Dynamic Fluid Policy for Online Resource Allocation: Logarithmic Correction and Constant Optimality Gap

Chen Chen[1] and Jiawei Zhang[2]

[1]New York University Shanghai
[2]Stern School of Business, New York University
cc8029@nyu.edu, jz31@stern.nyu.edu

September 29, 2026

Abstract

We study an online resource allocation problem with finitely many actions, random rewards, and bounded multidimensional random resource consumptions. We analyze the classic dynamic fluid policy, which re-solves a fluid relaxation at every state and uses the resulting optimal dual variable to approximate the continuation value. We show that, under standard non-degeneracy conditions, the value functions of the dynamic fluid policy and the optimal dynamic policy characterized by the Bellman equation admit the same asymptotic expansion: the fluid benchmark minus a common logarithmic correction term plus a uniformly bounded, policy-dependent remainder. Moreover, we explicitly characterize the coefficient of the logarithmic term. Consequently, the logarithmic term is a common correction to the fluid benchmark, and the dynamic fluid policy achieves a horizon-uniform constant optimality gap relative to the optimal online policy.

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

1 Introduction

In this paper, we study an 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 given by 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 focus on 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 nondegeneracy conditions, the dynamic fluid policy and its variants typically incur a $\Theta(\ln n)$ 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 $\Omega(\log T)$.

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: fluid benchmark minus the same 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 $\Theta(\ln n)$ loss relative to the offline optimal benchmark established by Bray (2025). The logarithmic correction $\Gamma \log n$ comes from the interaction between random resource consumption and the concavity of the fluid value function.

The proof of Theorem 4.1 is illustrated in Section 6. We first motivate the logarithmic correction in Section 6.1 through a second-order expansion of the fluid value function. This expansion identifies the one-period loss induced by random consumption of the binding resources. The leading coefficient of this loss is Γ, and summing the leading terms yields the logarithmic correction $\Gamma \log n$. Then, in Sections 6.2 and 6.3, we compare the value functions of the dynamic fluid policy and the optimal online policy with the fluid benchmark and demonstrate that, within a local region containing the initial state, they both deviate from the fluid benchmark by the same the logarithmic correction $\Gamma \log n$ term, up to a constant. The two comparisons use different trajectory arguments.

For the dynamic fluid policy, we show that the normalized state (i.e., the amount of inventory per remaining period) evolves as a martingale within a local region containing the initial state. Using a telescoping argument along the trajectory induced by the dynamic fluid policy, we decompose the gap between the fluid benchmark and the policy's value function into the cumulative residuals up to the time when the normalized state exits the local region and the terminal gap at the time of exit. A martingale concentration inequality shows that the expected remaining horizon at the time of exit is bounded by a constant, and hence so is the terminal gap. Moreover, the accumulated residuals equal the correction term $\Gamma \log n$ up to a constant. Consequently, the gap between the fluid benchmark and the policy's value function equals $\Gamma \log n$ up to a constant.

Unlike under the dynamic fluid policy, the trajectory of the normalized binding resources need not be a martingale under the optimal online policy. Using a telescoping argument along the trajectory induced by the optimal online policy, we express the gap between the fluid benchmark and the optimal value function as the sum of the accumulated one-period residuals along the optimal trajectory and a nonnegative terminal term. Thus, to lower-bound this gap, it suffices to derive a lower bound on the accumulated one-period residuals. We obtain such a bound by comparing these residuals with the corresponding one-period residuals under the static fluid policy, which is optimal to the fluid relaxation. This comparison shows that the accumulated one-period residuals are bounded below by the correction term $\Gamma \log n$ minus a constant. Consequently, the gap between the fluid benchmark and the optimal value function is at least $\Gamma \log n$ up to a constant.

In Section 5, we apply our results to several examples. For the one-dimensional stochastic knapsack problem, Lueker (1998) derives the same performance expansion with an $o(\log n)$ remainder. As a direct implication of our result, we sharpen the remainder to $O(1)$ and extend the expansion to more general online resource allocation problems. For the online linear programming problem studied therein, Bray (2025) establishes a tight $\Theta(\ln n)$ performance bound relative to the offline optimum under standard non-degeneracy conditions. Our model is more general. Under the additional assumption that the second-order derivatives of the fluid relaxation value function are locally Hölder continuous, we show that the dynamic fluid policy and the optimal online policy admit the same performance expansion up to an $O(1)$ remainder. Thus, the performance gap between the two policies is bounded by a constant.

The rest of the paper is organized as follows. Section 1.1 reviews the related literature. Section 2 formulates the problem and introduces the dynamic fluid policy. Section 3 states our non-degeneracy conditions. Section 4 presents the main theorem and interprets the coefficient of the logarithmic correction term. Section 5 illustrates several applications. Section 6 presents the proof of the main theorem. Finally, Section 7 concludes the paper.

1.1 Related Literature

Online resource allocation has been extensively studied in operations research and computer 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 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 $O(\ln n \ln\ln n)$ 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 $O(\ln n)$ loss relative to the fluid relaxation, improving to $O(1)$ in locally affine cases, and Bray (2025) establish a tight $\Theta(\ln n)$ 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 5.3 for further discussion.

Comparison with Wang and Wang (2022) Our work is also closely related to Wang and Wang (2022), who study a classic dynamic pricing 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 $\Omega(\log T)$. 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.

Beyond studying different problems, the analytical approaches are also substantially different.

Wang and Wang (2022) directly compare the stochastic inventory trajectories under the re-solving and optimal online policies, using a stopping-time argument that couples the two trajectories. In contrast, we analyze each policy separately by telescoping its one-period losses relative to the fluid benchmark along the trajectory induced by that policy.

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 (2026a) consider settings in which both rewards and resource consumption are continuously distributed and allow both primal degeneracy and dual non-uniqueness. Zhang (2026b) 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 $[m] \triangleq \{1, 2, \dots, m-1, m\}$ 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 and the Dynamic Fluid Policy

In this section, we formulate our online resource allocation problem (Section 2.1), introduce its fluid relaxation (Section 2.2), and describe the classic dynamic fluid policy analyzed in the paper (Section 2.3).

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, indexed by $0, 1, \ldots, L$, where action 0 represents rejection. We let $\mathcal{A} = \{0, 1, \ldots, L\}$ denote the action set. 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 $a \in \mathcal{A}$ 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*

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

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

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

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 $\mathcal{H}_{t-1}$ 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:

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

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 let $\mathcal{P}$ denote the set of all opportunity-dependent randomized controls, $p : \Theta \to \Delta(\mathcal{A})$. The fluid relaxation is defined as

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

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 D.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).

Remark 2.1 (Strong Duality and Depleted Resources)**.** We remark that the use of strong duality requires Slater's condition, which in turn requires $\boldsymbol{\rho} > \mathbf{0}$. However, if $\rho_i = 0$ for some resource i, then resource i is depleted. In this case, the original resource allocation problem is equivalent to a reduced problem obtained by removing all depleted resources and prohibiting any action a that consumes them. Strong duality continues to hold and an optimal dual solution exists for this reduced problem. Although the corresponding dual variable has fewer than m components, the omitted components correspond to the depleted resources, and they do not affect either the definition of the dynamic fluid policy in Section 2.3 or its subsequent analysis. We can extend the reduced dual variable to the original problem by setting the omitted components to zero.

2.3 The Dynamic Fluid Policy

In this section, we describe the dynamic fluid policy. At each state $(n, \mathbf{c})$, where n is the number of remaining periods and $\mathbf{c}$ is the remaining capacity, the policy solves the fluid relaxation dual problem (2) using the normalized remaining capacity $\boldsymbol{\rho} = \mathbf{c}/n$. It then uses the resulting optimal dual variable $\boldsymbol{\lambda}^*(\boldsymbol{\rho})$ as the shadow price of resource consumption. Formally, at state $(n, \mathbf{c})$, the dynamic fluid policy selects the action

$$a_n^{\mathrm{F}}(\mathbf{c}) \triangleq \underset{a \in \mathcal{A}(\mathbf{c}, \boldsymbol{\theta})}{\arg\max} \left\{ R_a - \boldsymbol{\lambda}^*(\mathbf{c}/n)^{\mathrm{T}} \mathbf{B}_a \right\} \tag{3}$$

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

3 The Non-Degeneracy Conditions

In this section, we state the non-degeneracy conditions underlying our main result (i.e., Theorem 4.1 in Section 4) in Assumptions 3.1–3.3. These assumptions are mild and, as we demonstrate in Section 5, 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 admit 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.

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 prove Lemma 3.1 in Section D.2. Lemma 3.1 shows that the optimal dual variables associated with the resource constraints indexed by S are zero. Consequently, letting $\mathbf{B}_{a,\mathsf{B}} \triangleq (\mathbf{B}_a)_{\mathsf{B}}$, at each state $(n, \mathbf{c})$ with $\boldsymbol{\rho} = \mathbf{c}/n \in \mathcal{U}$, the dynamic fluid policy selects the action

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

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

$$a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}}) \triangleq \underset{a \in \mathcal{A}}{\arg\max} \left\{ R_a - \boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})^{\mathrm{T}} \mathbf{B}_{a,\mathsf{B}} \right\} \tag{4}$$

using the smallest-index tie-breaking rule. Unlike the dynamic fluid policy defined in Section 2.3, this policy keeps the dual variable $\boldsymbol{\lambda}(\boldsymbol{\rho}_{\mathsf{B}})$ fixed throughout the horizon, rather than recomputing it in each period based on the remaining resource inventories. Recall also that the fluid relaxation (1) allows resource inventories to become negative. 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 D.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

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

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

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

where $a^{\mathrm{F}}(\mathbf{x})$ denotes the action with the highest dual-price score, as defined in (4).

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 $d_{a^{\mathrm{F}}(\mathbf{x}),b}^{\mathsf{B}}$ 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.

To conclude this section, we note that Assumption 3.1, together with Assumption 3.2, requires that at least one resource constraint is binding in the fluid relaxation (i.e., $\mathsf{B} \neq \varnothing$). If, instead, all resource constraints are slack, then $\phi^{\mathrm{F}} = \mathbb{E}[\max_a R_a]$, and the optimal policy in the fluid relaxation simply selects the action with the highest reward in each period. Under mild conditions, the gap between this greedy policy in the original problem and the fluid relaxation upper bound converges to zero as the time horizon $n \to \infty$, as shown in Section D.4. We prove Proposition 3.3 in Section D.4.

Proposition 3.3 (All-Slack Case). *Suppose Assumption 2.1 holds. For a generic opportunity, let* $a^{\mathrm{G}} \triangleq \arg\max_{a \in \mathcal{A}} R_a$ *with a certain tie-breaking rule. Consider the greedy policy that selects* a^{G} *when it is feasible and selects action* 0 *otherwise. Let* $V_n^{\mathrm{G}}(\mathbf{c})$ *denote its expected reward. Define* $\mu_j \triangleq \mathbb{E}\big[B_{a^{\mathrm{G}},j}\big]$ *for* $j \in [m]$*. Given a sequence of initial inventories* $(\mathbf{c}_n)_{n \geq 1}$*, let* $\boldsymbol{\rho}_n = (\rho_{nj})_{j \in [m]} \triangleq \mathbf{c}_n / n$ *and define* $\delta_n \triangleq \min_{j \in [m]}\{\rho_{nj} - \mu_j\}$ *as the minimum slackness across all resource constraints. If* $\delta_n \geq \kappa\sqrt{\ln n / n}$*, where* $\kappa > 0$ *is a constant depending only on the problem primitives, then*

$$n\, \mathbb{E}\Big[\max_a R_a\Big] - V_n^{\mathrm{G}}(\mathbf{c}_n) = o(1),$$

where $o(1)$ *denotes a term that converges to zero as* $n \to \infty$*, uniformly over all initial inventories satisfying the above slack-resource condition.*

Therefore, the all-slack regime is relatively straightforward. In this paper, we focus on the more interesting setting in which at least one resource constraint is binding.

4 The Main Result

In this section, we first define the logarithmic correction coefficient and then present the common expansion for the dynamic fluid policy and the optimal online policy, together with their constant optimality gap, in Theorem 4.1.

For any normalized state $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$, let $a^{\mathrm{F}}(\mathbf{x})$ denote the static fluid action in (4) 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

$$\begin{aligned} \Gamma(\mathbf{x}) &\triangleq -\frac{1}{2}\, \mathbb{E}\Big[\big(\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) - \mathbf{x}\big)^{\mathrm{T}}\, \nabla^2\phi_{\mathsf{B}}(\mathbf{x})\, \big(\mathbf{Y}_{\mathsf{B}}(\mathbf{x}) - \mathbf{x}\big)\Big] \\ &= -\frac{1}{2}\, \mathrm{tr}\big(\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\, \Sigma(\mathbf{x})\big)\,. \end{aligned} \tag{5}$$

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 (5), $\Gamma(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$. We now state our main result.

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*

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

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 $\mathbf{c}/n \in \mathcal{K}$,

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

Theorem 4.1 compares the values of the dynamic fluid policy and the optimal online policy with the deterministic fluid benchmark. Since they share the same leading fluid term and logarithmic correction, these terms cancel when the two values are compared, leaving a uniformly bounded optimality gap. Thus, the logarithmic term represents a common second-order correction to the fluid benchmark. Its coefficient, Γ, captures the interaction between the curvature of the fluid value function in the binding-resource directions and the random consumption of binding resources under the static fluid policy in the fluid relaxation.

In Section 5, we apply Theorem 4.1 to three specific online resource allocation problems. We prove Theorem 4.1 in Section 6.

5 Applications

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

5.1 The Uniform Multisecretary Model

In the uniform multisecretary problem (Section 3 of Bray 2025), 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 $\Theta(\ln n)$ 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 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 $V_n^{\mathrm{F}}(c)$ 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. We provide further details in Section E.1.

5.2 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]

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

[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. We provide further details in Section E.2.

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 $V_n^{\mathrm{F}}(c)$ 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. We provide further details in Section E.2.

5.3 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 $(R_1, \mathbf{B}_1) = (u, \mathbf{a})$.

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. We verify this implication in Section E.3. 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.

6 Proof of Theorem 4.1

In this section, we prove Theorem 4.1. We first explain the origin of the logarithmic correction in Section 6.1. The proof then proceeds in two parts. First, by analyzing the trajectory induced by the dynamic fluid policy, we show that its performance equals the fluid benchmark minus a logarithmic correction, up to a uniformly bounded remainder (Section 6.2). Second, by analyzing the trajectory induced by the optimal online policy, we show that its performance cannot exceed the same expression by more than a constant (Section 6.3).

Throughout this section, we fix a compact target set $\mathcal{K} \Subset \mathcal{U}$. Unless otherwise stated, all constants may depend on $\mathcal{K}$ and the model primitives, but are independent of the horizon n and the initial state $\mathbf{c}$, provided that $\mathbf{c}/n \in \mathcal{K}$.

6.1 One-Period Residual and Corrected Fluid Reference

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 corrected fluid reference takes the form

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

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 (4), 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

$$\mathbf{x}' \triangleq \frac{n\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1} = \mathbf{x} + \frac{\mathbf{x} - \mathbf{Y}_{\mathsf{B}}(\mathbf{x})}{n-1}. \tag{6}$$

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:

$$\begin{aligned}\delta_n^{\mathrm{F}}(\mathbf{x}) &= -\nabla\phi_{\mathsf{B}}(\mathbf{x})^{\mathrm{T}}\,\mathbb{E}[\mathbf{x}-\mathbf{Y}_{\mathsf{B}}(\mathbf{x})] - \frac{1}{2(n-1)}\mathbb{E}\big[(\mathbf{x}-\mathbf{Y}_{\mathsf{B}}(\mathbf{x}))^{\mathrm{T}}\nabla^2\phi_{\mathsf{B}}(\mathbf{x})(\mathbf{x}-\mathbf{Y}_{\mathsf{B}}(\mathbf{x}))\big] + O(n^{-1-\alpha}) \\ &= \frac{\Gamma(\mathbf{x})}{n-1} + O(n^{-1-\alpha}). \end{aligned} \tag{7}$$

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 A.4, we show that the expansion holds uniformly on a compact set containing $\mathcal{K}_{\mathsf{B}}$.

Choose n_0 sufficiently large so that the (7) holds for every $n \geq n_0$. Keeping $\mathbf{x}$ fixed and summing the one-period residuals then yields the logarithmic correction:

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

This calculation, however, provides only a frozen-state heuristic. Along an actual trajectory under either the dynamic fluid policy or the optimal online 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 the dynamic fluid policy, we show that replacing the values of Γ along the trajectory by its initial value changes the accumulated residual by only a bounded amount. For the optimal online policy, we similarly show that deviations from the optimal actions do not substantially alter the accumulated residual.

6.2 Comparing the Dynamic Fluid Policy with the Corrected Fluid Reference

In this section, we compare the value function $V_n^{\mathrm{F}}(\mathbf{c})$ of the dynamic fluid policy with the corrected fluid reference.

