid problem, as well as the fact that the consumption constraints stay binding if they were binding for the initial problem. These conditions, while simple to state, may not be easy to check in many applications. In §5 we provide a "tree" of sufficient conditions leading to 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 δ and K based on the primitives of the problem. In particular, we highlight the drivers for positive values of K, delineating when one should expect to obtain logarithmic versus constant reward loss. Next, in §6 we revisit the applications presented in §3 using the results from §5. For each problem, we provide problem-specific sufficient conditions for our assumptions to hold, and we derive 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.

5 Drivers of CE Heuristic Performance: Dual Problem and General Sufficient Conditions

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

We introduce a dual of Problem ($\mathcal{P}_{\text{FLUID}}$) 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 ($\mathcal{P}_{\text{FLUID}}$). Let $\bar{r} : \Theta \times \mathcal{A} \to \mathbb{R}_+$ denote the expected reward function, i.e., $\bar{r}(\theta, a) = \mathbb{E}_\epsilon\big[r(\theta, a, \epsilon)\big]$. In the same way, for each $l \in [L]$, let $\bar{y} : \Theta \times \mathcal{A} \to \mathbb{R}^L_+$ denote the expected resource consumption function, i.e., $\bar{y}(\theta, a) = \mathbb{E}_\epsilon\big[y(\theta, a, \epsilon)\big]$. Then, the Lagrangian function is given by

$$\begin{aligned}\mathcal{L}(\phi, \mu) &= \mathbb{E}_{\theta\sim p, a\sim\phi(\theta)}\big[\bar{r}(\theta, a)\big] + \mu^\top\left(\rho - \mathbb{E}_{\theta\sim p, a\sim\phi(\theta)}\big[\bar{y}(\theta, a)\big]\right)\\ &= \mathbb{E}_{\theta\sim p, a\sim\phi(\theta)}\Big[\bar{r}(\theta, a) - \mu^\top\bar{y}(\theta, a)\Big] + \mu^\top\rho.\end{aligned}$$

Define, for each $\theta \in \Theta$, the function $g_\theta : \mathbb{R}^L \to \mathbb{R}$, which captures the optimal opportunity-cost adjusted reward given a context, and is given by

$$g_\theta(\mu) = \sup_{a\in\mathcal{A}}\Big\{\bar{r}(\theta, a) - \mu^\top\bar{y}(\theta, a)\Big\}\,. \tag{3}$$

We will refer to $g_\theta(\mu)$ as the *context-dependent adjusted reward function*. We denote the set of maximizers associated with the function $g_\theta(\mu)$ by

$$\mathcal{A}^*_\theta(\mu) = \arg\max_{a\in\mathcal{A}}\Big\{\bar{r}(\theta, a) - \mu^\top\bar{y}(\theta, a)\Big\}\,, \tag{4}$$

and let the *adjusted reward function* to be $g(\mu) = \mathbb{E}_{\theta\sim p}\left[g_\theta(\mu)\right]$. The Lagrange dual function, for

fixed $\rho \geq 0$, is given by

$$\begin{aligned}\Psi_\rho(\mu) &= \sup_{\phi \in \Delta(\mathcal{A})} \mathcal{L}(\phi, \mu) = \mu^\top \rho + \mathbb{E}_{\theta \sim p}\left[\sup_{a \in \mathcal{A}} \left\{\bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a)\right\}\right] \\ &= \mu^\top \rho + \mathbb{E}_{\theta \sim p}\left[g_\theta(\mu)\right] = \mu^\top \rho + g(\mu)\,,\end{aligned}$$

where the second equality follows because the problem is separable over contexts. The dual problem of Problem ($\mathcal{P}_{\text{FLUID}}$) is then given by

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

In Figure 2, we provide an overview of the results in this Section, which highlight how a variety of assumptions naturally lead to Assumption 2. The arrows and associated results establish the sufficiency of the conditions from one block to another block. At a high level, we present two sets of conditions under which Assumption 2 holds and characterize the corresponding parameters. A first set of conditions is when the fluid problem ($\mathcal{P}_{\text{FLUID}}$) is a linear program, which corresponds the case of finite actions and contexts (§5.1); in this case, Assumption 2 holds with $K = 0$. A second set of conditions is on the dual of ($\mathcal{P}_{\text{FLUID}}$) (§5.2). These conditions apply to a variety of settings and the analysis links the values of K and δ to the primitives of the problem. In particular, we illustrate these conditions in two notable subfamilies of cases: a continuum of actions, and binary actions with a continuum of contexts.

5.1 When the fluid problem is a finite dimensional linear program

Suppose that the set of actions $\mathcal{A}$ is finite and the set of contexts Θ is finite. We will derive sufficient conditions for Assumption 2 to hold with $K = 0$. Examples of problems with finite set of contexts and actions are: network revenue management problems with finite customer classes, multi-secretary problems with finite types, choice-based revenue management problems, order fulfillment problems with discrete locations, online matching problems, among others.

Note that Problem ($\mathcal{P}_{\text{FLUID}}$), in this special case, can be written as a finite-dimensional linear programming problem as follows

$$\begin{aligned}\bar{J}(\rho) = \quad & \max_{\phi_\theta \in \Delta(\mathcal{A})} \quad \sum_{\theta \in \Theta} p_\theta \bar{r}_\theta^\top \phi_\theta \\ & \text{s.t} \quad \sum_{\theta \in \Theta} p_\theta \bar{y}_\theta \phi_\theta \leq \rho\,,\end{aligned} \tag{6}$$

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. Furthermore, the feasible set is non empty and compact, and therefore there exists an optimal solution.

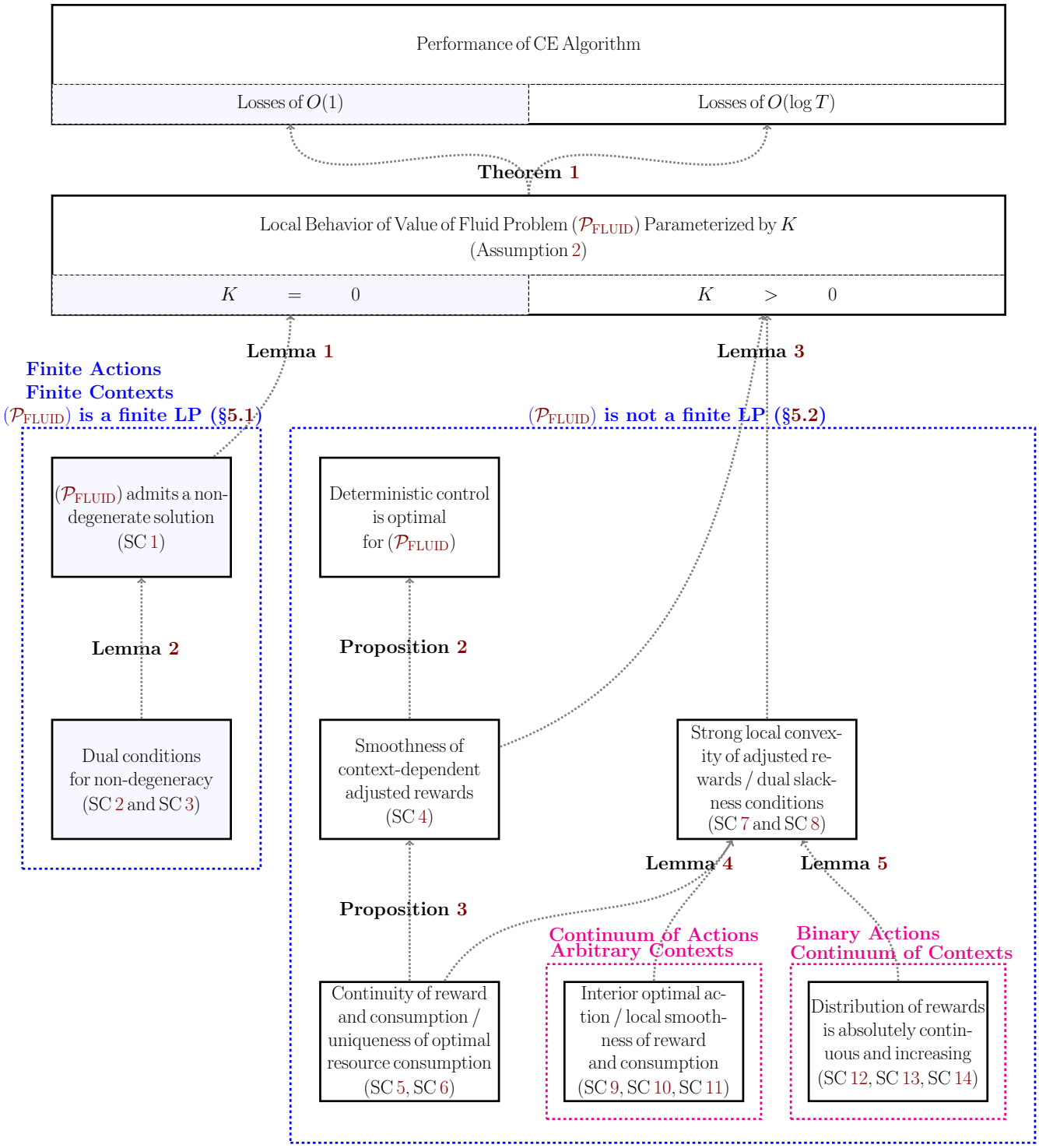


Figure 2: **Sufficient conditions and their implications on the performance of the CE heuristic**. Throughout, we assume that Assumption 1 holds.

The following is a common assumption in the literature associated with special cases of DRC^2 problems that guarantees that the CE policy has bounded reward loss (see, e.g., Jasin and Kumar 2012, Wu et al. 2015).

SC 1. *The primal problem (6) has a non-degenerate optimal solution for $\rho = \rho_1$, denoted by $\phi^*_{\rho_1}$.*

Considering the standard form of Problem (6) (see Appendix A.3 for details and the proof of Lemma 1 for an explicit matricial representation), let $B \in \mathbb{R}^{(L+|\Theta|)\times(L+|\Theta|)}$ be the corresponding optimal basis matrix and denote by $B^{-1}_{\rho_1}$ the submatrix of B^{-1} associated to the resource constraints. In the following lemma, we show that Assumption 2 holds under Assumption SC 1 and characterize the associated values of K and δ.

Lemma 1. *Suppose that Assumption SC 1 holds. Then Assumption 2 holds with $K = 0$ and $\delta = \min(\phi^*_{\min}, x^*_{\min})/\|B^{-1}_{\rho_1}\|$, where $\phi^*_{\min} = \min_{\theta\in\Theta, a\in\mathcal{A}}\{\phi^*_\theta(a) : \phi^*_\theta(a) > 0\}$, $x^*_{\min} = \min_{l\in[L]}\{x^*_l : x^*_l > 0\}$ and $x^*_l = \rho_l - \sum_{\theta\in\Theta} p_\theta \bar{y}_{l\theta}\phi^*_\theta$ is the slack of the l-th resource constraint.*

The proof is provided in Appendix C.1. A direct implication of Lemma 1 is that the reward loss if the action and context spaces are finite is on the order of $O(1)$.

Corollary 1. *Suppose that the sets of actions and contexts are finite and that Assumptions 1 and SC 1 hold. Then, the reward loss of the certainty equivalent heuristic is bounded by a constant for DRC^2 problems.*

Note that the bound on the reward loss is proportional to δ^{-2} and hence deteriorates when δ is small, which can occur if unconstrained resources are close to binding or if the fluid solution prescribes some actions to be taken with low probability.

Although the sufficient condition SC 1 is given on the primal problem (6), it is also possible to state sufficient conditions on the partial dual problem (5) to guarantee that Assumption 2 holds. 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 the duality gap is zero and $\bar{J}(\rho) = \inf_{\mu\in\mathbb{R}^L_+} \Psi_\rho(\mu)$.

It is well known that a sufficient condition to have a non-degenerate (and unique) solution of a linear problem is uniqueness and non-degeneracy of the dual problem (see, e.g., Bertsimas and Tsitsiklis 1997), and therefore in what follows we state sufficient conditions to guarantee the latter. More specifically, the condition SC 2 implies that the dual problem of (6) for $\rho = \rho_1$ has a unique solution, whereas from SC 2 and SC 3 we obtain that such optimal solution is not degenerate, concluding that under conditions SC 2 and SC 3, the problem (6) has a unique and non-degenerate optimal solution for $\rho = \rho_1$, which is stated in Lemma 2.

SC 2. *The dual problem (5) has a unique solution, μ^1, for $\rho = \rho_1$.*

SC 3. *The set of maximizers defined in* (4) *satisfies* $\sum_{\theta\in\Theta} |\mathcal{A}^*_\theta(\mu^1)| = L + |\Theta|$.

In the following lemma, we formalize that under Assumptions SC 2 and SC 3 there exists a unique and non-degenerate optimal solution of (6) for $\rho = \rho_1$. The proof is provided in Appendix C.2.

Lemma 2. *Under Assumptions SC 2 and SC 3, Assumption SC 1 holds.*

Geometric interpretation. We now provide a geometric interpretation for Assumption SC 2 as well as for the deterministic and dual functions for the case of finite actions and finite contexts.

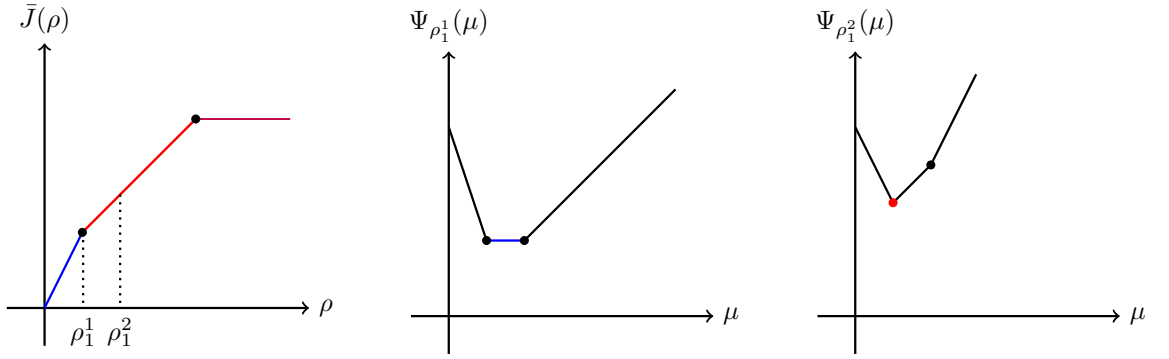
First, note that problem (6) is an LP and therefore the deterministic function $\bar{J}(\rho)$ is a concave piece-wise linear function (see Bertsimas and Tsitsiklis 1997 for more details). Moreover, due to the nature of the problem, it will be non-decreasing. In Figure 3 (a), the function $\bar{J}(\rho)$ is plotted for a problem with one resource and two contexts. Every optimal dual variable μ for $\Psi_\rho(\mu)$ gives a super-gradient to $\bar{J}(\rho)$. 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 ρ for which the same dual variable is optimal.

In Figure 3 (b) and (c), we plot the dual function $\Psi_\rho(\mu)$ as a function of μ for two different possible values of the parameter ρ. In Figure 3 (b), we take $\rho = \rho_1^1$, a value where $\bar{J}(\rho)$ has a kink. In this case, the dual problem admits an infinite number of solutions (flat blue segment in the figure) and every dual solution is a super-gradient of $\bar{J}(\rho_1)$. In Figure 3 (c), we plot the dual function at $\rho = \rho_1^2$, a value belonging to an interval where $\bar{J}(\rho)$ 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, Assumptions SC 2 and SC 3 are equivalently asking that the parameters ρ_1 lies in the interior of an interval where the deterministic function $\bar{J}(\rho)$ is smooth.

