Online Resource Allocation: Logarithmic Corrections and Constant Optimality Gaps

Chen Chen[1], Yilun Chen[2], Wenjia Wang[3], and Jiawei Zhang[4]

[1]New York University Shanghai
[2]School of Data Science, the Chinese University of Hong Kong, Shenzhen (CUHK Shenzhen)
[3]Department of Industrial Systems Engineering and Management, National University of Singapore
[4]Stern School of Business, New York University
cc8029@nyu.edu, chenyilun@cuhk.edu.cn, wenjiawang@nus.edu.sg, jz31@stern.nyu.edu

October 30, 2026

Abstract

We study online resource allocation with random rewards and bounded multidimensional resource consumptions under known arrival distributions. We show that, under standard non-degeneracy conditions, the value functions of both the optimal online policy and the dynamic fluid policy—which re-solves a fluid relaxation at every state and uses the resulting optimal dual variable to approximate the continuation value—admit the same performance expansion: the fluid benchmark minus a common logarithmic correction term, up to a uniformly bounded, policy-dependent remainder. Consequently, the dynamic fluid policy achieves a horizon-uniform constant optimality gap relative to the optimal online policy. We further show that, under an additional dual stability condition, a first-order policy that avoids repeated fluid optimization achieves the same performance guarantee.

Subject classifications: Online resource allocation, logarithmic correction, constant optimality gap, dynamic fluid policy, first-order policy.

1 Introduction

In this paper, we study a general online resource allocation problem in which a decision maker manages limited inventories of multiple resources over a finite horizon of n periods. In each period, an opportunity arrives according to a known distribution, and the decision maker selects an action from a finite set. Each action consumes a combination of resources and generates a reward. The objective is to select a sequence of actions that maximizes the expected total reward subject to the resource constraints. This problem can be formulated as a finite-horizon discrete-time dynamic program, with the state being the vector of remaining resources available.

Computing the optimal policy can be intractable due to the curse of dimensionality. Thus, much of the literature focuses on developing simple and computationally tractable policies with strong performance guarantees. A common and central heuristic is based on the deterministic fluid relaxation and the associated dual prices. The deterministic fluid relaxation simplifies the stochastic optimization problem by requiring that the capacity constraints hold only in expectation, and its optimal dual variables assign a price to each resource. In this paper, we study a classic policy, which we call the *dynamic fluid policy*. This policy re-solves the fluid relaxation and its optimal dual prices at each decision epoch, and selects the feasible action with the largest reward net of resource costs. This policy has been widely studied in the literature, including Lueker (1998), Bray (2025), and Chen and Wang (2025), and many other policies can be viewed as variants of it. Our question is how much reward the dynamic fluid policy loses relative to the optimal online policy characterized by the Bellman equation.

Instead of directly comparing a candidate policy with the optimal online policy, the literature primarily evaluates its performance against an optimal offline benchmark. The optimal offline benchmark allows the decision maker to observe the entire arrival sequence before selecting a sequence of actions and serves as an upper bound on the performance of the optimal online policy. For online resource allocation problems, when rewards and resource consumptions have finite support, a constant loss relative to this benchmark can be achieved (e.g., Arlotto and Gurvich 2019, Bumpensanti and Wang 2020, Vera and Banerjee 2021, and Vera et al. 2021). When rewards or resource consumptions are continuously distributed, under certain non-degeneracy conditions, the dynamic fluid policy and its variants typically incur a $\Theta(\log n)$ loss relative to this benchmark. Moreover, this logarithmic loss is order-optimal, as no non-anticipative policy can achieve a loss of a smaller order (Bray 2025).

To our knowledge, the only exception is Wang and Wang (2022). Wang and Wang (2022) study

a different dynamic pricing problem and analyze a re-solving heuristic that is analogous to our dynamic fluid policy. Under their non-degeneracy conditions, the re-solving heuristic achieves a constant loss relative to the optimal online policy, whereas the fluid benchmark can exceed the optimal online value by $\Omega(\log n)$.

Like Wang and Wang (2022), we measure the performance loss of the dynamic fluid policy relative to the optimal online policy, for the resource allocation problem. However, our performance characterization is more refined. Under standard non-degeneracy conditions, we show that the value functions of both the dynamic fluid policy and the optimal online policy share the same expansion: the fluid benchmark minus a common logarithmic correction term $\Gamma \log n$, up to a uniformly bounded remainder (Theorem 4.1). Consequently, the dynamic fluid policy incurs a constant optimality loss relative to the optimal online policy, in contrast to the order-optimal $\Theta(\log n)$ loss relative to the offline optimal benchmark established by Bray (2025). The logarithmic correction $\Gamma \log n$ captures the stochastic loss shared by all non-anticipative policies, including the optimal online policy, relative to the offline optimum. We explicitly characterize the coefficient Γ and show that it arises from the interaction between random resource consumption and the concavity of the fluid value function; see Section 4.1 for further intuition.

In Section 5, we further establish that, under an additional local dual stability condition, a first-order policy that avoid re-solving achieves the same performance guarantee as the dynamic fluid policy. This condition requires the gradient of the fluid relaxation's dual objective to be Lipschitz continuous in a neighborhood of the optimal dual variable. The first-order policy initializes the dual prices at zero and updates them via gradient descent in each period. Thus, under the local dual stability condition, re-solving is unnecessary to achieve a constant optimality gap relative to the optimal online policy.

The rest of the paper is organized as follows. Section 1.1 reviews the related literature. Section 2 formulates the problem. Section 3 introduces our non-degeneracy conditions. Section 4 analyzes the dynamic fluid policy and establishes its common performance expansion with the optimal online policy. Section 5 analyzes a first-order policy that avoids re-solving and establishes the same performance guarantee under an additional local dual stability condition. Finally, Section 6 concludes the paper.

1.1 Related Literature

Online resource allocation has been extensively studied in operations research and management science, with applications including network revenue management, online matching, and order

fulfillment, among others (Balseiro et al. 2023). A widely used approach to both policy design and performance analysis is based on deterministic fluid relaxations and the associated dual prices. Talluri and van Ryzin (1998) construct static bid-price controls based on deterministic fluid relaxation and establish an $O(\sqrt{T})$ loss relative to the fluid relaxation when demand and capacity grow proportionally. Much of the subsequent literature improves upon this baseline by periodically re-solving the fluid relaxation as time and remaining capacity evolve, thereby allowing the dual prices to adapt to the state. As discussed earlier, when rewards and resource consumptions have finite support, such re-solving policies can achieve a constant loss relative to the offline optimum. More recent work investigates beyond the finite-support setting.

Fluid-Based Policies under Non-Degeneracy Conditions When rewards or consumptions are continuously distributed, a constant loss relative to the fluid or offline optimal benchmark is no longer attainable. Nevertheless, logarithmic regret relative to these benchmarks can be achieved under suitable non-degeneracy or stability conditions on the deterministic fluid relaxation. These conditions take different forms, including local smoothness, strict complementarity, stability of the binding resource constraints, and uniqueness of the optimal dual solution. Li and Ye (2021) establish $O(\log n \log\log n)$ regret relative to the offline optimal benchmark under their non-degeneracy conditions. Under different sets of non-degeneracy assumptions, Balseiro et al. (2023) obtain an $O(\log n)$ loss relative to the fluid relaxation, improving to $O(1)$ in locally affine cases, and Bray (2025) establish a tight $\Theta(\log n)$ regret relative to the offline optimal benchmark. Finally, Chen and Wang (2025) impose regularity conditions on the request distributions rather than on the deterministic fluid relaxation. Their regularity conditions imply uniqueness of the optimal dual solution but allow primal degeneracy.

Our model is more general than the known-distribution setting in Bray (2025). Moreover, their non-degeneracy assumptions imply ours, except that we additionally require local Hölder continuity of the second-order derivatives of the fluid value function. We show that the dynamic fluid policy and the optimal online policy admit the same fluid-minus-logarithmic expansion up to an $O(1)$ remainder. Consequently, the dynamic fluid policy incurs only a constant loss relative to the optimal online policy. See Section 4.3.3 for further discussion.

Comparison with Wang and Wang (2022) Our work is also closely related to Wang and Wang (2022), who study a price-based network revenue management problem and analyze a re-solving heuristic that is analogous to our dynamic fluid policy. Under their non-degeneracy conditions, they

show that their re-solving heuristic achieves an $O(1)$ gap relative to the optimal online policy, whereas the fluid benchmark can exceed the optimal online value by $\Omega(\log n)$. We consider a distinct resource allocation problem and obtain a more refined performance characterization: the dynamic fluid policy and the optimal online policy admit the same performance expansion, $n\phi^{\mathrm{F}} - \Gamma \log n + O(1)$, from which an $O(1)$ optimality gap follows directly. We further establish the same performance guarantee for a first-order policy under an additional dual stability condition.

Finally, the analytical approaches are also substantially different. Wang and Wang (2022) directly compare the stochastic inventory processes under the re-solving and optimal online policies, using a coupling of their demand noises together with stopping-time arguments. In contrast, we analyze each policy separately by telescoping its one-period losses relative to the fluid benchmark along its own trajectory. This approach yields a general performance characterization theorem that bounds the gap between the value function of a policy π and the corrected fluid reference $n\phi^{\mathrm{F}} - \Gamma \log n$ in terms of its cumulative dual-price score loss up to a prescribed local exit (see Section B.3 for details). We then show that the cumulative score loss is uniformly bounded for both the dynamic fluid policy and the first-order policy, yielding the desired performance guarantees.

Beyond Non-Degeneracy Conditions Recently, several papers relax non-degeneracy assumptions in online resource allocation. Besbes et al. (2024) allow gaps in the support and low probability mass around these gaps. Jiang et al. (2025) study the network revenue management problem with discrete resource consumption and continuously distributed rewards without imposing non-degeneracy. Zhang (2026*a*) consider settings in which both rewards and resource consumption are continuously distributed and allow both primal degeneracy and dual non-uniqueness. Zhang (2026*b*) develop a Bellman certificate approach to derive lower bounds on the gap between the optimal offline and online values in the multi-secretary setting without non-degeneracy assumptions. These papers typically use the offline optimum as the benchmark. Extending our analysis and performance guarantees to more general settings without non-degeneracy requirements is a promising direction for future research.

1.2 Notation and Terminology

For any integer $m \in \mathbb{N}_+$, let $[m] \triangleq \{1, 2, \ldots, m-1, m\}$ denote the set of integers from 1 to m. For any real number $x \in \mathbb{R}$, let $(x)^+ \triangleq \max\{x, 0\}$. Given a vector $\mathbf{x} = (x_i)_{i\in[n]} \in \mathbb{R}^n$, we write $\mathbf{x} \geq \mathbf{0}$ if $x_i \geq 0$ for all $i \in [n]$. We define the harmonic numbers by $\mathsf{H}_k \triangleq \sum_{j=1}^{k} j^{-1}$, with $\mathsf{H}_0 \triangleq 0$. We use *a.s.* to denote almost surely. For two sets A and B, we write $A \Subset B$ if A is contained in B and

has positive distance from the boundary of B; equivalently, $A \subseteq B$ and $\operatorname{dist}(A, B^c) > 0$. Unless otherwise specified, $\|\cdot\|$ denotes the Euclidean norm for vectors and the induced operator norm (i.e., the spectral norm) for matrices.

2 Problem Formulation

In this section, we describe our online resource allocation problem in Section 2.1 and introduce its fluid relaxation in Section 2.2, which serves as the backbone of the policies analyzed in this paper.

2.1 The Online Resource Allocation Problem

We consider an online resource allocation problem over a finite horizon of $n \in \mathbb{N}_+$ periods, during which a decision maker collects rewards subject to resource constraints. Specifically, there are $m \in \mathbb{N}_+$ resources. The decision maker is initially endowed with an inventory of $c_i > 0$ units of resource i for each $i \in [m]$. We let $\mathbf{c} = (c_i)_{i\in[m]} \in \mathbb{R}_+^m$ denote the vector of initial inventories.

There are $L+1$ actions. We let $\mathcal{A} = \{0, 1, \ldots, L\}$ denote the action set, where action 0 represents rejection. In each period, an opportunity arrives and is characterized by a vector of action-specific rewards and resource consumptions $\boldsymbol{\theta} = (R_a, \mathbf{B}_a)_{a\in\mathcal{A}}$. Specifically, taking an action a yields a reward $R_a \in \mathbb{R}$ and consumes resources $\mathbf{B}_a = (B_{aj})_{j\in[m]} \in \mathbb{R}_+^m$, where B_{aj} denotes the amount of resource j consumed by action a. We assume that $R_0 = 0$ and $\mathbf{B}_0 = \mathbf{0}$, so rejecting an opportunity yields no reward and consumes no resources. The opportunities are drawn independently over time from a distribution Φ. Finally, we assume that the positive part of rewards is integrable and that the resource consumptions are uniformly bounded, as stated in Assumption 2.1.

Assumption 2.1. There exist constants $\bar{r} > 0$ and $\bar{b} > 0$ such that

$$\mathbb{E}\left[\max_{a\in\mathcal{A}} (R_a)^+\right] \le \bar{r} \quad \text{and} \quad \max_{a\in\mathcal{A}} \|\mathbf{B}_a\|_\infty \le \bar{b} \quad a.s.$$

In each period, the decision maker first observes the arriving opportunity and then chooses an action. An action a is feasible only if its resource consumption does not exceed the remaining capacity. Therefore, when the remaining capacity is $\mathbf{c} \in \mathbb{R}_+^m$ and the arriving opportunity is $\boldsymbol{\theta}$, the set of feasible actions is

$$\mathcal{A}(\mathbf{c}, \boldsymbol{\theta}) \triangleq \{a \in \mathcal{A} : \mathbf{B}_a \le \mathbf{c}\}.$$

Let Π denote the set of all non-anticipating policies, under which the action in period t depends only on the opportunity arriving in period t and the history up to period $t-1$, which consists

of all previously observed opportunities and actions. Let a_t^π denote the action taken in period t under a policy $\pi \in \Pi$. The decision maker's objective is to choose a policy $\pi \in \Pi$ that maximizes the expected total reward over the planning horizon, subject to the capacity constraints that the cumulative resource consumption does not exceed the initial inventory. The resulting stochastic optimization problem can be written as follows:

$$\begin{aligned} V_n^{\mathrm{OPT}}(\mathbf{c}) = \sup_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=1}^{n} R_{a_t^\pi}\right] \\ \text{s.t.} \quad & \sum_{t=1}^{n} \mathbf{B}_{a_t^\pi} \leq \mathbf{c}, \quad a.s. \end{aligned}$$

where $V_n^{\mathrm{OPT}}(\mathbf{c})$ denotes the optimal expected reward with n periods remaining and remaining capacity $\mathbf{c} \in \mathbb{R}_+^m$. The optimal value function satisfies the Bellman equation

$$V_n^{\mathrm{OPT}}(\mathbf{c}) = \mathbb{E}_{\boldsymbol{\theta} \sim \Phi}\left[\max_{a \in \mathcal{A}(\mathbf{c}, \boldsymbol{\theta})} \left\{ R_a + V_{n-1}^{\mathrm{OPT}}(\mathbf{c} - \mathbf{B}_a) \right\}\right]$$

with boundary conditions $V_0^{\mathrm{OPT}}(\mathbf{c}) = 0$ for any $\mathbf{c} \in \mathbb{R}_+^m$. Finally, we remark that both the classic network revenue management and stochastic matching problems can be formulated as instances of the the online resource allocation problem we study, thus illustrating the generality of our formulation (see Sections 3.3 and 3.5 of Balseiro et al. 2023).

2.2 Fluid Relaxation and Dual Price

In this section, we introduce the fluid relaxation of the online resource allocation problem. The fluid relaxation allows resource inventories to go negative and replaces the sample-path capacity constraints with expected capacity constraints; that is, it requires only that the expected total consumption of each resource not exceed its initial inventory.

Let $\boldsymbol{\rho} = \mathbf{c}/n \in \mathbb{R}_+^m$ denote the vector of normalized inventories, where each component represents the available inventory per remaining period. Let Θ denote the set of all possible opportunities $\boldsymbol{\theta}$, and $\mathcal{P}$ denote the set of all opportunity-dependent randomized controls, $p : \Theta \to \Delta(\mathcal{A})$. The fluid relaxation is defined as

$$\begin{aligned} \phi^{\mathrm{F}}(\boldsymbol{\rho}) = \sup_{p \in \mathcal{P}} \quad & \sum_{a \in \mathcal{A}} \mathbb{E}_{\boldsymbol{\theta} \sim \Phi}\Big[R_a \, p(\boldsymbol{\theta}, a)\Big] \\ \text{s.t.} \quad & \sum_{a \in \mathcal{A}} \mathbb{E}_{\boldsymbol{\theta} \sim \Phi}\Big[\mathbf{B}_a \, p(\boldsymbol{\theta}, a)\Big] \leq \boldsymbol{\rho}\,. \end{aligned} \tag{1}$$

The optimal value of (1), denoted by $\phi^{\mathrm{F}}(\boldsymbol{\rho})$, represents the maximum expected reward per period

achievable with normalized inventory $\boldsymbol{\rho}$. For a planning horizon of n periods with initial inventory $\mathbf{c} = n\boldsymbol{\rho}$, the following upper bound holds.

Lemma 2.1. *Let $\boldsymbol{\rho} = \mathbf{c}/n$ denote the vector of normalized inventories. Then*

$$V_n^{\mathrm{OPT}}(\mathbf{c}) \le V_n^{\mathrm{OFF}}(\mathbf{c}) \le n \cdot \phi^{\mathrm{F}}(\boldsymbol{\rho}),$$

where $V_n^{\mathrm{OFF}}(\mathbf{c})$ denotes the expected optimal reward when all n opportunities are observed before any action is selected, subject to the sample-path capacity constraints.

We provide a formal definition of the offline optimum and prove Lemma 2.1 in Appendix A.1.

2.2.1 The Fluid Relaxation Dual Problem

We now derive the dual of the fluid relaxation (1) by dualizing the capacity constraints with a dual variable $\boldsymbol{\lambda} \in \mathbb{R}^m_+$. The dual problem of (1) is

$$\phi^{\mathrm{F}}(\boldsymbol{\rho}) = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^m_+} \left\{ \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{\rho} + \mathcal{V}(\boldsymbol{\lambda}) \right\}, \tag{2}$$

where

$$\mathcal{V}(\boldsymbol{\lambda}) \triangleq \mathbb{E}\left[\max_{a \in \mathcal{A}} \left\{ R_a - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_a \right\} \right]$$

denotes the expected maximum net reward earned in a single period.

Since (1) is a linear program, strong duality results for linear optimization in vector spaces (see Theorem 1 in Section 8.6 of Luenberger 1997) imply that strong duality holds and that an optimal dual solution exists. Therefore, the optimal value of (2) is equal to $\phi^{\mathrm{F}}(\boldsymbol{\rho})$, and we let $\boldsymbol{\lambda}^*(\boldsymbol{\rho})$ denote an optimal solution to (2).

3 The Non-Degeneracy Conditions

In this section, we state the non-degeneracy conditions underlying our main results in Section 4 as Assumptions 3.1–3.3. These assumptions are mild and, as we demonstrate in Section 4.3, are satisfied by many notable online resource allocation problems.

In the remainder of the paper, given a vector $\boldsymbol{\rho} = (\rho_i)_{i \in [m]} \in \mathbb{R}^m_+$ and a subset $\mathsf{B} \subseteq [m]$, we let $\boldsymbol{\rho}_{\mathsf{B}} \triangleq (\rho_i)_{i \in \mathsf{B}}$ denote the restriction of $\boldsymbol{\rho}$ to the coordinates indexed by B. Similarly, for a set $\mathcal{U} \subseteq \mathbb{R}^m$, we define $\mathcal{U}_{\mathsf{B}} \triangleq \{\boldsymbol{\rho}_{\mathsf{B}} : \boldsymbol{\rho} \in \mathcal{U}\}$ as the projection of $\mathcal{U}$ onto the coordinates indexed by B.

We first impose Assumption 3.1, which requires that, in a neighborhood of the point of interest, the fluid value function admits a lower-dimensional representation (where the reduced coordinates will correspond to the binding resource constraints) and that this reduced function is locally smooth.

Assumption 3.1 (Local Smoothness)**.** There exist an open set $\mathcal{U} \subseteq (0,\infty)^m$, a nonempty subset $\mathsf{B} \subseteq [m]$, and a function $\phi_{\mathsf{B}} : \mathbb{R}^{|\mathsf{B}|} \to \mathbb{R}$ such that:

1. For every $\boldsymbol{\rho} \in \mathcal{U}$, $\phi^{\mathrm{F}}(\boldsymbol{\rho}) = \phi_{\mathsf{B}}(\boldsymbol{\rho}_{\mathsf{B}})$.
2. The function ϕ_{B} belongs to $C^{2,\alpha}(\mathcal{U}_{\mathsf{B}})$ for some Hölder exponent $\alpha \in (0,1]$. That is, ϕ_{B} is twice continuously differentiable on $\mathcal{U}_{\mathsf{B}}$, and, for every compact set $\mathcal{K}_{\mathsf{B}} \Subset \mathcal{U}_{\mathsf{B}}$, there exists a finite constant L such that

$$\left\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}) - \nabla^2\phi_{\mathsf{B}}(\mathbf{y})\right\| \leq L\left\|\mathbf{x}-\mathbf{y}\right\|^{\alpha}, \quad \forall\ \mathbf{x},\mathbf{y} \in \mathcal{K}_{\mathsf{B}}.$$

Assumption 3.1 implies that, for every $\boldsymbol{\rho} \in \mathcal{U}$, the optimal dual variable $\boldsymbol{\lambda}^*(\boldsymbol{\rho})$ is unique and admits the following characterization in Lemma 3.1. We prove Lemma 3.1 in Section A.2.

Lemma 3.1. *For every $\boldsymbol{\rho} \in \mathcal{U}$, the optimal dual variable $\boldsymbol{\lambda}^*(\boldsymbol{\rho})$ is unique. Moreover,*

$$\boldsymbol{\lambda}^*(\boldsymbol{\rho}) = \left(\boldsymbol{\lambda}\left(\boldsymbol{\rho}_{\mathsf{B}}\right), \mathbf{0}_{\mathsf{S}}\right),$$

where $\boldsymbol{\lambda}\left(\boldsymbol{\rho}_{\mathsf{B}}\right) \triangleq \nabla\phi_{\mathsf{B}}(\boldsymbol{\rho}_{\mathsf{B}})$ and $\mathsf{S} = \mathsf{B}^c$.

We next impose Assumption 3.2, which requires that the set of binding constraints in the fluid relaxation (1) remain unchanged throughout the neighborhood $\mathcal{U}$. Specifically, the constraints indexed by B are binding, whereas those indexed by S are strictly slack. We also assume that the expected consumption of the non-binding resources is a continuous function of $\boldsymbol{\rho}_{\mathsf{B}}$ on $\mathcal{U}_{\mathsf{B}}$.

Fix any $\boldsymbol{\rho} \in \mathcal{U}$. Consider a static fluid policy for the fluid relaxation (1) that always selects the action

$$a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}}) \triangleq \operatorname*{arg\,max}_{a\in\mathcal{A}} \left\{R_a - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}}\mathbf{B}_{a,\mathsf{B}}\right\} \tag{3}$$

using the smallest-index tie-breaking rule. For each resource $j \in [m]$, define

$$\mu_j(\boldsymbol{\rho}_{\mathsf{B}}) \triangleq \mathbb{E}\left[B_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}}),j}\right]$$

as the expected per-period consumption of resource j under the static fluid policy. Assumption 3.2 is stated as follows.