Proposition 6.1 (Performance of the Dynamic Fluid Policy). *Suppose Assumption 2.1 and Assumptions 3.1–3.3 hold. For every compact set $\mathcal{K} \Subset \mathcal{U}$, there exists a constant $C_{\mathrm{F}} < \infty$ such that, for all $n \geq 1$ and $\mathbf{c}/n \in \mathcal{K}$,*

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

The proof is based on comparing the performance of the dynamic fluid policy with the fluid benchmark along the trajectory induced by the policy. We outline the main argument in Section 6.2.1 and provide the complete proof in Appendix B.

6.2.1 Proof Sketch of Proposition 6.1

The proof has three steps. We first specify a local region and define a stopping time as the first time the trajectory induced by the dynamic fluid policy exits this region. Before stopping, the trajectory of the normalized binding resources is a martingale, and the trajectory of the normalized slack resources can be decomposed into a positive drift and a martingale. We then use a martingale concentration inequality to show that the remaining horizon at the stopping time has an exponentially decaying probability tail, which in turn implies that its expectation is uniformly bounded. Finally, using a telescoping argument, we decompose the initial gap $F_n(\mathbf{c}) - V_n^{\mathrm{F}}(\mathbf{c})$ into the residuals accumulated before stopping and the remaining gap at the stopping time. The latter is uniformly bounded by a constant, and the accumulated residuals yield the correction term $\Gamma(\mathbf{c}_{\mathsf{B}}/n)\,\mathsf{H}_{n-1}$. This establishes Proposition 6.1.

Local Region, Trajectory, and Stopping Time Fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. For a sufficiently large integer n_0, consider a horizon $n \geq n_0 + 1$. We terminate the analysis when n_0 periods remain and let $\bar{N} \triangleq n - n_0$ denote the maximum number of decisions made before this cutoff. This ensures that, prior to the cutoff, the remaining horizon is sufficiently large so that, by (6), each one-period transition changes the normalized state by only a small amount. Throughout the proof, we index time forward: We let k denote the number of decisions made since the beginning of the horizon, and $h_k \triangleq n - k$ denote the number of periods remaining after the k-th decision.

We now describe the states, actions, and transitions for $1 \leq k \leq \bar{N}$. Let $\mathbf{c}_0 = \mathbf{c} \in \mathbb{R}_+^m$ denote the initial inventory and $\mathbf{c}_k$ the remaining inventory after the k-th decision. In period $k \geq 1$, the dynamic fluid policy observes the independent opportunity $\boldsymbol{\theta}_k = (R_{a,k}, \mathbf{B}_{a,k})_{a\in\mathcal{A}}$, chooses the action $a_k \triangleq a^{\mathrm{F}}_{h_k+1}(\mathbf{c}_{k-1})$ defined in (3), collects a reward $\hat{R}_k \triangleq R_{a_k,k}$, and consumes a resource vector $\mathbf{Y}_k \triangleq \mathbf{B}_{a_k,k} \in \mathbb{R}_+^m$. The state variables satisfy

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

In addition, let $\{\mathcal{G}_k\}_{k\geq 0}$ denote the filtration of information, where $\mathcal{G}_k = \sigma\big(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\big)$ for $k \geq 1$ and $\mathcal{G}_0 = \{\varnothing, \Omega\}$.

If $\mathsf{S} \neq \varnothing$, compactness of set $\mathcal{K}$, continuity of the consumption functions $\mu_s(\boldsymbol{\rho}_\mathsf{B})$ for $s \in \mathsf{S}$, and strict slackness in Assumption 3.2 imply the existence of a uniform positive slack margin on $\mathcal{K}$:

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

Choose $r_0 > 0$ sufficiently small so that, for every $s \in \mathsf{S}$, $|\mu_s(\mathbf{x}) - \mu_s(\mathbf{x}_0)| \leq \underline{\delta}/4$ whenever $\|\mathbf{x} - \mathbf{x}_0\| \leq 2r_0$. We define the local region

$$\mathcal{D}(\mathbf{x}_0) = \Big\{ \boldsymbol{\rho} : \boldsymbol{\rho}_\mathsf{B} \in B(\mathbf{x}_0, r_0),\ \rho_s > \mu_s(\mathbf{x}_0) + \underline{\delta}/2 \text{ for every } s \in \mathsf{S} \Big\}.$$

When $\mathsf{S} = \varnothing$, the slack-resource conditions in this definition are omitted.

Define the stopping time as the first time the trajectory exits the region $\mathcal{D}(\mathbf{x}_0)$ or the remaining horizon reaches n_0:

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

Let $h_\tau \triangleq n - \tau$ denote the remaining horizon at stopping. In Lemma B.1 we show that, before stopping, every action is feasible; hence the dynamic fluid policy uses $a^{\mathrm{F}}_{h_k}(\mathbf{c}_k) = a^{\mathrm{F}}(\mathbf{x}_k)$. We also show that every one-period successor state has its normalized binding component $\boldsymbol{\rho}_\mathsf{B}$ in $B(\mathbf{x}_0, 2r_0)$ and its normalized slack components remain slack, i.e., $\rho_s > \mu_s(\boldsymbol{\rho}_\mathsf{B})$ for all $s \in \mathsf{S}$.

Bounding the Expected Remaining Horizon at Stopping We prove that the remaining horizon h_τ at the stopping time has an exponentially decaying tail:

$$\mathbb{P}(h_\tau \geq r) \leq C e^{-cr}, \qquad r > n_0,$$

for some constants $C < \infty$ and $c > 0$. This in turn implies that $\mathbb{E}[h_\tau] = O(1)$ uniformly over all initial states $\boldsymbol{\rho}_0 \in \mathcal{K}$ and horizons $n \geq n_0 + 1$.

We establish this result using a martingale concentration argument. Define

$$\Delta \mathbf{M}_{k+1} \triangleq \mathbb{1}\{k < \tau\} \, \frac{\mathbb{E}[\mathbf{Y}_{k+1} \mid \mathcal{G}_k] - \mathbf{Y}_{k+1}}{h_k - 1},$$

$$\mathbf{M}_k \triangleq \sum_{\ell=1}^{k} \Delta \mathbf{M}_\ell.$$

Since $\mathbb{E}[\Delta \mathbf{M}_{k+1} \mid \mathcal{G}_k] = \mathbf{0}$, the process $\{\mathbf{M}_k\}$ is a martingale with $\mathbf{M}_0 = \mathbf{0}$. On the event $\{k < \tau\}$, we have $a_{k+1} = a^{\mathrm{F}}(\mathbf{x}_k)$; hence, by Assumption 3.2,

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

Consequently (see Section B.2.1 for details),

$$\Delta\mathbf{M}_{k+1,\mathsf{B}} = \mathbf{x}_{(k+1)\wedge\tau} - \mathbf{x}_{k\wedge\tau}, \qquad \mathbf{M}_{k,\mathsf{B}} = \mathbf{x}_{k\wedge\tau} - \mathbf{x}_0,$$
$$\rho_{k\wedge\tau,s} - \rho_{0,s} = \sum_{\ell=0}^{(k\wedge\tau)-1} \frac{\rho_{\ell,s} - \mu_s(\mathbf{x}_\ell)}{h_\ell - 1} + M_{k,s} \geq M_{k,s}, \quad \forall\, s \in \mathsf{S}.$$

Moreover, martingale differences are orthogonal in expectation, i.e., for $\ell \neq j$,

$$\mathbb{E}\left[\langle \Delta\mathbf{M}_{\ell+1,\mathsf{B}}, \Delta\mathbf{M}_{j+1,\mathsf{B}}\rangle\right] = 0.$$

It follows that

$$\mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^2 = \sum_{\ell=0}^{k-1} \mathbb{E}\|\Delta\mathbf{M}_{\ell+1,\mathsf{B}}\|^2 \leq C^2 \sum_{u=h_k+1}^{n} u^{-2} \leq C^2 \int_{h_k}^{\infty} u^{-2}\,\mathrm{d}u = \frac{C^2}{h_k}, \tag{8}$$

where the first inequality follows from bounded consumption, which gives $|\Delta M_{k+1,j}| = O(1/h_k)$ for every resource j. We will later use (8) to control the effect of replacing the path-dependent coefficients $\Gamma(\mathbf{x}_k)$ appearing in $\delta^{\mathrm{F}}_{h_k}$ with the fixed coefficient $\Gamma(\mathbf{x}_0)$.

Fix an integer $n_0 < r < n$. By the definition of the stopping time τ and the region $\mathcal{D}(\mathbf{x}_0)$, an exit by time $n - r < \bar{N}$ can occur only if either the binding projection leaves $B(\mathbf{x}_0, r_0)$ or some slack coordinate falls to or below $\mu_s(\mathbf{x}_0) + \underline{\delta}/2$. Applying a martingale concentration inequality and then taking a union bound over the resource coordinates yields the desired exponentially decaying tail bound. See Lemma B.2 for details.

Telescoping the Fluid Residual We next decompose the difference $F_n(\mathbf{c}) - V^{\mathrm{F}}_n(\mathbf{c})$ into the residuals accumulated before stopping and the remaining difference at the stopping time.

Let $D_h(\mathbf{c}') \triangleq F_h(\mathbf{c}') - V^{\mathrm{F}}_h(\mathbf{c}')$. On event $\{k < \tau\}$, $a^{\mathrm{F}}_{h_k}(\mathbf{c}_k) = a^{\mathrm{F}}(\mathbf{x}_k)$. Hence, the definition of $\delta^{\mathrm{F}}_{h_k}$ and the Bellman equation for the dynamic fluid policy imply

$$F_{h_k}(\mathbf{c}_k) = \mathbb{E}\big[\hat{R}_{k+1} + F_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big] + \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k),$$
$$V^{\mathrm{F}}_{h_k}(\mathbf{c}_k) = \mathbb{E}\big[\hat{R}_{k+1} + V^{\mathrm{F}}_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big].$$

Subtracting the second identity from the first cancels the immediate reward and yields

$$D_{h_k}(\mathbf{c}_k) = \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k) + \mathbb{E}\big[D_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big] \quad \text{on} \quad \{k < \tau\}.$$

Define $Z_k \triangleq D_{h_{k\wedge\tau}}(\mathbf{c}_{k\wedge\tau})$ for $0 \le k \le \bar{N}$. We have

$$\mathbb{E}\big[Z_{k+1} - Z_k \mid \mathcal{G}_k\big] = \begin{cases} -\delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k), & k < \tau, \\ 0, & k \ge \tau. \end{cases}$$

Taking expectations and summing over $0 \le k < \bar{N}$ yields

$$\mathbb{E} Z_{\bar{N}} - Z_0 = \mathbb{E} D_{h_\tau}(\mathbf{c}_\tau) - D_n(\mathbf{c}) = -\mathbb{E}\sum_{k<\tau} \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k).$$

Consequently,

$$F_n(\mathbf{c}) - V^{\mathrm{F}}_n(\mathbf{c}) = D_n(\mathbf{c}) = \mathbb{E}\sum_{k<\tau} \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k) + \mathbb{E} D_{h_\tau}(\mathbf{c}_\tau). \tag{9}$$

Thus, the initial gap decomposes into the residuals accumulated before stopping and the remaining gap at the stopping time. Note that $0 \le \mathbb{E} D_{h_\tau}(\mathbf{c}_\tau) \le \bar{r}\,\mathbb{E} h_\tau = O(1)$. Therefore, it remains to analyze the accumulated residuals, corresponding to the first term on the right-hand side of (9), which we do next.

Controlling the Changing Coefficient and Completing the Proof We now control the cumulative effect of replacing the path-dependent coefficients $\Gamma(\mathbf{x}_k)$ appearing in $\delta^{\mathrm{F}}_{h_k}$ with the fixed coefficient $\Gamma(\mathbf{x}_0)$. This bounds the first term in (9) and completes the proof.

Substitute (7) into the first term in (9). The Taylor remainders are summable because $\alpha > 0$. To handle the leading terms, Lemma A.2 establishes that $\Gamma(\mathbf{x})$ is locally γ-Hölder continuous for some $\gamma > 0$. Combining this Hölder continuity with (8) and Jensen's inequality yields

$$\mathbb{E}\sum_{k<\tau} \frac{|\Gamma(\mathbf{x}_k) - \Gamma(\mathbf{x}_0)|}{h_k - 1} \le C \sum_{h>n_0} h^{-1-\gamma/2} < \infty.$$

See Section B.3 for details. Thus, replacing the path-dependent coefficients $\Gamma(\mathbf{x}_k)$ by $\Gamma(\mathbf{x}_0)$ changes the expected total by only a bounded amount. Moreover,

$$\sum_{k<\tau} \frac{1}{h_k - 1} = \mathsf{H}_{n-1} - \mathsf{H}_{h_\tau - 1},$$

and $\mathsf{H}_{h_\tau - 1} \le h_\tau$. Therefore,

$$\mathbb{E} \sum_{k<\tau} \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k) = \mathbb{E} \sum_{k<\tau} \frac{\Gamma(\mathbf{x}_k)}{h_k - 1} + O(1) = \Gamma(\mathbf{x}_0)\, \mathbb{E} \sum_{k<\tau} \frac{1}{h_k - 1} + O(1) = \Gamma(\mathbf{x}_0)\, \mathsf{H}_{n-1} + O(1).$$

This proves Proposition 6.1 for $n \ge n_0 + 1$. Enlarging the constant C_{F} to cover the finitely many smaller horizons completes the proof.

6.3 Comparing the Optimal Online Policy with the Corrected Fluid Reference

We next show that the value function $V_n^{\mathrm{OPT}}(\mathbf{c})$ of the optimal online policy cannot exceed the corrected fluid reference by more than a constant.

Proposition 6.2 (Performance of the Optimal Online Policy)**.** *Suppose Assumption 2.1 and Assumptions 3.1–3.3 hold. For every compact set $\mathcal{K} \Subset \mathcal{U}$, there exists a constant $C_{\mathrm{OPT}} < \infty$ such that, for all $n \ge 1$ and $\mathbf{c}/n \in \mathcal{K}$,*

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

The proof is based on comparing the performance of the optimal online policy with the fluid benchmark along the trajectory induced by the optimal online policy. We outline the main argument in Section 6.3.1 and provide the complete proof in Appendix C.

6.3.1 Proof Sketch of Proposition 6.2

Unlike under the dynamic fluid policy, the trajectory of the normalized binding resources is no longer a martingale under the optimal online policy. In the proof, we first express the gap $F_n - V_n^{\mathrm{OPT}}$ as the sum of the accumulated one-period residuals along the optimal trajectory and a nonnegative terminal term. Thus, to lower-bound this gap, it suffices to derive a lower bound on the accumulated one-period residuals. We obtain such a bound by comparing these residuals with the corresponding one-period residuals under the static fluid actions a^{F}. Finally, we control the cumulative effect of replacing the path-dependent coefficients Γ along the optimal trajectory with their initial value. This implies that the accumulated one-period residuals are bounded below by the correction term $\Gamma(\mathbf{c}_{\mathsf{B}}/n)\, \mathsf{H}_{n-1}$ minus a constant, thereby establishing Proposition 6.2.

The Optimal Trajectory Fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. For a sufficiently large integer n_0, consider a horizon $n \ge n_0 + 1$ and set $\bar{N} = n - n_0$. Throughout the proof, 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.

We now define the states, actions, and transitions under the optimal online policy for $1 \le k \le \bar{N}$. Let $\bar{\mathbf{c}}_0 = \mathbf{c} \in \mathbb{R}^m_+$ denote the initial inventory and $\bar{\mathbf{c}}_k$ the remaining inventory after the k-th decision. In period $k \ge 1$, upon observing the independent opportunity $\boldsymbol{\theta}_k = (R_{a,k}, \mathbf{B}_{a,k})_{a \in \mathcal{A}}$, the optimal online policy chooses

$$\bar{a}_k = \underset{a \in \mathcal{A}(\bar{\mathbf{c}}_{k-1}, \boldsymbol{\theta}_k)}{\arg\max} \Big\{ R_{a,k} + V^{\mathrm{OPT}}_{h_k}\big(\bar{\mathbf{c}}_{k-1} - \mathbf{B}_{a,k}\big) \Big\},$$

collects a reward $\bar{R}_k = R_{\bar{a}_k,k}$, and consumes a resource vector $\bar{\mathbf{Y}}_k = \mathbf{B}_{\bar{a}_k,k}$. The state variables evolve according to

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

Choose a fixed radius $r_0 > 0$ as specified in Appendix A.1 and define the stopping time

$$\bar{\tau} = \inf \Big\{ 0 \le k \le \bar{N} : \|\bar{\mathbf{x}}_k - \mathbf{x}_0\| \ge r_0 \text{ or } h_k \le n_0 \Big\}.$$

The cutoff $\bar{N}$ ensures that every increment $\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\|$ is sufficiently small before stopping, such that all stopped states $\bar{\mathbf{x}}_{k \wedge \bar{\tau}}$ and the line segments between consecutive states before stopping lie in the closed ball $\bar{B}(\mathbf{x}_0, 2r_0)$. See Section C.2 for further details.