5.2 Beyond finite dimensional linear programs as fluid problems

When either the set of actions or the set of contexts is a continuum, Problem ($\mathcal{P}_{\text{FLUID}}$) is not necessarily a finite dimensional linear program. Next, we explore these cases and the drivers of the parameters K and δ associated with Assumption 2, and in turn why these problems often lead to logarithmic reward loss. (As we will see, natural conditions on the primitives lead to Assumption 2 holding with $K > 0$.)

In particular, we first provide sufficient conditions to ensure that strong duality holds for the fluid problem and that the fluid problem admits an optimal deterministic solution. Then, we give a closed-form expression for the values of $K > 0$ and δ for which Assumption 2 holds. Finally, we illustrate our conditions on two prototypical examples: finite contexts with a continuum of actions and a continuum of contexts with binary actions.



(a) Function $\bar{J}(\rho)$

(b) Dual function at ρ_1^1. Multiple optimal dual solutions.

(c) Dual function at ρ_1^2. Unique optimal dual solution.

Figure 3: Function $\bar{J}$ and dual function Ψ_ρ for two different parameters.

5.2.1 Strong duality and deterministic controls

We next present conditions on the function g_θ defined in (3) that are sufficient to ensure that the fluid problem has zero duality gap for every positive parameter ρ and that the fluid problem admits an optimal deterministic solution. Our sufficient conditions require that the functions g_θ are differentiable and an optimal action exists for every context and dual variable.

SC 4. *Almost surely over $\theta \in \Theta$ we have that, for every $\mu \geq 0$, function $g_\theta(\mu)$ is differentiable in μ and the set of maximizers $\mathcal{A}_\theta^*(\mu)$ is non-empty.*

Proposition 2. *Under Assumptions 1 and SC 4, strong duality for ($\mathcal{P}_{\text{FLUID}}$) holds, i.e., $\bar{J}(\rho) = \inf_{\mu \in \mathbb{R}_+^L} \Psi_\rho(\mu)$ for all $\rho > 0$. Furthermore, $\bar{J}(\rho)$ admits a deterministic optimal solution for all ρ.*

A proof is provided in Appendix C.3. We first prove that the dual problem admits an optimal solution using the extreme value theorem. This follows because the dual objective is continous and, without loss, we can restrict dual variables to a compact set. We then prove that strong duality holds by first principles by constructing a primal-dual pair that satisfies complementary slackness and Lagrangian optimality. The result follows because the primal solution is, by construction, deterministic.

Proposition 2 implies that, under assumption SC 4, randomization in ($\mathcal{P}_{\text{FLUID}}$) is not needed. As a result, we can write $\bar{J}(\rho)$ as the following non-linear program:

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

where $a \in \mathcal{A}^{|\Theta|}$ is to be interpreted as a function that maps a context θ to an action a_θ. We remark here that in the literature, the fluid problem is sometimes formulated directly over deterministic

actions as stated above. The analyses of the induced certainty equivalent policies typically require concavity of the objective and convexity of the constraints (with respect to actions or appropriate transformations of actions); see, e.g., Gallego and Van Ryzin [1997], Maglaras and Meissner [2006], Jasin [2014]. Remarkably, Proposition 2 allows to recover the optimality of deterministic actions even when the problem is non-convex and allows to obtain strong guarantees on the performance of certainty equivalent controls in more generality.

The resulting non-linear program might be non-convex and, in general, challenging to solve. A straightforward corollary of Proposition 2 is that instead of 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 $\mu^* \in \arg\min_{\mu \in \mathbb{R}^L_+} \Psi_\rho(\mu)$. A primal control can be constructed by solving, for each context $\theta \in \Theta$, for an action $a^*_\theta \in \mathcal{A}^*_\theta(\mu^*) = \arg\max_{a \in \mathcal{A}} \left\{ \bar{r}(\theta, a) - {\mu^*}^\top \bar{y}(\theta, a) \right\}$. In other words, optimal actions maximize rewards minus the opportunity cost of consuming resources, where resources are priced according to the optimal dual variable μ^*. 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.

The assumptions presented above are stated in terms of $g_\theta(\mu)$, which are derived objects, and, in general, might not be easy to verify. We now present sufficient conditions on the primitives of the problem for Assumptions SC 4 to hold.

SC 5. *Almost surely over $\theta \in \Theta$, the expected reward function $\bar{r}(\theta, a)$ is upper-semicontinuous in a and the expected resource consumption function $\bar{y}(\theta, a)$ is continuous in a.*

SC 6. *Almost surely over $\theta \in \Theta$, the set of optimal actions $\mathcal{A}^*_\theta(\mu)$ is non-empty and the set of optimal resource consumptions $\{\bar{y}(\theta, a^*) : a^* \in \mathcal{A}^*_\theta(\mu)\}$ is a singleton for every μ.*

Proposition 3. *If the set of actions $\mathcal{A}$ is compact and conditions SC 5 and SC 6 hold, then condition SC 4 is fulfilled.*

The proof of Proposition 3 follows directly from Corollary 4 of Milgrom and Segal [2002]. When the set of actions is finite, the continuity conditions SC 5 trivially hold, and we only require that the set of optimal resource consumptions is a singleton. For the latter condition to hold, one needs that the set of contexts is uncountable. It follows from Proposition 3 that if $\mathcal{A}$ is compact and SC 5 and SC 6 hold, then there is an optimal deterministic solution to the fluid problem ($\mathcal{P}_{\text{FLUID}}$).

5.2.2 Dual-based sufficient conditions for Assumption 2

In addition to condition SC 4, we need to make another regularity assumption over the function g, stated below, to ensure Assumption 2 holds for these particular cases of DRC^2.

SC 7. *There exist positive real numbers κ and ν, with $\nu < \underline{\mu} = \min_{l\in\mathcal{C}} \mu^1_l$, such that for all $\mu \in \mathcal{N}(\mu^1, \nu) \cap \mathbb{R}^L_+$, $g(\mu)$ satisfies*

$$g(\mu) \geq g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \frac{\kappa}{2}\|\mu - \mu^1\|^2 . \tag{7}$$

In what follows we will refer to property (7) as g admitting a *κ-lower upward quadratic* (κ-LUQ) *envelope* in $\mathcal{N}(\mu^1, \nu)$. See Figure 4 (b) for an example of a function admitting a κ-LUQ and its envelope.[3] The next assumption provides a sufficient condition to guarantee that the resources that are not binding at the initial resource vector ρ_1 remain not binding if we locally perturb resource availability. The condition is stated in terms of the derivative of the dual function, but can be understood as requiring that resources that are not binding have sufficient "slack."

SC 8. *For each $\mu \in \mathcal{N}(\mu^1, \nu) \cap \mathbb{R}^L_+$ and $\rho_j \geq \rho_{j,1} - \nu\kappa/2$, it holds that $\frac{\partial g}{\partial \mu_j}(\mu) + \rho_j > 0$ for resources $j \in \mathcal{U}$, where $\mathcal{U} = \{j \in [L] : \mu^1_j = 0\}$ is the set of resources with zero initial dual variables.*

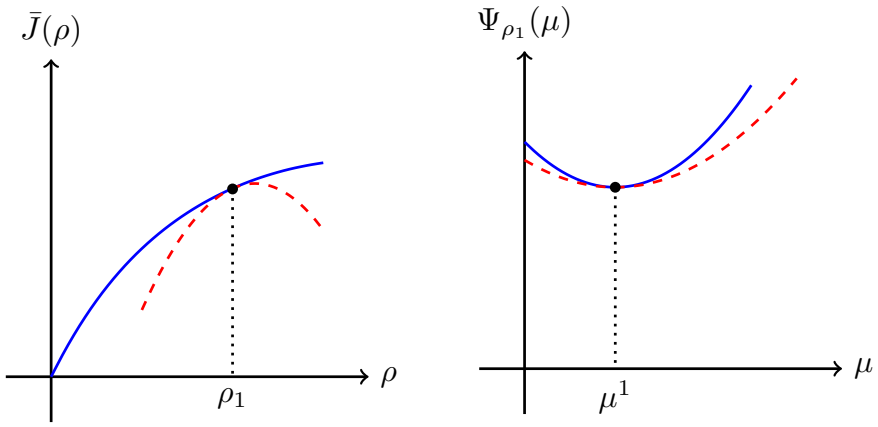
Lemma 3. *Suppose that Assumptions SC 4, SC 7, and SC 8 hold. Then Assumption 2 holds with $K = 1/\kappa$ and $\delta = (\nu\kappa)/2$.*

A proof is provided in Appendix C.4. To prove the result we need to show that the two conditions of Assumption 2 hold for $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$, i.e., $\bar{J}(\rho)$ admits a lower downward quadratic envelope and that all resources that are binding at ρ_1 remain binding at an optimal solutions in a neighborhood of ρ_1. Some intuition can be gleaned in light of well-known duality results between strong convexity and strong smoothness (see, e.g., Kakade et al. 2009). The proof of Lemma 3, however, is more delicate because our quadratic envelope condition is local. We prove the result by extending the envelope condition to hold globally over the dual domain and then optimizing over its convex envelope. For the second condition, we show that for every $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$ the dual optimal solutions are strictly positive for resources $\mathcal{C}$ that are binding at the initial resource vector and zero for the other resources that are not binding. The condition would then follow from complementary slackness.

We obtain the following result as a corollary.

Corollary 2. *Suppose that Assumptions 1, SC 4, SC 7, and SC 8 hold. Then, the reward loss of the certainty equivalent heuristic is of logarithmic order in T.*

[3] Admitting a κ-LUQ envelope is a weaker and local notion of the κ-strongly convex condition, which requires (7) to hold for every pair of dual variables μ, μ'.



(a) Deterministic proxy function in blue. Its 0.9-LDQ envelope in red.

(b) Lagrangian dual function in blue. Its 0.3-LUQ envelope in red.

Figure 4: Deterministic proxy $\bar{J}$ and its Lagrangian dual function, and their envelopes.

Geometric interpretation. The deterministic function $\bar{J}(\rho)$ can be easily shown to be concave and non-decreasing. Assumption SC 7 states that the dual function admits a κ-LUQ envelope in a neighborhood of the Lagrange multiplier μ^1. In Figure 4 (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 4 (a) and consists of $\bar{J}(\rho)$ admitting a K-LDQ envelope.

Recall that in the case of a finite set of contexts and actions (cf. §5.1), in contrast, it was not necessary to assume that the dual function admits a κ-LUQ envelope (condition SC 7). 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 context and action sets are finite because the dual problem is piece-wise linear.

5.2.3 Example 1: Continuum of actions

As a first example of the theory, we consider settings in which the set of actions $\mathcal{A}$ is a continuum (contexts can be either finite or not). Examples of applications with continuous actions include, for example, network dynamic pricing problems with finite segments and a continuum of prices, and dynamic bidding in repeated auctions with budgets. To simplify the exposition we assume that all resources are binding at the initial resource vector by imposing that $\mu^1 > 0$ and give sufficient conditions on the primitives for SC 7 to hold. In this case, we shall prove that $g_\theta(\mu)$ admits a lower upward quadratic envelope for each context θ and then simply take expectations over the contexts to establish that SC 7 holds. We do so by imposing envelope conditions on the reward and consumption functions.

For each $\theta \in \Theta$, let us denote by a_θ^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. These conditions need only hold almost surely over the distributions of contexts.

SC 9. *The feasible action a_θ^1 is interior. That is, there exists a positive number φ such that $\mathcal{N}(a_\theta^1, \varphi) \subseteq \mathcal{A}$ almost surely over θ in Θ.*

SC 10. *The expected reward function $\bar{r}(\theta, \cdot)$ admits a κ_r-LDQ envelope in $\mathcal{N}(a_\theta^1, \varphi)$. That is, almost surely over $\theta \in \Theta$,*

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

SC 11. *There exists a positive vector κ_y such that almost surely over $\theta \in \Theta$ and $a \in \mathcal{N}(a_\theta^1, \varphi)$ the consumption function $\bar{y}(\theta, \cdot)$ satisfies*

$$\bar{y}(\theta, a) \leq \bar{y}(\theta, a_\theta^1) + \nabla \bar{y}(\theta, a_\theta^1)(a - a_\theta^1) + \frac{\kappa_y}{2} \|a - a_\theta^1\|^2 , \tag{8}$$

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

We will refer to property (8) as the consumption function $\bar{y}(\theta, \cdot)$ admitting a *κ_y-upper upward quadratic* (κ_y-UUQ) *envelope* in $\mathcal{N}(a_\theta^1, \varphi)$.[4] Analogously to the relation made with the lower downward quadratic envelope notion, a sufficient condition for Assumption SC 11 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 7 to hold.

Lemma 4. *Suppose that Assumptions SC 5-SC 11 hold. Then, if $\mathcal{A}$ is compact, SC 7 holds with $\nu = \kappa\varphi/\sigma$ and $\kappa = \kappa_r + (\nu + \|\mu^1\|)\|\kappa_y\|$ where $\sigma = \inf_{\theta \in \Theta} \sigma_\theta$, with σ_θ the minimum singular value of $\nabla \bar{y}(\theta, a_\theta^1)$.*

A proof is provided in Appendix C.5. We obtain the following result as a corollary.

Corollary 3. *Suppose that there is a continuum set of actions. Furthermore, suppose that Assumption 1 and Assumptions SC 5-SC 11 hold. If the set of feasible actions $\mathcal{A}$ is compact and $\mu^1 > 0$, the reward loss of the certainty equivalent heuristic is of logarithmic order in T.*

[4] The κ_y-UUQ envelope condition is a weaker and local notion of the κ_y-strongly smooth condition for concave functions, which requires (8) to hold for every pair of actions a, a'.

5.2.4 Example 2: A continuum of contexts with binary actions

We now consider problems with a continuum of contexts and binary actions. These problems naturally emerge in a variety of applications, dating back to the early network revenue management formulation of Talluri and Van Ryzin [1998] in which the context is the fare offer of a customer, to the special case of multi-secretary problems and knapsack problems in which the context is the ability of the applicant and the type of items, respectively (see, e.g., Lueker 1998 and Bray 2022 for formulations with a continuum of contexts). To simplify the exposition, we restrict attention to the case of binary actions—similar results can be derived when multiple actions are available to the decision maker. Additionally, we assume that all resources are binding at the initial resource vector by imposing that $\mu^1 > 0$.

We denote the set of actions by $\mathcal{A} = \{0, 1\}$, i.e., the decision maker can either accept ($a = 1$) or reject ($a = 0$) each opportunity. We assume that the null action is $a = 0$ and to simplify the notation we denote by $\bar{r}(\theta) = \bar{r}(\theta, 1)$ and $\bar{y}(\theta) = \bar{y}(\theta, 1)$ the expected reward and resource consumption of accepting an opportunity, respectively. (The reward and resource consumption of rejecting are zero.) Resource consumption is non-negative, i.e., $\bar{y}(\theta) \geq 0$. Moreover, we have that $g_\theta(\mu) = \max\left(\bar{r}(\theta) - \mu^\top \bar{y}(\theta), 0\right)$. We assume that the distributions of contexts is such that only one action is optimal with probability one.