Assumption 3.2 (Fixed Binding Set)**.** For every $\boldsymbol{\rho} \in \mathcal{U}$, the corresponding static fluid policy evaluated at $\boldsymbol{\rho}_{\mathsf{B}}$ satisfies:

1. $\mu_j(\boldsymbol{\rho}_{\mathsf{B}}) = \rho_j$ for all $j \in \mathsf{B}$,
2. $\mu_j(\boldsymbol{\rho}_{\mathsf{B}}) < \rho_j$ for all $j \in \mathsf{S}$.

Moreover, for every $j \in \mathsf{S}$, the consumption function $\mu_j(\boldsymbol{\rho}_{\mathsf{B}})$ is continuous on $\mathcal{U}_{\mathsf{B}}$.

Assumption 3.2 implies that the expected per-period reward under the static fluid policy equals the fluid relaxation benchmark, as stated in Lemma 3.2. We prove Lemma 3.2 in Appendix A.3.

Lemma 3.2. *For every $\boldsymbol{\rho} \in \mathcal{U}$, we have $\mathbb{E}\Big[R_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})}\Big] = \phi^{\mathrm{F}}(\boldsymbol{\rho})$.*

Finally, we impose Assumption 3.3, which controls how often the action a^{F} encounters competing actions that are nearly indistinguishable in terms of their dual-price scores but differ substantially in their consumption of binding resources.

Assumption 3.3 (Consumption-Weighted Margin)**.** Define the dual-price score of an action a by

$$S_a(\boldsymbol{\rho}_{\mathsf{B}}) \triangleq R_a - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \mathbf{B}_{a,\mathsf{B}},$$

and the binding-resource consumption difference between two actions a and b by

$$d^{\mathsf{B}}_{ab} \triangleq \left\| \mathbf{B}_{a,\mathsf{B}} - \mathbf{B}_{b,\mathsf{B}} \right\|_1.$$

For every compact set $\mathcal{K}_{\mathsf{B}} \Subset \mathcal{U}_{\mathsf{B}}$, there exist constants $\beta_{\mathcal{K}} > 0$, $t_{\mathcal{K}} > 0$, and $C_{\mathcal{K}} < \infty$ such that

$$\sup_{\mathbf{x} \in \mathcal{K}_{\mathsf{B}}} \mathbb{E}\left[\max_{b \in \mathcal{A}} \left\{ d^{\mathsf{B}}_{a^{\mathrm{F}}(\mathbf{x}),b} \, \mathbb{1}\left(S_{a^{\mathrm{F}}(\mathbf{x})}(\mathbf{x}) - S_b(\mathbf{x}) \le t \, d^{\mathsf{B}}_{a^{\mathrm{F}}(\mathbf{x}),b} \right) \right\} \right] \le C_{\mathcal{K}} \, t^{\beta_{\mathcal{K}}}, \qquad 0 \le t \le t_{\mathcal{K}},$$

where $a^{\mathrm{F}}(\mathbf{x})$ denotes the action with the highest dual-price score, as defined in (3).

Because action a^{F} attains the highest dual-price score by definition, the difference $S_{a^{\mathrm{F}}}(\mathbf{x}) - S_b(\mathbf{x})$ measures how much more attractive action a^{F} is than a competing action b. Assumption 3.3 requires that, as the tolerance t decreases, the consumption-weighted frequency of such near-indifference events decays at a positive polynomial rate. The weight $d^{\mathsf{B}}_{a^{\mathrm{F}}(\mathbf{x}),b}$ measures the difference in binding-resource consumption between the two actions. Near-indifference is less consequential when the two actions consume similar bundles of binding resources, because switching between them has little effect on the future inventories of those resources.

4 Constant Optimality Gap for the Dynamic Fluid Policy

In this section, we show that the optimal online policy and the classic dynamic fluid policy share a common performance expansion: the fluid benchmark minus the logarithmic correction, up to a uniformly bounded, policy-dependent remainder. Consequently, the optimality gap between the two policies is uniformly bounded by a constant. In what follows, we first provide intuition for the logarithmic correction in Section 4.1, then define the dynamic fluid policy and analyze its performance in Section 4.2. Finally, in Section 4.3, we discuss the implications of our results for several specific online resource allocation problems.

For any normalized binding state $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$, let $a^{\mathrm{F}}(\mathbf{x})$ denote the static fluid action in (3) evaluated at $\boldsymbol{\rho}_{\mathsf{B}} = \mathbf{x}$. Let $\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) \triangleq \mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}$ denote its binding-resource consumption, and define its covariance matrix by $\Sigma(\mathbf{x}) \triangleq \mathbb{C}\mathrm{ov}(\mathbf{Y}_{\mathsf{B}}(\mathbf{x}))$, where the expectation is taken with respect to $\boldsymbol{\theta} \sim \Phi$. We define the logarithmic correction coefficient as

$$\begin{aligned} \Gamma(\mathbf{x}) &\triangleq -\frac{1}{2}\,\mathbb{E}\Big[\big(\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) - \mathbf{x}\big)^{\mathrm{T}}\,\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,\big(\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) - \mathbf{x}\big)\Big] \\ &= -\frac{1}{2}\,\mathrm{tr}\big(\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,\Sigma(\mathbf{x})\big)\,. \end{aligned} \tag{4}$$

Hence, $\Gamma(\mathbf{x})$ captures the interaction between the curvature of the fluid value function along the binding-resource directions and the randomness in binding-resource consumption under the static fluid policy. Since ϕ_{B} is concave on $\mathcal{U}_{\mathsf{B}}$ by the concavity of ϕ^{F}, its Hessian satisfies $\nabla^2\phi_{\mathsf{B}}(\mathbf{x}) \preceq 0$ for all $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$. Therefore, by (4), $\Gamma(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$.

4.1 Intuition for the Logarithmic Correction

Define the fluid relaxation upper bound by $F_n(\mathbf{c}) \triangleq n\phi^{\mathrm{F}}(\mathbf{c}/n)$ for $n \geq 1$ and let $F_0(\mathbf{c}) = 0$. We first provide intuition for why the logarithmic correction takes the form $\Gamma \log n$, which yields the following corrected fluid reference:

$$F_n(\mathbf{c}) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\mathsf{H}_{n-1} + O(1).$$

Suppose that $n \geq 2$ periods remain and that the normalized inventory of the binding resources is $\mathbf{x} \in \mathcal{K}_{\mathsf{B}}$. Let $a^{\mathrm{F}}(\mathbf{x})$ denote the static fluid action defined in (3), and recall that $\mathbf{Y}_{\mathsf{B}}(\mathbf{x})$ denotes its consumption of binding resources. After taking action $a^{\mathrm{F}}(\mathbf{x})$, the normalized binding-resource

inventory in the next period is

$$\mathbf{x}' \triangleq \frac{n\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1} = \mathbf{x} + \frac{\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1}.$$

Define the one-period fluid residual as

$$\delta_n^{\mathrm{F}}(\mathbf{x}) \triangleq n\phi_{\mathsf{B}}(\mathbf{x}) - \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})} + (n-1)\phi_{\mathsf{B}}(\mathbf{x}')\Big],$$

which compares the fluid benchmark with the expected one-period reward and fluid continuation value under the static action. By Assumption 3.2, $\mathbb{E}[\mathbf{Y}_{\mathsf{B}}(\mathbf{x})] = \mathbf{x}$, and hence $\mathbb{E}[\mathbf{x}'] = \mathbf{x}$. Moreover, Lemma 3.2 gives $\mathbb{E}\big[R_{a^{\mathrm{F}}(\mathbf{x})}\big] = \phi_{\mathsf{B}}(\mathbf{x})$. Therefore,

$$\delta_n^{\mathrm{F}}(\mathbf{x}) = (n-1)\left(\phi_{\mathsf{B}}(\mathbf{x}) - \mathbb{E}[\phi_{\mathsf{B}}(\mathbf{x}')]\right) \geq 0,$$

where the inequality follows from the concavity of ϕ_{B} and Jensen's inequality. Thus, the residual captures the effect of random consumption on the curved continuation fluid benchmark: although the static action earns the fluid reward in expectation, randomness in the successor state reduces the expected fluid continuation value relative to its value at the mean successor state.

A second-order Taylor expansion of $\phi_{\mathsf{B}}(\mathbf{x}')$ around $\mathbf{x}$ makes the magnitude of this effect explicit:

$$\begin{aligned}\delta_n^{\mathrm{F}}(\mathbf{x}) &= -\nabla\phi_{\mathsf{B}}(\mathbf{x})^{\mathrm{T}}\,\mathbb{E}[\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})] - \frac{1}{2(n-1)}\mathbb{E}\big[(\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x}))^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x})(\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x}))\big] + O(n^{-1-\alpha}) \\ &= \frac{\Gamma(\mathbf{x})}{n-1} + O(n^{-1-\alpha}). \end{aligned} \tag{5}$$

Here, the first equality follows from the local α-Hölder continuity of $\nabla^2\phi_{\mathsf{B}}$ in part 2 of Assumption 3.1, assuming that n is sufficiently large so that the line segment between $\mathbf{x}$ and $\mathbf{x}'$ lies in a compact subset of $\mathcal{U}_{\mathsf{B}}$. The second equality follows from $\mathbb{E}[\mathbf{Y}_{\mathsf{B}}(\mathbf{x})] = \mathbf{x}$ and the definition of $\Gamma(\mathbf{x})$. In Lemma B.5, we show that the expansion holds uniformly on a compact set containing $\mathcal{K}_{\mathsf{B}}$.

Choose n_0 sufficiently large so that (5) holds for every $n \geq n_0 + 1$. Keeping $\mathbf{x}$ fixed and summing the one-period residuals then yields the logarithmic correction:

$$\sum_{j=n_0+1}^{n} \delta_j^{\mathrm{F}}(\mathbf{x}) = \Gamma(\mathbf{x})\,\mathsf{H}_{n-1} + O(1).$$

This calculation, however, provides only a frozen-state heuristic. Along the actual trajectory of a policy, the normalized binding state $\mathbf{x}$ changes after each action, and hence so does $\Gamma(\mathbf{x})$. Therefore,

although the calculation identifies the source and order of the logarithmic correction, it does not by itself yield a valid Bellman telescope. Such a telescope must evaluate the state-dependent correction at the random successor state $\mathbf{x}'$, rather than repeatedly at the current state $\mathbf{x}$. Our subsequent analysis accounts for precisely this state dependence. For each policy analyzed below, we show that replacing the values of Γ along the trajectory with its initial value changes the accumulated residual by only a bounded amount.

4.2 Analysis of The Dynamic Fluid Policy

In this section, we describe the dynamic fluid policy and analyze its performance. Throughout Section 4.2, we index time forward: k denotes the number of decisions made since the beginning of the horizon, and $h_k \triangleq n - k$ denotes the number of periods remaining after the k-th decision. In addition, let $\mathbf{c}_k$ denote the remaining inventory after the k-th decision and $\boldsymbol{\rho}_k = \mathbf{c}_k/h_k$ denote the inventory per remaining period.

The dynamic fluid policy belongs to the family of *dual-price-based policies* described in Algorithm 1. In each period, a dual-price-based policy selects a feasible action that maximizes the reward net of resource costs evaluated at a dual-price vector $\boldsymbol{\lambda}_k \in \mathbb{R}^m_+$.

Algorithm 1: Dual-Price-Based Policy

Input: Initial inventory $\mathbf{c}$, horizon n, and a dual-price-selection rule.

Initialize $\mathbf{c}_0 = \mathbf{c}$.
for $k = 0, \ldots, n-1$ **do**
  Set $h_k = n - k$ and $\boldsymbol{\rho}_k = \mathbf{c}_k/h_k$.
  Compute the dual price $\boldsymbol{\lambda}_k \in \mathbb{R}^m_+$ using the specified rule.
  Observe the opportunity $\boldsymbol{\theta}_{k+1} = (R_{a,k+1}, \mathbf{B}_{a,k+1})_{a\in\mathcal{A}}$.
  Select $a_{k+1} = \arg\max_{a\in\mathcal{A}(\mathbf{c}_k,\boldsymbol{\theta}_{k+1})} \{R_{a,k+1} - \boldsymbol{\lambda}_k^{\mathrm{T}}\mathbf{B}_{a,k+1}\}$, breaking ties by the smallest index.
  Collect $\hat{R}_{k+1} = R_{a_{k+1},k+1}$ and set $\mathbf{Y}_{k+1} = \mathbf{B}_{a_{k+1},k+1}$.
  Update $\mathbf{c}_{k+1} = \mathbf{c}_k - \mathbf{Y}_{k+1}$.
end for

At each state $(h_k, \mathbf{c}_k)$, where h_k is the number of remaining periods and $\mathbf{c}_k$ is the remaining capacity, the dynamic fluid policy solves the fluid relaxation dual problem (2) using the normalized remaining capacity $\boldsymbol{\rho}_k = \mathbf{c}_k/h_k$. It then uses the resulting optimal dual variable $\boldsymbol{\lambda}^*(\boldsymbol{\rho}_k)$ as the dual price in Algorithm 1. Formally, at state $(h_k, \mathbf{c}_k)$, the dynamic fluid policy selects the action

$$a^{\mathrm{DF}}_{h_k}(\mathbf{c}_k) \triangleq \arg\max_{a\in\mathcal{A}(\mathbf{c}_k,\boldsymbol{\theta})} \left\{ R_a - \boldsymbol{\lambda}^*(\mathbf{c}_k/h_k)^{\mathrm{T}}\mathbf{B}_a \right\} \tag{6}$$

using the smallest-index tie-breaking rule. We let π^{DF} denote the dynamic fluid policy and $V_n^{\mathrm{DF}}(\mathbf{c})$ its expected total reward starting from state $(n, \mathbf{c})$.

When $\boldsymbol{\rho} = \mathbf{c}/n \in \mathcal{U}$, Lemma 3.1 shows that the optimal dual variables associated with the resource constraints indexed by S are zero. Consequently, letting $\mathbf{B}_{a,\mathsf{B}} \triangleq (\mathbf{B}_a)_{\mathsf{B}}$, the policy π^{DF} equivalently selects, at each state $(n, \mathbf{c})$ with $\boldsymbol{\rho} = \mathbf{c}/n \in \mathcal{U}$,

$$a_n^{\mathrm{DF}}(\mathbf{c}) = \arg\max_{a \in \mathcal{A}(\mathbf{c}, \boldsymbol{\theta})} \left\{ R_a - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \mathbf{B}_{a,\mathsf{B}} \right\}.$$

Moreover, when all actions are feasible at the state $(n, \mathbf{c})$, we have $a_n^{\mathrm{DF}}(\mathbf{c}) = a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$.

Theorem 4.1 establishes that the optimal online policy and the dynamic fluid policy share a common performance expansion: the fluid benchmark minus a logarithmic correction, up to a uniformly bounded remainder. Thus, the optimality gap between the two policies is uniformly bounded by a constant.

Theorem 4.1 (Common Expansion and Constant Optimality Gap). *Suppose Assumption 2.1 and Assumptions 3.1–3.3 hold. For every compact set $\mathcal{K} \Subset \mathcal{U}$, uniformly over states satisfying $\mathbf{c}/n \in \mathcal{K}$, we have*

$$V_n^{\mathrm{DF}}(\mathbf{c}) = n\,\phi^{\mathrm{F}}(\mathbf{c}/n) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\log n + O_{\mathcal{K}}(1)\,,$$
$$V_n^{\mathrm{OPT}}(\mathbf{c}) = n\,\phi^{\mathrm{F}}(\mathbf{c}/n) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\log n + O_{\mathcal{K}}(1)\,.$$

Here, $O_{\mathcal{K}}(1)$ denotes a remainder that is uniformly bounded over all $n \geq 1$ and all states with $\mathbf{c}/n \in \mathcal{K}$. Consequently, there exists a constant $C_{\mathcal{K}} < \infty$, depending only on the model primitives and the set $\mathcal{K}$, such that for every $n \geq 1$ and all states with $\mathbf{c}/n \in \mathcal{K}$,

$$0 \leq V_n^{\mathrm{OPT}}(\mathbf{c}) - V_n^{\mathrm{DF}}(\mathbf{c}) \leq C_{\mathcal{K}}.$$

We prove Theorem 4.1 in Section C.1 by establishing a stronger result. Specifically, for any policy π, define a stopping time as the first time its trajectory exits a local region or the remaining horizon reaches a predetermined integer. We define the cumulative dual-price score loss of policy π as its expected shortfall in dual-price score (defined in Assumption 3.3), evaluated at the state-dependent optimal dual prices and accumulated along its trajectory before stopping. In Theorem B.7, we establish that the performance gap between the value function of policy π and the corrected fluid reference $F_n(\mathbf{c}) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\mathsf{H}_{n-1} + O(1)$ is proportional to its cumulative dual-price score loss.

On the other hand, before stopping, every action is feasible, so the dynamic fluid policy π^{DF} selects the static fluid action $a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$ at the current binding state $\boldsymbol{\rho}_{\mathsf{B}}$, which maximizes the dual-price score. Hence, its cumulative dual-price score loss is zero. This implies that both the dynamic fluid policy and the optimal online policy achieve the corrected fluid reference up to uniformly bounded constants.

4.3 Applications to Specific Online Resource Allocation Problems

In this section, we apply Theorem 4.1 to three specific online resource allocation problems and demonstrate the new results it yields in these settings.

4.3.1 The Uniform Multisecretary Model

We first consider the uniform multisecretary problem (Section 3 of Bray 2025). In this problem, the decision maker has a single resource with an integer capacity $c \in \mathbb{N}_+$. In each period t, the decision maker observes a value V_t and either accepts it, consuming one unit of capacity, or rejects it. The values $\{V_t\}$ are i.i.d. standard uniform random variables. This problem is a special case of our online resource allocation problem with a single resource ($m = 1$), an action set $\mathcal{A} = \{0, 1\}$, and reward-consumption pairs $(R_0, B_0) = (0, 0)$ and $(R_1, B_1) = (V, 1)$ with $V \sim \mathrm{Unif}[0, 1]$. Propositions 1 and 2 of Bray (2025) establishes a tight $\Theta(\log n)$ lower bound on the performance gap of any policy relative to the offline optimal benchmark when the initial capacity is proportional to the horizon.

We note that on the set $\mathcal{U} = (0, 1)$, Assumption 2.1 and Assumptions 3.1 to 3.3 are satisfied with binding-resource set $\mathsf{B} = \{1\}$, fluid value function $\phi^{\mathrm{F}}(\rho) = \phi_{\mathsf{B}}(\rho) = \rho - \rho^2/2$, and optimal dual variable $\lambda^*(\rho) = 1 - \rho$. The dynamic fluid policy π^{DF} takes a simple threshold form: in each period with c units of remaining capacity and n periods remaining, it accepts the arrival if and only if $V > 1 - c/n$ and $c \geq 1$.

Theorem 4.1 implies that, for every compact set $\mathcal{K} \Subset (0, 1)$, uniformly over all states (c, n) satisfying $c/n \in \mathcal{K}$, both $V_n^{\mathrm{OPT}}(c)$ and $V_n^{\mathrm{DF}}(c)$ admit the expansion

$$n\left(\rho - \frac{1}{2}\rho^2\right) - \frac{1}{2}\rho(1-\rho)\log n + O_{\mathcal{K}}(1), \qquad \rho = c/n.$$

Therefore, the dynamic fluid policy has a uniformly bounded optimality gap relative to the optimal online policy.

4.3.2 The Stochastic Knapsack Problem of Lueker (1998)

In stochastic knapsack, the decision maker sequentially observes n i.i.d. nonnegative weight-value pairs (W, V) and must decide whether to accept each item before observing the next. The decision maker has a total capacity of $n\beta_0$. If the remaining capacity is c, an item can be accepted only if $W \leq c$. Acceptance yields reward V and consumes W units of capacity. This problem is a special case of our online resource allocation problem with one resource, action set $\mathcal{A} = \{0, 1\}$, and reward-consumption pairs $(R_0, B_0) = (0, 0)$ and $(R_1, B_1) = (V, W)$.

Lueker (1998) studies an OnLineGreedy policy that coincides with our dynamic fluid policy for the stochastic knapsack problem. Theorem 2 of Lueker (1998) shows that, under their Conditions 1–3, the value of the OnLineGreedy policy admits the following expansion:[1] [2]

$$n\,\phi^{\mathrm{F}}(\beta_0) - \Gamma(\beta_0)\log n + o(\log n)\,.$$

Thus, this policy differs from the offline optimal benchmark by $\Theta(\log n)$. Moreover, this order is tight: no non-anticipative policy can improve the gap by more than $o(\log n)$.

Lueker's Conditions 1–2 imply our Assumption 2.1 and Assumptions 3.1–3.3. Thus, our model is more general and our assumptions are weaker. Moreover, Theorem 4.1 yields a stronger result: both $V_n^{\mathrm{OPT}}(c)$ and $V_n^{\mathrm{DF}}(c)$ admit the expansion

$$n\,\phi^{\mathrm{F}}(\beta_0) - \Gamma(\beta_0)\log n + O(1)\,.$$

Consequently, the dynamic fluid policy has only a uniformly bounded performance gap relative to the optimal online policy.

4.3.3 The Online Linear Programming Problem of Bray (2025)

We next consider the online linear programming problem studied in Section 4 of Bray (2025). The decision maker begins with an inventory vector $n\boldsymbol{\beta} \in \mathbb{R}_+^m$ and sequentially observes n i.i.d. requests $(u_t, \mathbf{a}_t)$. Accepting request t yields utility $u_t \geq 0$ and consumes the resource vector $\mathbf{a}_t \geq \mathbf{0}$, whereas rejecting it yields no utility and consumes no resources. This problem is a special case of our problem with action set $\mathcal{A} = \{0, 1\}$ and reward-consumption pairs $(R_0, \mathbf{B}_0) = (0, \mathbf{0})$ and $(R_1, \mathbf{B}_1) = (u, \mathbf{a})$.

[1] As a notable example, in Lueker (1998), Conditions 1–3 hold when the values V and weights W are independent standard uniform random variables.

[2] We note that $\kappa(\beta_0)$ in Lueker's Theorem 2 equals our $\phi^{\mathrm{F}}(\beta_0)$ and $\sigma_0^2/(2\alpha_0)$ equals our $\Gamma(\beta_0)$. Thus, the first-order and logarithmic terms in Lueker's expansion coincide with those in Theorem 4.1.

In the known-distribution setting, Bray (2025) analyze the same policy (Algorithm 2 therein) as our dynamic fluid policy. Their Theorem 1 establishes an $O(\log n)$ gap for this policy relative to the offline optimum, and their Theorem 3 establishes an $\Omega(\log n)$ gap for the optimal online policy. Thus, the logarithmic order relative to the offline optimum is tight.

Bray's non-degeneracy conditions, stated in their Assumptions 1–6, imply ours in Section 3, except for the local α-Hölder condition on $\nabla^2\phi_{\mathsf{B}}$ required in part 2 of Assumption 3.1. Under this additional local Hölder condition, Theorem 4.1 further establishes that the dynamic fluid policy has a uniformly bounded optimality gap relative to the optimal online policy.

5 Constant Optimality Gap for a First-Order Policy

In this section, we show that, under additional regularity conditions, a first-order policy without re-solving achieves the same performance guarantee as the dynamic fluid policy π^{DF}. The first-order policy initializes the dual prices at zero and updates them using gradient descent. We denote this policy by π^{GD} and its value function by $V_n^{\mathrm{GD}}(\mathbf{c})$.