The Stopped Telescope For $h \ge 1$ and $\mathbf{c} \ge \mathbf{0}$ such that $\mathbf{x} = \mathbf{c}_{\mathsf{B}}/h \in \mathcal{U}_{\mathsf{B}}$, define $P_h(\mathbf{c}) = h\phi_{\mathsf{B}}(\mathbf{x})$. Whenever $P_h(\mathbf{c})$ is defined, we have $V^{\mathrm{OPT}}_h(\mathbf{c}) \le P_h(\mathbf{c})$, 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). At the initial state, $P_n(\mathbf{c}) = F_n(\mathbf{c})$.

For $0 \le k \le \bar{N}$, define

$$Z_k = P_{h_{k \wedge \bar{\tau}}}(\bar{\mathbf{c}}_{k \wedge \bar{\tau}}) - V^{\mathrm{OPT}}_{h_{k \wedge \bar{\tau}}}(\bar{\mathbf{c}}_{k \wedge \bar{\tau}}),$$

and define the conditional one-period difference along the optimal trajectory by

$$\delta^{\mathrm{OPT}}_k = P_{h_k}(\bar{\mathbf{c}}_k) - \mathbb{E}\Big[\bar{R}_{k+1} + P_{h_k - 1}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big], \quad k < \bar{\tau}, \tag{10}$$

with $\delta^{\mathrm{OPT}}_k = 0$ for $k \ge \bar{\tau}$. On the event $\{k < \bar{\tau}\}$, the definition of δ^{OPT}_k and the Bellman equation for the optimal policy imply

$$P_{h_k}(\bar{\mathbf{c}}_k) = \delta_k^{\mathrm{OPT}} + \mathbb{E}\Big[\bar{R}_{k+1} + P_{h_k-1}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big],$$
$$V_{h_k}^{\mathrm{OPT}}(\bar{\mathbf{c}}_k) = \mathbb{E}\Big[\bar{R}_{k+1} + V_{h_k-1}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big].$$

Subtracting the second identity from the first cancels the common reward term and yields

$$\mathbb{E}[Z_{k+1} - Z_k \mid \mathcal{G}_k] = -\delta_k^{\mathrm{OPT}}, \qquad 0 \le k < \bar{N}.$$

Taking expectations and telescoping over $0 \le k < \bar{N}$ yields

$$\mathbb{E}\sum_{k=0}^{\bar{N}-1} \delta_k^{\mathrm{OPT}} = \sum_{k=0}^{\bar{N}-1} \big(\mathbb{E}Z_k - \mathbb{E}Z_{k+1}\big) = Z_0 - \mathbb{E}Z_{\bar{N}}.$$

Since $Z_0 = F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c})$, $Z_{\bar{N}} = P_{h_{\bar{\tau}}}(\bar{\mathbf{c}}_{\bar{\tau}}) - V_{h_{\bar{\tau}}}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{\bar{\tau}}) \ge 0$, and $\delta_k^{\mathrm{OPT}} = 0$ for $k \ge \bar{\tau}$, the preceding equality gives

$$F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c}) = \mathbb{E}\sum_{k<\bar{\tau}} \delta_k^{\mathrm{OPT}} + \mathbb{E}\big[P_{h_{\bar{\tau}}}(\bar{\mathbf{c}}_{\bar{\tau}}) - V_{h_{\bar{\tau}}}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{\bar{\tau}})\big] \ge \mathbb{E}\sum_{k<\bar{\tau}} \delta_k^{\mathrm{OPT}}. \tag{11}$$

Therefore, by (11), to bound the gap $F_n - V_n^{\mathrm{OPT}}$ from below, it suffices to derive a lower bound on the accumulated one-period residuals.

A Lower Bound on the One-Period Residual δ_k^{OPT} In order to bound the accumulated one-period residuals in (11), we derive a lower bound on δ_k^{OPT} for $k < \bar{\tau}$. We do so by comparing δ_k^{OPT} with the fluid residual $\delta_{h_k}^{\mathrm{F}}$, whose behavior is characterized by (7).

Specifically, let $i = a^{\mathrm{F}}(\bar{\mathbf{x}}_k)$ denote the static fluid action and $a = \bar{a}_{k+1}$ the optimal feasible action. Define

$$g_{k+1} = S_i(\bar{\mathbf{x}}_k) - S_a(\bar{\mathbf{x}}_k) \ge 0, \qquad d_{k+1} = \|\mathbf{B}_{i,\mathsf{B}} - \mathbf{B}_{a,\mathsf{B}}\|_1,$$

where the dual-price score $S_{a'}(\mathbf{x})$ of an action a' is defined in Assumption 3.3. The quantity g_{k+1} is the realized score loss, and d_{k+1} is the corresponding difference in binding-resource consumption. Since i maximizes the dual-price score over all actions, $g_{k+1} \ge 0$. Before stopping (i.e., for $k < \bar{\tau}$), define

$$\ell_k = \mathbb{E}\big[g_{k+1} \mid \mathcal{G}_k\big], \qquad D_k = \mathbb{E}\big[d_{k+1} \mid \mathcal{G}_k\big], \qquad q = \mathbb{E}\sum_{k<\bar{\tau}} \ell_k, \tag{12}$$

and set $\ell_k = D_k = 0$ for $k \ge \bar{\tau}$.

By comparing the one-period residuals δ_k^{OPT} and $\delta_{h_k}^{\text{F}}$, we establish that, for $k < \bar{\tau}$,

$$\delta_k^{\text{OPT}} \geq \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \frac{1}{2}\ell_k - C_1\left(h_k^{-1-\alpha} + h_k^{-1-\beta_\star}\right) \tag{13}$$

for some constant $C_1 < \infty$, where $\beta_\star > 0$ is specified in Section A.1. Consequently, combining this bound with (11) and the definition of q in (12) yields

$$F_n(\mathbf{c}) - V_n^{\text{OPT}}(\mathbf{c}) \geq \mathbb{E}\sum_{k<\bar{\tau}} \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \frac{1}{2}q - C_2 \tag{14}$$

for some constant $C_2 < \infty$.

To prove (13), let $\mathbf{b} = \mathbf{B}_{i,\mathsf{B}}$ and $\mathbf{y} = \bar{\mathbf{Y}}_{k+1,\mathsf{B}}$ denote the binding-resource consumptions under actions $i = a^{\text{F}}(\bar{\mathbf{x}}_k)$ and $a = \bar{a}_{k+1}$, respectively. The corresponding successor states are

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

From the definitions of δ_k^{OPT} and $\delta_{h_k}^{\text{F}}$ we have:

$$\begin{aligned}
\delta_k^{\text{OPT}} - \delta_{h_k}^{\text{F}}(\bar{\mathbf{x}}_k) &= \mathbb{E}_k\big[R_i - R_a + (h_k - 1)\big\{\phi_{\mathsf{B}}(\mathbf{z}_i) - \phi_{\mathsf{B}}(\mathbf{z}_a)\big\}\big],\\
\ell_k &= \mathbb{E}_k\big[R_i - R_a + \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)^{\text{T}}(\mathbf{y} - \mathbf{b})\big],
\end{aligned}$$

where, for notational convenience, we let $\mathbb{E}_k \triangleq \mathbb{E}[\cdot \mid \mathcal{G}_k]$. Subtracting the second identity from the first and applying the fundamental theorem of calculus to ϕ_{B} along the line segment $\mathbf{z}_t = \mathbf{z}_a + t(\mathbf{z}_i - \mathbf{z}_a)$ with $t \in [0,1]$ yields

$$\delta_k^{\text{OPT}} - \delta_{h_k}^{\text{F}}(\bar{\mathbf{x}}_k) - \ell_k = \mathbb{E}_k \int_0^1 [\nabla\phi_{\mathsf{B}}(\mathbf{z}_t) - \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)]^{\text{T}}(\mathbf{y} - \mathbf{b})\,\mathrm{d}t.$$

Because the Hessian $\nabla^2\phi_{\mathsf{B}}$ is continuous and hence uniformly bounded on a compacted set by Assumption 3.1, we have

$$\left|\delta_k^{\text{OPT}} - \delta_{h_k}^{\text{F}}(\bar{\mathbf{x}}_k) - \ell_k\right| \leq \frac{C}{h_k - 1}\mathbb{E}_k\|\mathbf{y} - \mathbf{b}\|_1 = \frac{C}{h_k - 1}\,D_k$$

for some constant $C < \infty$. The preceding inequality together with Equation (7) implies

$$\delta_k^{\text{OPT}} \geq \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \ell_k - \frac{C}{h_k - 1}\,D_k - Ch_k^{-1-\alpha}$$

for some constant $C < \infty$. Finally, Assumption 3.3 allows us to bound the expected consumption difference D_k in terms of the expected score loss ℓ_k defined in (12), which yields (13). Further details are provided in Sections C.3 and C.4.

Controlling the Changing Coefficient and Completing the Proof As the final step, we control the cumulative effect of replacing the path-dependent coefficients $\Gamma(\bar{\mathbf{x}}_k)$ appearing in δ_k^{OPT} with the fixed coefficient $\Gamma(\mathbf{x}_0)$. To do so, as a preliminary step, we bound the γ-moment deviation of the stopped state $\bar{\mathbf{x}}_{k\wedge\bar{\tau}}$ from the initial state $\mathbf{x}_0$ (see Lemma C.3). Specifically, we show that, there exists a constant $C_3 < \infty$ such that, for any $k \leq \bar{N}$,

$$\mathbb{E}\Big(\big\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\big\|^\gamma\Big) \leq C_3\left(h_k^{-\gamma/2} + \big(q/h_k\big)^{p\gamma}\right),$$

where q is defined in (43) and $\gamma, p \in (0,1)$ are two constants specified in Section A.1 and Lemma C.1, respectively. Moreover, since $\Gamma(\mathbf{x})$ is locally γ-Hölder continuous by Lemma A.2, we have

$$\Gamma(\bar{\mathbf{x}}_k) \geq \Gamma(\mathbf{x}_0) - C\left\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\right\|^\gamma, \quad k < \bar{\tau},$$

for some constant $C < \infty$. Combining the preceding two inequalities, we establish that there exists a constant $C_4 < \infty$ such that

$$\mathbb{E}\sum_{k<\bar{\tau}} \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} \geq \Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1} - C_4\left(1 + q^{p\gamma}\right). \tag{15}$$

Further details are provided in Sections C.5 and C.6.

Combining (14) and (15), we obtain

$$F_n(\mathbf{c}) - V_n^{\text{OPT}}(\mathbf{c}) \geq \Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1} + \tfrac{1}{2}q - C\Big(1 + q^{p\gamma}\Big)$$

for some constant $C < \infty$. Since $0 < p\gamma < 1$, the linear term $q/2$ dominates the sublinear term $C\big(1 + q^{p\gamma}\big)$ for large values of q. Thus, $\frac{1}{2}q - C\big(1 + q^{p\gamma}\big)$ is uniformly bounded below over $q \geq 0$. Rearranging gives

$$V_n^{\text{OPT}}(\mathbf{c}) \leq F_n(\mathbf{c}) - \Gamma(\mathbf{c}_{\mathsf{B}}/n)\mathsf{H}_{n-1} + C_{\text{OPT}}$$

for some constant $C_{\text{OPT}} < \infty$. This completes the proof.

6.4 Completion of the Proof

Propositions 6.1 and 6.2, together with the ordering $V_n^{\mathrm{F}} \le V_n^{\mathrm{OPT}}$, imply that

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

for any initial state $(\mathbf{c}, n)$ with $\mathbf{c}/n \in \mathcal{K}$. Since $0 \le \mathsf{H}_{n-1} - \log n \le 1$ for $n \ge 1$, we have proved Theorem 4.1.

7 Conclusion

In this paper, we studied a general online resource allocation problem. As our main result, 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 logarithmic correction $\Gamma \log n$, up to uniformly bounded remainders. We analyze each policy separately by telescoping its one-period losses relative to the fluid benchmark along the trajectory induced by the policy and controlling the accumulated variation of the coefficient along that trajectory. The coefficient Γ captures the interaction between random resource consumption and the curvature of the fluid value function in the binding-resource directions. Consequently, the dynamic fluid policy has a horizon-uniform constant gap relative to the optimal online policy, in contrast to the order-optimal $\Theta(\ln n)$ gap relative to the offline optimum established in the literature.

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 Preliminaries for the Proof of Theorem 4.1

In this section, we develop the ingredients needed to prove Proposition 6.1 in Appendix B and Proposition 6.2 in Appendix C. Specifically, Section A.1 constructs buffer regions for the subsequent analysis, and Section A.2 establishes several auxiliary lemmas used in the subsequent proofs.

A.1 Notation and Buffer Regions

Fix the compact target set $\mathcal{K} \Subset \mathcal{U}$ from Section 6 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 $r > 0$, we denote the open and closed Euclidean balls 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 $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$,

$$\|\mathbf{x} - \mathbf{x}_0\| \le 2r_0 \quad \Longrightarrow \quad \big|\mu_s(\mathbf{x}) - \mu_s(\mathbf{x}_0)\big| \le \underline{\delta}/4, \ \forall\, s \in \mathsf{S}. \tag{16}$$

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 $\mathbf{x} \in \widetilde{\mathcal{K}}_{\mathsf{B}}$,

$$x_i \ge \underline{\rho}_{\mathsf{B}} \qquad \forall\, i \in \mathsf{B}, \qquad \text{and} \qquad \|\mathbf{x}\| \le \overline{\rho}_{\mathsf{B}}. \tag{17}$$

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

A.2 Useful Lemmas

In this section, we establish several auxiliary lemmas for Appendices B and C. 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.

A.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. In Lemma A.1, we show 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 (4) and $a^{\mathrm{F}}_n(\mathbf{c})$ denotes the action chosen by the dynamic fluid policy at state $(\mathbf{c}, n)$ as defined in (3). We now state Lemma A.1.

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

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

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 for* (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, $a^{\mathrm{F}}_n(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/n)$ almost surely.*

Proof. For every $\mathbf{x} \in \mathcal{U}_{\mathsf{B}}$, 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 action depends only on $\mathbf{x}$, Assumption 3.2 and Lemma 3.2 yield:

$$\mathbb{E}\Big[\mathbf{B}_{a^{\mathrm{F}}(\mathbf{x}),\mathsf{B}}\Big] = \mathbf{x}, \qquad \mathbb{E}\Big[B_{a^{\mathrm{F}}(\mathbf{x}),j}\Big] = \mu_j(\mathbf{x}) \quad \forall j \in \mathsf{S}, \qquad \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})}\Big] = \phi_{\mathsf{B}}(\mathbf{x}). \tag{18}$$

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 every $j \in \mathsf{S}$, (18) 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 solution. Let $\widetilde{\boldsymbol{\lambda}}$ be any dual optimum at $\boldsymbol{\rho}$. The strong duality and the optimality of $a^{\mathrm{F}}(\mathbf{x})$ imply:

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

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 it is the difference between a pointwise maximum and the score of one specific action a^{F}. Since their sum is zero, both terms must equal 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 slack coordinate. 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. When every action is feasible, (3) and (4) maximize the same objective with the same tie-breaking rule. Hence $a_n^{\mathrm{F}}(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/n)$ almost surely. □

A.2.2 Hölder Continuity of the Logarithmic Correction Coefficient $\Gamma(\mathbf{x})$

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

Lemma A.2 (Hölder Continuity of Γ). *There exist uniform constants $C_\Gamma, L_\Gamma < \infty$ 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 (5) 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 A.3, 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 $|\operatorname{tr}(AB)| \le d\,\|A\|\|B\|$ for any $A, B \in \mathbb{R}^{d\times d}$, it follows that

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

By Assumption 3.1, $\nabla^2\phi_{\mathsf{B}}$ is bounded and α-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. Moreover, by Lemma A.3, Σ is bounded and $\beta_\star$-Hölder continuous on $\widetilde{\mathcal{K}}_{\mathsf{B}}$. Hence, there exists a constant $C < \infty$ such that

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

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

Lemma A.3 (Hölder Continuity of the Covariance Matrix Σ)**.** *There exists a uniform constant $C < \infty$ such that, for every $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and all $\mathbf{x}, \mathbf{x}' \in \bar{B}(\mathbf{x}_0, 2r_0)$,*

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

Proof. Define the second-moment matrix

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

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 $\beta_\star \leq 1$. 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

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

Therefore,

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

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 $\big\{j = a \neq i\big\}$, the preceding inequality yields

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

Therefore,

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

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

$$\mathbb{E}\Big[d_{ij}^{\mathsf{B}} \, \mathbb{1}\big\{i \neq j\big\}\Big] \leq C_2 \, \|\mathbf{x} - \mathbf{x}'\|^{\beta_\star} \tag{19}$$

for some constant $C_2 < \infty$.