SC 12. *For every $\mu \geq 0$, we have that $\mathbb{P}_{\theta\sim p}\left\{\bar{r}(\theta) = \mu^\top \bar{y}(\theta)\right\} = 0$.*

In the case of a continuum of contexts with binary actions, condition SC 5 trivially holds. Moreover, under SC 12 we have that the set of maximizers $\mathcal{A}^*_\theta(\mu)$ is unique almost surely over the contexts θ and SC 6 holds. Therefore, by Proposition 3, we readily obtain that condition SC 4 is fulfilled. Thus, strong duality holds and the fluid problem admits an optimal deterministic control. The functions $g_\theta(\mu)$ are piecewise linear and, thus, do not admit lower upward quadratic envelopes as in the case of a continuum of actions. When there is a continuum of contexts, however, we can establish that the expected function $g(\mu) = \mathbb{E}_{\theta\sim p}\left[g_\theta(\mu)\right]$ admits a lower upward quadratic envelope because the convolution over contexts smoothes out kinks.

Our next assumption imposes that the density of rewards is bounded from below or, equivalently, its distribution is strictly increasing, which is required for contexts to act as mollifiers.

SC 13. *There exists a positive real number $\underline{p}$ such that for every measurable set $\mathcal{R} \subset [0, \bar{r}_\infty]$ and consumption vector $0 \leq y \leq \bar{y}_\infty$, we have that $\mathbb{P}_{\theta\sim p}\left\{\bar{r}(\theta) \in \mathcal{R} \mid \bar{y}(\theta) = y\right\} \geq \underline{p} \cdot |\mathcal{R}|$.*

Our final assumption imposes that there is enough variation in resource consumption. This condition is trivially satisfied when we have one resource. When we have multiple resources, it guarantees that the dual function is strongly convex by preventing it from being constant along fixed directions.

SC 14. *There exists some positive reals $\nu > \|\mu^1\|$ and λ such that the minimum eigenvalue of the positive definite matrix $\mathbb{E}_{\theta\sim p}\left[\bar{y}(\theta)\bar{y}(\theta)^\top \mathbf{1}\left\{y(\theta)^\top\mu^1 + \nu\|y(\theta)\| \leq \bar{r}_\infty\right\}\right] \in \mathbb{R}^{L\times L}$ is at least λ.*

We are now ready to provide sufficient conditions for Assumption SC 7 to hold.

Lemma 5. *Suppose that Assumptions SC 12-SC 14 hold. Then, SC 7 holds with $\kappa = \underline{p}\lambda$ and ν.*

A proof is provided in Appendix C.6. We obtain the following result as a corollary.

Corollary 4. *Suppose that there are two actions and the set of contexts is a continuum. Furthermore, suppose that Assumptions SC 12-SC 14 hold. If $\mu^1 > 0$, the reward loss of the certainty equivalent heuristic is of logarithmic order in T.*

6 CE Heuristic Performance: Corollaries Across Subclasses of Problems

In this section we revisit the applications discussed in §3. Using the results from §5, we discuss sufficient conditions for Assumption 2 together with the resulting performance characterization of the CE heuristic; the detailed sufficient conditions are presented in Appendix D. We then comment on the specific connections to the existing literature in the study of the CE heuristic.

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. For simplicity we assume for now that contexts are finite—similar results can be provided when there is a continuum of contexts.

When the set of prices is a continuum, in Appendix D.1, we present sufficient conditions on the primitives that allow to leverage Lemma 3 to ensure that Assumption 2 holds with values of $K > 0$, δ and ν. Therefore, we can use Corollary 2 to deduce that the revenue loss of the certainty equivalent heuristic is of order $O(\log T)$ 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 finite linear program and the sufficient conditions from Section 5.1 guarantee constant revenue loss for the CE heuristic. Conversely, if we have finite prices but a continuum contexts, then it is possible to guarantee $O(\log T)$ reward loss in a variety of cases by invoking results similar to those in Section 5.2.4.

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 presents 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 θ, 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 yields 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^2 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. In Appendix D.2, we characterize an optimal bidding strategy in terms of an optimal bidding function for the static auction without budget constraints. Moreover, we study the particular cases of second-price and first-price auctions, providing sufficient conditions on the primitives of the problem for conditions SC 4 and SC 7 to be satisfied, leading to a revenue loss of logarithmic order in T for these problems.

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^2 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 (binary). In the case of finite contexts, we present, in Appendix D.3, a closed-form expression for the value of δ based on the primitives of the problem (notice that here $K = 0$). When there is a continuum of contexts[5], one may use the analysis in §5.2.4 and the associated sufficient conditions to observe that the CE heuristic will guarantee logarithmic reward loss under a variety of settings.

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 with finite contexts (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.

Bray [2022] studies a NRM problem with a continuum of contexts and obtains logarithmic reward loss under some regularity conditions and assumptions that imply SC 4 and SC 7, namely the local strong convexity and smoothness of the adjusted reward function. The set of sufficient conditions in Bray [2022] are stated in terms of the Hessian matrix of the adjusted reward function, whereas the set of conditions in Section 5.2.4 are directly stated in terms of problem primitives. He considers a policy similar to the CE heuristic that in each period burns a certain amount of the resources that are not binding to prevent their inventory levels from drifting and make the sequence of remaining inventory follow a martingale. Lueker [1998] derives a logarithmic reward loss bound for online knapsack problems under conditions related to the ones we present. When specialized to the case of the multi-secretary problem his Condition 1 implies our Assumption 1; the first part of his Condition 2 implies that $\mu^1 > 0$ and the second part implies a local version of SC 13; and his Condition 3 implies SC 12.

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 [2020] 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).

[5] Interestingly, the early NRM formulation in Talluri and Van Ryzin [1998] considers a continuum of contexts.

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 we obtain a constant bound on the revenue loss of the CE heuristic. For completeness, we give a formulation for the fluid problem in Appendix D.4.

Connection with earlier analysis of the CE heuristic. Our formulation shows that a certainty equivalent heuristic admits strong performance guarantees under Assumption SC 1. This recovers a result from Jasin and Kumar [2012] for a slightly different model with consumer choice.

6.5 Online Matching

Here we revisit the problem presented in §3.5. In Appendix D.5, we conduct a similar analysis to the one for the NRM problem with finite contexts. That is, we interpret the value of δ for this class of problems, and we obtain a constant bound for the revenue loss.

Connection with earlier analysis of the CE heuristic. Although the online matching problem has been extensively studied under different settings (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 to document the performance of the CE heuristic for the online matching problem. Our analysis yields good performance bounds under the so-called "small bid assumption," which requires that the maximum possible resource consumption in a time period is small relative to the total amount of initial resources. This requirement is implicitly imposed by part 2 of Assumption 1, where we assume that resource consumption is uniformly bounded. While the CE policy still can be implemented when the small bid assumption does not hold, our performance bounds are not meaningful in such settings. In light of the recent results by Bumpensanti and Wang [2020] and Arlotto and Gurvich [2019], we expect that similar bounds might hold even in the absence of the small bid assumption. We refer the reader to Mehta [2013] for an overview of online matching problems and the small bid assumption.

6.6 Order Fulfillment

The order fulfillment problem was presented in §3.6. In Appendix D.6, we describe the fluid problem for this case. Here, Assumption 2 holds with $K = 0$ and we recover a constant revenue loss bound for the order fulfillment problem when the primal problem is non-degenerate.

Connection with earlier analysis of the CE heuristic. While there are related studies analyzing the performance of heuristics (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^2, 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^2 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-§5), which depend on whether the set of contexts or actions are finite or a continuum. We provide general conditions under which the CE heuristic guarantees constant or logarithmic reward loss. This leads us to recover a variety of existing results for some of the classical problems in DRC^2, 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^2 class. One of them is known as *hindsight optimum* (see, e.g., Reiman and Wang 2008, Bumpensanti and Wang 2020, Vera and Banerjee 2021), which is the problem obtained when all uncertainties are revealed in advance. This benchmark is usually considered in 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 contexts.) Another natural benchmark is the optimal value of the dynamic program, which was recently considered by Wang and Wang [2022] 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 reward loss across a broad family of problems.

A second important and related avenue is to study alternative heuristics at the 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 [2020] propose a heuristic that has a $O(1)$ revenue loss. The idea is to 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 DRC^2 problems that accommodates, e.g., a continuum of actions.

Performance bounds in the literature are typically instance dependent which makes it hard to directly compare different algorithms. For example, some algorithms might have better dependence on the length of the horizons but worse dependence on other important parameters such as the size of the action space or the number of resources. A third important direction is to investigate the impact of different parameters, either numerically or analytically, on the performance of the algorithms proposed in the literature.

A fourth direction pertains to further expanding the class of 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 reward loss 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 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.

Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014.

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. *Stochastic Systems*, 10(2):170–191, 2020.

Arash Asadpour, Xuan Wang, and Jiawei Zhang. Online resource allocation with limited flexibility. *Management Science*, 2019.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 207–216. IEEE, 2013.

Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. Resourceful contextual bandits. In *Conference on Learning Theory*, pages 1109–1134. PMLR, 2014.

Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, page to appear, 2022.

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.

Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

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.

Robert L. Bray. Logarithmic regret in multisecretary and online linear programming problems with continuous valuations. *working paper, Northwestern University*, 2022.

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.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

Pornpawee Bumpensanti and He Wang. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science*, 66(7):2993–3009, 2020.

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 Huseyin Topaloglu. *Revenue management and pricing analytics*, volume 209. Springer, 2019.

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.

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.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

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 ecommerce order fulfillment. *Operations Research*, 63(6):1336–1351, 2015.

Mark E. Johnson. *Multivariate Statistical Simulation*, pages 930–932. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

Sham Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. *Unpublished Manuscript, http://ttic. uchicago. edu/shai/papers/KakadeShalevTewari09. pdf*, 2(1), 2009.

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.

Robert Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 630–631. Citeseer, 2005.

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.

Xiaocheng Li and Yinyu Ye. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*, 2021.

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.

George S Lueker. Average-case analysis of off-line and on-line knapsack problems. *Journal of Algorithms*, 29(2):277–305, 1998.

Will Ma and David Simchi-Levi. Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. *Operations Research*, 68(6):1787–1803, 2020.

Constantinos Maglaras and Joern Meissner. Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing & Service Operations Management*, 8(2):136–148, 2006.

Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. Online optimization with uncertain information. *ACM Transactions on Algorithms (TALG)*, 8(1):1–29, 2012.

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.

Aranyak Mehta. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science*, 8(4):265–368, 2013.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2): 583–601, 2002.

Vahab S Mirrokni, Shayan Oveis Gharan, and Morteza Zadimoghaddam. Simultaneous approximations for adversarial and stochastic online budgeted allocation. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 1690–1701. SIAM, 2012.

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.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.

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. *Operations Research*, 2022.

Huasen Wu, Rayadurgam Srikant, Xin Liu, and Chong Jiang. Algorithms with logarithmic or sublinear regret for constrained contextual bandits. *Advances in Neural Information Processing Systems*, 28, 2015.

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

Santiago R. Balseiro[1], Omar Besbes[2], and Dana Pizarro[3]

Appendix

Table of Contents

[1]Columbia University, Graduate School of Business. Email: srb2155@columbia.edu.
[2]Columbia University, Graduate School of Business. Email: ob2105@columbia.edu.
[3]Universidad de O'Higgins, Institute of Engineering Sciences. Email: dana.pizarro@uoh.cl.

A Additional material

A.1 Idiosyncratic shock

The goal of this section is to show that we can assume, without loss of generality, that the probability distribution of the idiosyncratic shock is independent of the action and the context. To this end, we assume that for each context θ and each action a, we have a random shock $\epsilon_{\theta a}$ lying in a space $\mathcal{E}_{\theta a}$ and drawn from a distribution $f_{\theta a}$. We denote by $r(\theta, a, \epsilon_{\theta a})$ and $y(\theta, a, \epsilon_{\theta a})$ the reward and consumption functions, respectively. It is enough to show that there exists a random variable ϵ with probability distribution f, independent of θ and a, and functions $\bar{r}$ and $\bar{y}$ such that $\bar{r}(\theta, a, \epsilon)$ has the same distribution as $r(\theta, a, \epsilon_{\theta a})$ and $\bar{y}(\theta, a, \epsilon)$ has the same distribution as $y(\theta, a, \epsilon_{\theta a})$. We first introduce the following technical lemma.

Lemma A-1. *Let X be a random variable with distribution F. If $G : (0,1) \to \mathbb{R}$ is defined as the generalized inverse of F, that is $G(y) = \inf\{x : F(x) \geq y\}$, and $Y \sim U(0,1)$, then $Z = G(Y)$ has distribution F.*

Proof. We have to show that $Z \sim F$. Let x be such that $F(x) \in (0,1)$ and let $y \in (0,1)$. Note that $y \leq F(x)$ if and only if $G(y) \leq x$, where one of the directions holds because F is right continuous and the other by definition of the generalized inverse of F. Then,

$$\mathbb{P}(Z \leq x) = \mathbb{P}(G(Y) \leq x) = \mathbb{P}(Y \leq F(x)) = F(x),$$

where the first equality follows from the definition of Z, the second holds because of the observation above, and the last equality holds because $Y \sim U(0,1)$. □

Applying Lemma A-1 with $X = \epsilon_{\theta a}$ and $Y = \epsilon$ uniformly distributed in $(0,1)$, it follows that $G_{\theta a}(\epsilon) \sim f_{\theta a}$, where $G_{\theta a}$ is the generalized inverse of $f_{\theta a}$. Therefore, defining $\bar{r}(\theta, a, \epsilon) := r(\theta, a, G_{\theta a}(\epsilon))$ and $\bar{y}(\theta, a, \epsilon) := y(\theta, a, G_{\theta a}(\epsilon))$, we obtain that ϵ has a probability distribution independent of θ and a, and $\bar{r}(\theta, a, \epsilon)$ and $\bar{y}(\theta, a, \epsilon)$ has the same distributions as $r(\theta, a, \epsilon_{\theta a})$ and $y(\theta, a, \epsilon_{\theta a})$, respectively.

Above, we assumed that the idiosyncratic shock is univariate. If we assume that for each context θ and action a the random shock $\epsilon_{\theta a} = (\epsilon^1_{\theta a}, \ldots, \epsilon^D_{\theta a})$ is a D-dimensional random vector with joint distribution $f_{\theta a}$ and marginals $f^i_{\theta a}$, we can generate it by sampling sequentially from the conditional distributions. More specifically, we first generate the random variable X_1 with distribution function $f^1_{\theta a}$ by using the procedure described for the univariate case. Then, we consider the conditional distribution of $\epsilon^2_{\theta a}$ given that $X_1 = x_1$, denoted by $f^2_{\theta a}(\,\cdot\,|x_1)$ and we generate the random variable X_2 with distribution $f^2_{\theta a}(\,\cdot\,|x_1)$. Then, we consider the conditional distribution of $\epsilon^3_{\theta a}$ given that

$X_1 = x_1$ and $X_2 = x_2$, denoted by $f^3_{\theta a}(\,\cdot\,|x_1, x_2)$, and we generate the random variable X_3 with distribution $f^3_{\theta a}(\,\cdot\,|x_1, x_2)$, and so on and so forth. For more details on the algorithm to generate random vectors using conditional distributions we refer the reader to Johnson [2011].