For an opportunity $\boldsymbol{\theta}$ and a dual price $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, define the static fluid action and its expected resource consumption by

$$a(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \arg\max_{a \in \mathcal{A}} \big\{R_a - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_a\big\}, \qquad \mathbf{m}(\boldsymbol{\lambda}) = \mathbb{E}\big[\mathbf{B}_{a(\boldsymbol{\lambda}, \boldsymbol{\theta})}\big], \tag{7}$$

where ties are broken by the smallest index. Recall that the objective of the fluid relaxation dual problem (2) is

$$\mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}) \triangleq \boldsymbol{\rho}^{\mathrm{T}} \boldsymbol{\lambda} + \mathcal{V}(\boldsymbol{\lambda}).$$

The policy π^{GD} follows Algorithm 1 with the following dual-price update rule. Fix a step size $\eta > 0$ and initialize $\boldsymbol{\lambda}_0 = \mathbf{0}$. After period $k+1$, the policy updates the dual prices using the pre-decision inventory ratio $\boldsymbol{\rho}_k$ and the expected resource consumption $\mathbf{m}(\boldsymbol{\lambda}_k)$:

$$\boldsymbol{\lambda}_{k+1} = \Big(\boldsymbol{\lambda}_k - \eta\big\{\boldsymbol{\rho}_k - \mathbf{m}(\boldsymbol{\lambda}_k)\big\}\Big)^+, \tag{8}$$

where $\boldsymbol{\rho}_k - \mathbf{m}(\boldsymbol{\lambda}_k)$ is a subgradient of $\mathcal{L}_{\boldsymbol{\rho}_k}(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}_k$. To establish the performance guarantee of π^{GD}, we impose the following local dual stability condition.

Assumption 5.1 (Local Dual-Gradient Lipschitz Continuity)**.** For every compact set $\mathcal{K} \Subset \mathcal{U}$, there

exist constants $\varepsilon_{\mathcal{K}} > 0$ and $L_{\mathcal{K}} < \infty$ such that, for every $\boldsymbol{\rho} \in \mathcal{K}$,

$$\|\nabla \mathcal{V}(\mathbf{p}) - \nabla \mathcal{V}(\mathbf{q})\| \le L_{\mathcal{K}} \|\mathbf{p} - \mathbf{q}\|, \qquad \mathbf{p}, \mathbf{q} \in B\big(\boldsymbol{\lambda}^*(\boldsymbol{\rho}), \varepsilon_{\mathcal{K}}\big).$$

Under Assumption 5.1, Theorem 5.1 establishes that π^{GD} admits the same performance expansion and thus achieves the same performance guarantee as the dynamic fluid policy π^{DF}.

Theorem 5.1 (Performance of the First-Order Policy π^{GD}). *Suppose Assumptions 2.1, 3.1–3.3, and 5.1 hold. For every compact set $\mathcal{K} \Subset \mathcal{U}$, there exists a constant $\bar{\eta}_{\mathcal{K}} > 0$ such that, for any $\eta \in (0, \bar{\eta}_{\mathcal{K}}]$, the policy π^{GD} with step size η satisfies*

$$V_n^{\mathrm{GD}}(\mathbf{c}) = n\, \phi^{\mathrm{F}}(\mathbf{c}/n) - \Gamma(\mathbf{c}_{\mathsf{B}}/n) \log n + O_{\mathcal{K}}(1)$$

uniformly over states satisfying $\mathbf{c}/n \in \mathcal{K}$. Consequently,

$$0 \le V_n^{\mathrm{OPT}}(\mathbf{c}) - V_n^{\mathrm{GD}}(\mathbf{c}) \le C_{\mathcal{K},\eta}$$

for every $n \ge 1$ and all states with $\mathbf{c}/n \in \mathcal{K}$, where $C_{\mathcal{K},\eta} < \infty$ depends only on the model primitives, set $\mathcal{K}$, and step size $\eta \in (0, \bar{\eta}_{\mathcal{K}}]$.

We specify the constant $\bar{\eta}_{\mathcal{K}}$ and prove Theorem 5.1 in Appendix C.2. By Theorem B.7, the cumulative dual-price score loss of a policy π serves as a proxy for its performance gap relative to the dynamic fluid policy π^{DF}. In particular, if this cumulative loss is uniformly bounded, then π admits the same performance expansion and achieves the same performance guarantee as π^{DF}. In Appendix C.2, we establish such a uniform bound for π^{GD}, thereby proving the desired result.

The key argument proceeds in two steps. First, although the initial dual prices may lie outside the region where the dual objective is smooth, the global convexity of $\mathcal{V}(\boldsymbol{\lambda})$ and boundedness of its subgradients ensure that, for a sufficiently small fixed step size, the dual prices enter a suitable neighborhood of the optimal fluid price within a finite number of periods independent of the horizon. Thereafter, the local Lipschitz continuity of the dual gradient (Assumption 5.1) allows us to establish a contraction toward the current optimal fluid price $\boldsymbol{\lambda}_k^* = (\boldsymbol{\lambda}(\mathbf{x}_k), \mathbf{0}_{\mathsf{S}})$ under projected gradient updates. This contraction yields a uniform bound on the cumulative squared tracking error $\sum_k \|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_k^*\|^2$, which in turn bounds the cumulative dual-price score loss.

6 Conclusion

We have studied a general online resource allocation problem under known arrival distributions. We show that, under standard non-degeneracy conditions, the dynamic fluid policy and the optimal online policy admit the same performance expansion: the fluid benchmark minus a common logarithmic correction, up to uniformly bounded remainders. Consequently, the dynamic fluid policy achieves a horizon-uniform constant gap relative to the optimal online policy. Our analysis is based on a telescoping argument that connects policy performance to cumulative dual-price score loss before the policy's trajectory exits a local region. Finally, we show that, under an additional dual stability condition, a first-order policy achieves the same performance guarantee using gradient updates without re-solving the fluid problem.

References

Arlotto, A. and Gurvich, I. (2019), 'Uniformly bounded regret in the multisecretary problem', *Stochastic Systems* **9**(3), 231–260.

Balseiro, S. R., Besbes, O. and Pizarro, D. (2023), 'Survey of dynamic resource-constrained reward collection problems: Unified model and analysis', *Operations Research* **72**(5), 2168–2189.

Besbes, O., Kanoria, Y. and Kumar, A. (2024), 'Dynamic resource allocation: Algorithmic design principles and spectrum of achievable performances', *Operations Research* **73**(3), 1273–1288.

Bray, R. L. (2025), 'Logarithmic regret in multisecretary and online linear programs with continuous valuations', *Operations Research* **73**(4), 2188–2203.

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

Chen, Y. and Wang, W. (2025), 'Beyond non-degeneracy: Revisiting certainty equivalent heuristic for online linear programming', *arXiv preprint arXiv:2501.01716* .

Jiang, J., Ma, W. and Zhang, J. (2025), 'Degeneracy is OK: Logarithmic regret for network revenue management with indiscrete distributions', *Operations Research* **73**(6), 3405–3420.

Li, X. and Ye, Y. (2021), 'Online linear programming: Dual convergence, new algorithms, and regret bounds', *Operations Research* **70**(5), 2948–2966.

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

Luenberger, D. G. (1997), *Optimization by vector space methods*, John Wiley & Sons.

Roch, S. (2024), *Modern discrete probability: An essential toolkit*, Cambridge University Press.

Talluri, K. and van Ryzin, G. (1998), 'An analysis of bid-price controls for network revenue management', *Management Science* **44**(11-Part-1), 1577–1593.

Vera, A. and Banerjee, S. (2021), 'The bayesian prophet: A low-regret framework for online decision making', *Management Science* **67**(3), 1368–1391.

Vera, A., Banerjee, S. and Gurvich, I. (2021), 'Online allocation and pricing: Constant regret via bellman inequalities', *Operations Research* **69**(3), 821–840.

Wang, Y. and Wang, H. (2022), 'Constant regret resolving heuristics for price-based revenue management', *Operations Research* **70**(6), 3538–3557.

Zhang, J. (2026*a*), 'Online resource allocation with continuous random consumption: Regret under degeneracy', *arXiv preprint arXiv:2607.02196* .

Zhang, J. (2026*b*), 'Tight lower bounds for the multi-secretary problem via bellman certificates', *arXiv preprint arXiv:2607.02150* .

A Proofs in Sections 2 and 3

A.1 Further Details of Lemma 2.1

A.1.1 Definition of the Offline Optimal Value $V_n^{\mathrm{OFF}}(\mathbf{c})$

Let $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ be independent copies of $\boldsymbol{\theta} \sim \Phi$, where $\boldsymbol{\theta}_t = (R_{a,t}, \mathbf{B}_{a,t})_{a \in \mathcal{A}}$ records the action-specific rewards and resource consumptions in period t. The offline benchmark observes the entire sequence $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ before selecting an action in each period and is defined by

$$V_n^{\mathrm{OFF}}(\mathbf{c}) \triangleq \mathbb{E}\left[\max_{\substack{(a_1,\ldots,a_n)\in\mathcal{A}^n \\ \sum_{t=1}^n \mathbf{B}_{a_t,t} \le \mathbf{c}}} \sum_{t=1}^n R_{a_t,t}\right].$$

A.1.2 Proof of Lemma 2.1

Since the offline benchmark $V_n^{\mathrm{OFF}}(\mathbf{c})$ observes the entire sequence $\{\boldsymbol{\theta}_t\}_{t=1}^n$ before making any decisions, the value of perfect information implies that $V_n^{\mathrm{OPT}}(\mathbf{c}) \le V_n^{\mathrm{OFF}}(\mathbf{c})$. We next show that

$$V_n^{\mathrm{OFF}}(\mathbf{c}) \le n \cdot \phi^{\mathrm{F}}(\boldsymbol{\rho}),$$

that is, the fluid relaxation provides an upper bound on the offline optimal benchmark. This result is standard and can be established in several ways. For completeness, we provide a proof based on the dual formulation of the fluid relaxation. An alternative proof can be obtained from the primal formulation by following a similar argument as in the proof of Proposition 1 of Balseiro et al. (2023).

Fix a dual variable $\boldsymbol{\lambda} \in \mathbb{R}_+^m$. For every realization of the arriving sequence $\{\boldsymbol{\theta}_t\}$ and every feasible action sequence $(a_1, \ldots, a_n) \in \mathcal{A}^n$,

$$\begin{aligned}
\sum_{t=1}^n R_{a_t,t} &= \boldsymbol{\lambda}^{\mathrm{T}} \sum_{t=1}^n \mathbf{B}_{a_t,t} + \sum_{t=1}^n \big(R_{a_t,t} - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_{a_t,t}\big) \\
&\le \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{c} + \sum_{t=1}^n \big(R_{a_t,t} - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_{a_t,t}\big) \\
&\le \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{c} + \sum_{t=1}^n \max_{a \in \mathcal{A}} \Big\{ R_{a,t} - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_{a,t} \Big\},
\end{aligned}$$

where the first inequality follows from the capacity constraint $\sum_{t=1}^n \mathbf{B}_{a_t,t} \le \mathbf{c}$ and $\boldsymbol{\lambda} \ge \mathbf{0}$. Maximizing this pathwise inequality over all feasible action sequences and taking expectations gives

$$\begin{aligned}
V_n^{\mathrm{OFF}}(\mathbf{c}) &\le \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{c} + \sum_{t=1}^n \mathbb{E}\left[\max_{a \in \mathcal{A}} \Big\{ R_{a,t} - \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{B}_{a,t} \Big\}\right] \\
&= \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{c} + n\, \mathcal{V}(\boldsymbol{\lambda}).
\end{aligned}$$

The equality follows from the fact that the opportunities are identically distributed across time and the definition of $\mathcal{V}(\boldsymbol{\lambda})$. Since the bound holds for every $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, we have:

$$V_n^{\mathrm{OFF}}(\mathbf{c}) \le \inf_{\boldsymbol{\lambda} \in \mathbb{R}_+^m} \Big\{ \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{c} + n\, \mathcal{V}(\boldsymbol{\lambda}) \Big\}$$

$$= n \inf_{\boldsymbol{\lambda} \in \mathbb{R}_+^m} \Big\{ \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{\rho} + \mathcal{V}(\boldsymbol{\lambda}) \Big\}$$
$$= n\, \phi^{\mathrm{F}}(\boldsymbol{\rho})\,,$$

where the first equality uses $\mathbf{c} = n\boldsymbol{\rho}$ and the second equality follows from the strong duality between (1) and (2).

A.2 Proof of Lemma 3.1

Let $\boldsymbol{\lambda}^*$ be an optimal solution to the dual problem (2) at some $\boldsymbol{\rho} \in \mathcal{U}$. From (2) we have

$$\phi^{\mathrm{F}}(\boldsymbol{\rho}') \le \boldsymbol{\lambda}^{*\mathrm{T}} \boldsymbol{\rho}' + \mathcal{V}(\boldsymbol{\lambda}^*) = \phi^{\mathrm{F}}(\boldsymbol{\rho}) + \boldsymbol{\lambda}^{*\mathrm{T}} (\boldsymbol{\rho}' - \boldsymbol{\rho})\,, \qquad \forall\, \boldsymbol{\rho}' \in \mathbb{R}_+^m\,,$$

where the equality follows from the optimality of $\boldsymbol{\lambda}^*$ at $\boldsymbol{\rho}$. Hence, $\boldsymbol{\lambda}^*$ is a supergradient of the concave function ϕ^{F} at $\boldsymbol{\rho}$.

On the other hand, since $\phi^{\mathrm{F}}(\boldsymbol{\rho}) = \phi_{\mathsf{B}}(\boldsymbol{\rho}_{\mathsf{B}})$ and $\phi_{\mathsf{B}} \in C^{2,\alpha}(\mathcal{U}_{\mathsf{B}})$ by Assumption 3.1, the fluid value function ϕ^{F} is differentiable on $\mathcal{U}$, with gradient

$$\nabla \phi^{\mathrm{F}}(\boldsymbol{\rho}) = \big(\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}}), \mathbf{0}_{\mathsf{S}}\big).$$

Because a differentiable concave function admits a unique supergradient at every interior point of its domain, it follows that

$$\boldsymbol{\lambda}^* = \nabla \phi^{\mathrm{F}}(\boldsymbol{\rho}) = \big(\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}}), \mathbf{0}_{\mathsf{S}}\big).$$

Therefore, every optimal dual solution coincides with this vector, implying that the optimal dual solution is unique.

A.3 Proof of Lemma 3.2

For any $\boldsymbol{\rho} \in \mathcal{U}$, we have $\boldsymbol{\lambda}^*(\boldsymbol{\rho}) = (\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}}), \mathbf{0}_{\mathsf{S}})$ by Lemma 3.1. Therefore, from (2),

$$\begin{aligned}
\phi^{\mathrm{F}}(\boldsymbol{\rho}) &= \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \boldsymbol{\rho}_{\mathsf{B}} + \mathbb{E}\Big[\max_{a \in \mathcal{A}} \Big\{ R_a - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \mathbf{B}_{a,\mathsf{B}} \Big\}\Big] \\
&= \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \boldsymbol{\rho}_{\mathsf{B}} + \mathbb{E}\Big[R_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})} - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \mathbf{B}_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}}),\mathsf{B}}\Big] \\
&= \mathbb{E}\Big[R_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})}\Big],
\end{aligned}$$

where the first equality follows from the optimality of $\boldsymbol{\lambda}^*(\boldsymbol{\rho})$, the second equality follows from the definition of $a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$ in (3), and the third equality uses $\mathbb{E}\big[\mathbf{B}_{a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}}),\mathsf{B}}\big] = \boldsymbol{\rho}_{\mathsf{B}}$ from Assumption 3.2.

B Preliminaries for the Proofs in Sections 4 and 5

In this section, we develop preliminary results for the proofs in Sections 4 and 5. Specifically, Section B.1 constructs buffer regions, and Section B.2 establishes several auxiliary lemmas. Section B.3 establishes a performance characterization theorem (Theorem B.7) based on cumulative dual-price score loss, and Section B.4 provides its proof.

B.1 Notation and Buffer Regions

Fix a compact target set $\mathcal{K}$ and let $\mathcal{K}_{\mathsf{B}} \triangleq \{\boldsymbol{\rho}_{\mathsf{B}} : \boldsymbol{\rho} \in \mathcal{K}\} \Subset \mathcal{U}_{\mathsf{B}}$ denote its projection onto the binding coordinates. For $\mathbf{x}_0 \in \mathbb{R}^{|\mathsf{B}|}$ and $r > 0$, define the open and closed Euclidean balls of radius r centered at $\mathbf{x}_0$ by

$$B(\mathbf{x}_0, r) \triangleq \{\mathbf{x} \in \mathbb{R}^{|\mathsf{B}|} : \|\mathbf{x} - \mathbf{x}_0\| < r\},$$
$$\bar{B}(\mathbf{x}_0, r) \triangleq \{\mathbf{x} \in \mathbb{R}^{|\mathsf{B}|} : \|\mathbf{x} - \mathbf{x}_0\| \le r\}.$$

Choose $r_0 > 0$ sufficiently small such that

$$\widetilde{\mathcal{K}}_{\mathsf{B}} \triangleq \bigcup_{\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}} \Big\{\mathbf{x} : \big\|\mathbf{x} - \mathbf{x}_0\big\| \le 2r_0\Big\} \Subset \mathcal{U}_{\mathsf{B}}.$$

Let $\mathsf{S} \triangleq \mathsf{B}^c$ denote the set of slack resources. If $\mathsf{S} \neq \varnothing$, then, by the compactness of $\mathcal{K}$, the continuity of $\mu_j(\cdot)$ on $\mathcal{U}_{\mathsf{B}}$ for each $j \in \mathsf{S}$, and the strict slackness in Assumption 3.2, we have

$$\underline{\delta} \triangleq \inf_{\boldsymbol{\rho} \in \mathcal{K}} \min_{j \in \mathsf{S}} \Big\{\rho_j - \mu_j(\boldsymbol{\rho}_{\mathsf{B}})\Big\} > 0.$$

By further decreasing r_0, if necessary, the continuity of $\mu_j(\cdot)$ on $\mathcal{U}_{\mathsf{B}}$ for each $j \in \mathsf{S}$ ensures that, for every $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$,

$$\|\mathbf{x} - \mathbf{x}_0\| \le 2r_0 \quad \Longrightarrow \quad \big|\mu_s(\mathbf{x}) - \mu_s(\mathbf{x}_0)\big| \le \underline{\delta}/4, \quad \forall\, s \in \mathsf{S}. \tag{9}$$

We henceforth fix this choice of r_0. Since $\widetilde{\mathcal{K}}_{\mathsf{B}}$ is compact and contained in $(0,\infty)^{|\mathsf{B}|}$, there exist constants $\underline{\rho}_{\mathsf{B}} > 0$ and $\overline{\rho}_{\mathsf{B}} < \infty$ such that, for every $\mathbf{x} \in \widetilde{\mathcal{K}}_{\mathsf{B}}$,

$$x_i \ge \underline{\rho}_{\mathsf{B}} \qquad \forall\, i \in \mathsf{B}, \qquad \text{and} \qquad \|\mathbf{x}\| \le \overline{\rho}_{\mathsf{B}}. \tag{10}$$

Finally, let $\alpha \in (0,1]$ denote the Hölder exponent in part 2 of Assumption 3.1, and let $\beta_\star$, $C_\star$, and $t_\star$ be the constants specified in Assumption 3.3 for $\widetilde{\mathcal{K}}_{\mathsf{B}}$. Since replacing $t_\star$ and $\beta_\star$ by their respective minima with 1 preserves the bound in Assumption 3.3, we may assume without loss of generality that

$$t_\star \in (0,1] \qquad \text{and} \qquad \beta_\star \in (0,1].$$

We define $\gamma \triangleq \min\{\alpha, \beta_\star\} \in (0,1]$.

B.2 Useful Lemmas

In this section, we establish several auxiliary lemmas for the subsequent analysis. Throughout, all constants may depend on $\mathcal{K}$ and the model primitives, but are independent of a particular element in $\mathcal{K}$ and the horizon n.

B.2.1 Extension Along Slack Coordinates

Although Assumptions 3.1 and 3.2 are stated on the original neighborhood $\mathcal{U}$, the inventories per remaining period of slack resources may increase as the remaining horizon decreases. Therefore, the normalized state need not remain in $\mathcal{U}$ as it evolves. Lemma B.1 shows that, as long as the slack resources are strictly slack, the fluid value, optimal dual prices, and dual-price actions continue to depend only on the binding coordinates.

Recall that $a^{\mathrm{F}}(\mathbf{x})$ denotes the static fluid action defined in (3) and $a_n^{\mathrm{DF}}(\mathbf{c})$ denotes the action chosen by the dynamic fluid policy at state $(\mathbf{c}, n)$ defined in Section 4.2. We now state Lemma B.1.

Lemma B.1 (Extension Along Slack Coordinates). *Define the extended region*

$$\widetilde{\mathcal{U}} \triangleq \{\boldsymbol{\rho} : \boldsymbol{\rho}_{\mathsf{B}} \in \mathcal{U}_{\mathsf{B}},\ \rho_s > \mu_s(\boldsymbol{\rho}_{\mathsf{B}}) \text{ for every } s \in \mathsf{S}\}.$$

The following statements hold.

1. *For every $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$, the fluid relaxation (1) satisfies $\phi^{\mathrm{F}}(\boldsymbol{\rho}) = \phi_{\mathsf{B}}(\boldsymbol{\rho}_{\mathsf{B}})$, and the static action $a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$ is optimal to (1).*
2. *For every $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$, the fluid relaxation has the unique optimal dual variable $(\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}}), \mathbf{0}_{\mathsf{S}})$.*
3. *At any state $(\mathbf{c}, n)$ with $\mathbf{c}/n \in \widetilde{\mathcal{U}}$ at which all actions are feasible, $a_n^{\mathrm{DF}}(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/n)$ almost surely.*

Proof. For any $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$, there exists a reference point $\boldsymbol{\rho}^0(\mathbf{x}) \in \mathcal{U}$ such that $\boldsymbol{\rho}^0(\mathbf{x})_{\mathsf{B}} = \mathbf{x}$ by the definition of the projection $\mathcal{U}_{\mathsf{B}}$. Since the static fluid action a^{F} depends only on $\mathbf{x}$, Assumption 3.2 and Lemma 3.2 yield:

$$\mathbb{E}\Big[\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}\Big] = \mathbf{x}, \qquad \mathbb{E}\Big[B_{a^{\mathrm{F}}(\mathbf{x}),j}\Big] = \mu_j(\mathbf{x}) \quad \forall j \in \mathsf{S}, \qquad \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})}\Big] = \phi_{\mathsf{B}}(\mathbf{x}). \tag{11}$$

Proof of Part 1. Fix $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$ and let $\mathbf{x} = \boldsymbol{\rho}_{\mathsf{B}}$. Since $\mu_j(\mathbf{x}) < \rho_j$ for all $j \in \mathsf{S}$, (11) implies that $a^{\mathrm{F}}(\mathbf{x})$ is feasible to (1) at $\boldsymbol{\rho}$ and attains objective value $\phi_{\mathsf{B}}(\mathbf{x})$.

By Lemma 3.1, $(\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$ is dual optimal at $\boldsymbol{\rho}^0(\mathbf{x})$. Because $\boldsymbol{\rho}^0(\mathbf{x})_{\mathsf{B}} = \mathbf{x}$ and $\phi^{\mathrm{F}}(\boldsymbol{\rho}^0(\mathbf{x})) = \phi_{\mathsf{B}}(\mathbf{x})$, its dual objective value is

$$\boldsymbol{\lambda}(\mathbf{x})^{\mathrm{T}}\mathbf{x} + \mathcal{V}\big((\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})\big) = \phi_{\mathsf{B}}(\mathbf{x}).$$

Thus, $(\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$ has a dual objective value $\phi_{\mathsf{B}}(\mathbf{x})$. Since $a^{\mathrm{F}}(\mathbf{x})$ is primal feasible at $\boldsymbol{\rho}$ and attains the same value, strong duality implies that $a^{\mathrm{F}}(\mathbf{x})$ is optimal for (1), $(\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$ is dual optimal, and $\phi^{\mathrm{F}}(\boldsymbol{\rho}) = \phi_{\mathsf{B}}(\mathbf{x}) = \phi_{\mathsf{B}}(\boldsymbol{\rho}_{\mathsf{B}})$.