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

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

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

$$\left\| \mathbf{B}_{i,\mathsf{B}} \, \mathbf{B}_{i,\mathsf{B}}^{\mathrm{T}} - \mathbf{B}_{j,\mathsf{B}} \, \mathbf{B}_{j,\mathsf{B}}^{\mathrm{T}} \right\| \le C_3 \, d_{ij}^{\mathsf{B}} \,. \tag{20}$$

The random matrices inside the expectations defining $M_2(\mathbf{x})$ and $M_2(\mathbf{x}')$ coincide when $i = j$. Therefore, by (19) and (20), 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 $\widetilde{\mathcal{K}}_{\mathsf{B}}$. □

A.2.3 The One-Step Residual

Fix $\mathbf{x}_0 \in \mathcal{K}_{\mathsf{B}}$ and a normalized state $\mathbf{x} \in \bar{B}(\mathbf{x}_0, r_0)$. 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 sufficiently large n, the successor state $\mathbf{x}'$ and the line segment joining $\mathbf{x}$ and $\mathbf{x}'$ lie in $\bar{B}(\mathbf{x}_0, 2r_0)$. Define the one-step residual relative to the fluid benchmark by

$$\delta_n^{\mathrm{F}}(\mathbf{x}) \triangleq n\phi_{\mathsf{B}}(\mathbf{x}) - \mathbb{E}\Big[R_{a^{\mathrm{F}}(\mathbf{x})} + (n-1)\phi_{\mathsf{B}}(\mathbf{x}')\Big] \,. \tag{21}$$

Through a second-order Taylor expansion of $\delta_n^{\mathrm{F}}(\mathbf{x})$, Lemma A.4 shows that $\delta_n^{\mathrm{F}}(\mathbf{x})$ equals $\frac{\Gamma(\mathbf{x})}{n-1}$ plus a summable remainder.

Lemma A.4 (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 A.5,

$$\phi_{\mathsf{B}}(\mathbf{x}+\mathbf{h}) = \phi_{\mathsf{B}}(\mathbf{x}) + \boldsymbol{\lambda}(\mathbf{x})^{\mathrm{T}}\mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\,\nabla^2\phi_{\mathsf{B}}(\mathbf{x})\,\mathbf{h} + r_n(\mathbf{x},\mathbf{h})\,, \tag{22}$$

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

$$\Big\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}+t\mathbf{h}) - \nabla^2\phi_{\mathsf{B}}(\mathbf{x})\Big\| \le L \cdot t^{\alpha}\,\|\mathbf{h}\|^{\alpha}$$

for every $t \in [0,1]$. Therefore,

$$\begin{aligned}\big|r_n(\mathbf{x},\mathbf{h})\big| &\le \int_0^1 (1-t)\,\|\mathbf{h}\|^2\,\big\|\nabla^2\phi_{\mathsf{B}}(\mathbf{x}+t\mathbf{h}) - \nabla^2\phi_{\mathsf{B}}(\mathbf{x})\big\|\,\mathrm{d}t \\ &\le L\,\|\mathbf{h}\|^{2+\alpha}\int_0^1 (1-t)\,t^{\alpha}\,\mathrm{d}t \le C_2\,n^{-(2+\alpha)}\end{aligned} \tag{23}$$

for some constant $C_2 < \infty$.

Substitute (22) into (21). 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 (22) and (23) 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 A.2 therefore implies that $|\delta_n^{\mathrm{F}}(\mathbf{x})| \le C/n$ for some uniform constant $C < \infty$ and all sufficiently large n. □

Lemma A.5 (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,

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

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 Proof of Proposition 6.1

In this section, we prove Proposition 6.1. The proof is conducted by comparing the performance of the dynamic fluid policy with the fluid benchmark along the trajectory induced by the policy. We first specify a local region and define a stopping time as the first time the trajectory exits this region (Section B.1). We then use martingale concentration to show that the expected remaining horizon at the stopping time is bounded by a uniform constant (Section B.2). Finally, we complete the proof by telescoping the one-period residuals along the trajectory and bounding the terminal error (Section B.3).

Throughout the proof, we fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. All constants below are independent of n and uniform over $\boldsymbol{\rho}_0 \in \mathcal{K}$.

B.1 Local Region, Trajectory, and Stopping Time

Let r_0 denote the radius specified in Appendix A.1. Define the local region

$$\mathcal{D}(\mathbf{x}_0) = \Big\{\boldsymbol{\rho} : \boldsymbol{\rho}_{\mathsf{B}} \in B(\mathbf{x}_0, r_0),\ \rho_s > \mu_s(\mathbf{x}_0) + \underline{\delta}/2 \text{ for every } s \in \mathsf{S}\Big\}. \tag{24}$$

When $\mathsf{S} = \varnothing$, we omit the slack-resource conditions in (24). The initial state $\boldsymbol{\rho}_0$ belongs to $\mathcal{D}(\mathbf{x}_0)$ because its binding coordinates are given by $\mathbf{x}_0$ and

$$\rho_{0,s} - \mu_s(\mathbf{x}_0) \geq \underline{\delta}, \quad \forall\, s \in \mathsf{S}. \tag{25}$$

In Lemma B.1 we show that, when the remaining horizon is sufficiently large, the one-step successor state from any state in $\mathcal{D}(\mathbf{x}_0)$ remains in $\widetilde{\mathcal{U}}$, with its binding-resource coordinates contained in $B(\mathbf{x}_0, 2r_0)$.

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

1. *Every action $a \in \mathcal{A}$ is feasible at $\mathbf{c}$.*
2. *$\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$ and $a^{\mathrm{F}}_h(\mathbf{c}) = a^{\mathrm{F}}(\boldsymbol{\rho}_{\mathsf{B}})$ 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) $\boldsymbol{\rho}^a_{\mathsf{B}} \in B(\mathbf{x}_0, 2r_0)$ and, (ii) if $\mathsf{S} \neq \varnothing$, then $\rho^a_s - \mu_s(\boldsymbol{\rho}^a_{\mathsf{B}}) \geq \underline{\delta}/8$ for all $s \in \mathsf{S}$. Consequently, $\boldsymbol{\rho}^a \in \widetilde{\mathcal{U}}$.*

Proof. Choose an integer $n_0 \geq 2$ such that

$$(n_0 + 1)\underline{\rho}_{\mathsf{B}} \geq \bar{b}, \qquad \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{n_0} \leq r_0, \qquad \frac{\bar{b}}{n_0} \leq \frac{\underline{\delta}}{8}. \tag{26}$$

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

Proof of Part 1. It suffices to show that $\mathbf{c} \geq \bar{b}\mathbf{1}$. From (17), each binding capacity is at least $h\underline{\rho}_{\mathsf{B}} \geq \bar{b}$. For each slack resource $s \in \mathsf{S}$, its normalized capacity exceeds $\mu_s(\mathbf{x}_0) + \underline{\delta}/2 \geq \underline{\delta}/2$. Hence, by (26), each slack 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 $\mu_s(\boldsymbol{\rho}_{\mathsf{B}}) \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/4 < \rho_s$; the first inequality follows from (16) and the second from the definition of $\mathcal{D}(\mathbf{x}_0)$. Therefore, $\boldsymbol{\rho} \in \widetilde{\mathcal{U}}$. Part 3 of Lemma A.1, together with Part 1, implies that $a^{\mathrm{F}}_h(\mathbf{c}) = a^{\mathrm{F}}(\mathbf{c}_{\mathsf{B}}/h)$ almost surely.

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

$$\boldsymbol{\rho}^a = \boldsymbol{\rho} + \frac{\boldsymbol{\rho} - \mathbf{B}_a}{h-1}. \tag{27}$$

For the binding coordinates, (27) and (17) imply

$$\|\boldsymbol{\rho}_{\mathsf{B}}^a - \boldsymbol{\rho}_{\mathsf{B}}\| \le \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{h-1} \le \frac{\overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}}{n_0} \le r_0.$$

Therefore,

$$\|\boldsymbol{\rho}_{\mathsf{B}}^a - \mathbf{x}_0\| \le \|\boldsymbol{\rho}_{\mathsf{B}}^a - \boldsymbol{\rho}_{\mathsf{B}}\| + \|\boldsymbol{\rho}_{\mathsf{B}} - \mathbf{x}_0\| < 2r_0.$$

For each slack coordinate $s \in \mathsf{S}$, by (27), we have

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

The second inequality follows from $\boldsymbol{\rho} = \mathbf{c}/h \in \mathcal{D}(\mathbf{x}_0)$. The third inequality follows from $\mu_s(\boldsymbol{\rho}_{\mathsf{B}}^a) \le \mu_s(\mathbf{x}_0) + \underline{\delta}/4$, which holds because $\|\boldsymbol{\rho}_{\mathsf{B}}^a - \mathbf{x}_0\| < 2r_0$ and (16). The fourth inequality follows from the third inequality in (26). □

B.1.1 Definitions of Trajectory and Stopping Time

Fix a time horizon $n \ge n_0 + 1$. We terminate the analysis when n_0 periods remain and let $\bar{N} \triangleq n - n_0$ denote the maximum number of decisions made before this cutoff. Throughout the remainder of Section B, we index time forward: We let k denote the number of decisions made since the beginning of the horizon, and $h_k \triangleq n - k$ denote the number of periods remaining after the k-th decision.

We now define the states, actions, and transitions for $1 \le k \le \bar{N}$. Let $\mathbf{c}_0 = \mathbf{c} \in \mathbb{R}_+^m$ denote the initial inventory and $\mathbf{c}_k$ the remaining inventory after the k-th decision. In period $k \ge 1$, the dynamic fluid policy observes the independent opportunity $\boldsymbol{\theta}_k = (R_{a,k}, \mathbf{B}_{a,k})_{a \in \mathcal{A}}$, chooses the action $a_k \triangleq a_{h_k+1}^{\mathrm{F}}(\mathbf{c}_{k-1})$ defined in (3), collects a reward $\hat{R}_k \triangleq R_{a_k,k}$, and consumes a resource vector $\mathbf{Y}_k \triangleq \mathbf{B}_{a_k,k} \in \mathbb{R}_+^m$. The state variables satisfy

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

In addition, let $\{\mathcal{G}_k\}_{k \ge 0}$ represent the filtration of information, where $\mathcal{G}_k = \sigma\big(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\big)$ for $k \ge 1$

and $\mathcal{G}_0 = \{\varnothing, \Omega\}$. Hence, $\mathbf{c}_k$, $\boldsymbol{\rho}_k$, and $\mathbf{x}_k$ are $\mathcal{G}_k$-measurable, whereas $\boldsymbol{\theta}_{k+1}$ is independent of $\mathcal{G}_k$.

Define the stopping time τ by

$$\tau \triangleq \inf\left\{0 \le k \le \bar{N} : \boldsymbol{\rho}_k \notin \mathcal{D}(\mathbf{x}_0) \text{ or } h_k \le n_0\right\}.$$

That is, we stop on exit from the region $\mathcal{D}(\mathbf{x}_0)$ or when n_0 periods remain. Let $h_\tau \triangleq n - \tau$ denote the remaining horizon at stopping. By definition, $h_\tau \ge n_0 + 1$ only if the trajectory exits $\mathcal{D}(\mathbf{x}_0)$ within the first $\bar{N} - 1$ periods. If no such exit occurs before the cutoff, then $\tau = \bar{N}$ and $h_\tau = n_0$.

On event $\{k < \tau\}$, $h_k \ge n_0 + 1$ and $\boldsymbol{\rho}_k \in \mathcal{D}(\mathbf{x}_0)$. Hence, Lemma B.1 implies that $a_{k+1} = a^{\mathrm{F}}(\mathbf{x}_k)$. Moreover, since $\boldsymbol{\theta}_{k+1}$ is independent of $\mathcal{G}_k$, we have

$$\begin{aligned} \mathbb{E}\left[\mathbf{Y}_{k+1,\mathsf{B}} \mid \mathcal{G}_k\right] &= \mathbf{x}_k, \\ \mathbb{E}\left[Y_{k+1,s} \mid \mathcal{G}_k\right] &= \mu_s(\mathbf{x}_k), \quad s \in \mathsf{S}, \\ \mathbb{E}\left[\hat{R}_{k+1} \mid \mathcal{G}_k\right] &= \phi_{\mathsf{B}}(\mathbf{x}_k). \end{aligned} \tag{28}$$

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

$$\boldsymbol{\rho}_{k+1} = \frac{h_k \boldsymbol{\rho}_k - \mathbf{Y}_{k+1}}{h_k - 1} = \boldsymbol{\rho}_k + \frac{\boldsymbol{\rho}_k - \mathbf{Y}_{k+1}}{h_k - 1}, \qquad \mathbf{x}_{k+1} - \mathbf{x}_k = \frac{\mathbf{x}_k - \mathbf{Y}_{k+1,\mathsf{B}}}{h_k - 1}. \tag{29}$$

By Part 3 of Lemma B.1, every stopped binding state satisfies $\mathbf{x}_{k\wedge\tau} \in \bar{B}(\mathbf{x}_0, 2r_0)$.

B.2 Bounding the Expected Remaining Horizon at Stopping

In this section, we show that the remaining horizon h_τ at the stopping time has an exponentially decaying tail:

$$\mathbb{P}\{h_\tau \ge r\} \le Ce^{-cr},$$

which implies that $\mathbb{E}[h_\tau] = O(1)$. We also prove the second-moment bound $\mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^2 = O(1/h_k)$.

Lemma B.2 (Bounds on Second Moments and Early Exits). *There exist constants $C_M, C, c > 0$, depending only on $\mathcal{K}$ and the model primitives, such that, uniformly over $n \ge n_0 + 1$ and $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$,*

$$\mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^2 \le \frac{C_M^2}{h_k}, \qquad 0 \le k \le \bar{N}, \tag{30}$$

and, for every integer $r \ge n_0 + 1$,

$$\mathbb{P}(h_\tau \ge r) \le Ce^{-cr}.$$

Consequently,

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

B.2.1 Proof of Lemma B.2

We prove Lemma B.2 via martingale concentration. Specifically, we show that, before stopping, the trajectory of the normalized binding resources is a martingale, and the trajectory of the normalized slack resources can be decomposed into a positive predictable drift and a martingale. We then use a martingale concentration inequality to control the coordinatewise fluctuations of the binding and slack resources, which in turn proves Lemma B.2.

Martingale Construction Define the stopped increments and their partial sums by

$$\Delta\mathbf{M}_{k+1} \triangleq \mathbb{1}\{k<\tau\}\frac{\mathbb{E}[\mathbf{Y}_{k+1}\mid\mathcal{G}_k]-\mathbf{Y}_{k+1}}{h_k-1},$$

$$\mathbf{M}_k \triangleq \sum_{\ell=1}^{k}\Delta\mathbf{M}_\ell,$$

with $\mathbf{M}_0=\mathbf{0}$, where $\mathbf{Y}_{k+1}$ denotes the consumption of the dynamic fluid policy in period $k+1$. Since $\{k<\tau\}\in\mathcal{G}_k$,

$$\mathbb{E}[\Delta\mathbf{M}_{k+1}\mid\mathcal{G}_k] = \frac{\mathbb{1}\{k<\tau\}}{h_k-1}\Big(\mathbb{E}[\mathbf{Y}_{k+1}\mid\mathcal{G}_k]-\mathbb{E}[\mathbf{Y}_{k+1}\mid\mathcal{G}_k]\Big)=\mathbf{0}.$$

In addition, $\mathbf{M}_k$ is $\mathcal{G}_k$-measurable, so

$$\mathbb{E}[\mathbf{M}_{k+1}\mid\mathcal{G}_k]=\mathbf{M}_k+\mathbb{E}[\Delta\mathbf{M}_{k+1}\mid\mathcal{G}_k]=\mathbf{M}_k.$$

Therefore, $\{\mathbf{M}_k\}$ is a martingale. Denote $M_{k,j}=(\mathbf{M}_k)_j$ and set $C_0=2\bar{b}$. Since $0\leq Y_{k+1,j}$, $\mathbb{E}[Y_{k+1,j}\mid\mathcal{G}_k]\leq\bar{b}$ by bounded consumption,

$$|\Delta M_{k+1,j}| \leq \mathbb{1}\{k<\tau\}\frac{\bar{b}}{h_k-1}\leq\frac{C_0}{h_k},\qquad j\in[m], \tag{31}$$

where the second inequality uses $h_k\geq n_0+1\geq 2$ on $\{k<\tau\}$.