A.2 Approximate solution of Problem $\mathcal{P}_{\text{FLUID}}$

One of the goals of this paper is to study the performance of the certainty equivalent heuristic described in Section 4 for the class of DRC^2 problems. This heuristic uses, at each time period, an optimal solution of a fluid problem. However, for some applications the fluid problem could be computationally intractable and then it is important to develop guarantees when only an approximately optimal solution is available. We develop such guarantees in this section.

To this end, consider Algorithm 1 with $\phi^*_{\rho_t}$ a feasible solution of Problem $\mathcal{P}_{\text{FLUID}}$ with $\rho = \rho_t$ satisfying the second statement of Assumption 2 and such that $\sum_{\theta\in\Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon(r(\theta, a, \epsilon))\mathrm{d}\phi^*_{\rho_t} \geq \bar{J}(\rho_t) - \alpha_t$, that is, an α_t-approximate solution of Problem $\mathcal{P}_{\text{FLUID}}$ for $\rho = \rho_t$. Then, if we denote by $J^{\text{CE}}(C, T)$ the expected performance of Algorithm 1 and a^{CE}_t the action taken by the CE heuristic at time t, we have that

$$\begin{aligned} J^{\text{CE}}(C,T) &= \mathbb{E}\left(\sum_{t=1}^{T} r(\theta_t, a_t^{\text{CE}}, \epsilon_t)\right) \geq \mathbb{E}\left(\sum_{t=1}^{\tau} r(\theta_t, a_t^{\text{CE}}, \epsilon_t)\right) \\ &= \mathbb{E}\left(\sum_{t=1}^{\tau}\sum_{\theta\in\Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon(r(\theta, a, \epsilon))\mathrm{d}\phi^*_{\rho_t}\right) \geq \mathbb{E}\left(\sum_{t=1}^{\tau} \bar{J}(\rho_t)\right) - \sum_{t=1}^{T} \alpha_t, \end{aligned}$$

where τ is the first time that a resource is close to depletion or ρ_t leaves the ball defined in Assumption 2. The first inequality follows from dropping the performance after the stopping time τ, the second equality follows from the optional stopping theorem, and the last inequality because our controls are approximately optimal.

Therefore, it holds that $J^*(C,T) - J^{\text{CE}}(C,T) \leq J^*(C,T) - \mathbb{E}\left(\sum_{t=1}^{\tau} \bar{J}(\rho_t)\right) + \sum_{t=1}^{T} \alpha_t$, and applying the same arguments as in proof of Theorem 1 we obtain the same guarantee plus the sum of the approximation errors. Thus, it is enough that $\sum_{t=1}^{T} \alpha_T = O(1)$ or equivalently take $1/T$-approximate solutions of Problem $\mathcal{P}_{\text{FLUID}}$ in Algorithm 1 to have the same guarantee as in Theorem 1.

A.3 Finite set of actions

Problem 6 is a linear program. In particular, introducing the set of slack variables $\{x_1 \dots x_L\}$, the standard form is given by

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

B Proofs of Section 4

B.1 Proof of Proposition 1

Proof of Proposition 1. We divide the proof into three steps. First, we consider a dynamic feasible policy for the stochastic problem ($\mathcal{P}$) and we use it to define a static randomized policy. Then, we prove that this static policy is a feasible solution of the fluid problem for $\rho = C/T$ and that the value attained by this feasible solution is the same as the value of the stochastic problem, divided by T. Finally, we obtain the desired result by applying the argument to the optimal solution of ($\mathcal{P}$) and noting that ($\mathcal{P}_{\text{FLUID}}$) is a maximization problem.

Step 1. Let $\pi \in \Pi$ be a feasible policy for the stochastic problem ($\mathcal{P}$) and $a_t^\pi(\theta_t, \mathcal{H}_{t-1})$ the resulting actions, where $\mathcal{H}_{t-1}$ is the history up to (and including) time $t-1$. Let us define $\phi : \Theta \to \Delta(\mathcal{A})$ by $\phi_\theta(A) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{t-1} \left(\mathbf{1}_{\{a_t^\pi(\theta, \mathcal{H}_{t-1}) \in A\}} \right)$, where $\mathbb{E}_{t-1}(X)$ denotes the expectation of X with respect to the history up to time $t-1$.

Let us see that ϕ is well defined:

1. *Non negativity.* It follows directly from the definition that for every $\theta \in \Theta$ and $A \in \mathcal{A}$, $\phi_\theta(A) \geq 0$.
2. *Measure of $\mathcal{A}$.* Notice that $\phi_\theta(\mathcal{A}) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{t-1} \left(\mathbf{1}_{\{a_t^\pi(\theta, \mathcal{H}_{t-1}) \in \mathcal{A}\}} \right) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{t-1} (1) = 1$.
3. *$\sigma-$additivity.* Let $\{A_i : i \in I\}$ be a countable, pairwise disjoint collection of elements of $\mathcal{A}$, then

$$\begin{aligned}
\phi_\theta\left(\bigcup_{i\in I} A_i\right) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{t-1}\left(\mathbf{1}_{\{a_t^\pi(\theta,\mathcal{H}_{t-1})\in\cup_i A_i\}}\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{P}_{t-1}\left(a_t^\pi(\theta,\mathcal{H}_{t-1})\in\cup_i A_i\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{P}_{t-1}\left(\bigcup_i\{a_t^\pi(\theta,\mathcal{H}_{t-1})\in A_i\}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i\in I}\mathbb{P}_{t-1}\left(a_t^\pi(\theta,\mathcal{H}_{t-1})\in A_i\right)\\
&= \sum_{i\in I}\phi_\theta(A_i),
\end{aligned}$$

where the third equality holds because the elements are pairwise disjoint and the fourth equality because $\mathbb{P}_{t-1}$ is a probability measure.

Step 2. In this step, we will check that the probability distribution ϕ defined in Step 1 is a feasible solution of Problem $\mathcal{P}_{\text{FLUID}}$ for $\rho = C/T$ and that $J^\pi(C,T) = TJ^\phi(C/T)$, where $J^\phi(C/T)$ represents the value of the fluid problem for the feasible solution ϕ.

1. *Feasibility.* Given the probability distribution ϕ we need to prove that

$$\sum_{\theta\in\Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon\left(y_l(\theta,a,\epsilon)\right)\mathrm{d}\phi_\theta(a) \le C_l/T,\ \forall l\in[L].$$

Due to the fact that π is a feasible policy for problem ($\mathcal{P}$), it holds, almost surely, that

$$\sum_{t=1}^{T} y_l(\theta_t, a_t^\pi, \epsilon_t) \le C_l,\ \forall l\in[L].$$

Then, the inequality above also must hold in expectation, that is:

$$\sum_{t=1}^{T}\mathbb{E}_{\theta_t\sim p, a_t^\pi\sim\mathbb{P}_{t-1},\epsilon_t\sim f}\left(y_l(\theta_t,a_t^\pi,\epsilon_t)\right) \le C_l,\ \forall l\in[L].$$

Note that for each $l\in[L]$, we have that

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}_{\theta_t\sim p, a_t^\pi\sim\mathbb{P}_{t-1},\epsilon_t\sim f}\left(y_l(\theta_t,a_t^\pi,\epsilon_t)\right) &= \sum_{t=1}^{T}\sum_{\theta_t\in\Theta} p_{\theta_t}\int_{\mathcal{A}}\mathbb{E}_{\epsilon_t}\left(y_l(\theta_t,a,\epsilon_t)\right)\mathrm{d}\mathbb{P}_{t-1}\left(a_t^\pi(\theta_t,\mathcal{H}_{t-1})\in a\right)\\
&= \sum_{\theta\in\Theta} p_\theta\int_{\mathcal{A}}\mathbb{E}_\epsilon\left(y_l(\theta,a,\epsilon)\right)\mathrm{d}\left(\sum_{t=1}^{T}\mathbb{P}_{t-1}\left(a_t^\pi(\theta_t,\mathcal{H}_{t-1})\in a\right)\right)\\
&= T\sum_{\theta\in\Theta} p_\theta\int_{\mathcal{A}}\mathbb{E}_\epsilon\left(y_l(\theta,a,\epsilon)\right)\mathrm{d}\phi_\theta(a),
\end{aligned}$$

where the first equality follows from definition of expectation, the second equality holds because the distributions of contexts and shocks are independent and identically distributed and the last equality follows from the definition of ϕ.

Putting all together, we obtain that

$$T \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon \left(y_l(\theta, a, \epsilon) \right) \mathrm{d}\phi_\theta(a) \leq C_l, \ \forall l \in [L],$$

which is the desired inequality.

2. *Equality.* Notice that

$$J^\pi(C, T) = \sum_{t=1}^{T} \mathbb{E}_{\theta_t \sim p, a_t^\pi \sim \mathbb{P}_{t-1}, \epsilon_t \sim f} \left(r(\theta_t, a_t^\pi, \epsilon_t) \right).$$

And using the same arguments as before, we obtain

$$J^\pi(C, T) = T \sum_{\theta \in \Theta} p_\theta \int_{\mathcal{A}} \mathbb{E}_\epsilon \left(r(\theta, a, \epsilon) \right) \mathrm{d}\phi_\theta(a),$$

which is exactly $TJ^\phi(C/T)$.

Step 3. Fix $\delta > 0$ and let π be an $\delta-$approximately optimal solution of Problem $\mathcal{P}$, i.e., $J^\pi(C, T) \geq J^*(C, T) - \delta$. Applying Step 1 to π we define a feasible solution to the fluid problem ϕ. Therefore, we conclude that

$$J^*(C, T) \leq J^\pi(C, T) + \delta = TJ^\phi(C/T) + \delta \leq T\bar{J}(C/T) + \delta,$$

where the first equality follows from Step 2 because $J^\pi(C, T) = TJ^\phi(C/T)$, and the second inequality because, by Step 2 again, ϕ is a feasible solution of Problem $\mathcal{P}_{\text{FLUID}}$ for $\rho = C/T$. The result is obtained because δ was arbitrary. □

B.2 Proof of Theorem 1

The goal of this section is to prove Theorem 1, which is our main result regarding the performance of the heuristic CE for the set of DRC^2 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 desire result.

In what follows we will denote by y_t the resource consumption at time t if the decision maker follows the policy π^{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 the remaining horizon. More specifically, for each $t \in [T]$,

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

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

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

If $\|M_t\|$ is at most δ for all $t \in [T]$, we set $\tau_\delta = \infty$. On the other hand, we define τ_- 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 τ as the minimum between τ_δ and τ_-, and the number of periods T, i.e.,

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

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 τ_- are not constrained by resources and they are taken according to an optimal solution of the fluid 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 τ_δ and τ_- are stopping times with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 1}$, with $\mathcal{F}_t = \sigma(\theta_1, \ldots, \theta_t, a_1, \ldots, a_t, \epsilon_1, \ldots, \epsilon_t)$, the history up to the end of period t, and thus we obtain that τ 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 $\mathbb{E}\left(\mathbb{E}(y_{t+1}|\rho_{t+1}) - y_{t+1}|\mathcal{F}_t\right) = 0$, concluding that

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

Since $\{M_t\}_{t\geq 1}$ is a martingale and τ 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.

Lemma B-1. *Under Assumption 2, if t is at most τ_δ, it holds that:*

1. $(\rho_{t+1} - \rho_1)|_{\mathcal{C}} = M_t|_{\mathcal{C}}$ *and* $(\rho_{t+1} - \rho_1)|_{\mathcal{U}} \geq M_t|_{\mathcal{U}}$
2. $\|(\rho_t - \rho_1)|_{\mathcal{C}}\| < \delta$ *and* $(\rho_t - \rho_1)|_{\mathcal{U}} \geq -\mathbf{1}\delta$.

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 $\mathbb{E}(y_1|\rho_1)|_{\mathcal{C}} = \rho_1|_{\mathcal{C}}$ and $\mathbb{E}(y_1|\rho_1)|_{\mathcal{U}} \leq \rho_1|_{\mathcal{U}}$. Therefore, we can express $M_1|_{\mathcal{C}}$ and bound $M_1|_{\mathcal{U}}$ as follows:

$$M_1|_{\mathcal{C}} = \frac{(\mathbb{E}(y_1|\rho_1) - y_1)|_{\mathcal{C}}}{T-1} = \frac{\rho_1|_{\mathcal{C}} - y_1|_{\mathcal{C}}}{T-1}, \tag{B-1}$$

$$M_1|_{\mathcal{U}} = \frac{(\mathbb{E}(y_1|\rho_1) - y_1)|_{\mathcal{U}}}{T-1} \leq \frac{\rho_1|_{\mathcal{U}} - y_1|_{\mathcal{U}}}{T-1} \tag{B-2}$$

From the definition of ρ_1 and ρ_2, we have that $y_1 = \rho_1 T - \rho_2(T-1)$ and replacing in (B-1) and (B-2) follows that

$$M_1|_{\mathcal{C}} = (\rho_2 - \rho_1)|_{\mathcal{C}} \quad \text{and} \quad M_1|_{\mathcal{U}} \leq (\rho_2 - \rho_1)|_{\mathcal{U}},$$

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 $\|(\rho_{t+1} - \rho_1)|_{\mathcal{C}}\| < \delta$ and $(\rho_{t+1} - \rho_1)|_{\mathcal{U}} \geq -\mathbf{1}\delta$. Applying the induction hypothesis to t, it holds that $\|(\rho_{t+1} - \rho_1)|_{\mathcal{C}}\| = \|M_t|_{\mathcal{C}}\|$ and $(\rho_{t+1} - \rho_1)|_{\mathcal{U}} \geq M_t|_{\mathcal{U}}$. On the other

hand, $t < \tau_\delta$ and thus $\|M_t\| < \delta$, which implies that $\|M_t|_{\mathcal{C}}\| < \delta$ and $M_t|_{\mathcal{U}} \geq -\mathbf{1}\delta$, and the second statement follows.

In the remainder of the proof, we show that $(\rho_{t+2} - \rho_1)|_{\mathcal{C}} = M_{t+1}|_{\mathcal{C}}$ and $(\rho_{t+2} - \rho_1)|_{\mathcal{U}} \geq M_{t+1}|_{\mathcal{U}}$. Note that

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

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 $\|(\rho_s - \rho_1)|_{\mathcal{C}}\| < \delta$ for all $s \leq t+1$. Therefore we can apply Assumption 2 to the expression above obtaining that

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

On the other hand, $\mathbb{E}(y_s|\rho_s)|_{\mathcal{U}} \leq \rho_s|_{\mathcal{U}}$ and using again the expression above we conclude that

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

and the lemma follows. □

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 $\{\rho_{t\wedge\tau}|_{\mathcal{C}}\}_{t\geq 1}$ 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 τ is bounded by a constant. Specifically, the following result gives sufficient conditions on t to be a lower bound for the stopping time τ_-.