Proof of Part 2. It remains to show that $(\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$ is the unique optimal dual variable. Let $\widetilde{\boldsymbol{\lambda}}$ be any optimal dual variable at $\boldsymbol{\rho}$. The strong duality and the optimality of $a^{\mathrm{F}}(\mathbf{x})$ imply:

$$\begin{aligned} 0 &= \widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{\rho} + \mathcal{V}(\widetilde{\boldsymbol{\lambda}}) - \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})}\Big] \\ &= \sum_{i \in [m]} \widetilde{\lambda}_i \left(\rho_i - \mathbb{E}\big[B_{a^{\mathrm{F}}(\mathbf{x}),i}\big]\right) + \mathbb{E}\left[\max_{a \in \mathcal{A}} \left\{R_a - \widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{B}_a\right\} - \left(R_{a^{\mathrm{F}}(\mathbf{x})} - \widetilde{\boldsymbol{\lambda}}^{\mathrm{T}}\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x})}\right)\right]. \end{aligned}$$

The first term on the right-hand side is nonnegative because $\widetilde{\boldsymbol{\lambda}} \geq \mathbf{0}$ and $a^{\mathrm{F}}(\mathbf{x})$ is feasible at $\boldsymbol{\rho}$. The second term is also nonnegative because the pointwise maximum is at least the score of the specific action $a^{\mathrm{F}}(\mathbf{x})$. Since their sum is zero, both terms must be zero. In particular, since $\rho_s - \mu_s(\mathbf{x}) > 0$ for every $s \in \mathsf{S}$, we must have $\widetilde{\lambda}_s = 0$ for every $s \in \mathsf{S}$. Thus, $\widetilde{\boldsymbol{\lambda}} = (\widetilde{\boldsymbol{\lambda}}_{\mathsf{B}}, \mathbf{0}_{\mathsf{S}})$. Because $\boldsymbol{\rho}_{\mathsf{B}} = \boldsymbol{\rho}^0(\mathbf{x})_{\mathsf{B}} = \mathbf{x}$, this vector has the same dual objective value at $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^0(\mathbf{x})$ and is therefore dual optimal at $\boldsymbol{\rho}^0(\mathbf{x})$. Lemma 3.1 then yields $\widetilde{\boldsymbol{\lambda}} = (\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$.

Proof of Part 3. By part 2, the optimal dual variable at this state is $(\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}}), \mathbf{0}_{\mathsf{S}})$. When all actions are feasible, $a_n^{\mathrm{DF}}(\mathbf{c})$ and $a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/n)$ maximize the same objective with the same tie-breaking rule. Hence $a_n^{\mathrm{DF}}(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/n)$ almost surely. □

B.2.2 One-Step Containment and Local Feasibility

Fix a constant $r \in (0, r_0]$, where r_0 is the radius specified in Appendix B.1. For any $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$, define the local region

$$\mathcal{D}(\mathbf{x}_0) = \Big\{ \boldsymbol{\rho} : \boldsymbol{\rho}_{\mathsf{B}} \in B(\mathbf{x}_0, r),\ \rho_s > \mu_s(\mathbf{x}_0) + \underline{\delta}/2 \ \text{ for every } s \in \mathsf{S} \Big\}. \tag{12}$$

When $\mathsf{S} = \varnothing$, we omit the slack-resource conditions in (12). Lemma B.2 shows that, when the remaining horizon is large, the one-step successor of any state in $\mathcal{D}(\mathbf{x}_0)$ remains in $\widetilde{\mathcal{U}}$, with its binding-resource coordinates contained in $B(\mathbf{x}_0, 2r)$.

Lemma B.2 (One-Step Containment and Local Feasibility). *For any fixed radius $r \in (0, r_0]$, there exists an integer $n_0(r)$, depending only on $\mathcal{K}$, r, and the model primitives, such that the following statements hold for every remaining horizon $h \geq n_0(r) + 1$, $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$, and $\boldsymbol{\rho} = \mathbf{c}/h \in \mathcal{D}(\mathbf{x}_0)$:*

1. *Every action $a \in \mathcal{A}$ is feasible at $\mathbf{c}$.*
2. *$\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$ and $a_h^{\mathrm{DF}}(\mathbf{c}) = a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$ almost surely.*
3. *For any action $a \in \mathcal{A}$, the corresponding one-step successor ratio $\boldsymbol{\rho}^a \triangleq (\mathbf{c} - \mathbf{B}_a)/(h-1)$ satisfies (i) $\boldsymbol{\rho}^a_{\mathsf{B}} \in B(\mathbf{x}_0, 2r)$ and (ii) $\rho^a_s - \mu_s(\boldsymbol{\rho}^a_{\mathsf{B}}) \geq \underline{\delta}/8$ for any $s \in \mathsf{S}$. Thus, $\boldsymbol{\rho}^a \in \widetilde{\mathcal{U}}$.*

Proof. Choose an integer $n_0(r) \geq 1$ such that

$$(n_0(r)+1)\,\underline{\rho}_{\mathsf{B}} \geq \bar{b}, \qquad \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{n_0(r)} \leq r, \qquad \frac{\bar{b}}{n_0(r)} \leq \frac{\underline{\delta}}{8}. \tag{13}$$

The last condition is not needed if $\mathsf{S} = \varnothing$.

Proof of Part 1. It suffices to show that $\mathbf{c} \geq \bar{b}\mathbf{1}$. By (10), the capacity of each binding resource is at least $h\underline{\rho}_{\mathsf{B}} \geq \bar{b}$. For a slack resource $s \in \mathsf{S}$, its normalized capacity exceeds $\mu_s(\mathbf{x}_0) + \underline{\delta}/2 \geq \underline{\delta}/2$, so its capacity is at least $h\underline{\delta}/2 \geq 4\bar{b}$. Consequently, $\mathbf{c} \geq \bar{b}\mathbf{1}$ as desired.

Proof of Part 2. For every $s \in \mathsf{S}$, we have $\mu_s(\boldsymbol{\rho}_{\mathsf{B}}) \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/4 < \rho_s$, where the first inequality follows from (9) and the second from the definition of $\mathcal{D}(\mathbf{x}_0)$. Hence, $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$. Part 3 of Lemma B.1, together with part 1, implies that $a_h^{\mathrm{DF}}(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/h)$ almost surely.

Proof of Part 3. For every action $a \in \mathcal{A}$, we have

$$\boldsymbol{\rho}^a = \boldsymbol{\rho} + \frac{\boldsymbol{\rho} - \mathbf{B}_a}{h-1}. \tag{14}$$

For the binding coordinates, (14) and (10) imply

$$\|\boldsymbol{\rho}^a_{\mathsf{B}} - \boldsymbol{\rho}_{\mathsf{B}}\| \leq \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{h-1} \leq \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{n_0(r)} \leq r.$$

Therefore,

$$\|\boldsymbol{\rho}^a_{\mathsf{B}} - \mathbf{x}_0\| \leq \|\boldsymbol{\rho}^a_{\mathsf{B}} - \boldsymbol{\rho}_{\mathsf{B}}\| + \|\boldsymbol{\rho}_{\mathsf{B}} - \mathbf{x}_0\| < 2r.$$

For each slack coordinate $s \in \mathsf{S}$, from (14),

$$\begin{aligned}\rho_s^a \geq \rho_s - \frac{\bar{b}}{h-1} &> \mu_s(\mathbf{x}_0) + \frac{\underline{\delta}}{2} - \frac{\bar{b}}{h-1} \\ &\geq \mu_s(\boldsymbol{\rho}_{\mathsf{B}}^a) + \frac{\underline{\delta}}{4} - \frac{\bar{b}}{h-1} \geq \mu_s(\boldsymbol{\rho}_{\mathsf{B}}^a) + \frac{\underline{\delta}}{8}.\end{aligned}$$

The second inequality follows from $\boldsymbol{\rho} = \mathbf{c}/h \in \mathcal{D}(\mathbf{x}_0)$. The third inequality follows from $\mu_s(\boldsymbol{\rho}_{\mathsf{B}}^a) \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/4$ due to $\|\boldsymbol{\rho}_{\mathsf{B}}^a - \mathbf{x}_0\| < 2r$ and (9). The fourth inequality follows from the third inequality in (13). □

B.2.3 Hölder Continuity of the Coefficient $\Gamma(\mathbf{x})$

In this section, we establish that the logarithmic coefficient $\Gamma(\mathbf{x})$ is locally γ-Hölder continuous, where $\gamma \triangleq \min\{\alpha, \beta_\star\}$.

Lemma B.3 (Hölder Continuity of Γ). *There exist uniform constants C_Γ and L_Γ such that, for every $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and all $\mathbf{x}_1, \mathbf{x}_2 \in \bar{B}(\mathbf{x}_0, 2r_0)$, we have*

$$\big|\Gamma(\mathbf{x}_1)\big| \leq C_\Gamma, \qquad \big|\Gamma(\mathbf{x}_1) - \Gamma(\mathbf{x}_2)\big| \leq L_\Gamma \big\|\mathbf{x}_1 - \mathbf{x}_2\big\|^\gamma.$$

Proof. Recall from (4) that

$$\Gamma(\mathbf{x}) \triangleq -\frac{1}{2} \operatorname{tr}\Big(\nabla^2 \phi_{\mathsf{B}}(\mathbf{x})\, \Sigma(\mathbf{x})\Big).$$

Since $\nabla^2 \phi_{\mathsf{B}}$ is continuous on the compact set $\widetilde{\mathcal{K}}_{\mathsf{B}}$ by Assumption 3.1 and is therefore uniformly bounded, and $\Sigma(\mathbf{x})$ is uniformly bounded on $\widetilde{\mathcal{K}}_{\mathsf{B}}$ by Lemma B.4, it follows that $\Gamma(\mathbf{x})$ is uniformly bounded on $\widetilde{\mathcal{K}}_{\mathsf{B}}$.

We now prove the local γ-Hölder continuity. For two points $\mathbf{x}_1, \mathbf{x}_2 \in \bar{B}(\mathbf{x}_0, 2r_0)$, adding and subtracting $\nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_2)\Sigma(\mathbf{x}_1)$ inside the trace yields

$$\Gamma(\mathbf{x}_1) - \Gamma(\mathbf{x}_2) = -\frac{1}{2} \operatorname{tr}\Big(\Big\{\nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_1) - \nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_2)\Big\}\Sigma(\mathbf{x}_1) + \nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_2)\Big\{\Sigma(\mathbf{x}_1) - \Sigma(\mathbf{x}_2)\Big\}\Big).$$

Since $|\operatorname{tr}(AB)| \leq d\,\|A\|\|B\|$ for any $A, B \in \mathbb{R}^{d\times d}$, it follows that

$$|\Gamma(\mathbf{x}_1) - \Gamma(\mathbf{x}_2)| \leq \frac{|\mathsf{B}|}{2}\left\{\big\|\nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_1) - \nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_2)\big\|\,\big\|\Sigma(\mathbf{x}_1)\big\| + \big\|\nabla^2 \phi_{\mathsf{B}}(\mathbf{x}_2)\big\|\,\big\|\Sigma(\mathbf{x}_1) - \Sigma(\mathbf{x}_2)\big\|\right\}.$$

By Assumption 3.1, $\nabla^2 \phi_{\mathsf{B}}$ is bounded and α-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. Moreover, by Lemma B.4, Σ is bounded and $\beta_\star$-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. Hence, there exists a constant $C < \infty$ such that

$$\big|\Gamma(\mathbf{x}_1) - \Gamma(\mathbf{x}_2)\big| \leq C\left(\|\mathbf{x}_1 - \mathbf{x}_2\|^\alpha + \|\mathbf{x}_1 - \mathbf{x}_2\|^{\beta_\star}\right).$$

Consequently, $\Gamma(\mathbf{x})$ is locally γ-Hölder continuous with $\gamma \triangleq \min\big\{\alpha, \beta_\star\big\}$. □

Lemma B.4 (Hölder Continuity of the Covariance Matrix Σ). *There exists a constant $C < \infty$ such that, for every $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and all $\mathbf{x}, \mathbf{x}' \in \bar{B}(\mathbf{x}_0, 2r_0)$,*

$$\big\|\Sigma(\mathbf{x})\big\| \leq C, \qquad \big\|\Sigma(\mathbf{x}) - \Sigma(\mathbf{x}')\big\| \leq C\|\mathbf{x} - \mathbf{x}'\|^{\beta_\star}.$$

Proof. Define the second-moment matrix

$$M_2(\mathbf{x}) \triangleq \mathbb{E}\Big[\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}\,\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}^{\mathrm{T}}\Big].$$

By Assumption 3.2 part 1,

$$\Sigma(\mathbf{x}) = M_2(\mathbf{x}) - \mathbf{x}\mathbf{x}^{\mathrm{T}}.$$

First, Σ is bounded on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. This is because M_2 is bounded by Assumption 2.1 and compactness of $\widetilde{\mathcal{K}}_{\mathsf{B}}$ implies that $\mathbf{x}\mathbf{x}^{\mathrm{T}}$ is bounded.

Second, the map $\mathbf{x} \mapsto \mathbf{x}\mathbf{x}^{\mathrm{T}}$ is Lipschitz continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$ and hence $\beta_\star$-Hölder continuous because $\beta_\star \le 1$. Therefore, to establish that Σ is $\beta_\star$-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$, it suffices to show that M_2 is $\beta_\star$-Hölder continuous, which we do now.

Step 1. Preparation Fix $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and two points $\mathbf{x}, \mathbf{x}' \in \bar{B}(\mathbf{x}_0, 2r_0)$. Let $i = a^{\mathrm{F}}(\mathbf{x})$ and $j = a^{\mathrm{F}}(\mathbf{x}')$. Optimality of i at $\mathbf{x}$ and of j at $\mathbf{x}'$ yields

$$S_i(\mathbf{x}) - S_j(\mathbf{x}) \ge 0, \qquad S_j(\mathbf{x}') - S_i(\mathbf{x}') \ge 0.$$

Therefore,

$$\begin{aligned} 0 \le (S_i(\mathbf{x}) - S_j(\mathbf{x})) + (S_j(\mathbf{x}') - S_i(\mathbf{x}')) &= \Big(\boldsymbol{\lambda}(\mathbf{x}') - \boldsymbol{\lambda}(\mathbf{x})\Big)^{\mathrm{T}}\Big(\mathbf{B}_{i,\mathsf{B}} - \mathbf{B}_{j,\mathsf{B}}\Big) \\ &\le \big\|\boldsymbol{\lambda}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}')\big\|\,\big\|\mathbf{B}_{i,\mathsf{B}} - \mathbf{B}_{j,\mathsf{B}}\big\| \le C_1\,\|\mathbf{x} - \mathbf{x}'\|\,d_{ij}^{\mathsf{B}}\,, \end{aligned}$$

for a constant $C_1 < \infty$. Here, the second inequality is because $\boldsymbol{\lambda} = \nabla\phi_{\mathsf{B}}$ is continuously differentiable on $\widetilde{\mathcal{K}}_{\mathsf{B}} \Subset \mathcal{U}_{\mathsf{B}}$ by Assumption 3.1, and hence is Lipschitz continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$.

Note that on the event $\big\{j = a \ne i\big\}$, the preceding inequality yields

$$0 \le S_i(\mathbf{x}) - S_a(\mathbf{x}) \le C_1\,\|\mathbf{x} - \mathbf{x}'\|\,d_{ia}^{\mathsf{B}}\,.$$

Therefore,

$$d_{ij}^{\mathsf{B}}\,\mathbb{1}\big\{i \ne j\big\} \le \max_{a \ne i} d_{ia}^{\mathsf{B}}\,\mathbb{1}\big\{j = a\big\} \le \max_{a \in \mathcal{A}} \Big\{d_{ia}^{\mathsf{B}} \cdot \mathbb{1}\Big(0 \le S_i(\mathbf{x}) - S_a(\mathbf{x}) \le C_1\|\mathbf{x} - \mathbf{x}'\|\,d_{ia}^{\mathsf{B}}\Big)\Big\}\,.$$

Hence, whenever $C_1\|\mathbf{x} - \mathbf{x}'\| \le t_\star$, Assumption 3.3 implies

$$\mathbb{E}\Big[d_{ij}^{\mathsf{B}}\,\mathbb{1}\big\{i \ne j\big\}\Big] \le C_2\,\|\mathbf{x} - \mathbf{x}'\|^{\beta_\star} \tag{15}$$

for some constant $C_2 < \infty$.

Step 2. Hölder Continuity of the Second-Moment Matrix M_2 For any two vectors $u, v \in \mathbb{R}^d$,

$$\|uu^{\mathrm{T}} - vv^{\mathrm{T}}\| = \|u(u-v)^{\mathrm{T}} + (u-v)v^{\mathrm{T}}\| \le (\|u\| + \|v\|)\,\|u - v\|.$$

Hence, by bounded consumption, for some constant $C_3 < \infty$,

$$\big\|\mathbf{B}_{i,\mathsf{B}}\,\mathbf{B}_{i,\mathsf{B}}^{\mathrm{T}} - \mathbf{B}_{j,\mathsf{B}}\,\mathbf{B}_{j,\mathsf{B}}^{\mathrm{T}}\big\| \le C_3\,d_{ij}^{\mathsf{B}}\,. \tag{16}$$

The random matrices inside the expectations defining $M_2(\mathbf{x})$ and $M_2(\mathbf{x}')$ coincide when $i = j$. Therefore, by (15) and (16), whenever $C_1\|\mathbf{x} - \mathbf{x}'\| \le t_\star$,

$$\|M_2(\mathbf{x}) - M_2(\mathbf{x}')\| \le C_2\, C_3 \cdot \|\mathbf{x} - \mathbf{x}'\|^{\beta_\star}.$$

For the remaining case, $\|\mathbf{x} - \mathbf{x}'\| > r \triangleq t_\star / C_1$. By the boundedness of M_2,

$$\|M_2(\mathbf{x}) - M_2(\mathbf{x}')\| \le 2 \sup_{\mathbf{z} \in \widetilde{\mathcal{K}}_{\mathsf{B}}} \|M_2(\mathbf{z})\| \le \left(2 \sup_{\mathbf{z} \in \widetilde{\mathcal{K}}_{\mathsf{B}}} \|M_2(\mathbf{z})\| \cdot r^{-\beta_\star} \right) \|\mathbf{x} - \mathbf{x}'\|^{\beta_\star}.$$

Thus, M_2 is $\beta_\star$-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. □

B.2.4 The One-Step Residual

Fix $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and $\mathbf{x} \in \bar{B}(\mathbf{x}_0, r_0)$. Suppose that n periods remain and that the normalized inventory of the binding resources is $\mathbf{x}$. After taking action $a^{\mathrm{F}}(\mathbf{x})$, the normalized binding-resource inventory in the next period is

$$\mathbf{x}' \triangleq \frac{n\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1} = \mathbf{x} + \frac{\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1},$$

where $\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) \triangleq \mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}$ is the binding-resource consumption under action $a^{\mathrm{F}}(\mathbf{x})$, as defined before. For all $n \ge n_0(r_0)$, the successor state $\mathbf{x}'$ and the line segment joining $\mathbf{x}$ and $\mathbf{x}'$ lie in $\bar{B}(\mathbf{x}_0, 2r_0)$ by part 3 of Lemma B.2. Define the one-step residual relative to the fluid benchmark by

$$\delta_n^{\mathrm{F}}(\mathbf{x}) \triangleq n\phi_{\mathsf{B}}(\mathbf{x}) - \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})} + (n-1)\phi_{\mathsf{B}}(\mathbf{x}') \Big]. \tag{17}$$

Through a second-order Taylor expansion of $\delta_n^{\mathrm{F}}(\mathbf{x})$, Lemma B.5 shows that $\delta_n^{\mathrm{F}}(\mathbf{x})$ equals $\frac{\Gamma(\mathbf{x})}{n-1}$ plus a summable remainder.

Lemma B.5 (Second-Order Expansion of the One-Step Residual)**.** *Uniformly over any $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and $\mathbf{x} \in \bar{B}(\mathbf{x}_0, r_0)$,*

$$\delta_n^{\mathrm{F}}(\mathbf{x}) = \frac{\Gamma(\mathbf{x})}{n-1} + O(n^{-1-\alpha}).$$

In particular, there exists a constant $C < \infty$ such that $|\delta_n^{\mathrm{F}}(\mathbf{x})| \le C/n$ uniformly over $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and $\mathbf{x} \in \bar{B}(\mathbf{x}_0, r_0)$ for all sufficiently large n.

Proof. Define $\mathbf{h} \triangleq \frac{\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1}$. By bounded consumption and the compactness of $\widetilde{\mathcal{K}}_{\mathsf{B}}$, there exists a constant $C_1 < \infty$ such that $\|\mathbf{h}\| \le \frac{C_1}{n}$ uniformly over $\mathbf{x} \in \widetilde{\mathcal{K}}_{\mathsf{B}}$ for all sufficiently large n. Take n sufficiently large so that $C_1/n \le r_0$. Then, for every $t \in [0,1]$, $\|\mathbf{x} + t\mathbf{h} - \mathbf{x}_0\| \le 2r_0$, and hence the line segment $\{\mathbf{x} + t\mathbf{h} : t \in [0,1]\}$ is contained in $\bar{B}(\mathbf{x}_0, 2r_0)$. By Lemma B.6,

$$\phi_{\mathsf{B}}(\mathbf{x} + \mathbf{h}) = \phi_{\mathsf{B}}(\mathbf{x}) + \boldsymbol{\lambda}(\mathbf{x})^{\mathrm{T}}\mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\, \nabla^2\phi_{\mathsf{B}}(\mathbf{x})\, \mathbf{h} + r_n(\mathbf{x}, \mathbf{h}), \tag{18}$$

where

$$r_n(\mathbf{x}, \mathbf{h}) = \int_0^1 (1-t)\mathbf{h}^{\mathrm{T}} \Big[\nabla^2\phi_{\mathsf{B}}(\mathbf{x} + t\mathbf{h}) - \nabla^2\phi_{\mathsf{B}}(\mathbf{x}) \Big] \mathbf{h}\, \mathrm{d}t.$$

By the α-Hölder continuity of $\nabla^2\phi_{\mathsf{B}}$ in Assumption 3.1, there exists a constant $L < \infty$ such that

$$\Big\| \nabla^2\phi_{\mathsf{B}}(\mathbf{x} + t\mathbf{h}) - \nabla^2\phi_{\mathsf{B}}(\mathbf{x}) \Big\| \le L \cdot t^{\alpha}\, \|\mathbf{h}\|^{\alpha}, \qquad \forall\, t \in [0,1].$$

Therefore,

$$\begin{aligned}\left|r_n(\mathbf{x},\mathbf{h})\right| &\le \int_0^1 (1-t)\,\|\mathbf{h}\|^2\,\left\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}+t\mathbf{h})-\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\right\|\,\mathrm{d}t \\ &\le L\,\|\mathbf{h}\|^{2+\alpha}\int_0^1 (1-t)\,t^{\alpha}\,\mathrm{d}t \le C_2\,n^{-(2+\alpha)}\end{aligned} \tag{19}$$

for some constant $C_2 < \infty$.