Binding and Slack Coordinates On $\{k<\tau\}$, (28) and (29) imply

$$\Delta\mathbf{M}_{k+1,\mathsf{B}}=\frac{\mathbf{x}_k-\mathbf{Y}_{k+1,\mathsf{B}}}{h_k-1}=\mathbf{x}_{k+1}-\mathbf{x}_k.$$

On $\{k \geq \tau\}$, $\Delta\mathbf{M}_{k+1,\mathsf{B}} = 0$ by definition. Therefore,

$$\Delta\mathbf{M}_{k+1,\mathsf{B}} = \mathbf{x}_{(k+1)\wedge\tau} - \mathbf{x}_{k\wedge\tau}, \qquad \mathbf{M}_{k,\mathsf{B}} = \mathbf{x}_{k\wedge\tau} - \mathbf{x}_0.$$

If $\mathsf{S} \neq \varnothing$, then for each $s \in \mathsf{S}$ and on $\{k < \tau\}$, adding and subtracting $\mu_s(\mathbf{x}_k)$ in the slack-coordinate recursion in (29) gives

$$\begin{aligned} \rho_{k+1,s} - \rho_{k,s} &= \frac{\rho_{k,s} - \mu_s(\mathbf{x}_k)}{h_k - 1} + \frac{\mu_s(\mathbf{x}_k) - Y_{k+1,s}}{h_k - 1} \\ &= \frac{\rho_{k,s} - \mu_s(\mathbf{x}_k)}{h_k - 1} + \Delta M_{k+1,s}. \end{aligned} \tag{32}$$

The second equality follows from (28). By $\boldsymbol{\rho}_k \in \mathcal{D}(\mathbf{x}_0)$ and (16),

$$\rho_{k,s} - \mu_s(\mathbf{x}_k) > \mu_s(\mathbf{x}_0) + \underline{\delta}/2 - \big(\mu_s(\mathbf{x}_0) + \underline{\delta}/4\big) = \underline{\delta}/4 \geq 0.$$

Thus, summing (32) up to time $k \wedge \tau$ and using $\Delta M_{k+1,s} = 0$ on $\{k \geq \tau\}$, we obtain

$$\rho_{k\wedge\tau,s} - \rho_{0,s} = \sum_{\ell=0}^{(k\wedge\tau)-1} \frac{\rho_{\ell,s} - \mu_s(\mathbf{x}_\ell)}{h_\ell - 1} + M_{k,s} \geq M_{k,s}. \tag{33}$$

Bounding the Second Moments Set $C_M = \sqrt{|\mathsf{B}|}\, C_0$. By (31), $\|\Delta\mathbf{M}_{k+1,\mathsf{B}}\| \leq C_M/h_k$. Distinct martingale differences are orthogonal. Indeed, if $\ell < j$, since $\Delta\mathbf{M}_{\ell+1,\mathsf{B}}$ is $\mathcal{G}_j$-measurable, we have:

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

Therefore, expanding the squared norm and using the orthogonality of the increments gives

$$\begin{aligned} \mathbb{E}\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\|^2 &= \mathbb{E}\left\|\sum_{\ell=0}^{k-1} \Delta\mathbf{M}_{\ell+1,\mathsf{B}}\right\|^2 \\ &= \sum_{\ell=0}^{k-1} \mathbb{E}\|\Delta\mathbf{M}_{\ell+1,\mathsf{B}}\|^2 \\ &\leq C_M^2 \sum_{u=h_k+1}^{n} u^{-2} \leq C_M^2 \int_{h_k}^{\infty} u^{-2}\, \mathrm{d}u = C_M^2/h_k. \end{aligned}$$

This proves (30).

Coordinatewise Concentration Fix an integer $n_0 < r < n$ and let $t = n - r$. By the definition of the local region $\mathcal{D}(\mathbf{x}_0)$, an exit by time t can occur only if either the binding projection leaves $B(\mathbf{x}_0, r_0)$ or some slack coordinate falls to or below $\mu_s(\mathbf{x}_0) + \underline{\delta}/2$ despite its nonnegative drift. Since $\{\mathbf{M}_k\}$ remains constant after τ, any such deviation persists at time t. We first bound the coordinatewise tail probabilities of $\mathbf{M}_t$ and later use these bounds to control the event $\{h_\tau \geq r\}$.

For each coordinate j and time index $1 \leq \ell \leq t$, by (31),

$$M_{\ell,j} - M_{\ell-1,j} \in \left[-\frac{C_0}{h_{\ell-1}}, \frac{C_0}{h_{\ell-1}}\right].$$

This interval has a deterministic width $2C_0/h_{\ell-1}$. Since $h_{\ell-1} = n - \ell + 1$,

$$\sum_{\ell=1}^{t} \left(\frac{2C_0}{h_{\ell-1}}\right)^2 = 4C_0^2 \sum_{u=r+1}^{n} u^{-2} \leq 4C_0^2 \int_r^\infty u^{-2}\, \mathrm{d}u = \frac{4C_0^2}{r}.$$

Applying Lemma D.1 to $M_{k,j}$ and $-M_{k,j}$ and taking a union bound yields, for every $d > 0$,

$$\mathbb{P}(|M_{t,j}| \geq d) \leq 2\exp\left(-\frac{d^2 r}{2C_0^2}\right).$$

Bounding the Tail Probability and the Expected Remaining Horizon Since $\boldsymbol{\rho}_0 \in \mathcal{D}(\mathbf{x}_0)$ and $n \geq n_0 + 1$, we have $n_0 \leq h_\tau < n$. Hence, $\mathbb{P}(h_\tau \geq r) = 0$ for all $r \geq n$.

Now suppose $n_0 < r < n$ and set $t = n - r$. On event $\{h_\tau \geq r\}$, $\tau \leq t < \bar{N}$; hence, stopping must occur through an exit from $\mathcal{D}(\mathbf{x}_0)$. Since every increment after τ is zero,

$$\mathbf{M}_t = \mathbf{M}_\tau \quad \text{on} \;\; \{h_\tau \geq r\}.$$

If the exit occurs through the binding coordinates, then

$$\|\mathbf{M}_{t,\mathsf{B}}\| = \|\mathbf{M}_{\tau,\mathsf{B}}\| = \|\mathbf{x}_\tau - \mathbf{x}_0\| \geq r_0.$$

Consequently, at least one binding coordinate satisfies $|M_{t,j}| \geq \varsigma$, where $\varsigma \triangleq r_0/\sqrt{|\mathsf{B}|}$. Applying the coordinatewise concentration bound and a union bound yields

$$\mathbb{P}\big(h_\tau \geq r,\; \|\mathbf{x}_\tau - \mathbf{x}_0\| \geq r_0\big) \leq 2|\mathsf{B}| \exp\left(-\frac{\varsigma^2 r}{2C_0^2}\right).$$

If $\mathsf{S} \neq \varnothing$ and the exit occurs through a slack coordinate s, then $\rho_{\tau,s} \leq \mu_s(\mathbf{x}_0) + \underline{\delta}/2$. By (33) and

(25),

$$M_{t,s} = M_{\tau,s} \le \rho_{\tau,s} - \rho_{0,s} \le \big(\mu_s(\mathbf{x}_0) + \underline{\delta}/2\big) - \big(\mu_s(\mathbf{x}_0) + \underline{\delta}\big) = -\underline{\delta}/2.$$

Hence, for each $s \in \mathsf{S}$, the coordinatewise concentration bound gives

$$\mathbb{P}\big(h_\tau \ge r,\ \rho_{\tau,s} \le \mu_s(\mathbf{x}_0) + \underline{\delta}/2\big) \le 2\exp\left(-\frac{\underline{\delta}^2 r}{8C_0^2}\right).$$

Since every exit occurs through either the binding coordinates or a slack coordinate, a union bound over all coordinates yields $\mathbb{P}(h_\tau \ge r) \le 2me^{-cr}$, where

$$c = \begin{cases} \min\left\{\varsigma^2, \underline{\delta}^2/4\right\} / \left(2C_0^2\right), & \mathsf{S} \ne \varnothing, \\ \varsigma^2 / \left(2C_0^2\right), & \mathsf{S} = \varnothing. \end{cases}$$

Finally, since h_τ is integer-valued and $h_\tau \ge n_0$, the established tail bound yields

$$\mathbb{E}h_\tau = n_0 + \sum_{r=n_0+1}^{\infty} \mathbb{P}(h_\tau \ge r) \le n_0 + 2m \sum_{r=n_0+1}^{\infty} e^{-cr} < \infty.$$

B.3 Completing the Proof: Telescoping the Fluid Residual

In this section, we complete the proof of Proposition 6.1 using a telescoping analysis. Suppose first that $n \ge n_0 + 1$. Define

$$D_h(\mathbf{c}') \triangleq F_h(\mathbf{c}') - V_h^{\mathrm{F}}(\mathbf{c}'), \qquad h \ge 0.$$

Note that since $0 \le V_h^{\mathrm{F}}(\mathbf{c}') \le F_h(\mathbf{c}') \le h\bar{r}$, we have $0 \le D_h(\mathbf{c}') \le h\bar{r}$.

The Stopped Telescope On event $\{k < \tau\}$, from Lemma B.1, both the current state $\boldsymbol{\rho}_k$ and the successor state $\boldsymbol{\rho}_{k+1}$ remain in $\widetilde{\mathcal{U}}$, and $a^{\mathrm{F}}_{h_k}(\mathbf{c}_k) = a^{\mathrm{F}}(\mathbf{x}_k)$. Thus, the definition of $\delta^{\mathrm{F}}_{h_k}$ in (21) and the Bellman equation for the dynamic fluid policy give

$$\begin{aligned} F_{h_k}(\mathbf{c}_k) &= \mathbb{E}\big[\hat{R}_{k+1} + F_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big] + \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k), \\ V^{\mathrm{F}}_{h_k}(\mathbf{c}_k) &= \mathbb{E}\big[\hat{R}_{k+1} + V^{\mathrm{F}}_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big]. \end{aligned}$$

Subtracting the second identity from the first cancels the immediate reward and yields

$$D_{h_k}(\mathbf{c}_k) = \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k) + \mathbb{E}\big[D_{h_k-1}(\mathbf{c}_{k+1}) \mid \mathcal{G}_k\big] \quad \text{on} \quad \{k < \tau\}.$$

Define $Z_k \triangleq D_{h_{k\wedge\tau}}(\mathbf{c}_{k\wedge\tau})$ for $0 \le k \le \bar{N}$. Since $D_{h_k}(\mathbf{c}_k)$ and event $\{k < \tau\}$ are $\mathcal{G}_k$-measurable, we have

$$\mathbb{E}\big[Z_{k+1} - Z_k \mid \mathcal{G}_k\big] = \begin{cases} -\delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k), & k < \tau, \\ 0, & k \ge \tau. \end{cases}$$

Taking expectations and summing over $0 \le k < \bar{N}$ yields

$$\mathbb{E}Z_{\bar{N}} - Z_0 = \mathbb{E}D_{h_\tau}(\mathbf{c}_\tau) - D_n(\mathbf{c}) = -\mathbb{E}\sum_{k<\tau} \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k).$$

Consequently,

$$F_n(\mathbf{c}) - V^{\mathrm{F}}_n(\mathbf{c}) = D_n(\mathbf{c}) = \mathbb{E}\sum_{k<\tau} \delta^{\mathrm{F}}_{h_k}(\mathbf{x}_k) + \mathbb{E}D_{h_\tau}(\mathbf{c}_\tau). \tag{34}$$

Thus, the initial gap decomposes into the one-period residuals accumulated before stopping and the remaining gap at the stopping time.

Bounding the Terminal Error and Taylor Remainders From Lemma B.2,

$$0 \le \mathbb{E}D_{h_\tau}(\mathbf{c}_\tau) \le \bar{r}\,\mathbb{E}h_\tau = O(1). \tag{35}$$

For every state before stopping, Lemma A.4 gives

$$\delta^{\mathrm{F}}_h(\mathbf{x}) = \frac{\Gamma(\mathbf{x})}{h-1} + \epsilon_h(\mathbf{x}), \qquad |\epsilon_h(\mathbf{x})| \le C_\epsilon h^{-1-\alpha}, \tag{36}$$

for some uniform constant $C_\epsilon < \infty$. Since $\alpha > 0$, the Taylor remainders satisfy

$$\left|\mathbb{E}\sum_{k<\tau} \epsilon_{h_k}(\mathbf{x}_k)\right| \le C_\epsilon \sum_{h=n_0+1}^{\infty} h^{-1-\alpha} < \infty. \tag{37}$$

Coefficient Replacement and the Harmonic Sum Note that the residual in (34) involves $\Gamma(\mathbf{x}_k)$, whereas the expansion in Proposition 6.1 involves $\Gamma(\mathbf{x}_0)$. We now bound the cumulative effect of replacing the former by the latter using (30). From Lemma B.1, $\mathbf{x}_{k\wedge\tau} \in B(\mathbf{x}_0, 2r_0)$. Let L_Γ denote the uniform Hölder constant from Lemma A.2. Then,

$$\mathbb{1}\{k < \tau\}\big|\Gamma(\mathbf{x}_{k\wedge\tau}) - \Gamma(\mathbf{x}_0)\big| \le L_\Gamma \big\|\mathbf{x}_{k\wedge\tau} - \mathbf{x}_0\big\|^\gamma.$$

Since $u^{\gamma/2}$ is concave in u, from (30) and Jensen's inequality,

$$\mathbb{E}\big[\mathbb{1}\{k<\tau\}\,|\Gamma(\mathbf{x}_{k\wedge\tau})-\Gamma(\mathbf{x}_0)|\,\big] \le L_\Gamma\Big(\mathbb{E}\big\|\mathbf{x}_{k\wedge\tau}-\mathbf{x}_0\big\|^2\Big)^{\gamma/2} \le L_\Gamma\,C_M^\gamma\,h_k^{-\gamma/2}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}\sum_{k<\tau}\frac{|\Gamma(\mathbf{x}_k)-\Gamma(\mathbf{x}_0)|}{h_k-1} &= \sum_{k=0}^{\bar{N}-1}\frac{\mathbb{E}\big[\mathbb{1}\{k<\tau\}\,|\Gamma(\mathbf{x}_{k\wedge\tau})-\Gamma(\mathbf{x}_0)|\,\big]}{h_k-1}\\
&\le 2\sum_{k=0}^{\bar{N}-1}\frac{\mathbb{E}\big[\mathbb{1}\{k<\tau\}\,|\Gamma(\mathbf{x}_{k\wedge\tau})-\Gamma(\mathbf{x}_0)|\,\big]}{h_k}\\
&\le 2\,L_\Gamma\,C_M^\gamma\sum_{h=n_0+1}^{\infty}h^{-1-\gamma/2}<\infty,
\end{aligned}$$

where the first inequality uses $h_k-1\ge h_k/2$, because $h_k\ge n_0+1\ge 2$ on $\{k<\tau\}$. Consequently,

$$\mathbb{E}\sum_{k<\tau}\frac{\Gamma(\mathbf{x}_k)}{h_k-1} = \Gamma(\mathbf{x}_0)\,\mathbb{E}\sum_{k<\tau}\frac{1}{h_k-1}+O(1). \tag{38}$$

Note that

$$\sum_{k<\tau}\frac{1}{h_k-1}=\sum_{j=h_\tau}^{n-1}\frac{1}{j}=\mathsf{H}_{n-1}-\mathsf{H}_{h_\tau-1}.$$

Since $\mathsf{H}_{h_\tau-1}\le h_\tau$, Lemma B.2 and the boundedness of Γ from Lemma A.2 yield

$$\Gamma(\mathbf{x}_0)\,\mathbb{E}\sum_{k<\tau}\frac{1}{h_k-1}=\Gamma(\mathbf{x}_0)\mathsf{H}_{n-1}+O(1). \tag{39}$$

Wrap-Up Combining (34) – (39), for $n\ge n_0+1$, we have

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

Finally, for $1\le n\le n_0$, we have

$$0\le F_n(\mathbf{c})-V_n^{\mathrm{F}}(\mathbf{c})\le n_0\bar{r},\qquad 0\le\Gamma(\mathbf{c}_{\mathsf{B}}/n)\mathsf{H}_{n-1}\le C_\Gamma\mathsf{H}_{n_0},$$

where C_Γ is the uniform upper bound on Γ in Lemma A.2. Increasing C_{F} to cover these finitely many horizons completes the proof of Proposition 6.1.

C Proof of Proposition 6.2

In this section, we prove Proposition 6.2 by comparing the actions of the optimal online policy with the static fluid actions defined in (4) along the trajectory induced by the optimal policy. Assumption 3.3 allows us to control the resulting consumption differences through the loss in dual-price score.

We first express $F_n - V_n^{\mathrm{OPT}}$ as the sum of accumulated one-period residuals along the optimal trajectory and a nonnegative terminal term. Thus, a lower bound on $F_n - V_n^{\mathrm{OPT}}$ follows from a lower bound on the accumulated one-period residuals; see (42). We then lower-bound these residuals by comparing them with the corresponding one-period residuals under the static fluid actions (Lemma C.2). Finally, we control the cumulative effect of replacing the path-dependent coefficients Γ along the optimal trajectory with their initial value (Lemma C.4). Combining these bounds establishes Proposition 6.2.

Throughout the proof, we fix an initial state $\boldsymbol{\rho}_0 = \mathbf{c}/n \in \mathcal{K}$ and let $\mathbf{x}_0 = (\boldsymbol{\rho}_0)_{\mathsf{B}}$. All constants below are independent of n and uniform over $\boldsymbol{\rho}_0 \in \mathcal{K}$.

C.1 The Optimal Trajectory

Throughout Section C, 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.