Lemma B-2. *Assume that Assumption 1 and 2 hold, and define* $T^- = T + 1 - 2\frac{\bar{y}_\infty}{\underline{\rho}_1 - \delta}$, *where* $\underline{\rho}_1$ *is the smallest component of vector* ρ_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 τ_-, 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 $\|(\rho_s - \rho_1)|_{\mathcal{C}}\| < \delta$ and $(\rho_s - \rho_1)|_{\mathcal{U}} \geq -\mathbf{1}\delta$. In particular, $|(\rho_s - \rho_1)_l| < \delta \ \forall l \in \mathcal{C}$, obtaining

$$\rho_s > \rho_1 - \mathbf{1}\delta \, , \tag{B-3}$$

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 ρ_s, together with inequality (B-3); 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 τ_- is greater than t and the proof is completed. □

We next prove that the expected number of remaining periods after the stopping time τ 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,*

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

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

$$\mathbb{E}(\tau) = \sum_{t=1}^{T} \mathbb{P}(\tau_\delta \wedge \tau_- \geq t) \geq \sum_{t=1}^{T^- - 1} \mathbb{P}(\tau_\delta \wedge \tau_- \geq t) \tag{B-4}$$

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

On the other hand,

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

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-4) 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 $\sum_{t=1}^{T^- - 1} \mathbb{P}\left(\max_{s\in[t]} \|M_s\| \geq \delta\right)$, 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 $\mathbb{P}\left(\max_{s\in[t]} \|M_s\| \geq \delta\right) \leq 1$, 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

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

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

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

$$\mathbb{E}(T - \tau) < T - T^{-} + 1 + 14\frac{\bar{y}_{\infty}^2}{\delta^2} = \frac{2\bar{y}_{\infty}}{\underline{\rho}_1 - \delta} + 14\frac{\bar{y}_{\infty}^2}{\delta^2},$$

and the desire result is obtained. □

The next results shows that increasing the availability of a resource that is not binding should not change the optimal objective value of a the fluid problem.

Lemma B-4. *Under Assumption 2, if the consumption constraint corresponding to the resource i is not binding for the problem $(\mathcal{P}_{\text{FLUID}})$ with $\rho = \rho_1$, that is $i \in \mathcal{U}$, then $\left(\nabla \bar{J}(\rho_1)\right)_i = 0$.*

Proof of Lemma B-4. We take $i \in \mathcal{U}$ and we prove that there exists $\Delta \in \mathbb{R}_{++}$ such that $\bar{J}(\rho_1) = \bar{J}(\rho_1 + e_i\kappa)$ for all $\kappa \in [-\Delta, \Delta]$, where e_i denotes the i-th canonical vector of $\mathbb{R}^L$. We divide the proof into two parts: In the first part, we prove that there exists $\Delta \in \mathbb{R}_{++}$ such that $\bar{J}(\cdot)$ is constant in $[\rho_1 - e_i\Delta, \rho_1]$, whereas in the second part, we prove that increasing the i-th component of ρ_1, the optimal solution of problem $(\mathcal{P}_{\text{FLUID}})$ does not change and therefore the optimal value is constant.

Part 1. If we denote by ϕ^* the optimal solution of $(\mathcal{P}_{\text{FLUID}})$ with $\rho = \rho_1$, by hypothesis we know that $\bar{y}_i < (\rho_1)_i$, where $\bar{y}_i = \mathbb{E}_{\theta\sim p, a\sim\phi^*, \epsilon\sim f} y_i(\theta, a, \epsilon)$ is the expected resource consumption under ϕ^*. Thus, ϕ^* is also feasible of $(\mathcal{P}_{\text{FLUID}})$ with $\rho = \rho_1 - \xi e_i$ for all $\xi \in [0, (\rho_1)_i - \bar{y}_i]$. Moreover, the feasibility set of the latter problem is contained in the feasibility set of the former, and therefore ϕ^* is an optimal solution of $(\mathcal{P}_{\text{FLUID}})$ with $\rho = \rho_1 - \xi e_i$ for all $\xi \in [0, (\rho_1)_i - \bar{y}_i]$, obtaining that

$$\bar{J}(\rho_1) = \bar{J}(\rho_1 + e_i\xi) \text{ for all } \xi \in [\bar{y}_i - (\rho_1)_i, 0]. \tag{B-5}$$

Part 2. On the other hand, suppose that there exists $\kappa > 0$ such that every optimal solution ϕ^*_κ of $(\mathcal{P}_{\text{FLUID}})$ for $\rho_\kappa = \rho_1 + \kappa e_i$ satisfies $(\bar{y}_\kappa)_i > (\rho_1)_i$ where $(\bar{y}_\kappa)_i = \mathbb{E}_{\theta\sim p, a\sim\phi^*_\kappa, \epsilon\sim f} y_i(\theta, a, \epsilon)$ is the expected resource consumption under ϕ^*_κ. Fix such an optimal solution ϕ^*_κ. Then, we can take a number $\gamma \in (0, 1]$ such that $\hat{\rho} = \gamma\bar{y} + (1-\gamma)\rho_\kappa \leq \rho_1$ because by hypothesis we know that $\bar{y}_i < (\rho_1)_i$. Note that $\hat{\phi} = \gamma\phi^* + (1-\gamma)\phi^*_\kappa$ is feasible for $(\mathcal{P}_{\text{FLUID}})$ with $\rho = \rho_1$, since

$$\begin{aligned}
\mathbb{E}_{\theta\sim p, a\sim\hat{\phi}, \epsilon\sim f} y(\theta, a, \epsilon) &= \gamma\mathbb{E}_{\theta\sim p, a\sim\phi^*, \epsilon\sim f} y(\theta, a, \epsilon) + (1-\gamma)\mathbb{E}_{\theta\sim p, a\sim\phi^*_\kappa, \epsilon\sim f} y(\theta, a, \epsilon) \\
&= \gamma\bar{y} + (1-\gamma)\bar{y}_\kappa \leq \gamma\bar{y} + (1-\gamma)\rho_\kappa = \hat{\rho} \leq \rho_1\,,
\end{aligned}$$

because $\bar{y}_\kappa \leq \rho_\kappa$ since ϕ^*_κ is feasible for ρ_κ. Moreover,

$$\begin{aligned}\mathbb{E}_{\theta\sim p,a\sim\hat{\phi},\epsilon\sim f} r(\theta,a,\epsilon) &= \gamma\mathbb{E}_{\theta\sim p,a\sim\phi^*,\epsilon\sim f} r(\theta,a,\epsilon) + (1-\gamma)\mathbb{E}_{\theta\sim p,a\sim\phi^*_\kappa,\epsilon\sim f} r(\theta,a,\epsilon)\\ &> \gamma\mathbb{E}_{\theta\sim p,a\sim\phi^*,\epsilon\sim f} r(\theta,a,\epsilon) + (1-\gamma)\mathbb{E}_{\theta\sim p,a\sim\hat{\phi},\epsilon\sim f} r(\theta,a,\epsilon),\end{aligned}$$

where the inequality follows because $\hat{\phi}$ is feasible but not optimal for ($\mathcal{P}_{\text{FLUID}}$) with $\rho = \rho_\kappa$ (i.e., $\hat{\phi}$ cannot be optimal because it consumes less than ρ_1 resources). Because $\gamma > 0$, the latter expression implies that $\mathbb{E}_{\theta\sim p,a\sim\hat{\phi},\epsilon\sim f} r(\theta,a,\epsilon) > \mathbb{E}_{\theta\sim p,a\sim\phi^*,\epsilon\sim f} r(\theta,a,\epsilon)$, which contradicts that ϕ^* is optimal solution of ($\mathcal{P}_{\text{FLUID}}$) with $\rho = \rho_1$.

We therefore have that for every positive number κ, the optimal solution ϕ^*_κ satisfies $(\bar{y}_\kappa)_i < (\rho_1)_i$, and therefore must be also solution of ($\mathcal{P}_{\text{FLUID}}$) with $\rho = \rho_1$, obtaining that

$$\bar{J}(\rho_1) = \bar{J}(\rho_1 + e_i\kappa) \text{ for all } \kappa > 0. \tag{B-6}$$

Taking $\Delta = (\rho_1)_i - \bar{y}_i$, and using (B-5) and (B-6), we obtain the result. □

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

Proof of Theorem 1. We have to bound $J^*(C,T) - J^{\text{CE}}(C,T)$, which is upper bounded by $T\bar{J}(\rho_1) - J^{\text{CE}}(C,T)$ because $J^*(C,T) \leq T\bar{J}(C/T)$ by Proposition 1. Thus, it is enough to bound $T\bar{J}(\rho_1) - J^{\text{CE}}(C,T)$. By dividing the horizon from 1 to τ and from τ to T and dropping the performance of the CE policy after time τ, we obtain

$$T\bar{J}(\rho_1) - J^{\text{CE}}(C,T) \leq \underbrace{\mathbb{E}\left(\sum_{t=1}^{\tau}\bar{J}(\rho_1) - \sum_{t=1}^{\tau} r(\theta_t, a_t^{\text{CE}}, \epsilon_t)\right)}_{(A)} + \underbrace{\mathbb{E}\left(\sum_{t=\tau+1}^{T}\bar{J}(\rho_1)\right)}_{(B)}, \tag{B-7}$$

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 τ considering the policy given by the CE heuristic equals the expected reward until time τ of the fluid problem for $\rho = \rho_t$ at time t. Let us prove that

$$\mathbb{E}\left(\sum_{t=1}^{\tau} r(\theta_t, a_t^{\text{CE}}, \epsilon_t)\right) = \mathbb{E}\left(\sum_{t=1}^{\tau}\bar{J}(\rho_t)\right). \tag{B-8}$$

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

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

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 τ 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 $\mathbb{E}_{\theta,\epsilon}\left(r(\theta_t, a^{\text{CE}}, \epsilon_t)|\rho_t\right) = \bar{J}(\rho_t)$ for all $t \leq \tau$. Then, (B-8) follows from combining these equations.

Second, using (B-8) and applying Assumption 2 (the hypothesis is fulfilled because $t \leq \tau$ and then by Lemma B-1 we have $\|(\rho_t - \rho_1)|_{\mathcal{C}}\| < \delta$) we obtain

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

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

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

where we used that $\{\rho_{t\wedge\tau}|_{\mathcal{C}}\}_{t\geq 1}$ is a martingale and that $\left(\nabla\bar{J}(\rho_1)\right)_l = 0$ for all $l \in \mathcal{U}$ by Lemma B-4, 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

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

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 $\mathbb{E}\left(||M_t|_{\mathcal{C}}||^2\right)$. Because martingale increments are orthogonal (see e.g. Gut 2013, Chapter 10 Lemma 4.1), we have that

$$\mathbb{E}\left(\|M_t|_{\mathcal{C}}\|^2\right) = \mathbb{E}\left\|\sum_{s=1}^{t}\frac{(\mathbb{E}(y_s|\rho_s) - y_s)|_{\mathcal{C}}}{T-s}\right\|^2 = \sum_{s=1}^{t}\frac{1}{(T-s)^2}\mathbb{E}\left\|(\mathbb{E}(y_s|\rho_s) - y_s)|_{\mathcal{C}}\right\|^2. \tag{B-9}$$

Furthermore, by definition of ℓ^2-norm, it holds that

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

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

Using expression (B-9), 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 γ is the Euler–Mascheroni constant.

Putting all together we get

$$\mathbb{E}\left(\sum_{t=1}^{\tau}\left(\bar{J}(\rho_1)-\bar{J}(\rho_t)\right)\right)\leq\frac{K}{2}\bar{y}_2^2(\gamma+\log T)\leq K\bar{y}_2^2\log(T)\,,\tag{B-10}$$

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-7). Note that

$$\mathbb{E}\left(\sum_{t=\tau+1}^{T}\bar{J}(\rho_1)\right)=\mathbb{E}(T-\tau)\bar{J}(C/T)\leq\left[\frac{2\bar{y}_\infty}{\underline{\rho}_1-\delta}+\frac{14\bar{y}_\infty^2}{\delta^2}\right]\bar{J}(C/T)\,,\tag{B-11}$$

where the inequality follows from applying Lemma B-3.

Putting everything together. Using (B-10) together with (B-11) in (B-7) we get

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

and the result follows. □

C Proofs of Section 5

C.1 Proof of Lemma 1

Proof of Lemma 1. Let us prove that the first statement of Assumption 2 holds by showing that the function $\bar{J}(\cdot)$ is linear over the set $\mathcal{N}(\rho_1, \delta, \mathcal{C}) = \{\rho : \|(\rho - \rho_1)|_{\mathcal{C}}\| \leq \delta, (\rho - \rho_1)|_{\mathcal{U}} \geq -\delta \mathbf{1}\}$ with δ given in the statement of this result. To this end, let us prove it first for $\rho \in \mathcal{N}(\rho_1, \delta) = \{\rho : \|\rho - \rho_1\| \leq \delta\}$. Take $\rho \in \mathcal{N}(\rho_1, \delta)$. That is, $\rho = \rho_1 + \epsilon v$ for some v unitary vector and ϵ a positive real number smaller than δ.

In the standard form representation of the problem given in problem (A-1), the constraint matrix is $Q = \binom{Q_1}{Q_2} \in \mathbb{R}^{(L+|\Theta|)\times(|\mathcal{A}|\cdot|\Theta|+L)}$ where $Q_1 \in \mathbb{R}^{L\times(|\mathcal{A}|\cdot|\Theta|+L)}$ is the matrix associated to the resource constraints and $Q_2 \in \mathbb{R}^{|\Theta|\times(|\mathcal{A}|\cdot|\Theta|+L)}$ is the matrix associated to the constraints $\sum_{a\in\mathcal{A}} \phi_\theta(a) = 1$. The matrix Q_1 is obtained by horizontally stacking the matrices $p_\theta\ \bar{y}_\theta \in \mathbb{R}^{L\times|\mathcal{A}|}$ and the identity matrix $I \in \mathbb{R}^{L\times L}$. The matrix Q_2 is obtained by horizontally stacking the matrices $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{|\Theta|} \in \mathbb{R}^{|\Theta|\times|\mathcal{A}|}$ and the zero matrix $\mathbf{0} \in \mathbb{R}^{|\Theta|\times L}$, where $\mathbf{e}_\theta \in \mathbb{R}^{|\Theta|\times|\mathcal{A}|}$ is the matrix with all columns equal to the θ-th canonical vector of $\mathbb{R}^{|\Theta|}$. 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 $\mathbf{1} \in \mathbb{R}^{|\Theta|}$ and $\mathbf{0} \in \mathbb{R}^{|\Theta|}$ denote a vector of ones and zeros, with a proper size, respectively. In this notation, we can write the perturbed right-hand vector as $\xi + \epsilon u$. Let B be the submatrix of Q corresponding to the columns associated to the basic variables at an optimal solution for $\rho = \rho_1$, and $B_{\rho_1}^{-1}$ the submatrix of B^{-1} associated to the resource constraints. We can write the optimal basic variable vector as $\binom{\phi_B}{x_B} = B^{-1}\xi$.

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 E-3 we have that if $0 \leq \epsilon \leq \frac{\min(\phi^*_{\min}, x^*_{\min})}{\|B_{\rho_1}^{-1}\|}$, B is an optimal basis for the standard problem with right hand side $\xi + \epsilon u$ and therefore the optimal basic variable vector, namely $\binom{\phi_B}{x_B}$, can be computed as $B^{-1}(\xi + \epsilon u)$. Let us define $c \in \mathbb{R}^{|\mathcal{A}|\cdot|\Theta|+L}$ the objective function coefficient vector of problem (A-1). That is, the coefficient vector is obtained by vertically stacking the vectors $p_\theta \bar{r}_\theta \in \mathbb{R}^{|\mathcal{A}|}$ and the zero vector $\mathbf{0} \in \mathbb{R}^L$. Thus, calling c_B to the coefficient vector associated to the basic variables, it holds that

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

where the last equality follows because B is optimal basis of problem (A-1) and $\epsilon v = \rho - \rho_1$. We then have that $\bar{J}(\cdot)$ is linear over $\mathcal{N}(\rho_1, \delta)$. Notice that if we increase the components of ρ corresponding to the unbinding constraints, then the optimal basis does not change (see Part 2 in the proof of Lemma B-4) and therefore the equalities above still holds, obtaining that $\bar{J}(\cdot)$ is linear over $\mathcal{N}(\rho_1, \delta, \mathcal{C})$ and the first statement holds with $K = 0$.