Substitute (18) into (17). Note that $\mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})}\Big] = \phi_{\mathsf{B}}(\mathbf{x})$ by Lemma 3.2. Moreover, Assumption 3.2 gives $\mathbb{E}\big[\mathbf{Y}_{\mathsf{B}}(\mathbf{x})\big] = \mathbf{x}$ and hence $\mathbb{E}[\mathbf{h}] = 0$. Finally,

$$(n-1)^2\cdot\mathbb{E}\Big[\mathbf{h}^{\mathrm{T}}\,\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,\mathbf{h}\Big] = \mathbb{E}\Big[(\mathbf{x}-\mathbf{Y}_{\mathsf{B}}(\mathbf{x}))^{\mathrm{T}}\,\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,(\mathbf{x}-\mathbf{Y}_{\mathsf{B}}(\mathbf{x}))\Big] = \mathrm{tr}\Big(\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,\Sigma(\mathbf{x})\Big) = -2\Gamma(\mathbf{x}).$$

Combining these identities with (18) and (19) yields

$$\delta_n^{\mathrm{F}}(\mathbf{x}) = n\phi_{\mathsf{B}}(\mathbf{x}) - \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})} + (n-1)\,\phi_{\mathsf{B}}\,(\mathbf{x}+\mathbf{h})\Big] = \frac{\Gamma(\mathbf{x})}{n-1} + O\Big(n^{-1-\alpha}\Big).$$

The uniform bound on Γ in Lemma B.3 therefore implies that $|\delta_n^{\mathrm{F}}(\mathbf{x})| \le C/n$ for some uniform constant $C < \infty$ and all sufficiently large n. □

Lemma B.6 (Taylor Expansion). *Suppose that function $\phi : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable (i.e., $\phi \in C^2$). Then, for any $x, h \in \mathbb{R}^d$,*

$$\phi(x+h) = \phi(x) + \nabla\phi(x)^{\mathrm{T}}h + \frac{1}{2}h^{\mathrm{T}}\nabla^2\phi(x)h + \int_0^1 (1-t)h^{\mathrm{T}}\Big[\nabla^2\phi(x+th) - \nabla^2\phi(x)\Big]h\,\mathrm{d}t.$$

Proof. Define the one-dimensional function $g(t) = \phi(x+th)$ for $t \in [0,1]$. By the chain rule,

$$g'(t) = \nabla\phi(x+th)^{\mathrm{T}}h, \qquad g''(t) = h^{\mathrm{T}}\nabla^2\phi(x+th)h.$$

Applying the fundamental theorem of calculus twice gives

$$\begin{aligned}g(1) - g(0) = \int_0^1 g'(s)\,\mathrm{d}s &= g'(0) + \int_0^1 \big(g'(s) - g'(0)\big)\,\mathrm{d}s \\ &= g'(0) + \int_0^1\int_0^s g''(t)\,\mathrm{d}t\,\mathrm{d}s \\ &= g'(0) + \int_0^1 (1-t)\,g''(t)\,\mathrm{d}t,\end{aligned}$$

where the last equality follows from changing the integration order. Therefore,

$$\phi(x+h) = \phi(x) + \nabla\phi(x)^{\mathrm{T}}h + \int_0^1 (1-t)h^{\mathrm{T}}\nabla^2\phi(x+th)h\,\mathrm{d}t. \tag{20}$$

Adding and subtracting $\nabla^2\phi(x)$ inside the integral and using $\int_0^1(1-t)\,\mathrm{d}t = \frac{1}{2}$ yields the desired result. □

B.3 A Performance Characterization Theorem

In this section, we establish a performance characterization theorem (Theorem B.7) for any policy π. The theorem shows that the performance gap between the value function of π and the corrected fluid reference $F_n(\mathbf{c}) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\mathsf{H}_{n-1} + O(1)$ is proportional to the policy's cumulative dual-price score loss along its trajectory up to a prescribed local exit. In what follows, we first define the cumulative dual-price score loss and then state the characterization theorem. The proof is deferred to Section B.4.

Fix a compact target set $\mathcal{K} \Subset \mathcal{U}$. Let r_0 be the radius specified in Appendix B.1 and $n_0 = n_0(r_0)$ the integer specified in Lemma B.2. Consider a feasible, non-anticipating policy π that selects actions with nonnegative rewards. Let $V_n^\pi(\mathbf{c})$ denote its value function with n periods remaining and remaining capacity $\mathbf{c} \in \mathbb{R}_+^m$. Fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. Throughout Section B.3, we index time forward: k denotes the number of decisions made since the beginning of the horizon, and $h_k \triangleq n - k$ denotes the number of periods remaining after the k-th decision.

Let $\mathbf{c}_0 = \mathbf{c}$ denote the initial inventory and $\mathbf{c}_k$ the remaining inventory after the k-th decision. In period $k \geq 1$, policy π observes the independent opportunity $\boldsymbol{\theta}_k = (R_{a,k}, \mathbf{B}_{a,k})_{a\in\mathcal{A}}$, chooses an action a_k^π, collects a reward $\hat{R}_k \triangleq R_{a_k^\pi,k} \geq 0$, and consumes a resource vector $\mathbf{Y}_k \triangleq \mathbf{B}_{a_k^\pi,k} \in \mathbb{R}_+^m$. The state variables evolve according to

$$\mathbf{c}_k = \mathbf{c}_{k-1} - \mathbf{Y}_k, \qquad \boldsymbol{\rho}_k = \mathbf{c}_k/h_k, \qquad \mathbf{x}_k = (\boldsymbol{\rho}_k)_{\mathsf{B}}.$$

In addition, let ξ be a policy randomization seed independent of the entire opportunity sequence. Define the information filtration by $\mathcal{G}_0 = \sigma(\xi)$ and $\mathcal{G}_k = \sigma(\xi, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)$ for $k \geq 1$. For a deterministic policy, ξ may be taken to be constant. Hence, $\mathbf{c}_k$, $\boldsymbol{\rho}_k$, and $\mathbf{x}_k$ are $\mathcal{G}_k$-measurable, whereas $\boldsymbol{\theta}_{k+1}$ is independent of $\mathcal{G}_k$.

Fix a time horizon $n \geq n_0 + 1$. Let $\bar{N} = n - n_0$ and define the prescribed stopping time by

$$\tau_\pi = \inf\big\{0 \leq k \leq \bar{N} : \|\mathbf{x}_k - \mathbf{x}_0\| \geq r_0 \text{ or } h_k \leq n_0\big\}.$$

The choice of n_0 ensures that the stopped binding state $\mathbf{x}_{k\wedge\tau_\pi}$ lies in $\bar{B}(\mathbf{x}_0, 2r_0)$.

For $k < \tau_\pi$, define the conditional dual-price score loss in period $k+1$ by

$$\ell_k^\pi \triangleq \mathbb{E}\Big[S_{a^{\mathrm{F}}(\mathbf{x}_k)}(\mathbf{x}_k) - S_{a_{k+1}^\pi}(\mathbf{x}_k) \mid \mathcal{G}_k\Big] \geq 0, \tag{21}$$

where S_a denotes the dual-price score defined in Assumption 3.3. Since the static fluid action a^{F} maximizes the dual-price score over all actions, we have $\ell_k^\pi \geq 0$. Define the expected cumulative score loss before stopping by

$$Q_\pi(n, \mathbf{c}) \triangleq \mathbb{E}\left[\sum_{k=0}^{\tau_\pi - 1} \ell_k^\pi\right]. \tag{22}$$

We now state the performance characterization theorem in Theorem B.7.

Theorem B.7 (Performance Characterization via Cumulative Score Loss). *Suppose Assumption 2.1 and Assumptions 3.1–3.3 hold. Fix a compact set $\mathcal{K} \Subset \mathcal{U}$ and the prescribed stopping rule above. There exist constants $C_U, C_L < \infty$ such that every feasible, non-anticipating policy π that selects an action with a nonnegative reward in each period satisfies*

$$\tfrac{1}{2}Q_\pi(n, \mathbf{c}) - C_U \leq F_n(\mathbf{c}) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\,\mathsf{H}_{n-1} - V_n^\pi(\mathbf{c}) \leq C_L\big(1 + Q_\pi(n, \mathbf{c})\big) \tag{23}$$

for all $n \geq n_0 + 1$ and $\mathbf{c}/n \in \mathcal{K}$.

We prove Theorem B.7 in Appendix B.4. In the proof, we use a telescoping argument along the trajectory of policy π to decompose the initial gap $V_n^\pi(\mathbf{c}) - F_n(\mathbf{c})$ into the residuals accumulated before stopping and the remaining gap at the stopping time. The latter is bounded by a constant multiple of the expected remaining horizon at stopping, $\mathbb{E}h_{\tau_\pi}$, which is in turn bounded by a constant multiple of the cumulative score loss Q_π. We further show that the accumulated residuals equal the logarithmic correction $\Gamma(\mathbf{c}_\mathsf{B}/n)\,\mathsf{H}_{n-1}$ plus the cumulative score loss Q_π up to an error term that is sublinear in Q_π. Combining these bounds yields the desired result.

B.4 Proof of Theorem B.7

B.4.1 Decomposing Performance Gap via Telescoping

We first decompose the gap $F_n(\mathbf{c}) - V_n^\pi(\mathbf{c})$ into the expected cumulative one-period residuals along the trajectory induced by policy π before stopping and the gap at stopping.

For any remaining horizon $h \geq 1$ and inventory vector $\mathbf{c} \geq \mathbf{0}$ such that $\mathbf{c}_\mathsf{B}/h \in \mathcal{U}_\mathsf{B}$, define

$$P_h(\mathbf{c}) \triangleq h\phi_\mathsf{B}(\mathbf{c}_\mathsf{B}/h).$$

Whenever $P_h(\mathbf{c})$ is defined,

$$0 \leq V_h^{\mathrm{OPT}}(\mathbf{c}) \leq P_h(\mathbf{c}) \leq h\bar{r}.$$

The second inequality follows because $P_h(\mathbf{c})$ is the fluid-relaxation upper bound when every resource $j \in \mathsf{S}$ is slack in the fluid relaxation (1).

Fix $n \geq n_0 + 1$ and write $\tau = \tau_\pi$ for brevity. Define the one-period residual before stopping by

$$\delta_k^\pi = P_{h_k}(\mathbf{c}_k) - \mathbb{E}_k\big[\hat{R}_{k+1} + P_{h_k-1}(\mathbf{c}_{k+1})\big], \qquad k < \tau,$$

where $\mathbb{E}_k \triangleq \mathbb{E}\big[\cdot \mid \mathcal{G}_k\big]$ denotes conditional expectation given $\mathcal{G}_k$. Summing these residuals and taking expectations yields

$$\mathbb{E}\sum_{k<\tau} \delta_k^\pi = P_n(\mathbf{c}) - \mathbb{E}P_{h_\tau}(\mathbf{c}_\tau) - \mathbb{E}\sum_{j=1}^{\tau} \hat{R}_j.$$

Since $V_n^\pi(\mathbf{c}) = \mathbb{E}\sum_{j=1}^{\tau} \hat{R}_j + \mathbb{E}V_{h_\tau}^\pi(\mathbf{c}_\tau)$ and $F_n(\mathbf{c}) = P_n(\mathbf{c})$, we obtain the decomposition

$$F_n(\mathbf{c}) - V_n^\pi(\mathbf{c}) = \mathbb{E}\sum_{k<\tau} \delta_k^\pi + \mathbb{E}\big[P_{h_\tau}(\mathbf{c}_\tau) - V_{h_\tau}^\pi(\mathbf{c}_\tau)\big]. \tag{24}$$

In Section B.4.5, we establish two-sided bounds on both terms to complete the proof. As preparation, we first develop the necessary lemmas in Sections B.4.2 to B.4.4.

B.4.2 Bounding Consumption Differences and One-Period Residuals via Score Loss

Let $i = a^{\mathrm{F}}(\mathbf{x}_k)$ be the static fluid action defined in (3), and let $a = a_{k+1}^\pi$ be the actual action taken by policy π. For $k < \tau$, define

$$\ell_k = \mathbb{E}_k\big[S_i(\mathbf{x}_k) - S_a(\mathbf{x}_k)\big] \geq 0, \qquad D_k = \mathbb{E}_k\big\|\mathbf{B}_{i,\mathsf{B}} - \mathbf{Y}_{k+1,\mathsf{B}}\big\|_1, \qquad q = \mathbb{E}\sum_{k<\tau} \ell_k, \tag{25}$$

and set $\ell_k = D_k = 0$ for $k \geq \tau$. Since i maximizes the dual-price score over all actions, $\ell_k \geq 0$.

Lemma B.8 bounds the consumption difference D_k in terms of the score loss ℓ_k and controls the deviation of the one-period residual δ_k^π from $\ell_k + \Gamma(\mathbf{x}_k)/(h_k - 1)$.

Lemma B.8. *There exists a constant $C < \infty$, uniform over feasible policies π, horizons $n \geq n_0 + 1$, and initial states $\boldsymbol{\rho}_0 \in \mathcal{K}$, such that*

$$D_k \leq C\ell_k^\theta, \qquad \theta = \frac{\beta_\star}{1+\beta_\star}, \qquad 0 \leq k \leq \bar{N}. \tag{26}$$

Moreover, the one-period residual satisfies

$$\left|\delta_k^\pi - \ell_k - \frac{\Gamma(\mathbf{x}_k)}{h_k - 1}\right| \leq \frac{CD_k}{h_k - 1} + C(h_k - 1)^{-1-\alpha}, \qquad k < \tau. \tag{27}$$

Proof of (26). Define $d_k \triangleq \|\mathbf{B}_{i,\mathsf{B}} - \mathbf{Y}_{k+1,\mathsf{B}}\|_1$ and $g_k \triangleq S_i(\mathbf{x}_k) - S_a(\mathbf{x}_k) \geq 0$. We have

$$d_k = d_k \mathbb{1}\{g_k \leq t d_k\} + d_k \mathbb{1}\{g_k > t d_k\} \leq d_k \mathbb{1}\{g_k \leq t d_k\} + \frac{g_k}{t}. \tag{28}$$

Let $\beta_\star$, $C_\star$, and $t_\star$ be the constants specified in Section B.1. For any $0 < t \leq t_\star$, we have

$$\mathbb{E}_k\Big[d_k \mathbb{1}\{g_k \leq t d_k\}\Big] \leq \mathbb{E}_k\left[\max_{b \in \mathcal{A}} \left\{d_{ib}^{\mathsf{B}}\, \mathbb{1}\{S_i(\mathbf{x}_k) - S_b(\mathbf{x}_k) \leq t d_{ib}^{\mathsf{B}}\}\right\}\right] \leq C_\star t^{\beta_\star},$$

where the second inequality follows from Assumption 3.3. Thus, taking conditional expectations $\mathbb{E}_k[\cdot]$ on both sides of (28) and noting that $D_k = \mathbb{E}_k d_k$ and $\ell_k = \mathbb{E}_k g_k$ yield

$$D_k \leq C_\star t^{\beta_\star} + \frac{\ell_k}{t}, \qquad 0 < t \leq t_\star. \tag{29}$$

Set $\ell_\star \triangleq t_\star^{1+\beta_\star}$ and $d_{\max} \triangleq |\mathsf{B}|\bar{b}$. We consider three cases. If $\ell_k = 0$, letting $t \to 0$ in (29) yields $D_k = 0$. If $0 < \ell_k \leq \ell_\star$, choosing $t = \ell_k^{\frac{1}{1+\beta_\star}}$ yields

$$D_k \leq C_\star \ell_k^{\frac{\beta_\star}{1+\beta_\star}} + \ell_k^{1-\frac{1}{1+\beta_\star}} = (C_\star + 1)\,\ell_k^\theta.$$

Finally, if $\ell_k > \ell_\star$, bounded consumption implies

$$D_k \leq d_{\max} \leq d_{\max} \ell_\star^{-\theta} \ell_k^\theta.$$

Thus, choosing $C = \max\{C_\star + 1, d_{\max}\ell_\star^{-\theta}\}$ covers all cases and yields $D_k \leq C\ell_k^\theta$ for all $k < \bar{N}$. □

Proof of (27). Fix a period $k < \tau$ and define $\mathbf{u}_i \triangleq \mathbf{x}_k - \mathbf{B}_{i,\mathsf{B}}$ and $\mathbf{u}_a \triangleq \mathbf{x}_k - \mathbf{Y}_{k+1,\mathsf{B}}$. The successor states under actions i and a are, respectively,

$$\mathbf{z}_i = \mathbf{x}_k + \frac{\mathbf{u}_i}{h_k - 1}, \qquad \mathbf{z}_a = \mathbf{x}_k + \frac{\mathbf{u}_a}{h_k - 1}.$$

Bounded consumption and the compactness of $\widetilde{\mathcal{K}}_{\mathsf{B}}$ imply that $\|\mathbf{u}_i\|, \|\mathbf{u}_a\| \leq K_0 \triangleq \overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}$, where $\overline{\rho}_{\mathsf{B}}$ is defined in (10).

Before stopping, we have $\mathbf{x}_k \in B(\mathbf{x}_0, r_0)$, and $\mathbf{z}_i, \mathbf{z}_a \in \bar{B}(\mathbf{x}_0, 2r_0)$ by part 3 of Lemma B.2. By convexity, the line segments joining $\mathbf{x}_k$ to $\mathbf{z}_i$ and $\mathbf{z}_a$ also lie in $\bar{B}(\mathbf{x}_0, 2r_0)$. Applying the Taylor

expansion used in (18) and (19) gives

$$(h_k-1)\phi_{\mathsf{B}}(\mathbf{z}_a) = (h_k-1)\phi_{\mathsf{B}}\left(\mathbf{x}_k + \frac{\mathbf{u}_a}{h_k-1}\right) = (h_k-1)\phi_{\mathsf{B}}(\mathbf{x}_k) + \boldsymbol{\lambda}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{u}_a + \frac{\mathbf{u}_a^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\mathbf{u}_a}{2(h_k-1)} + r_k,$$
$$|r_k| \le C(h_k-1)\left\|\frac{\mathbf{u}_a}{h_k-1}\right\|^{2+\alpha} \le C(h_k-1)^{-1-\alpha}.$$

Conditional on $\mathcal{G}_k$, the state $\mathbf{x}_k$ is fixed, and the next opportunity follows the original distribution Φ. Hence,

$$\mathbb{E}_k[S_i(\mathbf{x}_k)] = \mathbb{E}_k\big[R_i - \boldsymbol{\lambda}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{B}_{i,\mathsf{B}}\big] = \phi_{\mathsf{B}}(\mathbf{x}_k) - \boldsymbol{\lambda}(\mathbf{x}_k)^{\mathrm{T}}\mathbf{x}_k,$$

where the second equality follows from Assumption 3.2 and Lemma 3.2. Consequently,

$$\ell_k = \mathbb{E}_k[S_i(\mathbf{x}_k)] - \mathbb{E}_k[S_a(\mathbf{x}_k)] = \phi_{\mathsf{B}}(\mathbf{x}_k) - \mathbb{E}_k\hat{R}_{k+1} - \boldsymbol{\lambda}(\mathbf{x}_k)^{\mathrm{T}}\mathbb{E}_k\mathbf{u}_a.$$

Combining this identity with the Taylor expansion above, the definition of δ_k^π, and the fact that $P_{h_k}(\mathbf{c}_k) = h_k\phi_{\mathsf{B}}(\mathbf{x}_k)$ and $P_{h_k-1}(\mathbf{c}_{k+1}) = (h_k-1)\phi_{\mathsf{B}}(\mathbf{z}_a)$ yields

$$\delta_k^\pi = \ell_k - \frac{\mathbb{E}_k[\mathbf{u}_a^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\mathbf{u}_a]}{2(h_k-1)} - \mathbb{E}_k r_k. \tag{30}$$

By the definition of Γ in (4), $\Gamma(\mathbf{x}_k) = -\frac{1}{2}\mathbb{E}_k\big[\mathbf{u}_i^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\mathbf{u}_i\big]$. Hence, replacing $\mathbf{u}_a$ with $\mathbf{u}_i$ in the quadratic term of (30) yields $\Gamma(\mathbf{x}_k)/(h_k-1)$. To bound the resulting error, since $\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)$ is symmetric and $\mathbf{u}_a - \mathbf{u}_i = \mathbf{B}_{i,\mathsf{B}} - \mathbf{Y}_{k+1,\mathsf{B}}$, we have

$$\begin{aligned}\left|\mathbf{u}_a^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\mathbf{u}_a - \mathbf{u}_i^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\mathbf{u}_i\right| &= \left|(\mathbf{u}_a-\mathbf{u}_i)^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)(\mathbf{u}_a+\mathbf{u}_i)\right| \\ &\le 2K_0\big\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\big\|\big\|\mathbf{B}_{i,\mathsf{B}} - \mathbf{Y}_{k+1,\mathsf{B}}\big\|_1.\end{aligned}$$

Taking conditional expectations and using (30) and the definition of D_k, we obtain

$$\left|\delta_k^\pi - \ell_k - \frac{\Gamma(\mathbf{x}_k)}{h_k-1}\right| \le \frac{K_0\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)\|D_k}{h_k-1} + |\mathbb{E}_k r_k|.$$

Since $\nabla^2\phi_{\mathsf{B}}(\mathbf{x}_k)$ is uniformly bounded on $\bar{B}(\mathbf{x}_0, 2r_0) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$ and $|\mathbb{E}_k r_k| \le C(h_k-1)^{-1-\alpha}$, this proves (27). □

B.4.3 Bounding Expected Remaining Horizon via Score Loss

We next bound the expected remaining horizon in terms of the score loss. In particular, we show that $\mathbb{E}h_\tau \le C(1+q)$ for some constant $C < \infty$, where q is defined in (25). We also establish score-loss bounds on the accumulation of the first term in (27) before stopping and on the γ-moment deviation $\mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^\gamma$. The latter bound will be used to control the error from replacing the path-dependent coefficients $\Gamma(\mathbf{x}_k)$ in (27) with the fixed coefficient $\Gamma(\mathbf{x}_0)$.

Define $L_k = \sum_{j<k\wedge\tau}\ell_j$ for $0 \le k \le \bar{N}$. Hence, $L_k \le L_\tau = \sum_{j<\tau}\ell_j$ and $q = \mathbb{E}L_\tau$. The following lemma summarizes these results.