We now define the states, actions, and transitions under the optimal online policy for $1 \le k \le \bar{N}$. Let $\bar{\mathbf{c}}_0 = \mathbf{c} \in \mathbb{R}_+^m$ denote the initial inventory and $\bar{\mathbf{c}}_k$ the remaining inventory after the k-th decision. In period $k \ge 1$, upon observing the independent opportunity $\boldsymbol{\theta}_k = (R_{a,k}, \mathbf{B}_{a,k})_{a\in\mathcal{A}}$, the optimal online policy chooses

$$\bar{a}_k = \operatorname*{arg\,max}_{a\in\mathcal{A}(\bar{\mathbf{c}}_{k-1},\boldsymbol{\theta}_k)} \Big\{ R_{a,k} + V_{h_k}^{\mathrm{OPT}}\big(\bar{\mathbf{c}}_{k-1} - \mathbf{B}_{a,k}\big) \Big\},$$

collects a reward $\bar{R}_k = R_{\bar{a}_k,k}$, and consumes a resource vector $\bar{\mathbf{Y}}_k = \mathbf{B}_{\bar{a}_k,k}$. The state variables satisfy

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

Let $\{\mathcal{G}_k\}_{k\ge0}$ denote the information filtration, where $\mathcal{G}_k = \sigma\big(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_k\big)$ for $k \ge 1$ and $\mathcal{G}_0 = \{\varnothing,\Omega\}$.

Finally, for $h \ge 1$ and $\mathbf{c} \ge \mathbf{0}$ with $\mathbf{x} = \mathbf{c}_{\mathsf{B}}/h \in \mathcal{U}_{\mathsf{B}}$, define

$$P_h(\mathbf{c}) = h\phi_{\mathsf{B}}(\mathbf{x}).$$

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

$$0 \le V_h^{\mathrm{OPT}}(\mathbf{c}) \le P_h(\mathbf{c}) \le h\bar{r}. \tag{40}$$

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).

C.2 The Stopping Time and the Telescoping Identity

Let r_0 and $\widetilde{\mathcal{K}}_{\mathsf{B}}$ be the radius and compact neighborhood, respectively, defined in Appendix A.1. Define

$$K_0 = \overline{\rho}_{\mathsf{B}} + \sqrt{|\mathsf{B}|}\,\bar{b}, \qquad M_H = \sup_{\mathbf{z} \in \widetilde{\mathcal{K}}_{\mathsf{B}}} \|\nabla^2 \phi_{\mathsf{B}}(\mathbf{z})\|, \qquad L = 2\max\big\{M_H K_0, 1\big\}.$$

Choose a fixed integer $n_0 \ge 2$ satisfying (26) and $n_0 \ge L/t_\star$. Fix a time horizon $n \ge n_0 + 1$ and set $\bar{N} = n - n_0$. Define the stopping time

$$\bar{\tau} = \inf\Big\{0 \le k \le \bar{N} : \|\bar{\mathbf{x}}_k - \mathbf{x}_0\| \ge r_0 \text{ or } h_k \le n_0\Big\}.$$

Before stopping, $h_k \ge n_0 + 1$ and $\bar{\mathbf{x}}_k \in B(\mathbf{x}_0, r_0)$. In addition,

$$\bar{\mathbf{x}}_{k+1} = \frac{h_k \bar{\mathbf{x}}_k - \bar{\mathbf{Y}}_{k+1,\mathsf{B}}}{h_k - 1} = \bar{\mathbf{x}}_k + \frac{\bar{\mathbf{x}}_k - \bar{\mathbf{Y}}_{k+1,\mathsf{B}}}{h_k - 1}, \qquad \|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\| \le \frac{K_0}{h_k - 1} \le r_0,$$

where the norm bound follows from (17) and (26). Hence, for every $0 \le u \le 1$,

$$\|\bar{\mathbf{x}}_k + u(\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k) - \mathbf{x}_0\| \le \|\bar{\mathbf{x}}_k - \mathbf{x}_0\| + u\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\| < 2r_0.$$

Therefore, all stopped states $\bar{\mathbf{x}}_{k \wedge \bar{\tau}}$ and the line segments between consecutive states before stopping lie in the closed ball $\bar{B}(\mathbf{x}_0, 2r_0) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}} \Subset \mathcal{U}_{\mathsf{B}}$.

For $0 \le k \le \bar{N}$, define

$$Z_k = P_{h_{k\wedge\bar{\tau}}}(\bar{\mathbf{c}}_{k\wedge\bar{\tau}}) - V^{\mathrm{OPT}}_{h_{k\wedge\bar{\tau}}}(\bar{\mathbf{c}}_{k\wedge\bar{\tau}}),$$

and define the conditional one-period difference along the optimal trajectory by

$$\delta_k^{\mathrm{OPT}} = P_{h_k}(\bar{\mathbf{c}}_k) - \mathbb{E}\Big[\bar{R}_{k+1} + P_{h_k - 1}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big], \quad k < \bar{\tau}, \tag{41}$$

with $\delta_k^{\mathrm{OPT}} = 0$ for $k \ge \bar{\tau}$. On the event $\{k < \bar{\tau}\}$, the definition of δ_k^{OPT} and the Bellman equation for the optimal policy imply

$$\begin{aligned}
P_{h_k}(\bar{\mathbf{c}}_k) &= \delta_k^{\mathrm{OPT}} + \mathbb{E}\Big[\bar{R}_{k+1} + P_{h_k-1}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big],\\
V_{h_k}^{\mathrm{OPT}}(\bar{\mathbf{c}}_k) &= \mathbb{E}\Big[\bar{R}_{k+1} + V_{h_k-1}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{k+1}) \mid \mathcal{G}_k\Big].
\end{aligned}$$

Subtracting the second identity from the first cancels the common reward term and yields

$$Z_k = \delta_k^{\mathrm{OPT}} + \mathbb{E}[Z_{k+1} \mid \mathcal{G}_k] \qquad \text{on } \{k < \bar{\tau}\}.$$

On the event $\{k \geq \bar{\tau}\}$, we have $Z_{k+1} = Z_k$ and $\delta_k^{\mathrm{OPT}} = 0$. Since Z_k and $\{k < \bar{\tau}\}$ are $\mathcal{G}_k$-measurable, the two cases give

$$\mathbb{E}[Z_{k+1} - Z_k \mid \mathcal{G}_k] = -\delta_k^{\mathrm{OPT}}, \qquad 0 \leq k < \bar{N}.$$

Taking expectations and telescoping over $0 \leq k < \bar{N}$ yields

$$\begin{aligned}
\mathbb{E}\sum_{k=0}^{\bar{N}-1} \delta_k^{\mathrm{OPT}} = \sum_{k=0}^{\bar{N}-1} \big(\mathbb{E}Z_k - \mathbb{E}Z_{k+1}\big) &= Z_0 - \mathbb{E}Z_{\bar{N}},\\
Z_0 &= F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c}),\\
Z_{\bar{N}} &= P_{h_{\bar{\tau}}}(\bar{\mathbf{c}}_{\bar{\tau}}) - V_{h_{\bar{\tau}}}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{\bar{\tau}}).
\end{aligned}$$

The second identity follows from $P_n(\mathbf{c}) = F_n(\mathbf{c})$, and the third follows from $\bar{\tau} \leq \bar{N}$. Since $\delta_k^{\mathrm{OPT}} = 0$ for $k \geq \bar{\tau}$, rearranging gives

$$F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c}) = \mathbb{E}\sum_{k<\bar{\tau}} \delta_k^{\mathrm{OPT}} + \mathbb{E}\big[P_{h_{\bar{\tau}}}(\bar{\mathbf{c}}_{\bar{\tau}}) - V_{h_{\bar{\tau}}}^{\mathrm{OPT}}(\bar{\mathbf{c}}_{\bar{\tau}})\big] \geq \mathbb{E}\sum_{k<\bar{\tau}} \delta_k^{\mathrm{OPT}}, \tag{42}$$

where the inequality follows from (40).

C.3 From Score Loss to Consumption Differences

Let $i = a^{\mathrm{F}}(\bar{\mathbf{x}}_k)$ denote the static fluid action and $a = \bar{a}_{k+1}$ the optimal feasible action. Define

$$g_{k+1} = S_i(\bar{\mathbf{x}}_k) - S_a(\bar{\mathbf{x}}_k) \geq 0, \qquad d_{k+1} = \|\mathbf{B}_{i,\mathsf{B}} - \mathbf{B}_{a,\mathsf{B}}\|_1,$$

where the dual-price score $S_{a'}(\mathbf{x})$ of an action a' is defined in Assumption 3.3. The quantity g_{k+1} is the realized score loss, and d_{k+1} is the corresponding difference in binding-resource consumption. Since i maximizes the dual-price score over all actions, $g_{k+1} \geq 0$.

Before stopping (i.e., for $k < \bar{\tau}$), define

$$\ell_k = \mathbb{E}\big[g_{k+1} \mid \mathcal{G}_k\big], \qquad D_k = \mathbb{E}\big[d_{k+1} \mid \mathcal{G}_k\big], \qquad q = \mathbb{E}\sum_{k<\bar{\tau}} \ell_k. \tag{43}$$

Set $\ell_k = D_k = 0$ for $k \geq \bar{\tau}$. In Lemma C.1, we bound the expected consumption difference D_k in terms of the expected score loss ℓ_k.

Lemma C.1. *Let*

$$p = \frac{\beta_\star}{1+\beta_\star} \in (0,1), \qquad \frac{1}{1-p} = 1 + \beta_\star.$$

Suppose that $k < \bar{N}$. Then

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

Moreover, there exists a constant $C < \infty$ such that

$$D_k \leq C\ell_k^p.$$

The constants $\beta_\star$, $C_\star$, and $t_\star$ are specified in Section A.1.

Proof. Suppose $0 < t \leq t_\star$. Define

$$J_t(\bar{\mathbf{x}}_k) = \max_{b\in\mathcal{A}} \left\{ d_{ib}^{\mathsf{B}}\, \mathbb{1}\big\{S_i(\bar{\mathbf{x}}_k) - S_b(\bar{\mathbf{x}}_k) \leq t d_{ib}^{\mathsf{B}}\big\} \right\}.$$

We have

$$\begin{aligned} d_{k+1} &= d_{k+1}\mathbb{1}\big\{g_{k+1} \leq t d_{k+1}\big\} + d_{k+1}\mathbb{1}\big\{g_{k+1} > t d_{k+1}\big\} \\ &\leq J_t(\bar{\mathbf{x}}_k) + \frac{g_{k+1}}{t}. \end{aligned}$$

By the uniform bound in Assumption 3.3, $\mathbb{E}\big[J_t(\bar{\mathbf{x}}_k) \mid \mathcal{G}_k\big] \leq C_\star t^{\beta_\star}$. Therefore,

$$D_k \leq C_\star t^{\beta_\star} + \frac{\ell_k}{t}, \qquad 0 < t \leq t_\star. \tag{44}$$

Now set $\ell_\star = t_\star^{1+\beta_\star}$ and $d_{\max} = |\mathsf{B}|\bar{b}$. If $\ell_k = 0$, letting $t \to 0$ in (44) yields $D_k = 0$. If $0 < \ell_k \leq \ell_\star$, choosing $t = \ell_k^{1/(1+\beta_\star)}$ yields

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

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

$$D_k \le d_{\max} \le d_{\max} \ell_\star^{-p} \ell_k^p.$$

Thus, one common constant $C = \max\{C_\star + 1, d_{\max}\ell_\star^{-p}\}$ covers all three cases and yields $D_k \le C\ell_k^p$ for all $k < \bar{N}$. □

C.4 Bounding δ_k^{OPT} by Comparing the One-Period Residuals

From (42), to prove Proposition 6.2, we need to derive a lower bound on δ_k^{OPT} for $k < \bar{\tau}$. In this section, we do so by comparing δ_k^{OPT} with the fluid residual $\delta_{h_k}^{\mathrm{F}}$, whose behavior is characterized by Lemma A.4. This comparison yields the following lower bound on δ_k^{OPT}.

Lemma C.2. *For $k < \bar{\tau}$, the one-period residual δ_k^{OPT} satisfies*

$$\delta_k^{\mathrm{OPT}} \ge \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \frac{1}{2}\ell_k - C_1\left(h_k^{-1-\alpha} + h_k^{-1-\beta_\star}\right)$$

for some constant $C_1 < \infty$. Consequently, combining this bound with (42) *and the definition of q in* (43) *yields*

$$F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c}) \ge \mathbb{E}\sum_{k<\bar{\tau}} \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \frac{1}{2}q - C_2$$

for some constant $C_2 < \infty$.

Proof. Let $i = a^{\mathrm{F}}(\bar{\mathbf{x}}_k)$ and $a = \bar{a}_{k+1}$ be the static fluid action and the optimal feasible action, respectively, and let $\mathbf{b} = \mathbf{B}_{i,\mathsf{B}}$ and $\mathbf{y} = \bar{\mathbf{Y}}_{k+1,\mathsf{B}}$ denote their respective binding-resource consumptions. The successor states under these two actions are

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

From (21) and (41) we have:

$$\begin{aligned} \delta_k^{\mathrm{OPT}} - \delta_{h_k}^{\mathrm{F}}(\bar{\mathbf{x}}_k) &= \mathbb{E}_k\big[R_i - R_a + (h_k - 1)\{\phi_{\mathsf{B}}(\mathbf{z}_i) - \phi_{\mathsf{B}}(\mathbf{z}_a)\}\big], \\ \ell_k &= \mathbb{E}_k\big[R_i - R_a + \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)^{\mathrm{T}}(\mathbf{y} - \mathbf{b})\big], \end{aligned} \tag{45}$$

where ℓ_k is defined in (43), and, for notational convenience, we write $\mathbb{E}_k \triangleq \mathbb{E}[\cdot \mid \mathcal{G}_k]$. For $0 \le t \le 1$, define $\mathbf{z}_t = \mathbf{z}_a + t(\mathbf{z}_i - \mathbf{z}_a)$. By the fundamental theorem of calculus and $(h_k - 1)(\mathbf{z}_i - \mathbf{z}_a) = \mathbf{y} - \mathbf{b}$,

$$\begin{aligned}(h_k - 1)\big\{\phi_{\mathsf{B}}(\mathbf{z}_i) - \phi_{\mathsf{B}}(\mathbf{z}_a)\big\} &= (h_k - 1)\int_0^1 \nabla\phi_{\mathsf{B}}(\mathbf{z}_t)^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_a)\,\mathrm{d}t \\ &= \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)^{\mathrm{T}}(\mathbf{y} - \mathbf{b}) + \int_0^1 [\nabla\phi_{\mathsf{B}}(\mathbf{z}_t) - \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)]^{\mathrm{T}}(\mathbf{y} - \mathbf{b})\,\mathrm{d}t.\end{aligned}$$

Subtracting the second identity in (45) from the first therefore yields

$$\delta_k^{\mathrm{OPT}} - \delta_{h_k}^{\mathrm{F}}(\bar{\mathbf{x}}_k) - \ell_k = \mathbb{E}_k \int_0^1 [\nabla\phi_{\mathsf{B}}(\mathbf{z}_t) - \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)]^{\mathrm{T}}(\mathbf{y} - \mathbf{b})\,\mathrm{d}t.$$

We now bound the integrand. Since $\mathbf{z}_t$ is a convex combination of the two successor states, the analysis in Section C.2 implies $\|\mathbf{z}_t - \bar{\mathbf{x}}_k\| \le K_0/(h_k - 1)$. Moreover, the line segment joining $\bar{\mathbf{x}}_k$ and $\mathbf{z}_t$ lies in $\bar{B}(\mathbf{x}_0, 2r_0) \subseteq \widetilde{\mathcal{K}}_{\mathsf{B}}$. Integrating the Hessian along this segment gives

$$\begin{aligned}\nabla\phi_{\mathsf{B}}(\mathbf{z}_t) - \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k) &= \int_0^1 \nabla^2\phi_{\mathsf{B}}\big(\bar{\mathbf{x}}_k + u(\mathbf{z}_t - \bar{\mathbf{x}}_k)\big)(\mathbf{z}_t - \bar{\mathbf{x}}_k)\,\mathrm{d}u, \\ \|\nabla\phi_{\mathsf{B}}(\mathbf{z}_t) - \nabla\phi_{\mathsf{B}}(\bar{\mathbf{x}}_k)\| &\le M_H\|\mathbf{z}_t - \bar{\mathbf{x}}_k\| \le \frac{M_H K_0}{h_k - 1}.\end{aligned}$$

Consequently,

$$\big|\delta_k^{\mathrm{OPT}} - \delta_{h_k}^{\mathrm{F}}(\bar{\mathbf{x}}_k) - \ell_k\big| \le \frac{M_H K_0}{h_k - 1}\mathbb{E}_k\|\mathbf{y} - \mathbf{b}\|_1 = \frac{M_H K_0}{h_k - 1}\,D_k.$$

Since $\bar{\mathbf{x}}_k \in B(\mathbf{x}_0, r_0)$ on $\{k < \bar{\tau}\}$, the preceding inequality and Lemma A.4 imply

$$\delta_k^{\mathrm{OPT}} \ge \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \ell_k - \frac{M_H K_0}{h_k - 1}\,D_k - C h_k^{-1-\alpha} \tag{46}$$

for some constant $C < \infty$.