We proved that taking $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$ the optimal basis does not change and therefore the second statement follows directly because the constraints are binding for $\rho_1|_{\mathcal{C}}$.

Therefore, we conclude that Assumption 2 holds for $\delta = \frac{\min(\phi^*_{\min}, x^*_{\min})}{\|B^{-1}_{\rho_1}\|}$, and $K = 0$. □

C.2 Proof of Lemma 2

Proof of Lemma 2. We want to prove that the problem (6) has a unique non-degenerate optimal solution for $\rho = \rho_1$. To this end, we will prove that its dual problem has a unique non-degenerate optimal solution.

Note that if we denote by $z_\theta \in \mathbb{R}$ to the dual variable associated to the constraint $\sum_a \phi_\theta(a) = 1$, the dual problem of (6) can be written as follows

$$\begin{aligned} \min_{\mu, z_\theta} \quad & \rho_1^\top \mu + \sum_{\theta \in \Theta} z_\theta \\ \text{s.t} \quad & p_\theta \mu^\top \bar{y}_\theta(a) + z_\theta \geq p_\theta \bar{r}_\theta(a) \quad \forall a \in \mathcal{A} \quad \forall \theta \in \Theta\,, \\ & \mu \geq 0\,. \end{aligned} \tag{C-1}$$

From the constraints, for each $\theta \in \Theta$, the dual variable z_θ should satisfy $z_\theta \geq p_\theta(\bar{r}_\theta(a) - \mu^\top \bar{y}_\theta(a)) \ \forall a \in \mathcal{A}$, and therefore $z_\theta \geq p_\theta \max_{a \in \mathcal{A}}\{\bar{r}_\theta(a) - \mu^\top \bar{y}_\theta(a)\}$. Due to (C-1) being a minimization problem, we conclude that the optimal value for the dual variable z_θ is given by $z^1_\theta = p_\theta \max_{a \in \mathcal{A}}\{\bar{r}_\theta(a) - \mu^\top \bar{y}_\theta(a)\}$.

The dual problem is then equivalent to the following problem

$$\min_{\mu \in \mathbb{R}^L_+} \quad \rho^\top \mu + \sum_{\theta \in \Theta} p_\theta \max_{a \in \mathcal{A}}\{\bar{r}_\theta(a) - \mu^\top \bar{y}_\theta(a)\}\,,$$

which has a unique solution μ^1 by Assumption SC 2, and therefore we obtain that the dual problem (C-1) has a unique solution for $\rho = \rho_1$. It remains to show that such solution (μ^1, z^1) is not degenerate. To show that, we will prove that there is at most $L + |\Theta|$ (number of variables) active constraints. Note that the number of constraints in problem (C-1) is $L + |\Theta| \times |\mathcal{A}|$. However, by Assumption SC 3 we know that for each $\theta \in \Theta$, there exists a unique $a^*_\theta \in \mathcal{A}$ such that $z^1_\theta = p_\theta(\bar{r}_\theta(a^*_\theta) - \mu^\top \bar{y}_\theta(a^*_\theta))$ and thus $p_\theta {\mu^1}^\top \bar{y}_\theta(a) + z^1_\theta > p_\theta \bar{r}_\theta(a) \ \ \forall a \in \mathcal{A} \setminus \{a^*_\theta\}$. We conclude that exactly $|\Theta|$ constraints from the first set are binding and therefore there are at most $L + |\Theta|$ active

constraints, obtaining the desired result. □

C.3 Proof of Proposition 2

Proof of Proposition 2. We divide the proof into two steps. First, we prove that an optimal solution to the dual problem exists, namely μ^*. Then, we define ϕ^* properly and we apply Proposition 5.1.5 in Bertsekas [1997] to prove that ϕ^* is primal solution and μ^* is in fact a Lagrangian multiplier and that therefore there is no duality gap, obtaining the desired result.

Step 1. Note that for each $\theta \in \Theta$, g_θ is convex because it is defined as the supremum of a family of linear functions and therefore the dual problem is a convex problem. Then, the function g is convex because convex combinations of convex functions are convex.

To prove the existence of optimal dual solution μ^*, we first prove Ψ_ρ is 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. Continuity of the dual function follows because it is differentiable (as we argue in the step 2 below). 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 $\bar{\mu}_\rho = \bar{r}_\infty / \underline{\rho}$, where $\underline{\rho} = \min_{l \in [L]} \rho_l$ and $\bar{r}_\infty$ is the positive real number provided by Assumption1.1. 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 μ greater than $\bar{\mu}_\rho$. Then, we have

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

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 μ and ρ, the third inequality holds because L_1 contains at least one element and $\rho_l \geq \underline{\rho}$, and the last one follows because $\Psi_\rho(0) = \mathbb{E}_{\theta \sim p}\left[\max_{a \in \mathcal{A}} \bar{r}(\theta, a)\right]$ 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 μ^* optimal dual solution satisfying $\mu^* \in [0, \bar{\mu}_\rho]^L$.

Step 2. Given $\rho > 0$, take μ^* 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 4. Let us now show that (ϕ^*, μ^*) 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 the gradient of $\bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a)$ with respect to μ exists and is given by $-\bar{y}(\theta, a)$. Because the value function $g_\theta(\mu)$ is differentiable and achieved for some action a^*_θ, we have from the envelope theorem applied to g_θ (see, e.g., Theorem 1 in Milgrom and Segal 2002), that $\nabla g_\theta(\mu) = -\bar{y}(\theta, a^*_\theta)$. The last statement holds almost surely over $\theta \in \Theta$. Because $g_\theta(\mu)$ is convex, we obtain from Theorem 7.46 of Shapiro et al. [2009] that $g(\mu) = \mathbb{E}_{\theta\sim p}[g_\theta(\mu)]$ is differentiable with gradient $\nabla g(\mu) = -\mathbb{E}_{\theta\sim p}[\bar{y}(\theta, a^*_\theta)]$. Therefore, Ψ_ρ is differentiable and its gradient evaluated at μ^* is given by

$$\nabla \Psi_\rho(\mu^*) = \rho - \mathbb{E}_{\theta\sim p}[\bar{y}(\theta, a^*_\theta)]\,. \tag{C-2}$$

   Because μ^* 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

$$\nabla \Psi_\rho(\mu^*)^\top (\mu - \mu^*) \geq 0,\ \forall \mu \in \mathbb{R}^L_+\,. \tag{C-3}$$

   Letting $\mu_l \to \infty$, we obtain that $\nabla \Psi_\rho(\mu^*) \geq 0$, which, in turn, implies that $\mathbb{E}_{\theta\sim p}[\bar{y}(\theta, a^*_\theta)] \leq \rho$ by (C-2). 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-3), 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

$$\begin{aligned}
\arg\max_{\phi\in\Phi} \mathcal{L}(\phi, \mu^*) &= \arg\max_{\phi\in\Phi} \left\{ \mu^* \rho + \mathbb{E}_{\theta\sim p, a\sim\phi(\theta)} \left[\bar{r}(\theta, a) - {\mu^*}^\top \bar{y}(\theta, a) \right] \right\} \\
&= \left\{ \arg\max_{\phi_\theta \in \Delta(\mathcal{A})} \mathbb{E}_{a\sim\phi_\theta} \left[\bar{r}(\theta, a) - {\mu^*}^\top \bar{y}(\theta, a) \right] \right\}_{\theta\in\Theta} \\
&= \left\{ \arg\max_{a\in\mathcal{A}} g_\theta(\mu^*) \right\}_{\theta\in\Theta},
\end{aligned}$$

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

Therefore, the four conditions hold and the proof is complete. □

C.4 Proof of Lemma 3

Proof of Lemma 3. We have to show that if $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C}) = \{\rho : \|(\rho - \rho_1)|_{\mathcal{C}}\| \leq \delta, (\rho - \rho_1)|_{\mathcal{U}} \geq -\delta \mathbf{1}\}$, with $\delta = (\nu\kappa)/2$ then it holds that

1. $\bar{J}(\rho) \geq \bar{J}(\rho_1) + \nabla \bar{J}(\rho_1)(\rho - \rho_1) - \frac{K}{2}\|\rho|_{\mathcal{C}} - \rho_1|_{\mathcal{C}}\|^2$,
2. $\left(\mathbb{E}_{\theta \sim p, a \sim \phi^*(\theta)}\left[\bar{y}(\theta, a)\right]\right)|_{\mathcal{C}} = \rho|_{\mathcal{C}}$.

We divide the proof into five parts: First, we extend the strong convexity lower bound of g in SC 7 to the entire domain. Then, we use a claim (that we prove in Part 4) to show the second statement of Assumption 2. In the third part of the proof, we show that given $\delta \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$, the optimal solution of (5) is contained in $\mathcal{N}(\mu^1, \nu) \cap \mathbb{R}^L_+$. In the last part, we use the preliminary results showed along the proof to prove the first statement of Assumption 2.

Part 1. We first extend the strong convexity lower bound of g_θ to the entire domain. By SC 7, g admits a κ-LUQ envelope in $I^\nu = \mathcal{N}(\mu^1, \nu) \cap \mathbb{R}^L_+$. Then, for all $\mu \in I^\nu$.

$$g(\mu) \geq g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \frac{\kappa}{2}\|\mu - \mu^1\|^2 \,. \tag{C-4}$$

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

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

which can be reordered to give

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

where the second inequality follows by (C-4) 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 $\mathcal{N}(\mu^1, \nu)$. Combining both cases we obtain that

$$g(\mu) \geq g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \kappa \ell(\mu - \mu^1)\,, \tag{C-5}$$

where

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

Part 2. Let us now see the second statement. We will actually prove a stronger result. Let μ be any optimal solution of (5) for a fixed ρ, i.e., $\mu \in \arg\min_{\mu \in \mathbb{R}^L_+} \Psi_\rho(\mu)$ with $\Psi_\rho(\mu) = \rho^\top \mu + g(\mu)$ the Lagrange dual function. We shall show that $\mu_i > 0\ \forall i \in \mathcal{C}$ and $\mu_i = 0\ \forall i \in \mathcal{U}$ for every optimal solution when the resource vector ρ belongs to $\mathcal{N}(\rho_1, \delta, \mathcal{C})$. In the latter we denote by $\mathcal{C} = [L] \setminus \mathcal{U} = \{j \in [L] : \mu^1_j > 0\}$ the resources with positive initial dual variables.

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

$$\mu_i \left(\rho - \mathbb{E}_{\theta \sim p, a \sim \phi(\theta)} \left[\bar{y}(\theta, a)\right]\right)_j = 0,$$

where μ is the optimal solution of (5). Therefore, the previous claim would imply for all $j \in \mathcal{C}$ and $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$ then

$$\mathbb{E}_{\theta \sim p, a \sim \phi(\theta)} \left[\bar{y}_j(\theta, a)\right] = \rho_j,$$

where ϕ^* is an optimal solution of the fluid problem when the resource vector is ρ, and thus the second statement of Assumption 2 would follow.

In the rest of this part, we use the bound obtained in the first part of the proof to bound the Lagrange dual function. We prove the claim in Part 4.

For all $\mu \geq 0$, we have from (C-5) that

$$\begin{aligned} \Psi_\rho(\mu) &\geq \rho^\top \mu + g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \kappa \ell(\mu - \mu^1) \\ &= \rho^\top \mu^1 + g(\mu^1) + (\mu - \mu^1)^\top \left(\nabla g(\mu^1) + \rho\right) + \kappa \ell(\mu - \mu^1) \\ &= \Psi_\rho(\mu^1) + \kappa \ell(\mu - \mu^1) + (\mu - \mu^1)^\top \left(g(\mu^1) + \rho\right)\,, \end{aligned}$$

where we used that $\Psi_\rho(\mu^1) = \rho^\top \mu^1 + g(\mu^1)$. Because μ^1 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

$$(\bar{\mu} - \mu^1)^\top \nabla \Psi_\rho(\mu^1) \geq 0,\ \forall \bar{\mu} \in \mathbb{R}^L_+\,,$$

and applying it for $\bar{\mu} = \mu$ and using that $\nabla \Psi_\rho(\mu^1) = \nabla g(\mu^1) + \rho_1$ we obtain

$$(\mu - \mu^1)^\top \left(\nabla g(\mu^1) + \rho\right) \geq (\mu - \mu^1)^\top (\rho - \rho_1)\,.$$

Putting all together we conclude that

$$\Psi_\rho(\mu) \geq \Psi_\rho(\mu^1) + \kappa \ell(\mu - \mu^1) + (\mu - \mu^1)^\top (\rho - \rho_1)\,. \tag{C-6}$$

Part 3. We now prove that μ, the optimal solution of (5) for a fixed $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$, belongs to the truncated ball $I^\nu = \mathcal{N}(\mu^1, \nu) \cap \mathbb{R}^L_+$. To this end, we show that $\min_{\mu \notin I^\nu} \Psi_\rho(\mu) > \Psi_{\rho_1}(\mu^1)$. To do so, using (C-6), it is sufficient to prove that

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

Note that it is enough to consider ρ with $\|(\rho_1 - \rho)\| < (\kappa\nu)/2$. In fact, if we have $\bar{\rho} = \rho + \xi$, where $\rho \in \mathcal{N}(\rho_1, (\kappa\nu)/2)$ and $\xi \in \mathbb{R}^L_+$ with $\xi_i = 0$ for all $i \in \mathcal{C}$ and $\|\xi\| > 0$, we obtain

$$(\mu^1 - \mu)^\top (\rho_1 - \bar{\rho}) = (\mu^1 - \mu)^\top (\rho_1 - \rho - \xi) = (\mu^1 - \mu)^\top (\rho_1 - \rho) + \sum_{1 \in \mathcal{U}} \mu_i \xi_i \geq (\mu^1 - \mu)^\top (\rho_1 - \rho)\,,$$

where we use that $\mu^1_i = 0$ for all $i \in \mathcal{U}$.

Let us show now that given ρ such that $\|(\rho_1 - \rho)\| < (\kappa\nu)/2$, it holds that $(I) > 0$. Take $\mu \notin I^\nu$. We then have $\|\mu - \mu^1\| > \nu$ and, in this case, $\ell(z) = \nu\|z\|/2$. Using Cauchy-Schwartz we obtain

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

where the third inequality follows because $\|(\rho_1 - \rho)\| < (\kappa\nu)/2$. We conclude that μ belongs to I^ν.

Part 4. In this part we prove the claim of Part 2, that is, that $\mu_i > 0\ \forall i \in \mathcal{C}$ and $\mu_i = 0\ \forall i \in \mathcal{U}$ for every optimal dual solution when the resource vector ρ belongs to $\mathcal{N}(\rho_1, \delta, \mathcal{C})$.