Lemma B.9. *There exists a constant $C < \infty$, uniform over feasible policies π, horizons $n \ge n_0 + 1$, and initial states $\boldsymbol{\rho}_0 \in \mathcal{K}$, such that the following bounds hold. First, the weighted cumulative*

consumption discrepancies satisfy

$$\sum_{j<k\wedge\tau} \frac{D_j}{h_j - 1} \le C h_k^{-\theta} L_k^{\theta}, \qquad 0 \le k \le \bar{N}. \tag{31}$$

Second, letting $\gamma \triangleq \min\{\alpha, \beta_\star\} \in (0, 1]$ *as in Section B.1, the stopped inventory ratio satisfies*

$$\mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^{\gamma} \le C\left(h_k^{-\gamma/2} + q^{\theta\gamma} h_k^{-\theta\gamma}\right). \tag{32}$$

Finally, the expected remaining horizon at stopping satisfies

$$\mathbb{E}h_\tau \le C(1 + q). \tag{33}$$

Proof. We first decompose the trajectory $\{\mathbf{x}_k\}$ into predictable drift and martingale noise. Let $\mathbf{m}_j = \mathbb{E}_j[\mathbf{Y}_{j+1,\mathsf{B}}]$ denote the conditional mean consumption of binding resources under policy π. Then

$$\mathbf{x}_{j+1} - \mathbf{x}_j = \frac{\mathbf{x}_j - \mathbf{Y}_{j+1,\mathsf{B}}}{h_j - 1} = \frac{\mathbf{m}_j - \mathbf{Y}_{j+1,\mathsf{B}}}{h_j - 1} + \frac{\mathbf{x}_j - \mathbf{m}_j}{h_j - 1}.$$

The first term has conditional mean zero and the second is the predictable drift. Define

$$\mathbf{M}_k = \sum_{j=0}^{k-1} \mathbb{1}_{\{j<\tau\}} \frac{\mathbf{m}_j - \mathbf{Y}_{j+1,\mathsf{B}}}{h_j - 1},$$
$$\boldsymbol{A}_k = \sum_{j=0}^{k-1} \mathbb{1}_{\{j<\tau\}} \frac{\mathbf{x}_j - \mathbf{m}_j}{h_j - 1}.$$

Therefore, $\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0 = \mathbf{M}_k + \boldsymbol{A}_k$. The indicator in each increment is $\mathcal{G}_j$-measurable, so $\mathbf{M}_k$ is a martingale with $\mathbf{M}_0 = \mathbf{0}$. Define martingale difference

$$\Delta\mathbf{M}_{j+1} \triangleq \mathbf{M}_{j+1} - \mathbf{M}_j = \mathbb{1}_{\{j<\tau\}} \frac{\mathbf{m}_j - \mathbf{Y}_{j+1,\mathsf{B}}}{h_j - 1}.$$

Bounded consumption implies $\|\Delta\mathbf{M}_{j+1}\| \le C/(h_j - 1)$. Moreover, martingale differences at distinct times are orthogonal in expectation: for $i < j$, since $\Delta\mathbf{M}_{i+1}$ is $\mathcal{G}_j$-measurable,

$$\mathbb{E}\langle \Delta\mathbf{M}_{i+1}, \Delta\mathbf{M}_{j+1}\rangle = \mathbb{E}\langle \Delta\mathbf{M}_{i+1}, \mathbb{E}[\Delta\mathbf{M}_{j+1} \mid \mathcal{G}_j]\rangle = 0.$$

Hence,

$$\mathbb{E}\|\mathbf{M}_k\|^2 = \sum_{j=0}^{k-1} \mathbb{E}\|\Delta\mathbf{M}_{j+1}\|^2 \le C \sum_{s=h_k}^{n-1} s^{-2} \le C/h_k. \tag{34}$$

Proof of (31) From (26) we have

$$\sum_{j<k\wedge\tau} \frac{D_j}{h_j - 1} \le C \sum_{j<k\wedge\tau} \frac{\ell_j^{\theta}}{h_j - 1} \le C \left(\sum_{j<k\wedge\tau} \ell_j\right)^{\theta} \left(\sum_{s=h_k}^{n-1} s^{-1/(1-\theta)}\right)^{1-\theta} \le C h_k^{-\theta} L_k^{\theta},$$

where the second inequality follows from Hölder's inequality to ℓ_j^θ and $(h_j-1)^{-1}$ with conjugate exponents $1/\theta$ and $1/(1-\theta)$. For the third inequality, note that $1/(1-\theta)=1+\beta_\star$ and $\beta_\star(1-\theta)=\theta$. An integral comparison yields

$$\sum_{s=h_k}^{n-1} s^{-1/(1-\theta)} \le \int_{h_k-1}^{\infty} t^{-1-\beta_\star}\,\mathrm{d}t = \frac{(h_k-1)^{-\beta_\star}}{\beta_\star} \le Ch_k^{-\beta_\star},$$
$$\left(\sum_{s=h_k}^{n-1} s^{-1/(1-\theta)}\right)^{1-\theta} \le Ch_k^{-\beta_\star(1-\theta)} = Ch_k^{-\theta}.$$

Proof of (32) Since $\mathbf{x}_j = \mathbb{E}_j\big[\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}_j),\mathsf{B}}\big]$ for $j<\tau$, we have

$$\begin{aligned}\|\mathbf{x}_j-\mathbf{m}_j\| = \big\|\mathbb{E}_j\big(\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}_j),\mathsf{B}}-\mathbf{Y}_{j+1,\mathsf{B}}\big)\big\| &\le \mathbb{E}_j\|\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}_j),\mathsf{B}}-\mathbf{Y}_{j+1,\mathsf{B}}\| \\ &\le \mathbb{E}_j\|\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}_j),\mathsf{B}}-\mathbf{Y}_{j+1,\mathsf{B}}\|_1 = D_j, \qquad j<\tau,\end{aligned}$$

where the first inequality follows from the conditional Jensen's inequality. Therefore,

$$\|\boldsymbol{A}_k\| \le \sum_{j<k\wedge\tau} \frac{D_j}{h_j-1} \le Ch_k^{-\theta}L_\tau^\theta, \tag{35}$$

where the second inequality follows from (31) and $L_k\le L_\tau$. As a result,

$$\mathbb{E}\|\mathbf{x}_{k\wedge\tau}-\mathbf{x}_0\| \le \mathbb{E}\|\mathbf{M}_k\|+\mathbb{E}\|\mathbf{A}_k\| \le (\mathbb{E}\|\mathbf{M}_k\|^2)^{1/2} + Ch_k^{-\theta}\mathbb{E}\big(L_\tau^\theta\big) \le Ch_k^{-1/2} + Ch_k^{-\theta}(\mathbb{E}L_\tau)^\theta,$$

where the last inequality follows from (34) and Jensen's inequality. Finally, since $0<\gamma\le 1$, Jensen's inequality for the concave function t^γ, together with $(a+b)^\gamma\le a^\gamma+b^\gamma$, yields

$$\mathbb{E}\Big(\big\|\mathbf{x}_{k\wedge\tau}-\mathbf{x}_0\big\|^\gamma\Big) \le \Big(\mathbb{E}\big\|\mathbf{x}_{k\wedge\tau}-\mathbf{x}_0\big\|\Big)^\gamma \le Ch_k^{-\gamma/2} + Ch_k^{-\theta\gamma}q^{\theta\gamma}.$$

This proves (32).

Proof of (33) Since $n\ge n_0+1$, we have $n_0\le h_\tau<n$. Hence, $\mathbb{P}(h_\tau\ge r)=0$ for all $r\ge n$.

Fix an integer $u\in[n_0+1:n-1]$ and set $k=n-u<\bar{N}$. On $\{h_\tau\ge u\}$, we have $\tau\le k<\bar{N}$, so stopping must have been triggered by an exit from the binding-state local region $B(\mathbf{x}_0,r_0)$ rather than by reaching the cutoff $\bar{N}$. Consequently, $\mathbf{x}_{k\wedge\tau}=\mathbf{x}_\tau$, $\mathbf{M}_k=\mathbf{M}_\tau$, and

$$r_0 \le \|\mathbf{x}_{k\wedge\tau}-\mathbf{x}_0\| \le \|\mathbf{M}_k\|+\|\mathbf{A}_k\| \le \|\mathbf{M}_k\| + Cu^{-\theta}L_\tau^\theta,$$

where the last inequality follows from (35). Therefore, either $\|\mathbf{M}_k\|\ge r_0/2$ or $L_\tau\ge au$, where $a=(r_0/(2C))^{1/\theta}>0$. We next bound the first event by applying a concentration inequality to each martingale coordinate.

For each $i\in\mathsf{B}$, let $\Delta M_{j+1,i}=M_{j+1,i}-M_{j,i}$. Since $0\le Y_{j+1,i}\le\bar{b}$, we have

$$\Delta M_{j+1,i} \in \left[\frac{\mathbb{1}_{\{j<\tau\}}(m_{j,i}-\bar{b})}{h_j-1}, \frac{\mathbb{1}_{\{j<\tau\}}m_{j,i}}{h_j-1}\right].$$

Both endpoints are $\mathcal{G}_j$-measurable, and the interval width is at most $\bar{b}/(h_j-1)$. The sum of squared

width bounds through time $k = n - u$ satisfies

$$\sum_{j=0}^{k-1} \frac{\bar{b}^2}{(h_j - 1)^2} = \bar{b}^2 \sum_{s=u}^{n-1} s^{-2} \le \bar{b}^2 (u^{-2} + u^{-1}) \le \frac{2\bar{b}^2}{u}.$$

Apply the Azuma–Hoeffding inequality in Lemma B.10 yields, for any $z > 0$,

$$\mathbb{P}(M_{k,i} \ge z),\ \mathbb{P}(-M_{k,i} \ge z) \le \exp\left(-\frac{z^2 u}{\bar{b}^2}\right).$$

If $\|\mathbf{M}_k\| \ge r_0/2$, at least one of its $|\mathsf{B}|$ coordinates has absolute value at least $r_0/\left(2\sqrt{|\mathsf{B}|}\right)$. A union bound over the coordinates therefore gives

$$\mathbb{P}(\|\mathbf{M}_k\| \ge r_0/2) \le 2|\mathsf{B}| \exp\left(-\frac{r_0^2 u}{4|\mathsf{B}|\bar{b}^2}\right) \le Ce^{-cu}$$

for some constants $c, C < \infty$. Therefore,

$$\mathbb{P}(h_\tau \ge u) \le Ce^{-cu} + \mathbb{P}(L_\tau \ge au), \qquad u \ge n_0 + 1.$$

For any $v \ge 0$, at most v/a positive integers u satisfy $au \le v$. Hence,

$$\sum_{u=n_0+1}^{n} \mathbb{P}(L_\tau \ge au) = \sum_{u=n_0+1}^{n} \mathbb{E}\Big[\mathbf{1}_{\{au \le L_\tau\}}\Big] = \mathbb{E}\left[\sum_{u=n_0+1}^{n} \mathbf{1}_{\{au \le L_\tau\}}\right] \le \frac{\mathbb{E}L_\tau}{a} = \frac{q}{a}.$$

Since h_τ is integer-valued and $h_\tau \ge n_0$, we have

$$\mathbb{E}h_\tau = n_0 + \sum_{u=n_0+1}^{n} \mathbb{P}(h_\tau \ge u) \le n_0 + C \sum_{u=n_0+1}^{\infty} e^{-cu} + a^{-1}q \le C(1+q).$$

This proves (33). □

Lemma B.10 (Maximal Azuma-Hoeffding Inequality; Theorem 3.2.1 of Roch 2024)**.** *Let $(Z_t)_{t\in\mathbb{Z}_+}$ be a martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$. Suppose there exist predictable processes $(A_t)_{t\ge1}$ and $(B_t)_{t\ge1}$, with $A_t, B_t \in \mathcal{F}_{t-1}$, and constants $0 < c_t < \infty$ such that, for all $t \ge 1$, almost surely,*

$$A_t \le Z_t - Z_{t-1} \le B_t \qquad \text{and} \qquad B_t - A_t \le c_t.$$

Then, for every $\beta > 0$ and $t \ge 1$,

$$\mathbb{P}\left(Z_t - Z_0 \ge \beta\right) \le \mathbb{P}\left(\max_{0\le i\le t}(Z_i - Z_0) \ge \beta\right) \le \exp\left(-\frac{2\beta^2}{\sum_{i=1}^{t} c_i^2}\right).$$

B.4.4 Bounding the Effect of Coefficient Variation

We next use (32) to bound the cumulative effect of replacing the path-dependent coefficients $\Gamma(\mathbf{x}_k)$ with the fixed coefficient $\Gamma(\mathbf{x}_0)$.

Lemma B.11. *Let $\gamma \triangleq \min\{\alpha, \beta_\star\} \in (0,1]$ as defined in Section B.1. There exists a constant $C < \infty$, uniform over feasible policies π, horizons $n \geq n_0 + 1$, and initial states $\boldsymbol{\rho}_0 \in \mathcal{K}$, such that*

$$\left| \mathbb{E} \sum_{k<\tau} \frac{\Gamma(\mathbf{x}_k)}{h_k - 1} - \Gamma(\mathbf{x}_0)\big(\mathsf{H}_{n-1} - \mathsf{H}_{n_0-1}\big) \right| \leq C \left(n_0^{-\gamma/2} + q^{\theta\gamma} n_0^{-\theta\gamma} \right). \tag{36}$$

Proof. Let C_Γ and L_Γ be the upper bound and Hölder constant, respectively, in Lemma B.3, and set $L_* \triangleq \max\left\{L_\Gamma, C_\Gamma r_0^{-\gamma}\right\}$. Define

$$\widehat{\Gamma}_k \triangleq \begin{cases} \Gamma(\mathbf{x}_k), & k < \tau, \\ 0, & k \geq \tau. \end{cases}$$

We first show that

$$\left|\widehat{\Gamma}_k - \Gamma(\mathbf{x}_0)\right| \leq L_* \left\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\right\|^\gamma, \qquad k < \bar{N}. \tag{37}$$

Proof of (37). If $k < \tau$, then $\mathbf{x}_k \in B(\mathbf{x}_0, r_0)$. (37) follows directly from the Hölder continuity in Lemma B.3. If $\tau \leq k < \bar{N}$, stopping must occur through an exit, so $\|\mathbf{x}_\tau - \mathbf{x}_0\| \geq r_0$. Consequently, $\left|\widehat{\Gamma}_k - \Gamma(\mathbf{x}_0)\right| = \left|\Gamma(\mathbf{x}_0)\right| \leq C_\Gamma \leq L_* \left\|\mathbf{x}_\tau - \mathbf{x}_0\right\|^\gamma$, which proves (37). □

Since

$$\sum_{k=0}^{\bar{N}-1} \frac{1}{h_k - 1} = \sum_{s=n_0}^{n-1} \frac{1}{s} = \mathsf{H}_{n-1} - \mathsf{H}_{n_0-1},$$

we have

$$\begin{aligned} \left| \mathbb{E} \sum_{k<\tau} \frac{\Gamma(\mathbf{x}_k)}{h_k - 1} - \Gamma(\mathbf{x}_0)\big(\mathsf{H}_{n-1} - \mathsf{H}_{n_0-1}\big) \right| = \left| \mathbb{E} \sum_{k=0}^{\bar{N}-1} \frac{\widehat{\Gamma}_k - \Gamma(\mathbf{x}_0)}{h_k - 1} \right| &\leq \sum_{k=0}^{\bar{N}-1} \frac{\mathbb{E}\left|\widehat{\Gamma}_k - \Gamma(\mathbf{x}_0)\right|}{h_k - 1} \\ &\leq \sum_{k=0}^{\bar{N}-1} \frac{C\left(h_k^{-\gamma/2} + q^{\theta\gamma} h_k^{-\theta\gamma}\right)}{h_k - 1}, \end{aligned}$$

where the second inequality follows from (37) and (32). For either $a = \gamma/2$ or $a = \theta\gamma$, we have $a > 0$ and hence

$$\sum_{h=n_0+1}^{n} \frac{h^{-a}}{h-1} \leq 2 \sum_{h=n_0+1}^{\infty} h^{-1-a} \leq 2 \int_{n_0}^{\infty} t^{-1-a}\, \mathrm{d}t = \frac{2 n_0^{-a}}{a}.$$

Therefore,

$$\left| \mathbb{E} \sum_{k<\tau} \frac{\Gamma(\mathbf{x}_k)}{h_k - 1} - \Gamma(\mathbf{x}_0)\big(\mathsf{H}_{n-1} - \mathsf{H}_{n_0-1}\big) \right| \leq C \left(n_0^{-\gamma/2} + q^{\theta\gamma} n_0^{-\theta\gamma} \right)$$

for some constant $C < \infty$. □

B.4.5 Completing the Proof

Applying (31) at $k = \bar{N}$ and using Jensen's inequality, we obtain

$$\mathbb{E} \sum_{k<\tau} \frac{D_k}{h_k - 1} \leq C n_0^{-\theta} \mathbb{E} L_\tau^\theta \leq C n_0^{-\theta} q^\theta.$$

Combining this with (27) and $\sum_{s=n_0}^{\infty} s^{-1-\alpha} \le C n_0^{-\alpha}$ yields

$$\mathbb{E}\sum_{k<\tau}\left|\delta_k^\pi - \ell_k - \frac{\Gamma(\mathbf{x}_k)}{h_k - 1}\right| \le C n_0^{-\theta} q^\theta + C n_0^{-\alpha}.$$

Together with (24) and (36), this gives

$$F_n(\mathbf{c}) - \Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1} - V_n^\pi(\mathbf{c}) = q + \mathcal{T}_\pi + e_\pi, \qquad |e_\pi| \le C(1+q^\theta), \tag{38}$$

where $\mathcal{T}_\pi \triangleq \mathbb{E}\big[P_{h_\tau}(\mathbf{c}_\tau) - V_{h_\tau}^\pi(\mathbf{c}_\tau)\big]$ denotes the performance gap at stopping. Moreover,

$$0 \le \mathcal{T}_\pi \le \bar{r}\,\mathbb{E}h_\tau \le C(1+q),$$

where the last inequality follows from (33).

For the upper bound, $q^\theta \le 1+q$ and the upper bound on $\mathcal{T}_\pi$ yield the desired result. For the lower bound, since $\mathcal{T}_\pi \ge 0$ and

$$\frac{q}{2} + e_\pi \ge \frac{q}{2} - C(1+q^\theta) \ge -C_U$$

for some constant $C_U > 0$ uniformly over $q \ge 0$, we obtain the desired lower bound. This completes the proof of Theorem B.7.

C Proofs in Sections 4 and 5

C.1 Proof of Theorem 4.1

We first prove the performance expansion for the dynamic fluid policy π^{DF}. By Theorem B.7, it suffices to show that, for any compact set $\mathcal{K} \Subset \mathcal{U}$, the cumulative dual-price score loss $Q_{\pi^{\mathrm{DF}}}(n, \mathbf{c})$ is uniformly bounded over all horizons n and initial states satisfying $\mathbf{c}/n \in \mathcal{K}$.

Throughout Section C.1, we write $\tau = \tau_{\pi^{\mathrm{DF}}}$ for brevity, where $\tau_{\pi^{\mathrm{DF}}}$ is the stopping time defined in Section B.3. Let a_k denote the action taken by π^{DF} in period k.

In the proof, we introduce an auxiliary stopping time σ that will also monitor slack inventories. Before σ, all actions are feasible, and the dynamic fluid action selects $a_{k+1} = a^{\mathrm{F}}(\mathbf{x}_k)$, so its score loss ℓ_k is zero for all $k < \sigma$. Lemma C.1 bounds the cumulative score loss before τ in terms of that before σ and the expected remaining horizon $\mathbb{E}h_\sigma$.

Lemma C.1 (Cumulative Score Losses between Stopping Times). *Fix a feasible, non-anticipative policy π that selects actions with nonnegative rewards, and let $\sigma, \tau \le n$ be two stopping times. Then*

$$\mathbb{E}\sum_{k<\tau}\ell_k \le \mathbb{E}\sum_{k<\sigma}\ell_k + C\,\mathbb{E}h_\sigma$$

for some uniform constant $C < \infty$.

Proof. The result follows from $(\tau - \sigma)^+ \le n - \sigma = h_\sigma$ and the fact that ℓ_k is uniformly bounded: $0 \le \ell_k \le \bar{r} + \left(\sup_{\mathbf{x}\in\widetilde{\mathcal{K}}_{\mathsf{B}}} \|\nabla\phi_{\mathsf{B}}(\mathbf{x})\|\right)\sqrt{|\mathsf{B}|}\,\bar{b}$. □

C.1.1 Definition of the Auxiliary Stopping Time σ

Fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. Let r_0 be the radius specified in Appendix B.1 and $n_0 = n_0(r_0)$ the integer specified in Lemma B.2. Let $\mathcal{D}(\mathbf{x}_0)$ be the local region defined in (12)

using radius r_0. The initial state $\boldsymbol{\rho}_0$ belongs to $\mathcal{D}(\mathbf{x}_0)$ because its binding coordinates are $\mathbf{x}_0$ and

$$\rho_{0,s} - \mu_s(\mathbf{x}_0) \geq \underline{\delta}, \quad \forall\, s \in \mathsf{S}. \tag{39}$$

Fix a time horizon $n \geq n_0 + 1$ and let $\bar{N} = n - n_0$. Define the stopping time σ by

$$\sigma \triangleq \inf\big\{0 \leq k \leq \bar{N} : \boldsymbol{\rho}_k \notin \mathcal{D}(\mathbf{x}_0) \text{ or } h_k \leq n_0\big\}.$$

Hence, we stop on exit from $\mathcal{D}(\mathbf{x}_0)$ or when n_0 periods remain. On $\{k < \sigma\}$, we have $h_k \geq n_0+1$ and $\boldsymbol{\rho}_k \in \mathcal{D}(\mathbf{x}_0)$; therefore, Lemma B.2 implies that $a_{k+1} = a^{\mathrm{F}}(\mathbf{x}_k)$. Moreover, since $\boldsymbol{\theta}_{k+1}$ is independent of $\mathcal{G}_k$,

$$\mathbb{E}\big[\mathbf{Y}_{k+1,\mathsf{B}} \mid \mathcal{G}_k\big] = \mathbf{x}_k, \quad \mathbb{E}\big[Y_{k+1,s} \mid \mathcal{G}_k\big] = \mu_s(\mathbf{x}_k),\ s \in \mathsf{S}, \quad \mathbb{E}\big[\hat{R}_{k+1} \mid \mathcal{G}_k\big] = \phi_{\mathsf{B}}(\mathbf{x}_k). \tag{40}$$

Finally, the capacity recursion gives $h_{k+1} = h_k - 1$ and

$$\boldsymbol{\rho}_{k+1} = \frac{h_k \boldsymbol{\rho}_k - \mathbf{Y}_{k+1}}{h_k - 1} = \boldsymbol{\rho}_k + \frac{\boldsymbol{\rho}_k - \mathbf{Y}_{k+1}}{h_k - 1}, \qquad \mathbf{x}_{k+1} - \mathbf{x}_k = \frac{\mathbf{x}_k - \mathbf{Y}_{k+1,\mathsf{B}}}{h_k - 1}. \tag{41}$$

C.1.2 Bounding the Expected Remaining Horizon at Stopping

Let $h_\sigma \triangleq n - \sigma$ denote the remaining horizon at stopping. We show that h_σ has an exponentially decaying tail, which implies that $\mathbb{E}[h_\sigma] = O(1)$.

Lemma C.2 (Bound on the Expected Remaining Horizon). *There exist constants $C, c > 0$, depending only on $\mathcal{K}$ and the model primitives, such that, uniformly over $n \geq n_0 + 1$ and $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$,*

$$\mathbb{P}(h_\sigma \geq u) \leq Ce^{-cu}, \qquad u \geq n_0 + 1.$$

Consequently,

$$\sup_{n \geq n_0,\, \boldsymbol{\rho}_0 \in \mathcal{K}} \mathbb{E}[h_\sigma] \leq n_0 + C \sum_{u=n_0+1}^{\infty} e^{-cu} < \infty.$$

Proof. Let $\mathbf{m}_j = \mathbb{E}[\mathbf{Y}_{j+1} \mid \mathcal{G}_j]$. Define $\mathbf{M}_0 = \mathbf{0}$ and, for $1 \leq k \leq \bar{N}$,

$$\mathbf{M}_k = \sum_{j=0}^{k-1} \mathbb{1}\{j < \sigma\} \frac{\mathbf{m}_j - \mathbf{Y}_{j+1}}{h_j - 1}.$$

Since $\mathbb{1}\{j < \sigma\} \in \mathcal{G}_j$, $\{\mathbf{M}_k\}$ is a martingale. By the first identity in (40) and the binding-resource recursion in (41), we have

$$\mathbf{x}_{k\wedge\sigma} - \mathbf{x}_0 = \mathbf{M}_{k,\mathsf{B}}.$$

For each slack coordinate $s \in \mathsf{S}$, the second identity in (40) yields

$$m_{j,s} = \mu_s(\mathbf{x}_j) \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/4 < \rho_{j,s}, \quad j < \sigma,$$

where the first inequality follows from (9) and the second inequality from the definition of $\mathcal{D}(\mathbf{x}_0)$ in (12) and $j < \sigma$. Summing the slack-resource recursion in (41) therefore gives

$$\rho_{k\wedge\sigma,s} - \rho_{0,s} = \sum_{j<k} \mathbb{1}\{j < \sigma\} \frac{\rho_{j,s} - m_{j,s}}{h_j - 1} + M_{k,s} \geq M_{k,s}\,.$$

Together with (39), this implies that an exit through coordinate s requires $M_{\sigma,s} \le -\underline{\delta}/2$.