Now let $L \triangleq 2\max\{M_H K_0, 1\}$ as defined in Section C.2, and set $t = L/(h_k - 1)$. Since $n_0 \ge L/t_\star$ by Section C.2, we have $0 < t \le t_\star$ for $k < \bar{\tau}$. Therefore, Lemma C.1 yields

$$\begin{aligned}\frac{M_H K_0}{h_k - 1}\,D_k &\le \frac{M_H K_0}{h_k - 1}\,C_\star\left(\frac{L}{h_k - 1}\right)^{\beta_\star} + \frac{M_H K_0}{h_k - 1}\frac{\ell_k}{t} \\ &= \left(M_H K_0 C_\star L^{\beta_\star}\right)(h_k - 1)^{-1-\beta_\star} + \frac{M_H K_0}{L}\ell_k \\ &\le \frac{1}{2}\ell_k + C' h_k^{-1-\beta_\star}\end{aligned}$$

for some constant $C' < \infty$. The second inequality follows from $M_H K_0/L \le 1/2$ and $h_k/(h_k - 1) \le 2$

since $h_k \geq n_0 + 1 \geq 2$ for $k < \bar{\tau}$. Together with (46), by taking $C_1 = \max\{C, C'\}$, we have

$$\delta_k^{\mathrm{OPT}} \geq \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} + \frac{1}{2}\ell_k - C_1 \left(h_k^{-1-\alpha} + h_k^{-1-\beta_\star} \right) \quad k < \bar{\tau}.$$

Finally, since $h_k = n - k \geq n_0 + 1$ before stopping, the two remainders are uniformly summable along every sample path:

$$\sum_{k < \bar{\tau}} \left(h_k^{-1-\alpha} + h_k^{-1-\beta_\star} \right) \leq \sum_{h = n_0 + 1}^{\infty} \left(h^{-1-\alpha} + h^{-1-\beta_\star} \right) < \infty.$$

This yields the second inequality in Lemma C.2. $\square$

C.5 Moment Bound for the Stopped-State Deviation

In Section C.6, we bound the cumulative effect of replacing the path-dependent coefficients $\Gamma(\bar{\mathbf{x}}_k)$ appearing in δ_k^{OPT} with the fixed coefficient $\Gamma(\mathbf{x}_0)$. As a preliminary step, we first bound the γ-moment deviation $\mathbb{E}\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\|^\gamma$ of the stopped state $\bar{\mathbf{x}}_{k\wedge\bar{\tau}}$ from the initial state $\mathbf{x}_0$, as stated in Lemma C.3.

Lemma C.3 (Bounding γ-Moment Deviation). *There exists a constant $C_3 < \infty$ such that, for any $k \leq \bar{N}$,*

$$\mathbb{E}\Big(\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\|^\gamma \Big) \leq C_3 \left(h_k^{-\gamma/2} + \big(q/h_k\big)^{p\gamma} \right),$$

where p is defined in Lemma C.1 and q is defined in (43).

Proof. Let $\mathbf{b}_{k+1} \triangleq \mathbf{B}_{a^{\mathrm{F}}(\bar{\mathbf{x}}_k),\mathsf{B}}$ denote the consumption of the static fluid action $a^{\mathrm{F}}(\bar{\mathbf{x}}_k)$ at state $\bar{\mathbf{x}}_k$. We have

$$\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k = \frac{\bar{\mathbf{x}}_k - \bar{\mathbf{Y}}_{k+1,\mathsf{B}}}{h_k - 1} = \frac{\bar{\mathbf{x}}_k - \mathbf{b}_{k+1}}{h_k - 1} + \frac{\mathbf{b}_{k+1} - \bar{\mathbf{Y}}_{k+1,\mathsf{B}}}{h_k - 1}.$$

The first term has conditional mean zero by (18), and the second captures the realized consumption difference controlled in Appendix C.3. Define the stopped increments by

$$\Delta\bar{\mathbf{M}}_{k+1} = \begin{cases} \big(\bar{\mathbf{x}}_k - \mathbf{b}_{k+1}\big)/\big(h_k - 1\big), & k < \bar{\tau}, \\ \mathbf{0}, & k \geq \bar{\tau}, \end{cases}$$

$$\Delta\bar{\mathbf{A}}_{k+1} = \begin{cases} \big(\mathbf{b}_{k+1} - \bar{\mathbf{Y}}_{k+1,\mathsf{B}}\big)/\big(h_k - 1\big), & k < \bar{\tau}, \\ \mathbf{0}, & k \geq \bar{\tau}. \end{cases}$$

Define $\bar{\mathbf{M}}_k = \sum_{j<k} \Delta\bar{\mathbf{M}}_{j+1}$ and $\bar{\mathbf{A}}_k = \sum_{j<k} \Delta\bar{\mathbf{A}}_{j+1}$. Summing the stopped increments gives, for $0 \le k \le \bar{N}$,

$$\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0 = \bar{\mathbf{M}}_k + \bar{\mathbf{A}}_k. \tag{47}$$

Bounding the Moment of $\bar{\mathbf{M}}_k$ On the event $\{k < \bar{\tau}\}$, we have

$$\mathbb{E}[\mathbf{b}_{k+1} \mid \mathcal{G}_k] = \bar{\mathbf{x}}_k, \qquad \mathbb{E}[\Delta\bar{\mathbf{M}}_{k+1} \mid \mathcal{G}_k] = \frac{\bar{\mathbf{x}}_k - \mathbb{E}[\mathbf{b}_{k+1} \mid \mathcal{G}_k]}{h_k - 1} = \mathbf{0},$$

so $\{\bar{\mathbf{M}}_k\}$ is a martingale. For $\ell < j$, the martingale difference $\Delta\bar{\mathbf{M}}_{\ell+1}$ is $\mathcal{G}_j$-measurable. Hence,

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

Before stopping, each coordinate of $\mathbf{b}_{k+1}$ and its conditional mean $\bar{\mathbf{x}}_k$ lies in $[0, \bar{b}]$. Therefore,

$$\|\Delta\bar{\mathbf{M}}_{k+1}\| \le \frac{\sqrt{|\mathsf{B}|}\,\bar{b}}{h_k - 1}.$$

Expanding the squared norm and using the orthogonality of the martingale differences, we obtain

$$\begin{aligned} \mathbb{E}\|\bar{\mathbf{M}}_k\|^2 &= \sum_{j=0}^{k-1} \mathbb{E}\|\Delta\bar{\mathbf{M}}_{j+1}\|^2 \\ &\le |\mathsf{B}|\,\bar{b}^2 \sum_{j=0}^{k-1} (h_j - 1)^{-2} = |\mathsf{B}|\,\bar{b}^2 \sum_{s=h_k}^{n-1} s^{-2} \le C/h_k. \end{aligned} \tag{48}$$

For the last inequality, since $h_k \ge n_0 + 1 \ge 2$, $\sum_{s=h_k}^{n-1} s^{-2} \le \int_{h_k-1}^{\infty} t^{-2}\,\mathrm{d}t = 1/(h_k - 1) \le 2/h_k$.

Bounding the Moment of $\bar{\mathbf{A}}_k$ By the definition of D_k in (43),

$$\mathbb{E}\big[\|\Delta\bar{\mathbf{A}}_{k+1}\| \mid \mathcal{G}_k\big] \le \frac{D_k}{h_k - 1}.$$

Set $u_k \triangleq \mathbb{E}\ell_k$ for $0 \le k < \bar{N}$. By the definitions of ℓ_k and q in (43),

$$\sum_{k=0}^{\bar{N}-1} u_k = \mathbb{E}\sum_{k=0}^{\bar{N}-1} \ell_k = \mathbb{E}\sum_{k<\bar{\tau}} \ell_k = q.$$

From Lemma C.1, we have

$$\mathbb{E}\|\bar{\mathbf{A}}_k\| \le \sum_{j<k} \mathbb{E}\|\Delta\bar{\mathbf{A}}_{j+1}\| \le \sum_{j<k} \frac{\mathbb{E}D_j}{h_j-1} \le C\sum_{j<k} \frac{\mathbb{E}\,\ell_j^p}{h_j-1} \le C\sum_{j<k} \frac{u_j^p}{h_j-1},$$

where the last inequality follows from the concavity of t^p and Jensen's inequality.

For $k \ge 1$, apply Hölder's inequality to u_j^p and $(h_j-1)^{-1}$ with conjugate exponents $1/p$ and $1/(1-p)$ gives

$$\sum_{j<k} \frac{u_j^p}{h_j-1} \le \left(\sum_{j<k} u_j\right)^p \left(\sum_{j<k} (h_j-1)^{-1/(1-p)}\right)^{1-p} \le q^p \left(\sum_{s=h_k}^{n-1} s^{-1/(1-p)}\right)^{1-p}.$$

Since $1/(1-p) = 1+\beta_\star$ and $\beta_\star(1-p) = p$, an integral comparison yields

$$\sum_{s=h_k}^{n-1} s^{-1/(1-p)} \le \int_{h_k-1}^{\infty} t^{-1-\beta_\star}\,\mathrm{d}t = \frac{(h_k-1)^{-\beta_\star}}{\beta_\star} \le C h_k^{-\beta_\star},$$

$$\left(\sum_{s=h_k}^{n-1} s^{-1/(1-p)}\right)^{1-p} \le C h_k^{-\beta_\star(1-p)} = C h_k^{-p}.$$

Combining the above inequalities and noting that $\bar{\mathbf{A}}_0 = \mathbf{0}$, we obtain

$$\mathbb{E}\|\bar{\mathbf{A}}_k\| \le C\left(\frac{q}{h_k}\right)^p, \qquad 0 \le k \le \bar{N}. \tag{49}$$

Bounding the Moment of Deviation Using (47) – (49), we have

$$\mathbb{E}\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\| \le \mathbb{E}\|\bar{\mathbf{M}}_k\| + \mathbb{E}\|\bar{\mathbf{A}}_k\| \le (\mathbb{E}\|\bar{\mathbf{M}}_k\|^2)^{1/2} + \mathbb{E}\|\bar{\mathbf{A}}_k\| \le C\big(h_k^{-1/2} + (q/h_k)^p\big).$$

Finally, since $0 < \gamma \le 1$, Jensen's inequality applied to the concave function t^γ, together with $(a+b)^\gamma \le a^\gamma + b^\gamma$, yields

$$\mathbb{E}\Big(\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\|^\gamma\Big) \le \Big(\mathbb{E}\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\|\Big)^\gamma \le C\Big(h_k^{-\gamma/2} + (q/h_k)^{p\gamma}\Big). \qquad \square$$

C.6 Controlling Coefficient Variation and Completing the Proof

We now use the moment bound established in Lemma C.3 to control the cumulative effect of replacing the path-dependent coefficients $\Gamma(\bar{\mathbf{x}}_k)$ appearing in δ_k^{OPT} with the fixed coefficient $\Gamma(\mathbf{x}_0)$. The resulting bound is stated in Lemma C.4, whose proof is deferred to the end of this section. We then use Lemma C.4 to complete the proof of Proposition 6.2.

Lemma C.4. *There exists a constant $C_4 < \infty$ such that*

$$\mathbb{E}\sum_{k<\bar{\tau}} \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} \geq \Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1} - C_4\left(1 + q^{p\gamma}\right).$$

Combining Lemmas C.2 and C.4, we obtain

$$F_n(\mathbf{c}) - V_n^{\mathrm{OPT}}(\mathbf{c}) \geq \Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1} + \tfrac{1}{2}q - C\left(1 + q^{p\gamma}\right)$$

for some constant $C < \infty$. To complete the proof of Proposition 6.2, it remains to show that $\frac{1}{2}q - C\big(1 + q^{p\gamma}\big)$ is uniformly bounded below over $q \geq 0$, which we do now.

Set $\eta = p\gamma \in (0,1)$ and $q_* = \max\big\{1, (4C)^{1/(1-\eta)}\big\}$. First suppose $q \geq q_*$. Then $q^{1-\eta} \geq 4C$, so $Cq^\eta \leq q/4$. Therefore,

$$\tfrac{1}{2}q - Cq^\eta \geq q/4 \geq 0.$$

If instead $0 \leq q < q_*$, then $Cq^\eta \leq Cq_*^\eta$, and thus

$$\tfrac{1}{2}q - Cq^\eta \geq -Cq_*^\eta.$$

Combining the two cases, there exists a constant $C_{\mathrm{OPT}} < \infty$ such that

$$V_n^{\mathrm{OPT}}(\mathbf{c}) \leq F_n(\mathbf{c}) - \Gamma(\mathbf{x}_0)\mathsf{H}_{n-1} + C_{\mathrm{OPT}}, \qquad n \geq n_0 + 1.$$

For $1 \leq n \leq n_0$, we have $V_n^{\mathrm{OPT}}(\mathbf{c}) \leq F_n(\mathbf{c}) \leq n_0\bar{r}$, and $|\Gamma(\mathbf{x}_0)\,\mathsf{H}_{n-1}| \leq C_\Gamma \mathsf{H}_{n_0-1}$ by Lemma A.2. Increasing C_{OPT} to cover these finitely many horizons completes the proof of Proposition 6.2.

Proof of Lemma C.4. Let constants C_Γ and L_Γ be the uniform upper bound and Hölder constant in Lemma A.2, and set

$$\bar{C} \triangleq \max\left\{L_\Gamma, C_\Gamma r_0^{-\gamma}\right\}.$$

Define

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

We first show that

$$\hat{\Gamma}_k \geq \Gamma(\mathbf{x}_0) - \bar{C}\left\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\right\|^\gamma, \quad 0 \leq k < \bar{N}. \tag{50}$$

Proof of Inequality (50). If $k < \bar{\tau}$, then $\bar{\mathbf{x}}_k$ and $\mathbf{x}_0$ are in $\bar{B}(\mathbf{x}_0, 2r_0)$ by Section C.2. Hence,

$$\hat{\Gamma}_k = \Gamma(\bar{\mathbf{x}}_k) \geq \Gamma(\mathbf{x}_0) - L_\Gamma \left\|\bar{\mathbf{x}}_k - \mathbf{x}_0\right\|^\gamma \geq \Gamma(\mathbf{x}_0) - \bar{C} \left\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\right\|^\gamma.$$

Otherwise, $\bar{\tau} \leq k < \bar{N}$, which implies $h_{\bar{\tau}} \geq n_0 + 1$. Hence, stopping must occur through early exit, so $\|\bar{\mathbf{x}}_{\bar{\tau}} - \mathbf{x}_0\| \geq r_0$. Therefore,

$$\Gamma(\mathbf{x}_0) - \bar{C} \left\|\bar{\mathbf{x}}_{k\wedge\bar{\tau}} - \mathbf{x}_0\right\|^\gamma = \Gamma(\mathbf{x}_0) - \bar{C} \left\|\bar{\mathbf{x}}_{\bar{\tau}} - \mathbf{x}_0\right\|^\gamma \leq C_\Gamma - \bar{C} r_0^\gamma \leq 0 = \hat{\Gamma}_k. \qquad \square$$

Taking expectations in (50) and applying Lemma C.3 yield

$$\mathbb{E}\hat{\Gamma}_k \geq \Gamma(\mathbf{x}_0) - C\left(h_k^{-\gamma/2} + (q/h_k)^{p\gamma}\right). \tag{51}$$

Note that

$$\sum_{k=0}^{\bar{N}-1} \frac{1}{h_k - 1} = \sum_{s=n_0}^{n-1} \frac{1}{s} = \mathsf{H}_{n-1} - \mathsf{H}_{n_0-1}. \tag{52}$$

Moreover, for either $a = \gamma/2$ or $a = p\gamma$, we have $a > 0$ and hence

$$\sum_{h=n_0+1}^{n} \frac{h^{-a}}{h-1} \leq 2 \sum_{h=n_0+1}^{\infty} h^{-1-a} \leq 2 \int_{n_0}^{\infty} t^{-1-a}\, \mathrm{d}t = \frac{2n_0^{-a}}{a}. \tag{53}$$

Combining (51) – (53), we have

$$\begin{aligned}
\mathbb{E} \sum_{k<\bar{\tau}} \frac{\Gamma(\bar{\mathbf{x}}_k)}{h_k - 1} &= \sum_{k<\bar{N}} \frac{\mathbb{E}\hat{\Gamma}_k}{h_k - 1} \\
&\geq \Gamma(\mathbf{x}_0)\left(\mathsf{H}_{n-1} - \mathsf{H}_{n_0-1}\right) - C \sum_{h=n_0+1}^{n} \frac{h^{-\gamma/2} + (q/h)^{p\gamma}}{h-1} \\
&\geq \Gamma(\mathbf{x}_0)\mathsf{H}_{n-1} - C_4\left(1 + q^{p\gamma}\right)
\end{aligned}$$

for some constant $C_4 < \infty$, where the first inequality follows from (51) and (52), and the second inequality from (53) and the uniform bound $|\Gamma(\mathbf{x}_0)\, \mathsf{H}_{n_0-1}| \leq C_\Gamma\, \mathsf{H}_{n_0-1}$. $\square$

D Proofs and Additional Details for Sections 2 and 3

D.1 Further Details of Lemma 2.1

D.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

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

D.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

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

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:

$$\begin{aligned} V_n^{\mathrm{OFF}}(\mathbf{c}) &\le \inf_{\boldsymbol{\lambda}\in\mathbb{R}_+^m}\left\{\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{c} + n\,\mathcal{V}(\boldsymbol{\lambda})\right\} \\ &= n \inf_{\boldsymbol{\lambda}\in\mathbb{R}_+^m}\left\{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\rho} + \mathcal{V}(\boldsymbol{\lambda})\right\} \\ &= n\,\phi^{\mathrm{F}}(\boldsymbol{\rho})\,, \end{aligned}$$

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

D.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.