- Case $\mathcal{U}$. We first handle the resources that are unconstrained at the initial optimal dual variable. We will prove that if ρ is suitably chosen, then the optimal solution of the dual problem satisfies $\mu_i = 0$ for all $i \in \mathcal{U}$. Recall that the first order conditions are given by

$$\nabla \Psi_\rho(\mu)^\top (\bar{\mu} - \mu) \geq 0,\ \forall \bar{\mu} \in \mathbb{R}^L_+\,, \tag{C-7}$$

  where $\nabla \Psi_\rho(\mu) = \rho + \nabla g(\mu)$. By Part 3, we know that $\mu \in I^\nu$ and thus we are under the

hypothesis of Assumption SC 8, obtaining $\rho_j + \frac{\partial g}{\partial \mu_i}(\mu) > 0$ for all $i \in \mathcal{U}$. We then conclude that $\mu_i = 0$ for all $i \in \mathcal{U}$. Otherwise, if there exists $j \in \mathcal{U}$ with $\mu_j > 0$, we can take $\bar{\mu}_l = \mu_l \mathbf{1}_{\{l \neq j\}} \in \mathbb{R}^L_+$, contradicting (C-7).

- Case $\mathcal{C}$. We now handle the resources that are constrained at the initial optimal dual variable. Notice that since $\nu < \underline{\mu} = \min_{i \in \mathcal{C}} \mu^1_i$ and $\mu^1|_{\mathcal{C}} > 0$, it holds that for all $\bar{\mu} \in I^\nu$, $\bar{\mu}|_{\mathcal{C}}$ is strictly positive. On the other hand, by Part 3 we know that the optimal dual solution μ belongs to I^ν, concluding that $\mu|_{\mathcal{C}} > 0$, and the claim is proved.

Part 5. It remains to prove the first statement of Assumption 2. To this end, we use the lower bound C-5 on g obtained in Part 1 of the proof, together with the claim proved in Part 4.

The function $\ell : \mathbb{R}^L \to \mathbb{R}$ in (C-5) 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 $f^* : \mathbb{R}^L \to \mathbb{R}$ 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 $\ell(z) \geq f^*(z)$.) Combining this with (C-5) we obtain that

$$g(\mu) \geq g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1)\,. \tag{C-8}$$

Using this lower bound on g to bound Ψ_ρ we have

$$\begin{aligned}
\bar{J}(\rho) &= \min_{\{\mu \geq 0\colon\ \mu|_{\mathcal{U}} = 0\}} \rho^\top \mu + g(\mu) \\
&\geq \min_{\{\mu \geq 0\colon\ \mu|_{\mathcal{U}} = 0\}} \left\{ \rho^\top \mu + g(\mu^1) + \nabla g(\mu^1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1) \right\} \\
&= \rho_1^\top \mu^1 + g(\mu^1) + (\mu^1)^\top (\rho - \rho_1) + \min_{\{\mu \geq 0\colon\ \mu|_{\mathcal{U}} = 0\}} \left\{ (\rho - \rho_1)^\top (\mu - \mu^1) + \kappa f^*(\mu - \mu^1) \right\} \\
&= \bar{J}(\rho_1) + \nabla \bar{J}(\rho_1)^\top (\rho - \rho_1) + \underbrace{\min_{\{z \geq -\mu^1,\ z|_{\mathcal{U}} = 0\}} \left\{ \hat{\rho}^\top z + \kappa f^*(z) \right\}}_{(E)},
\end{aligned}$$

where the second equality holds because for all constrained resources $i \in \mathcal{C}$ we have by the first-order conditions of the dual problem that $\frac{\partial g}{\partial \mu_i}(\mu^1) + \rho_{1,i} = 0$ and for all unconstrained resources $i \in \mathcal{U}$ we have $\mu_i = \mu^1_i = 0$, and the last equality follows from performing the change of variables $\mu - \mu^1 = z$, defining the vector $\hat{\rho} \in \mathbb{R}^L_+$ as $\hat{\rho}|_{\mathcal{C}} = (\rho - \rho_1)|_{\mathcal{C}}$ and $\hat{\rho}|_{\mathcal{U}} = \mathbf{0}$, and because $\bar{J}(\rho_1) = \rho_1^\top \mu^1 + g(\mu^1)$

together with $\nabla \bar{J}(\rho_1) = \mu^1$ from the envelope theorem. Note that envelope theorem applies to J because both g and J—in a neighborhood of ρ_1—are continuously differentiable (it follows from Assumption SC 4, 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

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

where the first inequality follows from relaxing the constraints $z \geq -\mu^1$ and $z|_{\mathcal{U}=0}$, 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 E-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 $\|\hat{\rho}\| = \|(\rho - \rho_1)|_{\mathcal{C}}\| \leq \nu\kappa/2$, we have

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

Putting it all together, we conclude that for $\rho \in \mathcal{N}(\rho_1, \delta, \mathcal{C})$ with $\delta = (\nu\kappa)/2$,

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

The result follows. □

C.5 Proof of Lemma 4

Proof of Lemma 4. We will show that under assumptions SC 5-SC 11, $g(\mu)$ admits a κ-LUQ envelope in $\mathcal{N}(\mu^1, \nu)$ for $\nu = \kappa\varphi/\sigma$ and $\kappa = \kappa_r + (\nu + \|\mu^1\|)\|\kappa_y\|$. Fix a context $\theta \in \Theta$ for which the assumptions hold.

From SC 10, $\bar{r}(\theta, \cdot)$ admits a κ_r-LDQ envelope in $\mathcal{N}(a^1_\theta, \varphi)$. That is, for all $\theta \in \Theta$,

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

On the other hand, from SC 11, $\bar{y}(\theta, \cdot)$ admits a κ_y-UUQ in $\mathcal{N}(a^1_\theta, \varphi)$. That is, for all $\theta \in \Theta$,

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

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

$$\begin{aligned}
&\bar{r}(\theta, a) - \mu^\top \bar{y}(\theta, a) \\
&\quad \geq \bar{r}(\theta, a_\theta^1) - \mu^\top \bar{y}(\theta, a_\theta^1) + (\nabla \bar{r}(\theta, a_\theta^1) - \nabla \bar{y}(\theta, a_\theta^1)^\top \mu)^\top (a - a_\theta^1) - \frac{\kappa_r + \mu^\top \kappa_y}{2} \|a - a_\theta^1\|^2 \\
&\quad = g_\theta(\mu^1) + \nabla g_\theta(\mu^1)^\top (\mu - \mu^1) + (\nabla \bar{y}(\theta, a_\theta^1)^\top (\mu^1 - \mu))^\top (a - a_\theta^1) - \frac{\kappa_r + \mu^\top \kappa_y}{2} \|a - a_\theta^1\|^2, \quad \text{(C-9)}
\end{aligned}$$

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_θ (by assumption SC 9, a_θ^1 is interior), and because $\nabla g_\theta(\mu^1) = -\bar{y}(\theta, a_\theta^1)$ from the envelope theorem applied to g_θ (using compactness of $\mathcal{A}$ and Assumptions SC 5-SC 6 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-9), 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 E-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 E-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 σ 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 by taking expectations over contexts and using that the assumptions hold almost surely over $\theta \in \Theta$ together with the fact that $g_\theta(\mu)$ is uniformly bounded from below (by zero). □

C.6 Proof of Lemma 5

Proof. By Danskin's theorem we have that the value function $g_\theta(\mu)$ is differentiable whenever the set of optimal actions is unique, which, in light of SC 12, happens almost surely over the contexts θ. Because $g_\theta(\mu)$ is convex, we obtain from Theorem 7.46 of Shapiro et al. [2009] that $g(\mu) = \mathbb{E}_{\theta\sim p}[g_\theta(\mu)]$ is differentiable with gradient

$$\nabla g(\mu) = -\mathbb{E}_{\theta\sim p}\left[\bar{y}(\theta)\mathbf{1}\left\{\bar{r}(\theta) \geq \mu^\top \bar{y}(\theta)\right\}\right],$$

because action $a = 1$ is optimal whenever $\bar{r}(\theta) \geq \mu^\top\bar{y}(\theta)$. We prove SC 7 holds by showing that

$$\big(\nabla g(\mu) - \nabla g(\mu^1)\big)^\top(\mu - \mu^1) \geq \kappa\|\mu - \mu^1\|^2, \tag{C-10}$$

for every $\mu \geq 0$ with $\|\mu - \mu^1\| \leq \nu$. To see that the latter condition implies the result, write $\ell(\mu) = g(\mu) - (\kappa/2)\mu^\top\mu$ and note that condition (C-10) is equivalent to $\big(\nabla\ell(\mu) - \nabla\ell(\mu^1)\big)^\top(\mu - \mu^1) \geq 0$. Now consider the function $h(\alpha) = \ell\big(\mu^1 + \alpha(\mu - \mu^1)\big)$ with $h'(\alpha) = \nabla\ell\big(\mu^1 + \alpha(\mu - \mu^1)\big)^\top(\mu - \mu^1)$. We obtain that

$$\begin{aligned}\ell(\mu) = h(1) &= h(0) + \int_0^1 h'(\alpha)d\alpha \\ &= \ell(\mu^1) + \int_0^1 \nabla\ell\big(\mu^1 + \alpha(\mu - \mu^1)\big)^\top(\mu - \mu^1)d\alpha \geq \ell(\mu^1) + \nabla\ell(\mu^1)^\top(\mu - \mu^1),\end{aligned}$$

where the inequality follows from using (C-10) with $\mu = \mu^1 + \alpha(\mu - \mu^1)$. The claim follows.

Using our formula for the gradient, we obtain that

$$\big(\nabla g(\mu) - \nabla g(\mu^1)\big)^\top(\mu - \mu^1) = \mathbb{E}_{\theta\sim p}\left[D(\theta)\right].$$

where

$$D(\theta) := \bar{y}(\theta)^\top(\mu - \mu^1)\left(1\left\{\bar{r}(\theta) \geq \bar{y}(\theta)^\top\mu^1\right\} - \mathbf{1}\left\{\bar{r}(\theta) \geq \bar{y}(\theta)^\top\mu\right\}\right).$$

Let $M(\theta) = \max\left(\bar{y}(\theta)^\top \mu^1, \bar{y}(\theta)^\top \mu\right)$ and $m(\theta) = \min\left(\bar{y}(\theta)^\top \mu^1, \bar{y}(\theta)^\top \mu\right)$. Then, we can write the term inside the expectation as

$$D(\theta) = (M(\theta) - m(\theta))\mathbf{1}\left\{\bar{r}(\theta) \in [m(\theta), M(\theta))\right\},$$

which follows simply by using that $M(\theta) = \bar{y}(\theta)^\top \mu^1$ and $m(\theta) = \bar{y}(\theta)^\top \mu$ when $\bar{y}(\theta)^\top \mu^1 \geq \bar{y}(\theta)^\top \mu$ and viceversa. Because $D(\theta) \geq 0$, we obtain by restricting to the event $M(\theta) \leq \bar{r}_\infty$ and taking expectations that

$$\begin{aligned}
\left(\nabla g(\mu) - \nabla g(\mu^1)\right)^\top (\mu - \mu^1) &= \mathbb{E}_{\theta\sim p}\left[D(\theta)\right] \\
&\geq \mathbb{E}_{\theta\sim p}\left[D(\theta)\mathbf{1}\left\{M(\theta) \leq \bar{r}_\infty\right\}\right] \\
&= \mathbb{E}_{\theta\sim p}\left[(M(\theta) - m(\theta))\mathbb{P}\left\{\bar{r}(\theta) \in [m(\theta), M(\theta)) \mid \bar{y}(\theta)\right\}\mathbf{1}\left\{M(\theta) \leq \bar{r}_\infty\right\}\right] \\
&\geq \underline{p}\mathbb{E}_{\theta\sim p}\left[(M(\theta) - m(\theta))^2\mathbf{1}\left\{M(\theta) \leq \bar{r}_\infty\right\}\right] \\
&= \underline{p}(\mu - \mu^1)^\top \mathbb{E}_{\theta\sim p}\left[y(\theta)y(\theta)^\top \mathbf{1}\left\{M(\theta) \leq \bar{r}_\infty\right\}\right](\mu - \mu^1) \\
&\geq \underline{p}(\mu - \mu^1)^\top \mathbb{E}_{\theta\sim p}\left[y(\theta)y(\theta)^\top \mathbf{1}\left\{y(\theta)^\top \mu^1 + \nu\|y(\theta)\| \leq \bar{r}_\infty\right\}\right](\mu - \mu^1) \\
&\geq \underline{p}\lambda\|\mu - \mu^1\|^2,
\end{aligned}$$

where the second equality follows from conditioning on $\bar{y}(\theta)$ and using that $M(\theta)$ and $m(\theta)$ are measurable with respect to $\bar{y}(\theta)$, the second inequality from using SC 13 together with the fact that $[m(\theta), M(\theta)) \subseteq [0, \bar{r}_\infty]$, the third equality because

$$(M(\theta) - m(\theta))^2 = \left(\bar{y}(\theta)^\top(\mu - \mu^1)\right)^2 = (\mu - \mu^1)^\top y(\theta)y(\theta)^\top(\mu - \mu^1)$$

and using the linearity of expectations, the third inequality because

$$\begin{aligned}
M(\theta) &= \max\left(\bar{y}(\theta)^\top \mu^1, \bar{y}(\theta)^\top \mu\right) \\
&\leq \bar{y}(\theta)^\top \mu^1 + \left|\bar{y}(\theta)^\top(\mu - \mu^1)\right| \\
&\leq \bar{y}(\theta)^\top \mu^1 + \|\bar{y}(\theta)\| \cdot \|\mu - \mu^1\| \\
&\leq \bar{y}(\theta)^\top \mu^1 + \nu\|\bar{y}(\theta)\|
\end{aligned}$$

from Cauchy-Schwartz and using that $\|\mu - \mu^1\| \leq \nu$, and the last inequality from SC 14. The result follows with $\kappa = \underline{p}\lambda$. □

D Complement to Section 6: CE Heuristic Performance Across Subclasses of Problems

In this section, we complement the exposition in §6 by providing the detailed sufficient conditions on the primitives in each problem class.

D.1 Network Dynamic Pricing

For simplicity we assume that contexts are finite—similar results can be provided when there is a continuum of contexts.

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:

$$\begin{aligned} \bar{J}(\rho) = \quad & \max_{\phi \in \Phi} \quad \sum_{\theta \in \Theta} p_\theta \mathbb{E}_{a \sim \phi(\theta)} \left[\bar{r}(\theta, a) \right] \\ & \text{s.t} \quad \sum_{\theta \in \Theta} p_\theta \mathbb{E}_{a \sim \phi(\theta)} \left[Q_\theta \bar{D}(\theta, a) \right] \leq \rho. \end{aligned} \tag{D-11}$$

In this problem, we will assume, as is commonly done in the literature, that there is a continuum of actions (we comment on the case of finite actions later). We map conditions SC 5-SC 11 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 4 to hold.

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

In particular, under the conditions above, from Lemma 3, 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 σ is a lower bound on the minimum singular value of $Q_\theta \nabla \bar{D}(\theta, a_\theta^1)$.

Therefore, we can use Corollary 2 to deduce that the revenue loss of the certainty equivalent heuristic is of order $O(\log T)$ in this case. 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.