Fix an integer $u \ge n_0 + 1$ and set $k = n - u$. On $\{h_\sigma \ge u\}$, $\sigma \le k < \bar{N}$, so stopping must occur through an exit from $\mathcal{D}(\mathbf{x}_0)$ via either a binding or a slack coordinate. Since $\mathbf{M}_k = \mathbf{M}_\sigma$ for $k \ge \sigma$, we have

$$\big\{h_\sigma \ge u\big\} \subseteq \big\{\|\mathbf{M}_{k,\mathsf{B}}\| \ge r_0\big\} \cup \bigcup_{s \in \mathsf{S}} \big\{M_{k,s} \le -\underline{\delta}/2\big\}.$$

By bounded consumption, each coordinate increment $M_{j+1,i} - M_{j,i}$ lies in a predictable interval of width at most $\bar{b}/(h_j - 1)$, as in the proof of (33). The sum of the squared interval widths satisfies

$$\sum_{j=0}^{k-1} \frac{\bar{b}^2}{(h_j - 1)^2} = \bar{b}^2 \sum_{s=u}^{n-1} s^{-2} \le \bar{b}^2 \left(u^{-2} + \int_u^\infty t^{-2}\, \mathrm{d}t \right) \le 2\bar{b}^2/u.$$

Consequently, Lemma B.10 gives $\mathbb{P}(M_{k,i} \ge z), \mathbb{P}(-M_{k,i} \ge z) \le \exp\big(-z^2 u/\bar{b}^2\big)$ for $z > 0$. Moreover, $\|\mathbf{M}_{k,\mathsf{B}}\| \ge r_0$ implies $|M_{k,i}| \ge r_0/\sqrt{m}$ for some $i \in \mathsf{B}$. Applying a union bound over the binding and slack coordinates therefore gives

$$\mathbb{P}(h_\sigma \ge u) \le 2|\mathsf{B}| \exp\left(-\frac{r_0^2 u}{m\bar{b}^2}\right) + |\mathsf{S}| \exp\left(-\frac{\underline{\delta}^2 u}{4\bar{b}^2}\right) \le Ce^{-cu}$$

for some uniform constants $C, c > 0$, completing the proof. □

C.1.3 Completing the Proof of Theorem 4.1

By Theorem B.7, Lemmas C.1 and C.2, and the fact that $\mathbb{E}\sum_{k<\sigma} \ell_k = 0$, there exists a constant $C_{\mathrm{DF}} < \infty$, depending only on the model primitives and set $\mathcal{K}$, such that

$$\big|F_n(\mathbf{c}) - \Gamma(\mathbf{c}_\mathsf{B}/n)\, \mathsf{H}_{n-1} - V_n^{\mathrm{DF}}(\mathbf{c})\big| \le C_{\mathrm{DF}}$$

for all $n \ge 1$ and $\mathbf{c}/n \in \mathcal{K}$.

It remains to establish the corresponding bound for the optimal online policy. Specifically, we show that there exists a constant $C_{\mathrm{OPT}} < \infty$, depending only on the model primitives and set $\mathcal{K}$, such that

$$\big|F_n(\mathbf{c}) - \Gamma(\mathbf{c}_\mathsf{B}/n)\, \mathsf{H}_{n-1} - V_n^{\mathrm{OPT}}(\mathbf{c})\big| \le C_{\mathrm{OPT}}$$

for all $n \ge 1$ and $\mathbf{c}/n \in \mathcal{K}$. To see this, the first inequality in (23) holds for every policy π and yields

$$V_n^{\mathrm{OPT}}(\mathbf{c}) \le F_n(\mathbf{c}) - \Gamma(\mathbf{c}_\mathsf{B}/n)\, \mathsf{H}_{n-1} + C_U.$$

For the lower bound, we have

$$V_n^{\mathrm{OPT}}(\mathbf{c}) \ge V_n^{\mathrm{DF}}(\mathbf{c}) \ge F_n(\mathbf{c}) - \Gamma(\mathbf{c}_\mathsf{B}/n)\, \mathsf{H}_{n-1} - C_{\mathrm{DF}}.$$

Thus, choosing $C_{\mathrm{OPT}} = \max\{C_U, C_{\mathrm{DF}}\}$ completes the proof.

C.2 Proof of Theorem 5.1

In this section, we prove Theorem 5.1. According to Theorem B.7, it suffices to show that, for any compact set $\mathcal{K} \Subset \mathcal{U}$, the cumulative dual-price score loss $Q_{\pi^{\mathrm{GD}}}(n, \mathbf{c})$ is uniformly bounded over all horizons n and initial states satisfying $\mathbf{c}/n \in \mathcal{K}$. As in the proof of Theorem 4.1, we introduce an

auxiliary stopping time σ that also monitors slack inventories. By Lemma C.1,

$$Q_{\pi^{\mathrm{GD}}}(n, \mathbf{c}) \le \mathbb{E} \sum_{k<\sigma} \ell_k + C\, \mathbb{E} h_\sigma.$$

In what follows, we establish uniform bounds on both terms on the right-hand side, thereby completing the proof.

C.2.1 Notation and Definition of Local Neighborhoods

Let $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ denote the initial state, $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$, and let $\widetilde{\mathcal{K}}_{\mathsf{B}}$ be the compact set defined in Section B.1. Define the uniform Hessian bound

$$M_H = 1 \vee \sup_{\mathbf{x} \in \widetilde{\mathcal{K}}_{\mathsf{B}}} \big\| \nabla^2 \phi_{\mathsf{B}}(\mathbf{x}) \big\|.$$

By Assumption 5.1, there exist constants $L \ge 1$ and $\varepsilon > 0$ such that $\nabla \mathcal{V}$ is L-Lipschitz continuous on $B\big(\boldsymbol{\lambda}^*(\boldsymbol{\rho}_0), \varepsilon\big)$ for every $\boldsymbol{\rho}_0 \in \mathcal{K}$. Choose $\zeta > 0$ sufficiently small that

$$8\zeta < \varepsilon, \qquad 8\zeta / M_H \le r_0, \qquad 8L\zeta \le \underline{\delta}/8 \quad \text{if } \mathsf{S} \neq \varnothing, \tag{42}$$

where $\underline{\delta}$ is the slack margin defined in Appendix B.1.

In addition, choose $0 < r \le \min\big\{r_0, \zeta/(8M_H)\big\}$. Since $r \le r_0$, we have $\bar{B}(\mathbf{x}_0, 2r) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$. The uniform Hessian bound gives

$$\|\boldsymbol{\lambda}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \le \zeta/8 \quad \text{on } \bar{B}\big(\mathbf{x}_0, r\big), \qquad \|\boldsymbol{\lambda}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \le \zeta/4 \quad \text{on } \bar{B}\big(\mathbf{x}_0, 2r\big). \tag{43}$$

Finally, note that $-\mathbf{m}(\mathbf{p}) \in \partial \mathcal{V}(\mathbf{p})$ for every $\mathbf{p} \in \mathbb{R}^m$, with $\nabla \mathcal{V}(\mathbf{p}) = -\mathbf{m}(\mathbf{p})$ whenever $\mathcal{V}$ is differentiable. Let $\boldsymbol{\lambda}_0^* \triangleq \boldsymbol{\lambda}^*(\boldsymbol{\rho}_0) = (\boldsymbol{\lambda}(\mathbf{x}_0), \mathbf{0}_{\mathsf{S}})$. Since $m_s(\boldsymbol{\lambda}_0^*) = \mu_s(\mathbf{x}_0)$, Assumption 5.1 and (42) imply that, for every $\mathbf{p} \in B(\boldsymbol{\lambda}_0^*, 8\zeta)$ and $s \in \mathsf{S}$,

$$|m_s(\mathbf{p}) - \mu_s(\mathbf{x}_0)| \le L \|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \le \underline{\delta}/8. \tag{44}$$

C.2.2 Finite-Step Entry from Initial Price

We first show that, although the initial dual price $\boldsymbol{\lambda}_0 = \mathbf{0}$ may lie outside the region where the dual objective is smooth, global convexity of $\mathcal{V}(\boldsymbol{\lambda})$ ensures that the dual prices enter a suitable neighborhood of the optimal fluid price within a deterministic number k_0 of periods.

Define the entry-region radius $r_{\mathrm{ent}} \triangleq \zeta/16$ and set

$$R_\rho \triangleq \sup_{\boldsymbol{\rho}_0 \in \mathcal{K}} \|\boldsymbol{\rho}_0\|, \qquad R_\lambda \triangleq \sup_{\boldsymbol{\rho}_0 \in \mathcal{K}} \|\boldsymbol{\lambda}_0^*\|, \qquad G \triangleq R_\rho + 1 + \sqrt{m}\, \bar{b}.$$

As preparation, in Lemma C.3, we first prove a uniform linear lower bound on the growth of the dual objective $\mathcal{L}_{\boldsymbol{\rho}_0}$ away from its minimizer.

Lemma C.3. *There exists a constant $\Delta \in (0, G r_{\mathrm{ent}}]$ such that, for every $\boldsymbol{\rho}_0 \in \mathcal{K}$ and $\mathbf{p} \ge \mathbf{0}$ with $d = \|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \ge r_{\mathrm{ent}}$, we have*

$$\mathcal{L}_{\boldsymbol{\rho}_0}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*) \ge \frac{\Delta}{r_{\mathrm{ent}}} d.$$

Proof. Define

$$\Delta \triangleq \min_{\substack{\boldsymbol{\rho}_0 \in \mathcal{K},\ \mathbf{p} \geq \mathbf{0} \\ \|\mathbf{p}-\boldsymbol{\lambda}_0^*\| = r_{\mathrm{ent}}}} \left\{ \mathcal{L}_{\boldsymbol{\rho}_0}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*) \right\} > 0.$$

For any $\mathbf{p} \geq \mathbf{0}$ with $d = \|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \geq r_{\mathrm{ent}}$, define

$$\mathbf{q} \triangleq \boldsymbol{\lambda}_0^* + \frac{r_{\mathrm{ent}}}{d}(\mathbf{p} - \boldsymbol{\lambda}_0^*) = \left(1 - \frac{r_{\mathrm{ent}}}{d}\right) \boldsymbol{\lambda}_0^* + \frac{r_{\mathrm{ent}}}{d}\mathbf{p}.$$

Then $\mathbf{q} \geq \mathbf{0}$ and $\|\mathbf{q} - \boldsymbol{\lambda}_0^*\| = r_{\mathrm{ent}}$. By convexity of $\mathcal{L}_{\boldsymbol{\rho}_0}(\cdot)$,

$$\mathcal{L}_{\boldsymbol{\rho}_0}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*) \geq \frac{d}{r_{\mathrm{ent}}} \left\{ \mathcal{L}_{\boldsymbol{\rho}_0}(\mathbf{q}) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*) \right\} \geq \frac{\Delta}{r_{\mathrm{ent}}} d.$$

Finally, since $\boldsymbol{\rho}_0 - \mathbf{m}(\mathbf{p}) \in \partial L_{\boldsymbol{\rho}_0}(\mathbf{p})$, the subgradient inequality implies that, for any $\mathbf{p} \geq \mathbf{0}$ with $\|\mathbf{p} - \boldsymbol{\lambda}_0^*\| = r_{\mathrm{ent}}$,

$$\mathcal{L}_{\boldsymbol{\rho}_0}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*) \leq \left\{ \boldsymbol{\rho}_0 - \mathbf{m}(\mathbf{p}) \right\}^{\mathrm{T}} (\mathbf{p} - \boldsymbol{\lambda}_0^*) \leq G r_{\mathrm{ent}}. \tag{45}$$

Hence, $\Delta \leq G r_{\mathrm{ent}}$. □

Define

$$\bar{\eta}_{\mathcal{K}} \triangleq \min \left\{ (2L)^{-1}, \frac{\Delta}{4G^2} \right\} \tag{46}$$

and choose a step size $\eta \in (0, \bar{\eta}_{\mathcal{K}}]$. Lemma C.4 shows that the dual price $\boldsymbol{\lambda}_k$ enter a neighborhood of the optimal fluid price $\boldsymbol{\lambda}_k^* = (\boldsymbol{\lambda}(\mathbf{x}_k), \mathbf{0}_{\mathsf{S}})$ within a deterministic number k_0 of periods.

Lemma C.4 (Finite-Step Entry). *For every step size $\eta \in (0, \bar{\eta}_{\mathcal{K}}]$, there is an integer k_0, independent of the horizon n, such that, for all sufficiently large n, every sample path satisfies*

$$\|\boldsymbol{\lambda}_{k_0} - \boldsymbol{\lambda}_0^*\| \leq \zeta/8, \qquad \|\boldsymbol{\lambda}_{k_0} - (\boldsymbol{\lambda}(\mathbf{x}_{k_0}), \mathbf{0}_{\mathsf{S}})\| \leq \zeta/4.$$

Moreover, $\sup_{k \leq k_0} \|\boldsymbol{\rho}_k - \boldsymbol{\rho}_0\| \leq C/n$ *for some constant* $C < \infty$*, uniformly over all initial states* $\boldsymbol{\rho}_0$*.*

Proof. Fix η and let Δ be as in Lemma C.3. Define

$$k_0 = \left\lceil \frac{R_\lambda^2}{\eta \Delta} \right\rceil + 1.$$

For $n \geq 2k_0 + 1$, the capacity recursion and bounded consumption give

$$\boldsymbol{\rho}_k - \boldsymbol{\rho}_0 = \frac{k\boldsymbol{\rho}_0 - \sum_{j=1}^k \mathbf{Y}_j}{n-k}, \qquad \sup_{k \leq k_0} \|\boldsymbol{\rho}_k - \boldsymbol{\rho}_0\| \leq E_n \triangleq \frac{2k_0(R_\rho + \sqrt{m}\,\bar{b})}{n}. \tag{47}$$

Take n sufficiently large that $E_n \leq 1$. Let $\boldsymbol{g}_k \triangleq \boldsymbol{\rho}_k - \boldsymbol{m}(\boldsymbol{\lambda}_k)$. Then

$$\|\mathbf{g}_k\| \leq \|\boldsymbol{\rho}_0\| + \|\boldsymbol{\rho}_k - \boldsymbol{\rho}_0\| + \|\mathbf{m}(\boldsymbol{\lambda}_k)\| \leq G, \qquad k \leq k_0.$$

Furthermore, since $\boldsymbol{\lambda}_{k+1} = (\boldsymbol{\lambda}_k - \eta \mathbf{g}_k)^+$, the nonexpansiveness of projection gives

$$\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\| \leq \|(\boldsymbol{\lambda}_k - \eta \boldsymbol{g}_k) - \boldsymbol{\lambda}_k\| = \eta \|\boldsymbol{g}_k\| \leq \eta G, \qquad k \leq k_0. \tag{48}$$

Proving the First Bound Let $d_k \triangleq \|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_0^*\|$. For $k \le k_0$, we have

$$\begin{aligned} d_{k+1}^2 \le \|\boldsymbol{\lambda}_k - \eta \mathbf{g}_k - \boldsymbol{\lambda}_0^*\|^2 &= d_k^2 - 2\eta(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_0^*)^{\mathrm{T}}(\boldsymbol{\rho}_k - \boldsymbol{\rho}_0) - 2\eta(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_0^*)^{\mathrm{T}}(\boldsymbol{\rho}_0 - \mathbf{m}(\boldsymbol{\lambda}_k)) + \eta^2\|\mathbf{g}_k\|^2 \\ &\le d_k^2 + 2\eta d_k \|\boldsymbol{\rho}_k - \boldsymbol{\rho}_0\| - 2\eta\Big(\mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_k) - \mathcal{L}_{\boldsymbol{\rho}_0}(\boldsymbol{\lambda}_0^*)\Big) + \eta^2 G^2, \end{aligned} \tag{49}$$

where the second inequality follows from the subgradient inequality (the first inequality in (45) with $\mathbf{p} = \boldsymbol{\lambda}_k$) and $\|\mathbf{g}_k\| \le G$. Choose n sufficiently large that

$$E_n \le \min\left\{1, \frac{\Delta}{4r_{\mathrm{ent}}}, \frac{r}{4}\right\}. \tag{50}$$

We prove the recursive bound

$$d_{k+1}^2 \le \max\Big\{\frac{25}{16}r_{\mathrm{ent}}^2, d_k^2 - \eta\Delta\Big\}, \qquad k \le k_0. \tag{51}$$

Proof of (51). Suppose $d_k \ge r_{\mathrm{ent}}$. Then, Lemma C.3 and (49) imply

$$d_{k+1}^2 \le d_k^2 + 2\eta d_k E_n - 2\eta\frac{\Delta}{r_{\mathrm{ent}}}d_k + \eta^2 G^2 \le d_k^2 - \frac{3\eta\Delta}{2r_{\mathrm{ent}}}d_k + \frac{\eta\Delta}{4} \le d_k^2 - \eta\Delta,$$

where the second inequality follows from $E_n \le \Delta/(4r_{\mathrm{ent}})$ and $\eta \le \Delta/(4G^2)$, and the third inequality from $d_k \ge r_{\mathrm{ent}}$. Otherwise, $d_k < r_{\mathrm{ent}}$. By the triangle inequality and (48),

$$d_{k+1} \le d_k + \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\| \le d_k + \eta G < \frac{5}{4}r_{\mathrm{ent}},$$

where the third inequality follows from $\eta G \le r_{\mathrm{ent}}/4$, which holds because $\eta \le \Delta/(4G^2)$ and $\Delta \le G r_{\mathrm{ent}}$ by Lemma C.3. □

By (51) and induction, as well as the initial condition $d_0 = \|\boldsymbol{\lambda}_0^*\| \le R_\lambda$, we obtain

$$d_k^2 \le \max\Big\{\frac{25}{16}r_{\mathrm{ent}}^2, R_\lambda^2 - k\eta\Delta\Big\}, \qquad k \le k_0.$$

Since $k_0 > R_\lambda^2/(\eta\Delta)$, it follows that $d_{k_0} \le (5/4)r_{\mathrm{ent}} < 2r_{\mathrm{ent}} = \zeta/8$.

Proving the Second Bound Note that $\|\boldsymbol{x}_{k_0} - \boldsymbol{x}_0\| \le \|\boldsymbol{\rho}_{k_0} - \boldsymbol{\rho}_0\| \le E_n \le r/4$. Hence, the line segment joining $\mathbf{x}_0$ and $\mathbf{x}_{k_0}$ lies in $\bar{B}\big(\mathbf{x}_0, r/4\big) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$. Since $\boldsymbol{\lambda}(\boldsymbol{x}) = \nabla\phi_{\mathsf{B}}(\boldsymbol{x})$, letting $\boldsymbol{h} \triangleq \boldsymbol{x}_{k_0} - \boldsymbol{x}_0$, we obtain

$$\|\boldsymbol{\lambda}(\boldsymbol{x}_{k_0}) - \boldsymbol{\lambda}(\boldsymbol{x}_0)\| = \Big\|\int_0^1 \nabla^2\phi_{\mathsf{B}}(\boldsymbol{x}_0 + t\boldsymbol{h})\,\boldsymbol{h}\,dt\Big\| \le \int_0^1 M_H\|\boldsymbol{h}\|\,dt \le M_H E_n \le \zeta/32,$$

where the last inequality follows from $E_n \le r/4 \le \zeta/(32M_H)$. Consequently,

$$\|\boldsymbol{\lambda}_{k_0} - (\boldsymbol{\lambda}(\mathbf{x}_{k_0}), \mathbf{0}_{\mathsf{S}})\| \le \|\boldsymbol{\lambda}_{k_0} - \boldsymbol{\lambda}_0^*\| + \|\boldsymbol{\lambda}(\mathbf{x}_{k_0}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \le \zeta/8 + \zeta/32 \le \zeta/4. \qquad \square$$

C.2.3 Contraction at a Fixed Inventory State

We next establish a local contraction property of the projected gradient update toward the optimal fluid price when the inventory ratio is held fixed. This property, formalized in Lemma C.5, is used

in Section C.2.4 to prove the price-tracking result in Lemma C.6, which accounts for changes in the optimal fluid price as inventory evolves.

Lemma C.5 (Local Price Contraction at a Fixed Inventory State)**.** *Consider a normalized inventory* $\boldsymbol{\rho}$ *satisfying* $\mathbf{x} \triangleq \boldsymbol{\rho}_{\mathsf{B}} \in \bar{B}(\mathbf{x}_0, 2r) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$ *and* $\rho_s > \mu_s(\mathbf{x}_0) + \underline{\delta}/2$ *for all* $s \in \mathsf{S}$*. Let* $\boldsymbol{\lambda}^* \triangleq \boldsymbol{\lambda}^*(\boldsymbol{\rho}) = (\boldsymbol{\lambda}(\mathbf{x}), \mathbf{0}_{\mathsf{S}})$*. For every* $\mathbf{p} \geq \mathbf{0}$ *with* $\|\mathbf{p} - \boldsymbol{\lambda}^*\| \leq \zeta/2$ *and every step size* $\eta \in (0, \bar{\eta}_{\mathcal{K}}]$*, the projected gradient update* $\mathbf{p}^+ = \big(\mathbf{p} - \eta \nabla \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p})\big)^+$ *satisfies*

$$\|\mathbf{p}^+ - \boldsymbol{\lambda}^*\| \leq q \, \|\mathbf{p} - \boldsymbol{\lambda}^*\|, \qquad q = \big(1 + \eta/M_H\big)^{-1/2} < 1.$$

Proof. We first prove that, for every $\mathbf{p} \geq \mathbf{0}$ with $\|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \leq 4\zeta$,

$$\mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*) \geq \frac{1}{2M_H} \|\mathbf{p} - \boldsymbol{\lambda}^*\|^2. \tag{52}$$

Proof of (52). Define $\mathbf{d} = (\boldsymbol{\lambda}(\mathbf{x}) - \mathbf{p}_{\mathsf{B}})/M_H$, $\mathbf{x}_t = \mathbf{x} + t\mathbf{d}$ for $t \in [0,1]$, and $\mathbf{u} = \mathbf{x} + \mathbf{d}$. By (43),

$$\begin{aligned} M_H \|\mathbf{d}\| &= \|\boldsymbol{\lambda}(\mathbf{x}) - \mathbf{p}_{\mathsf{B}}\| \\ &\leq \|\boldsymbol{\lambda}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| + \|\boldsymbol{\lambda}(\mathbf{x}_0) - \mathbf{p}_{\mathsf{B}}\| \leq \zeta/4 + 4\zeta = 17\zeta/4, \\ \|\mathbf{x}_t - \mathbf{x}_0\| &\leq t\|\mathbf{d}\| + \|\mathbf{x} - \mathbf{x}_0\| \leq \frac{17\zeta}{4M_H} + 2r < \frac{8\zeta}{M_H} \leq r_0, \end{aligned}$$

where the last inequality follows from (42) and the choice of r. As a result, $\mathbf{x}_t \in \bar{B}\big(\mathbf{x}_0, r_0\big) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$ for all $t \in [0,1]$. Moreover,

$$\|\boldsymbol{\lambda}(\mathbf{u}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \leq \|\boldsymbol{\lambda}(\mathbf{u}) - \boldsymbol{\lambda}(\mathbf{x})\| + \|\boldsymbol{\lambda}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \leq M_H\|\mathbf{d}\| + \zeta/4 < 8\zeta,$$

where the second inequality follows from the uniform Hessian bound M_H and (43).

For $\mathbf{w} \in \{\mathbf{x}, \mathbf{u}\}$, (44) implies $\mu_s(\mathbf{w}) \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/8$. Hence, $\rho_s - \mu_s(\mathbf{w}) > 3\underline{\delta}/8$. Thus, $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$, and Lemma B.1 implies $\mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*) = \phi_{\mathsf{B}}(\mathbf{x})$. On the other hand, evaluating the maximum defining $\mathcal{V}(\mathbf{p})$ at $a^{\mathrm{F}}(\mathbf{u})$ and applying (11) yield

$$\mathcal{V}(\mathbf{p}) \geq \phi_{\mathsf{B}}(\mathbf{u}) - \mathbf{p}_{\mathsf{B}}^{\top}\mathbf{u} - \sum_{s \in \mathsf{S}} p_s \mu_s(\mathbf{u}).$$

As a result,

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*) &= \boldsymbol{\rho}^{\mathrm{T}}\mathbf{p} + \mathcal{V}(\mathbf{p}) - \phi_{\mathsf{B}}(\mathbf{x}) \\ &\geq \phi_{\mathsf{B}}(\mathbf{u}) - \phi_{\mathsf{B}}(\mathbf{x}) - \mathbf{p}_{\mathsf{B}}^{\mathrm{T}}\mathbf{d} + \sum_{s \in \mathsf{S}} p_s\big(\rho_s - \mu_s(\mathbf{u})\big) \\ &\geq (\boldsymbol{\lambda}(\mathbf{x}) - \mathbf{p}_{\mathsf{B}})^{\mathrm{T}}\mathbf{d} - \frac{M_H}{2}\|\mathbf{d}\|^2 + \frac{3\underline{\delta}}{8}\sum_{s \in \mathsf{S}} p_s \\ &= \frac{\|\mathbf{p}_{\mathsf{B}} - \boldsymbol{\lambda}_{\mathsf{B}}^*\|^2}{2M_H} + \frac{3\underline{\delta}}{8}\sum_{s \in \mathsf{S}} p_s \geq \frac{\|\mathbf{p} - \boldsymbol{\lambda}^*\|^2}{2M_H}. \end{aligned}$$

The second inequality follows from the Taylor expansion along the segment joining $\mathbf{x}$ and $\mathbf{u}$ (cf. (20)) and the uniform Hessian bound M_H. The second equality follows from the definition of $\mathbf{d}$. To justify the third inequality, note that $\boldsymbol{\lambda}_{0,\mathsf{S}}^* = \boldsymbol{\lambda}_{\mathsf{S}}^* = \mathbf{0}$ and $0 \leq p_s \leq 4\zeta$ by $\|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \leq 4\zeta$. Hence,

$3\underline{\delta} p_s/8 \geq p_s^2/(2M_H)$ for every $s \in \mathsf{S}$, because

$$\frac{p_s}{M_H} \leq \frac{4\zeta}{M_H} \leq \frac{\underline{\delta}}{16LM_H} \leq \frac{\underline{\delta}}{16} < \frac{3\underline{\delta}}{4},$$

where the second inequality follows from the last inequality in (42). □

By (43), $\|\mathbf{p} - \boldsymbol{\lambda}_0^*\| \leq \|\mathbf{p} - \boldsymbol{\lambda}^*\| + \|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}_0^*\| \leq 3\zeta/4$. Note that

$$\nabla \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*) = \boldsymbol{\rho} - \mathbf{m}(\boldsymbol{\lambda}^*) = (\mathbf{0}_{\mathsf{B}}, \boldsymbol{\rho}_{\mathsf{S}} - \boldsymbol{\mu}_{\mathsf{S}}(\mathbf{x})), \qquad \boldsymbol{\lambda}^* = \big(\boldsymbol{\lambda}^* - \eta \nabla \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*)\big)^+.$$

Since $\|\mathbf{p} - \boldsymbol{\lambda}^*\| \leq \zeta/2$, projection nonexpansiveness and Assumption 5.1 imply

$$\|\mathbf{p}^+ - \boldsymbol{\lambda}^*\| \leq \big\|\mathbf{p} - \boldsymbol{\lambda}^* - \eta\big(\nabla \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}) - \nabla \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*)\big)\big\| \leq (1 + \eta L)\, \|\mathbf{p} - \boldsymbol{\lambda}^*\| \leq \zeta.$$

Consequently, $\|\mathbf{p}^+ - \boldsymbol{\lambda}_0^*\| \leq 5\zeta/4 < 4\zeta$, so (52) applies at $\mathbf{p}^+$. Moreover, since both $\mathbf{p}$ and $\mathbf{p}^+$ lie in $B(\boldsymbol{\lambda}_0^*, 4\zeta)$, so does the line segment joining them.

Let $\mathbf{g} \triangleq \nabla \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p})$. The projected gradient update can be written as

$$\mathbf{p}^+ = \operatorname*{arg\,min}_{\mathbf{z} \geq \mathbf{0}} \frac{1}{2}\|\mathbf{z} - (\mathbf{p} - \eta \mathbf{g})\|^2.$$

Since $\boldsymbol{\lambda}^* \geq \mathbf{0}$ is feasible, the first-order optimality condition yields

$$\big(\mathbf{p}^+ - (\mathbf{p} - \eta \mathbf{g})\big)^{\mathrm{T}}(\boldsymbol{\lambda}^* - \mathbf{p}^+) \geq 0, \qquad \eta \mathbf{g}^{\mathrm{T}}(\mathbf{p}^+ - \boldsymbol{\lambda}^*) \leq (\mathbf{p} - \mathbf{p}^+)^{\mathrm{T}}(\mathbf{p}^+ - \boldsymbol{\lambda}^*).$$

The second inequality follows by expanding and rearranging the first. By convexity of $\mathcal{L}_{\boldsymbol{\rho}}(\cdot)$ and the L-Lipschitz continuity of its gradient along the segment joining $\mathbf{p}$ and $\mathbf{p}^+$, we have

$$\mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}) - \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*) \leq \mathbf{g}^{\mathrm{T}}(\mathbf{p} - \boldsymbol{\lambda}^*), \qquad \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}^+) - \mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}) \leq \mathbf{g}^{\mathrm{T}}(\mathbf{p}^+ - \mathbf{p}) + \frac{L}{2}\|\mathbf{p}^+ - \mathbf{p}\|^2.$$

Adding the two inequalities therefore yields

$$\begin{aligned} 2\eta\left\{\mathcal{L}_{\boldsymbol{\rho}}(\mathbf{p}^+) - \mathcal{L}_{\boldsymbol{\rho}}(\boldsymbol{\lambda}^*)\right\} &\leq 2(\mathbf{p} - \mathbf{p}^+)^{\mathrm{T}}(\mathbf{p}^+ - \boldsymbol{\lambda}^*) + \eta L \|\mathbf{p}^+ - \mathbf{p}\|^2 \\ &= \|\mathbf{p} - \boldsymbol{\lambda}^*\|^2 - \|\mathbf{p}^+ - \boldsymbol{\lambda}^*\|^2 - (1 - \eta L)\|\mathbf{p}^+ - \mathbf{p}\|^2, \end{aligned}$$

where the equality follows from the identity

$$\|\mathbf{p} - \boldsymbol{\lambda}^*\|^2 = \|(\mathbf{p} - \mathbf{p}^+) + (\mathbf{p}^+ - \boldsymbol{\lambda}^*)\|^2 = \|\mathbf{p} - \mathbf{p}^+\|^2 + \|\mathbf{p}^+ - \boldsymbol{\lambda}^*\|^2 + 2(\mathbf{p} - \mathbf{p}^+)^{\mathrm{T}}(\mathbf{p}^+ - \boldsymbol{\lambda}^*).$$

Since $\eta L \leq 1$, the last term is nonpositive and can be dropped from the upper bound. Applying (52) at $\mathbf{p}^+$ then gives $(1 + \eta/M_H)\|\mathbf{p}^+ - \boldsymbol{\lambda}^*\|^2 \leq \|\mathbf{p} - \boldsymbol{\lambda}^*\|^2$, which establishes the desired result. □

C.2.4 Tracking the Optimal Dual Price and Bounding Cumulative Score Loss

Define

$$M = \sup_{\mathbf{x} \in \widetilde{\mathcal{K}}_{\mathsf{B}}} \|\mathbf{x}\| + \sqrt{m}\,\bar{b}, \qquad C_1 = M_H M.$$

Choose an integer $n_0 \geq n_0(r)$ such that

$$\frac{C_1}{n_0} \leq \frac{(1-q)\zeta}{2}, \tag{53}$$

and an integer $N_0 \geq 2k_0 + n_0 + 1$ such that (50) holds for all $n \geq N_0$.

Fix $n \geq N_0$ and let $\bar{N} = n - n_0$. Consider the region $\mathcal{D}(\mathbf{x}_0)$ defined in (12), with the radius r chosen in Section C.2.1, and define the auxiliary stopping time

$$\sigma \triangleq \inf\big\{k_0 \leq k \leq \bar{N} : \boldsymbol{\rho}_k \notin \mathcal{D}(\mathbf{x}_0) \text{ or } h_k \leq n_0\big\}.$$

By (47) and (50), for every $0 \leq k \leq k_0$,

$$\|\mathbf{x}_k - \mathbf{x}_0\| \leq r/4, \qquad \rho_{k,s} \geq \mu_s(\mathbf{x}_0) + 7\underline{\delta}/8, \quad s \in \mathsf{S}. \tag{54}$$

Indeed, when $\mathsf{S} \neq \varnothing$, the definition of $\underline{\delta}$ gives $\rho_{0,s} \geq \mu_s(\mathbf{x}_0) + \underline{\delta}$, and the third inequality in (42) implies $E_n \leq r/4 \leq \underline{\delta}/(2048LM_H) \leq \underline{\delta}/8$. Thus, the initial phase remains in $\mathcal{D}(\mathbf{x}_0)$, implying $k_0 < \sigma \leq \bar{N}$. Moreover, for every $k < \sigma$, $\boldsymbol{\rho}_k \in \mathcal{D}(\mathbf{x}_0)$ and $h_k \geq n_0 + 1 \geq n_0(r) + 1$. By Lemma B.2, all actions are feasible before σ, and $\mathbf{x}_{k\wedge\sigma} \in \bar{B}\big(\mathbf{x}_0, 2r\big)$.

Let $\boldsymbol{\lambda}_k^* = (\boldsymbol{\lambda}(\mathbf{x}_k), \mathbf{0}_{\mathsf{S}})$ denote the optimal fluid price at state $\mathbf{x}_k$, and define the tracking error $e_k = \|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_k^*\|$. Lemma C.6 establishes a contraction recursion for the tracking error with a one-period perturbation of order $(h_k - 1)^{-1}$, which in turn implies a uniform bound on $\sum_{k=k_0}^{\sigma-1} e_k^2$. We defer its proof to the end of this subsection.

Lemma C.6 (Bound on Cumulative Squared Tracking Error). *For $k_0 \leq k \leq \sigma$, we have $\boldsymbol{\lambda}_k \in B(\boldsymbol{\lambda}_0^*, \zeta)$ and $e_k \triangleq \|\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_k^*\| \leq \zeta/2$. Moreover,*

$$e_{k+1} \leq q e_k + \frac{C_1}{h_k - 1}, \qquad k_0 \leq k < \sigma, \tag{55}$$

and the cumulative squared tracking error satisfies, on every sample path,

$$\sum_{k=k_0}^{\sigma-1} e_k^2 \leq C_E \triangleq \frac{(\zeta/4 + \sqrt{2}\, C_1)^2}{(1-q)^2},$$

where q is the constant specified in Lemma C.5.

We next use the cumulative squared tracking error bound to control the score loss before σ. For $k_0 \leq k < \sigma$, all actions are feasible, so (40) gives

$$\mathbb{E}\big[\hat{R}_{k+1} - \boldsymbol{\lambda}_k^{\mathrm{T}} \mathbf{Y}_{k+1} \mid \mathcal{G}_k\big] = \mathcal{V}(\boldsymbol{\lambda}_k), \qquad \mathbb{E}\big[\mathbf{Y}_{k+1} \mid \mathcal{G}_k\big] = \mathbf{m}(\boldsymbol{\lambda}_k).$$

Therefore,

$$\begin{aligned}
\ell_k &= \mathcal{V}(\boldsymbol{\lambda}_k^*) - \mathbb{E}\big[\hat{R}_{k+1} - (\boldsymbol{\lambda}_k^*)^{\mathrm{T}} \mathbf{Y}_{k+1} \mid \mathcal{G}_k\big] \\
&= \mathcal{V}(\boldsymbol{\lambda}_k^*) - \mathcal{V}(\boldsymbol{\lambda}_k) + (\boldsymbol{\lambda}_k^* - \boldsymbol{\lambda}_k)^{\mathrm{T}} \mathbf{m}(\boldsymbol{\lambda}_k) \\
&= \mathcal{V}(\boldsymbol{\lambda}_k^*) - \mathcal{V}(\boldsymbol{\lambda}_k) - \nabla\mathcal{V}(\boldsymbol{\lambda}_k)^{\mathrm{T}} (\boldsymbol{\lambda}_k^* - \boldsymbol{\lambda}_k) \leq \frac{L}{2}\, e_k^2,
\end{aligned}$$

where the third equality uses $\nabla\mathcal{V} = -\mathbf{m}$. To justify the inequality, note that $\boldsymbol{\lambda}_k^*, \boldsymbol{\lambda}_k \in B(\boldsymbol{\lambda}_0^*, \zeta)$ by (43) and Lemma C.6, hence so does the line segment joining them. The inequality follows from integrating the L-Lipschitz gradient along this segment.

For the initial phase $k < k_0$, we apply the uniform bound on ℓ_k:

$$0 \le \ell_k \le \bar{r} + M_\lambda \sqrt{m}\, \bar{b},$$

where M_λ is a uniform bound on $\|\boldsymbol{\lambda}(\mathbf{x})\|$ and may be taken as $R_\lambda + \zeta/8$. Combining the bounds for the two phases yields

$$L_\sigma \triangleq \sum_{k<\sigma} \ell_k \le k_0\big(\bar{r} + M_\lambda \sqrt{m}\, \bar{b}\big) + \frac{L}{2} C_E \triangleq C_Q. \tag{56}$$

Proof of Lemma C.6. We prove by induction that $\boldsymbol{\lambda}_k \in B(\boldsymbol{\lambda}_0^*, \zeta)$ and $e_k \le \zeta/2$ for $k_0 \le k \le \sigma$. These bounds hold at k_0 by Lemma C.4. Suppose they hold at some $k < \sigma$. By Lemma C.5,

$$\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k^*\| \le q e_k.$$

Moreover, as in the proof of part 3 of Lemma B.2, (14) and the definition of M yield

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le \frac{\|\mathbf{x}_k\| + \sqrt{m}\, \bar{b}}{h_k - 1} \le \frac{M}{h_k - 1}.$$

Thus, the triangle inequality and the uniform bound M_H on $\|\nabla \boldsymbol{\lambda}(\mathbf{x})\| = \|\nabla^2 \phi_{\mathsf{B}}(\mathbf{x})\|$ imply

$$e_{k+1} \le q e_k + \|\boldsymbol{\lambda}_{k+1}^* - \boldsymbol{\lambda}_k^*\| \le q e_k + \frac{M_H M}{h_k - 1} = q e_k + \frac{C_1}{h_k - 1}.$$

This proves (55). By (53) and (43),

$$e_{k+1} \le q\zeta/2 + C_1/n_0 \le \zeta/2, \qquad \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_0^*\| \le e_{k+1} + \|\boldsymbol{\lambda}(\mathbf{x}_{k+1}) - \boldsymbol{\lambda}(\mathbf{x}_0)\| \le \zeta/2 + \zeta/4 < \zeta.$$

This completes the induction, including at the stopping state.

To bound the cumulative squared tracking error, fix a sample path and collect the errors into the vector $\mathbf{e} = (e_{k_0}, \ldots, e_{\sigma-1})$. We separate the initial error, the preceding errors, and the one-period perturbations by defining

$$\begin{aligned} \mathbf{a} &= (e_{k_0}, 0, \ldots, 0), \\ \mathbf{b} &= (0, e_{k_0}, \ldots, e_{\sigma-2}), \\ \mathbf{d} &= \left(0, \frac{C_1}{h_{k_0} - 1}, \ldots, \frac{C_1}{h_{\sigma-2} - 1}\right). \end{aligned}$$

By (55), $\mathbf{0} \le \mathbf{e} \le \mathbf{a} + q\mathbf{b} + \mathbf{d}$. The triangle inequality thus gives

$$\begin{aligned} S \triangleq \|\mathbf{e}\| &\le \|\mathbf{a} + q\mathbf{b} + \mathbf{d}\| \le \|\mathbf{a}\| + q\|\mathbf{b}\| + \|\mathbf{d}\| \\ &= e_{k_0} + q\left(\sum_{k=k_0}^{\sigma-2} e_k^2\right)^{1/2} + C_1\left(\sum_{k=k_0}^{\sigma-2} (h_k - 1)^{-2}\right)^{1/2} \\ &\le \zeta/4 + qS + \sqrt{2}\, C_1. \end{aligned}$$

Rearranging yields $S^2 \le C_E$ on every sample path, completing the proof. □

C.2.5 Bounding the Expected Remaining Horizon and Completing the Proof

In this section, we show that the expected remaining horizon at stopping, $\mathbb{E}h_\sigma$, is uniformly bounded. Together with (56), Lemma C.1, and Theorem B.7, this bound completes the proof of Theorem 5.1.

To bound $\mathbb{E}h_\sigma$, we adapt the stopped-martingale argument of Lemma C.2, starting at k_0. Let $\mathbf{m}_j \triangleq \mathbb{E}[\mathbf{Y}_{j+1} \mid \mathcal{G}_j]$ and define

$$\mathbf{M}_k \triangleq \sum_{j=k_0}^{k-1} \mathbb{1}_{\{j<\sigma\}} \frac{\mathbf{m}_j - \mathbf{Y}_{j+1}}{h_j - 1}, \qquad k_0 \le k \le \bar{N}.$$

Then $\{\mathbf{M}_k\}$ is a martingale. The binding-resource recursion in (41) yields

$$\mathbf{x}_{k\wedge\sigma} - \mathbf{x}_0 = (\mathbf{x}_{k_0} - \mathbf{x}_0) + \mathbf{M}_{k,\mathsf{B}} + \sum_{j=k_0}^{k\wedge\sigma-1} \frac{\mathbf{x}_j - \mathbf{m}_{j,\mathsf{B}}}{h_j - 1}.$$

Thus,

$$\begin{aligned} \|\mathbf{x}_{k\wedge\sigma} - \mathbf{x}_0\| &\le \|\mathbf{x}_{k_0} - \mathbf{x}_0\| + \|\mathbf{M}_{k,\mathsf{B}}\| + \sum_{j=k_0}^{k\wedge\sigma-1} \frac{\|\mathbf{x}_j - \mathbf{m}_{j,\mathsf{B}}\|}{h_j - 1} \\ &\le r/4 + \|\mathbf{M}_{k,\mathsf{B}}\| + L\sqrt{C_E} \left(\sum_{j=k_0}^{k-1} (h_j - 1)^{-2} \right)^{1/2} \\ &\le r/4 + \|\mathbf{M}_{k,\mathsf{B}}\| + L\sqrt{2C_E/h_k}. \end{aligned} \tag{57}$$

For the second inequality, note that $\|\mathbf{x}_{k_0} - \mathbf{x}_0\| \le r/4$ by (54). Moreover, for $k_0 \le j < \sigma$, both $\boldsymbol{\lambda}_j^*$ and $\boldsymbol{\lambda}_j$ lie in $B(\boldsymbol{\lambda}_0^*, \zeta)$ by (43) and Lemma C.6. Hence,

$$\|\mathbf{x}_j - \mathbf{m}_{j,\mathsf{B}}\| = \|\mathbf{m}_{\mathsf{B}}(\boldsymbol{\lambda}_j^*) - \mathbf{m}_{\mathsf{B}}(\boldsymbol{\lambda}_j)\| \le \|\mathbf{m}(\boldsymbol{\lambda}_j^*) - \mathbf{m}(\boldsymbol{\lambda}_j)\| \le L\|\boldsymbol{\lambda}_j^* - \boldsymbol{\lambda}_j\| = Le_j,$$

where the second inequality follows from $\nabla\mathcal{V} = -\mathbf{m}$ and Assumption 5.1. Therefore,

$$\begin{aligned} \sum_{j=k_0}^{k\wedge\sigma-1} \frac{\|\mathbf{x}_j - \mathbf{m}_{j,\mathsf{B}}\|}{h_j - 1} \le L \sum_{j=k_0}^{k\wedge\sigma-1} \frac{e_j}{h_j - 1} &\le L \left(\sum_{j=k_0}^{k\wedge\sigma-1} e_j^2 \right)^{1/2} \left(\sum_{j=k_0}^{k\wedge\sigma-1} \frac{1}{(h_j - 1)^2} \right)^{1/2} \\ &\le L\sqrt{C_E} \left(\sum_{j=k_0}^{k-1} \frac{1}{(h_j - 1)^2} \right)^{1/2}, \end{aligned}$$

where the second inequality is Cauchy–Schwarz and the third inequality follows from Lemma C.6.

For each slack coordinate $s \in \mathsf{S}$, by (44),

$$m_{j,s} \le \mu_s(\mathbf{x}_0) + \underline{\delta}/8 < \mu_s(\mathbf{x}_0) + \underline{\delta}/2 < \rho_{j,s} \quad k_0 \le j < \sigma,$$

where the last inequality follows from the definition of $\mathcal{D}(\mathbf{x}_0)$ in (12) and $j < \sigma$. The slack-resource recursion in (41) therefore yields

$$\rho_{k\wedge\sigma,s} - \rho_{k_0,s} = \sum_{j=k_0}^{k\wedge\sigma-1} \frac{\rho_{j,s} - m_{j,s}}{h_j - 1} + M_{k,s} \ge M_{k,s}.$$

Since $\rho_{k_0,s} \geq \mu_s(\mathbf{x}_0) + 7\underline{\delta}/8$ by (54), an exit through coordinate s requires $M_{\sigma,s} \leq -3\underline{\delta}/8$.

Choose

$$u_* = \max\left\{n_0 + 1, \left\lceil \frac{128L^2C_E}{9r^2} \right\rceil\right\},$$

so that $L\sqrt{2C_E/u} \leq 3r/8$ for all $u \geq u_*$. Fix an integer $u \geq u_*$ and set $k = n - u$. On $\{h_\sigma \geq u\}$, we have $\sigma \leq k < \bar{N}$, so stopping must occur through an exit from $\mathcal{D}(\mathbf{x}_0)$ via either a binding or a slack coordinate. An exit through a binding coordinate implies $\|\mathbf{x}_\sigma - \mathbf{x}_0\| \geq r$, which, together with (57), requires $\|\mathbf{M}_{\sigma,\mathsf{B}}\| \geq 3r/8$. Since $\mathbf{M}_k = \mathbf{M}_\sigma$ for $k \geq \sigma$, we obtain

$$\big\{h_\sigma \geq u\big\} \subseteq \big\{\|\mathbf{M}_{k,\mathsf{B}}\| \geq 3r/8\big\} \cup \bigcup_{s \in \mathsf{S}} \big\{M_{k,s} \leq -3\underline{\delta}/8\big\}.$$

Applying the coordinatewise concentration bounds and a union bound as in the proof of Lemma C.2 gives $\mathbb{P}(h_\sigma \geq u) \leq Ce^{-cu}$ for some uniform constants $C, c > 0$. Consequently,

$$\mathbb{E}h_\sigma \leq u_* + C\sum_{u=u_*}^{\infty} e^{-cu} < \infty.$$