D.2 Dynamic Bidding in Repeated Auctions

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: $\bar{q}(a) = \mathbb{E}_{\epsilon\sim f}[q(a,\epsilon)]$, $\bar{m}(a) = \mathbb{E}_{\epsilon\sim f}[m(a,\epsilon)]$. For simplicity, we assume that the set of values Θ is finite, but our results hold when the set of values is a continuum. For the particular setting described in §3.2, the fluid problem ($\mathcal{P}_{\text{FLUID}}$) is equivalent to the following problem:

$$\begin{aligned} \bar{J}(\rho) = \quad & \max_{\phi\in\Phi} \quad \sum_{\theta\in\Theta} p_\theta \mathbb{E}_{a\sim\phi(\theta)}\left[\theta\bar{q}(a) - \bar{m}(a))\right] \\ & \text{s.t} \quad \sum_{\theta\in\Theta} p_\theta \mathbb{E}_{a\sim\phi(\theta)}\left[\bar{m}(a)\right] \leq \rho\,. \end{aligned} \tag{D-12}$$

For each value $\theta \in \Theta$, let $g_\theta(\mu) = \max_{a\in\mathcal{A}}\{\theta\bar{q}(a) - (\mu+1)\bar{m}(a)\}$. Assumption SC 4 requires that $g_\theta(\mu)$ is differentiable in μ and that $g_\theta(\mu)$ is achieved by an action. Under these conditions, Proposition 2 implies that strong duality holds and the Problem (D-12) 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 $\beta : \Theta \to \mathcal{A}$. That is, given an advertiser with valuation θ, 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.

Proposition D-1. *Under Assumption SC 4, an optimal solution of (D-12) is to bid $\beta(\theta/(1+\mu^*))$ when the value is θ, where μ^* is the optimal solution of the dual problem of (D-12).*

Proof. Recall that there exist μ^* optimal dual solution satisfying $\mu^* \in [0,\bar{\mu}]$ (see Step 1 in the proof of Proposition 2). 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 μ^* 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 μ^*

is optimal dual solution, we have that

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

where the derivative of Ψ_ρ is given by

$$\Psi'_\rho(\mu) = \rho + \sum_{\theta \in \Theta} p_\theta g'_\theta(\mu) = \rho - \sum_{\theta \in \Theta} p_\theta \bar{m} \left(\beta \left(\frac{\theta}{1+\mu} \right) \right) . \tag{D-13}$$

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 (D-13) 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 $\Psi'_\rho(\mu^*)(-\nu)$ and non-negative, obtaining $\Psi'_\rho(\mu^*) = 0$, and primal feasibility and complementary slackness hold.

3. *Lagrangian optimality.* Note that

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

where the second equality holds because we can separate the problem for each θ. 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. □

If in addition Assumptions SC 7 and SC 8 hold, from Lemma 3 we obtain that 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 4 and SC 7 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 ϵ to a random variable capturing the maximum bid of the competitors and take $\mathcal{E} = \mathbb{R}_+$. Again, we assume

ϵ is distributed according to f, with density function f'. While the maximum competing bid ϵ is assumed to be independent of the values θ, our results can easily incorporate correlation (see Appendix A.1). Without loss, 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 ξ-Lipschitz continuous with respect to a^1_θ in $\mathcal{N}(a^1_\theta, \varphi)$, i.e., $|f'(a) - f'(a^1_\theta)| \leq \xi |a - a^1_\theta|$ for all $a \in \mathcal{N}(a^1_\theta, \varphi)$.

In the next Lemma, we show that these conditions are sufficient to apply Proposition 3 and Lemma 4 and, in turn, Corollary 3 holds. This leads to a revenue loss of logarithmic order in T.

Lemma D-1. *If f absolutely continuous and strictly increasing and f' is locally ξ-Lipschitz continuous in $\mathcal{N}(a^1_\theta, \varphi)$, then conditions SC 5-SC 11 hold.*

Proof. Let us see that conditions SC 5- SC 11 hold.

- **Condition SC 5:** By hypothesis f is absolutely continuous and therefore both $\bar{q}$ and $\bar{m}$ are continuous.
- **Condition SC 6:** For each $\theta \in \Theta$, let us define the function $G_\theta : \mathcal{A} \to \mathbb{R}$ by $G_\theta(a) = \theta f(a) - \int_0^a x \, \mathrm{d}f(x)$. Consider $\theta \in (0, \Theta_{\max})$. Note that $G'_\theta(a) = (\theta - a) f'(a)$. Because $\theta > 0$, then $\lim_{a \searrow 0} G'_\theta(a) > 0$ and $\beta(\theta) \neq 0$. Because $\theta \neq \Theta_{\max}$, we also have $\lim_{a \nearrow \Theta_{\max}} G'_\theta(a) < 0$. Therefore, $\beta(\theta) \in \arg\max_a G_\theta(a)$ is interior and $\beta(\theta)$ satisfies the first-order condition

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

  Therefore, as the cumulative distribution function f is strictly increasing, the unique optimum is to bid truthfully, as it is known in the literature. We conclude that SC 6 holds because the optimal auction is interior almost surely.
- **Condition SC 9:** Due to the truthfulness property of the second price auction, from Proposition D-1, it follows that $a^1_\theta = \theta/(1+\mu^1)$ which belongs to $(0, \Theta_{\max})$ due to the fact that the bid is positive, and thus condition SC 9 holds.
- **Conditions SC 10 and SC 11:** Note first that $\bar{r}(\theta, a) = \theta f(a) - \bar{m}(\theta, a)$ and $\bar{y}(\theta, a) = \bar{m}(a)$, because f is absolutely continuous. Then, it is enough to show the conditions hold for

$h(\theta,\cdot)=\theta f(\cdot)$ and $\bar{m}(a)=\int_0^a x\,\mathrm{d}f(x)$. Specifically, we will show that if the density function f' is locally ξ-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 $\bar{m}'(a)=af'(a)$ 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 ξ-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 φ. We now show that $\bar{m}'(a)$ is locally Lipschitz continuous:

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

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. □

First-price auctions. In a first-price auction, the winner is the highest bidder but pays his bid. Again, we reduce the definition of ϵ to a random variable capturing the maximum bid of the competitors. We assume ϵ 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 $m(a,\epsilon)=a1_{\{a\geq\epsilon\}}$, 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_θ^1 maximizing $\theta\bar{q}(a)-(1+\mu^1)\bar{m}(a)$ is interior. That is, there exists a positive number φ 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_θ^1 in $\mathcal{N}(a_\theta^1, \varphi)$.

Moreover, we show in the next Lemma that if we have $\mu^1 > 0$, then assumptions SC 5-SC 11 hold and, therefore, by Corollary 3 we obtain a revenue loss of logarithmic order in T.

Lemma D-2. *If f absolutely continuous, $M(a) = a + f(a)/f'(a)$ strictly increasing, the bid a_θ^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 η, conditions SC 5-SC 11 hold.*

Proof. As in the lemma for second-price auctions, condition SC 5 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

$$f'(\beta(\theta))\theta - f(\beta(\theta)) - \beta(\theta)f'(\beta(\theta)) = 0. \tag{D-14}$$

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 SC 6 holds.

Note that condition SC 9 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. □

D.3 Network Revenue Management

Here, in contrast to the two problems exposed before, the set of actions is finite (binary).

In the case of finite customer classes (contexts), Problem (6) can be written as follows

$$\begin{aligned} \bar{J}(\rho) = \max_{\phi_\theta(1)\in[0,1]} \quad & \sum_{\theta\in\Theta} p_\theta \; r_\theta \; \phi_\theta(1) \\ \text{s.t} \quad & \sum_{\theta\in\Theta} p_\theta \; Q_\theta \; \phi_\theta(1) \le \rho \,. \end{aligned} \tag{D-15}$$

Because the set of actions consists of $\mathcal{A} = \{0, 1\}$, it is enough to consider decision variables $\phi_\theta(1)$ for all $\theta \in \Theta$ because $\phi_\theta(0) = 1-\phi_\theta(1)$. Removing the variable $\phi_\theta(0)$ from the fluid problem requires a slight change in the statement of Lemma 1. In the standard form representation of the problem, the constraint matrix is $Q = (\tilde{Q}, I) \in \mathbb{R}^{(L+|\Theta|)\times(L+2|\Theta|)}$ where $\tilde{Q} \in \mathbb{R}^{(L+|\Theta|)\times|\Theta|}$ is the constraint matrix associated to the decision variables $(\phi_\theta(1))_{\theta\in\Theta}$ and the identity matrix $I \in \mathbb{R}^{(L+|\Theta|)\times(L+|\Theta|)}$ is associated to the slack variables of the resource constraints and the constraints $\phi_\theta(1) \le 1$. The θ-th column of $\tilde{Q}$ consists of the vector $\binom{p_\theta Q_\theta}{e_\theta}$, where $e_\theta \in \mathbb{R}^{|\Theta|}$ is the canonical vector. Let B be the submatrix of Q corresponding to the columns associated to the basic variables at an optimal solution, and $B_{\rho_1}^{-1}$ the submatrix of B^{-1} associated to the resource constraints. Then, under Assumption SC 1, Assumption 2 holds with $K = 0$ and $\delta = \min\{\phi^*_{\min}, x^*_{\min}\}/\|B_{\rho_1}^{-1}\|$, where $x^*_{\min} = \min_{l\in[L]}\{x^*_l : x^*_l > 0\}$ and $x^*_l = \rho_l - \sum_{\theta\in\Theta} p_\theta Q_{\theta l}\phi^*_\theta(1)$ is the slack of the l-th resource constraint. In the definition of $\phi^*_{\min}$ we now take into account how close the controls are to both zero and 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.

D.4 Choice-Based Network Revenue Management

In this case, the set of actions is finite and the fluid problem can be expressed as

$$\begin{aligned} \bar{J}(\rho) = \max_{\phi\in\Phi} \quad & \sum_{\theta\in\Theta}\sum_{a\in\mathcal{A}} \phi_\theta(a) p_\theta \sum_{n\in[N]} f_n g^n_{\theta a} \\ \text{s.t} \quad & \sum_{\theta\in\Theta}\sum_{a\in\mathcal{A}} \phi_\theta(a) p_\theta \sum_{n\in[N]} Q^n_\theta g^n_{\theta a} \le \rho. \end{aligned} \tag{D-16}$$

We can define the vector of variables and the associated matrix involved in the constraints of problem (D-16) by setting each column to $p_\theta \sum_{n\in[N]} Q^n_\theta g^n_{\theta a}$. Then, under Assumption SC 1, we can apply Lemma 1 and Assumption 2 holds for $K = 0$ and δ 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.

D.5 Online Matching

First, note that the fluid problem is given by

$$\begin{aligned} \bar{J}(\rho) = \max_{\phi \geq 0} \quad & \sum_{\theta \in \Theta} \sum_{a \in [L]} \phi_\theta(a) p_\theta f_{\theta a} \\ \text{s.t} \quad & \sum_{\theta \in \Theta} p_\theta \operatorname{diag}(Q_\theta) \phi_\theta \leq \rho \\ & \sum_{a \in [L]} \phi_\theta(a) \leq 1 \quad \forall \theta \in \Theta \,, \end{aligned} \tag{D-17}$$

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

Here, as in the network revenue management problem, we need a slight change in the statement of Lemma 1 because we removed the decision variable associated to the action $a = 0$. In the standard form representation of the problem, the constraint matrix is $Q = \binom{Q_1}{Q_2} \in \mathbb{R}^{(L+|\Theta|) \times (L|\Theta| + L + |\Theta|)}$ where $Q_1 \in \mathbb{R}^{L \times (L|\Theta| + L + |\Theta|)}$ is the matrix associated to the resource constraints and $Q_2 \in \mathbb{R}^{|\Theta| \times (L|\Theta| + L + |\Theta|)}$ is the matrix associated to the constraints $\sum_{a \in [L]} \phi_\theta(a) \leq 1$. The matrix Q_1 is obtained by horizontally stacking the matrices $\operatorname{diag}(p_\theta Q_\theta) \in \mathbb{R}^{L \times L}$, the identity matrix $I \in \mathbb{R}^{L \times L}$, and the zero matrix $\mathbf{0} \in \mathbb{R}^{L \times |\Theta|}$. The matrix $Q_2 \in \mathbb{R}^{|\Theta| \times (L|\Theta| + L + |\Theta|)}$ is obtained by horizontally stacking the matrices $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{|\Theta|}, \mathbf{0} \in \mathbb{R}^{|\Theta| \times L}$ and the identity matrix $I \in \mathbb{R}^{|\Theta|}$, where $\mathbf{e}_\theta \in \mathbb{R}^{|\Theta| \times L}$ is the matrix with all columns equal to the θ-th canonical vector of $\mathbb{R}^{|\Theta|}$. Let B be the submatrix of Q corresponding to the columns associated to the basic variables at an optimal solution, and $B_{\rho_1}^{-1}$ the submatrix of B^{-1} associated to the resource constraints. Then, if the problem (D-17) has a non-degenerate optimal solution for $\rho = \rho_1$, Assumption 2 holds with $K = 0$ and $\delta = \min\{\phi^*_{\min}, x^*_{\min}\} / \|B_{\rho_1}^{-1}\|$, where $\phi^*_{\min} = \min_{\theta \in \Theta, a \in [L]} \{\phi^*_\theta(a) : \phi^*_\theta(a) > 0\} \wedge \min_{\theta \in \Theta} \left\{1 - \sum_{a \in [L]} \phi^*_\theta(a) : \sum_{a \in [L]} \phi^*_\theta(a) < 1\right\}$, $x^*_{\min} = \min_{l \in [L]} \{x^*_l : x^*_l > 0\}$ and $x^*_l = \rho_l - \sum_{\theta \in \Theta} p_\theta Q_{\theta l} \phi^*_\theta(l)$ is the slack of the l-th resource constraint, and the constant bound for the revenue loss is obtained.

D.6 Order Fulfillment

In this case, the fluid problem can be expressed as follows

$$\begin{aligned} \bar{J}(\rho) = \max_{\phi \in \Phi} \quad & \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} p_\theta \phi_\theta(a) \sum_{l \in \theta} \sum_{n \in N} f_{ln} 1_{\{a_l = n\}} \\ \text{s.t} \quad & \sum_{\theta \in \Theta} \sum_{a \in \mathcal{A}} p_\theta \phi_\theta(a) 1_{\{a_l = n\}} \leq \rho_{ln}. \end{aligned} \tag{D-18}$$

As in the previous problems, we can write the constraints of problem (D-18) in matrix form and,

thus, obtain an expression for δ involved in the assumption. We do not give an explicit formula for δ here to avoid introducing more notation. Moreover, if problem (D-18) has a non-degenerate optimal solution for $\rho = \rho_1$, Assumption 2 holds with $K = 0$ and we recover a constant revenue loss bound for the order fulfillment problem.

E Auxiliary Results

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

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

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

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 exists 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

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

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

$$\int_0^T a\exp(-b(T-t)) \wedge 1 \,\mathrm{d}t \leq \frac{1}{b}(\log a + 1). \qquad \square$$

Lemma E-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*

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

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

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

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

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 $\tilde{x} = \arg\max_{\|x\|\le 1}\{z^\top x\}$, 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 $\|x\| = \min(\varphi, \|z\|_*)\|\tilde{x}\| \le \varphi$ since $\|\tilde{x}\| \le 1$. Therefore,

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

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

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

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

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* ϵ *a real parameter. Define* $x^*_{\min} = \min\{x^*_i : x^*_i > 0\}$, *where* x^* *denotes a non-degenerate optimal solution of problem* (E-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* (E-1) *for all* $0 \le \epsilon \le \delta$.

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 (E-1), that is, $x = (x_B, \mathbf{0})$ with $x_B = A_B^{-1}(\xi + \epsilon u)$ is an optimal solution for Problem (E-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 $j \in \{1, \ldots, m\}$ 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

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

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 δ. The proof is completed. □