D.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),

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

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 (4), 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.

D.4 Proof of Proposition 3.3

In the proof, we set $\kappa = 2\bar{b}$, where $\bar{b}$ is the constant from Assumption 2.1. For each period t, let $\mathbf{Y}_t = (Y_{tj})_{j\in[m]} \triangleq \mathbf{B}_{a_t^{\mathrm{G}},t} \in \mathbb{R}_+^m$ denote the resource-consumption vector associated with action a^{G} in period t. Then, $\mathbb{E}[\mathbf{Y}_t] = \boldsymbol{\mu}$ for every t.

To prove the performance gap, define the safety threshold by

$$h_n \triangleq n\delta_n - \bar{b}.$$

For each $j \in [m]$, define the cumulative consumption deviation by

$$S_{kj} \triangleq \sum_{t=1}^{k}\left\{Y_{tj} - \mu_j\right\}.$$

Finally, define the corresponding stopping time by

$$\tau_n \triangleq \min\Big\{k \in \{1,\ldots,n-1\} : S_{kj} > h_n \text{ for some } j \in [m]\Big\},$$

with the convention $\tau_n = n$ if the set is empty.

Note that the constrained and unconstrained greedy policies take the same action and receive the same reward up to period τ_n. To see this, suppose that the two policies have agreed during the first $k < \tau_n$ periods. For each $j \in [m]$, the remaining capacity satisfies

$$c_{nj} - \sum_{t=1}^{k} Y_{tj} = (n-k)\mu_j + n\left(\rho_{nj} - \mu_j\right) - S_{kj} \geq (n-k)\mu_j + n\delta_n - S_{kj} \geq \bar{b}\,,$$

where the first inequality uses the definition of δ_n and the second uses $S_{kj} \le h_n$. By Assumption 2.1, every action consumes at most $\bar{b}$ units of each resource. Thus, every action is feasible in period $k+1$, so the constrained greedy policy also selects a^{G}_{k+1}.

Since the constrained and unconstrained greedy policies agree up to period τ_n, and the one-period loss is at most $(\max_{a\in\mathcal{A}} R_a)^+$, whose expectation is at most $\bar{r}$ under Assumption 2.1, we have

$$0 \le n\,\mathbb{E}\Big[\max_a R_a\Big] - V^{\mathrm{G}}_n(\mathbf{c}_n) \le \bar{r}\cdot\mathbb{E}\,(n-\tau_n)^+ \le n\bar{r}\cdot\mathbb{P}\{\tau_n < n\}.$$

Moreover, the tail probability can be bounded as:

$$\begin{aligned}
\mathbb{P}\{\tau_n < n\} &= \mathbb{P}\left\{\cup_{j\in[m]}\left\{\max_{k\le n-1} S_{kj} > h_n\right\}\right\}\\
&\le \sum_{j\in[m]}\mathbb{P}\left\{\max_{k\le n-1} S_{kj} > h_n\right\}\\
&\le m\exp\left(-\frac{2h_n^2}{n\bar{b}^2}\right)\\
&\le \frac{m}{n^2}.
\end{aligned}$$

The first inequality follows from the union bound. The second inequality follows from the maximal Azuma-Hoeffding inequality (see Lemma D.1) because, for each resource $j\in[m]$, the deviations $Y_{tj}-\mu_j$ are independent and have mean zero, with $Y_{tj}-\mu_j \in [-\mu_j, \bar{b}-\mu_j]$. The third inequality follows from the definition of h_n, the choice $\kappa = 2\bar{b}$, and the bound $(2\sqrt{n\ln n}-1)^2 \ge n\ln n$ for every $n\in\mathbb{N}_+$. Combining this tail estimate with the performance bound gives

$$0 \le n\mathbb{E}\left[\max_{a\in\mathcal{A}} R_a\right] - V^{\mathrm{G}}_n(\mathbf{c}_n) \le \frac{m\bar{r}}{n} = o(1).$$

This proves the proposition.

Lemma D.1 (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 $t\ge1$, 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 $t \geq 1$,

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

E Proofs and Additional Details for Section 5

E.1 Further Details on the Uniform Multisecretary Model in Section 5.1

We restrict attention to $\rho \in \mathcal{U} \triangleq (0,1)$. We first derive the fluid value function $\phi^{\mathrm{F}}(\rho)$. For any dual variable $\lambda \geq 0$, we have:

$$\mathcal{V}(\lambda) = \mathbb{E}[(V-\lambda)^+] = \begin{cases} \frac{1}{2}(1-\lambda)^2 & 0 \leq \lambda \leq 1 \\ 0 & \lambda \geq 1. \end{cases}$$

Thus,

$$\phi^{\mathrm{F}}(\rho) = \min_{\lambda\geq 0}\{\lambda\rho + \mathcal{V}(\lambda)\} = \min_{\lambda\in[0,1]}\left\{\lambda\rho + \frac{1}{2}(1-\lambda)^2\right\} = \rho - \frac{1}{2}\rho^2,$$

where the optimal dual variable is $\lambda^*(\rho) = 1-\rho$.

We next verify that Assumption 2.1 and Assumptions 3.1 to 3.3 hold. Assumption 2.1 holds with $\bar{r} = \bar{b} = 1$. Because ϕ^{F} is C^∞, Assumption 3.1 holds with binding-resource set $\mathsf{B} = \{1\}$ and $\phi_{\mathsf{B}}(\rho) = \phi^{\mathrm{F}}(\rho)$. It remains to verify Assumptions 3.2 and 3.3.

For Assumption 3.2, the static fluid action in (4) is

$$a^{\mathrm{F}}(\rho) = \mathbb{1}\{V > 1-\rho\}.$$

Direct integration yields

$$\mathbb{E}\left[B_{a^{\mathrm{F}}(\rho)}\right] = \mathbb{P}(V > 1-\rho) = \rho.$$

Therefore, Assumption 3.2 holds.

For Assumption 3.3, the only top-versus-competitor gap is $S_{a^{\mathrm{F}}}(\rho) - S_{1-a^{\mathrm{F}}}(\rho) = |V-(1-\rho)|$, and its weight is $d_{01} = 1$. Thus, uniformly over $\rho \in \mathcal{K}$,

$$\mathbb{E}\left[d_{a^{\mathrm{F}}(\rho),1-a^{\mathrm{F}}(\rho)}\mathbb{1}\left\{S_{a^{\mathrm{F}}(\rho)}(\rho) - S_{1-a^{\mathrm{F}}(\rho)}(\rho) \leq t\, d_{a^{\mathrm{F}}(\rho),1-a^{\mathrm{F}}(\rho)}\right\}\right] = \mathbb{P}(|V-(1-\rho)| \leq t) \leq 2t.$$

Therefore, Assumption 3.3 holds with $\beta_{\mathcal{K}} = 1$ and $C_{\mathcal{K}} = 2$.

Finally, we derive the logarithmic coefficient in Theorem 4.1. Note that $B_{a^{\mathrm{F}}(\rho)} = a^{\mathrm{F}}(\rho)$, which is Bernoulli with parameter ρ. Therefore,

$$\Sigma(\rho) = \mathbb{V}\mathrm{ar}\left[B_{a^{\mathrm{F}}(\rho)}\right] = \rho(1-\rho), \quad \Gamma(\rho) = \frac{1}{2}\rho(1-\rho).$$

E.2 Further Details on the Stochastic Knapsack Problem in Section 5.2

Let W and V denote the weight and profit of an item, respectively. For a threshold $m \geq 0$, define

$$F_a(m) \triangleq \mathbb{E}[W\mathbb{1}\{V \geq mW\}], \quad F_c(m) \triangleq \mathbb{E}[V\mathbb{1}\{V \geq mW\}].$$

Therefore, $F_a(m)$ is the expected capacity consumed by accepting all items with a profit-to-weight ratio of at least m, and $F_c(m)$ is the corresponding expected reward.

Condition 1 of Lueker (1998) assumes that, for some finite d, $0 \leq W, V \leq d$ with probability one. Condition 2 assumes that there is an $m_0 > 0$ satisfying $\beta_0 = F_a(m_0)$ and that, on a neighborhood N_m of m_0, the derivative $f_a \triangleq F_a'$ exists, is strictly negative, and is Lipschitz continuous.

Since $f_a(m_0) < 0$ and f_a is Lipschitz continuous, we may shrink N_m so that there exist constants $0 < \underline{\alpha} \leq \overline{\alpha} < \infty$ such that

$$\underline{\alpha} \leq -f_a(m) \leq \overline{\alpha}, \qquad m \in N_m.$$

Thus, function F_a is strictly decreasing on N_m. Choose an open neighborhood $\mathcal{U}$ of β_0 such that $\overline{\mathcal{U}} \Subset F_a(N_m)$, and define

$$m_\rho \triangleq F_a^{-1}(\rho), \qquad \alpha_\rho \triangleq -f_a(m_\rho) > 0, \qquad \rho \in \mathcal{U}.$$

After shrinking $\mathcal{U}$ if necessary, the compact set $\{m_\rho : \rho \in \overline{\mathcal{U}}\}$ lies in the interior of N_m.

In Lemma E.1, we show that the boundary events are immaterial under Lueker's Conditions 1 and 2; specifically, changing the acceptance rule from $V \geq mW$ to $V > mW$ changes neither expected consumption nor expected reward.

Lemma E.1. *Under Conditions 1 and 2 of Lueker (1998), for any $m \in N_m$,*

$$W\mathbb{1}\{V = mW\} = 0, \;\; V\mathbb{1}\{V = mW\} = 0, \;\; W^2\,\mathbb{1}\{V = mW\} = 0, \quad a.s.$$

Lemma E.1 implies that $\frac{d}{d\lambda}\mathbb{E}[(V - \lambda W)^+] = -F_a(\lambda)$, as stated in Lemma E.2.

Lemma E.2. *For any $\lambda \in N_m$,*

$$\frac{d}{d\lambda}\mathbb{E}[(V-\lambda W)^+] = -\mathbb{E}\big[W\mathbb{1}\{V \geq \lambda W\}\big] = -F_a\big(\lambda\big).$$

The rest of this section is organized as follows. In Section E.2.1, we derive the fluid relaxation and characterize the optimal dual variable for the stochastic knapsack problem. Section E.2.2 shows that our dynamic fluid policy coincides with the OnLineGreedy policy in Lueker (1998). Section E.2.3 verifies that Conditions 1 and 2 of Lueker (1998) imply our Assumption 2.1 and Assumptions 3.1 to 3.3. Section E.2.4 demonstrates that the first-order and logarithmic terms in Lueker's expansion coincide with ours in Theorem 4.1. Finally, Section E.2.5 analyzes the special case with uniformly distributed weights and values.

Proof of Lemma E.1. For every $m \in N_m$ and every $h > 0$ such that $m + h \in N_m$, we have

$$F_a(m) - F_a(m+h) = \mathbb{E}\big[W\mathbb{1}\{mW \leq V < (m+h)W\}\big].$$

Letting $h \downarrow 0$ yields

$$\begin{aligned}\mathbb{E}[W\mathbb{1}\{V = mW\}] &= \lim_{h\downarrow 0}\mathbb{E}\big[W\mathbb{1}\{mW \leq V < (m+h)W\}\big] \\ &= \lim_{h\downarrow 0}\{F_a(m) - F_a(m+h)\} = 0,\end{aligned}$$

where the first equality follows from the dominated convergence theorem and the second equality from the continuity of F_a. Moreover, since $W \leq d$ almost surely by Condition 1, we have

$$0 \leq \mathbb{E}\big[W^2\mathbb{1}\{V = mW\}\big] \leq d \cdot \mathbb{E}[W\mathbb{1}\{V = mW\}] = 0\,.$$

This proves $\mathbb{E}[W^2\mathbb{1}\{V = mW\}] = 0$. Finally, because $V = mW$ on the boundary event, we have

$$\mathbb{E}[V\mathbb{1}\{V = mW\}] = m \cdot \mathbb{E}[W\mathbb{1}\{V = mW\}] = 0.$$

Since $W\mathbb{1}\{V = mW\}$, $V\mathbb{1}\{V = mW\}$, and $W^2\mathbb{1}\{V = mW\}$ are nonnegative and have zero expectations, they must each be zero almost surely. $\square$

Proof of Lemma E.2. Fix $\lambda \in N_m$. For every nonzero h,

$$\Big|\big(V-(\lambda+h)W\big)^+ - \big(V-\lambda W\big)^+\Big| \leq |h|W.$$

Dividing by $|h|$ gives

$$\left| \frac{\big(V - (\lambda + h)W\big)^+ - \big(V - \lambda W\big)^+}{h} \right| \leq W.$$

Because $0 \leq W \leq d$ almost surely, the dominated convergence theorem gives

$$\begin{aligned}
\frac{d}{d\lambda}\mathbb{E}[(V - \lambda W)^+] &= \lim_{h \to 0} \mathbb{E}\left\{ \frac{\big(V - (\lambda + h)W\big)^+ - \big(V - \lambda W\big)^+}{h} \right\} \\
&= \mathbb{E}\left\{ \lim_{h \to 0} \frac{\big(V - (\lambda + h)W\big)^+ - \big(V - \lambda W\big)^+}{h} \right\} \\
&= \mathbb{E}\left[\frac{d}{d\lambda}(V - \lambda W)^+ \right] = -\mathbb{E}\big[W \mathbb{1}\{V \geq \lambda W\}\big] = -F_a\big(\lambda\big).
\end{aligned}$$

The third and fourth equalities follow from the identity

$$\frac{d}{d\lambda}(V - \lambda W)^+ = -W\mathbb{1}\{V \geq \lambda W\}, \;\; a.s.$$

Indeed, the derivative exists almost surely because $W\mathbb{1}\{V = \lambda W\} = 0$ almost surely by Lemma E.1.

□

E.2.1 Fluid Relaxation and Optimal Dual Variable

Fix $\rho \in \mathcal{U}$. Because the action set is binary, an opportunity-dependent randomized control can be represented by a measurable function $x : \mathbb{R}_+^2 \to [0, 1]$, where $x(W, V)$ is the conditional probability of accepting an item with weight W and profit V. Specializing the fluid relaxation in (1) to this accept–reject problem gives

$$\begin{aligned}
\phi^{\mathrm{F}}(\rho) = \sup_{x : \mathbb{R}_+^2 \to [0,1]} \quad & \mathbb{E}[V x(W, V)] \\
\text{subject to} \quad & \mathbb{E}[W x(W, V)] \leq \rho.
\end{aligned}$$

Let $\lambda \geq 0$ be a multiplier for the expected-capacity constraint. The dual problem in (2) becomes

$$\phi^{\mathrm{F}}(\rho) = \inf_{\lambda \geq 0} \big\{ \lambda\rho + \mathbb{E}[(V - \lambda W)^+] \big\} = \inf_{\lambda \geq 0} g_\rho(\lambda),$$

where $g_\rho(\lambda) \triangleq \lambda\rho + \mathbb{E}[(V - \lambda W)^+]$. Note that the function g_ρ is convex. Moreover, for $\lambda \in N_m$, Lemma E.2 yields

$$g_\rho'(\lambda) = \rho - F_a(\lambda).$$

By the first-order condition, the optimal dual variable $\lambda^*(\rho)$ is uniquely given by

$$\lambda^*(\rho) = F_a^{-1}(\rho) = m_\rho > 0. \tag{54}$$

Substituting this minimizer into the dual objective yields

$$\begin{aligned} \phi^{\mathrm{F}}(\rho) = g_\rho(m_\rho) &= m_\rho \rho + \mathbb{E}[(V - m_\rho W)^+] \\ &= m_\rho F_a(m_\rho) + F_c(m_\rho) - m_\rho F_a(m_\rho) \\ &= F_c(m_\rho), \end{aligned} \tag{55}$$

where the second equality uses $F_a(m_\rho) = \rho$, the fact that $\mathbb{E}[(V - m_\rho W)^+] = \mathbb{E}[(V - m_\rho W)\,\mathbb{1}\{V > m_\rho W\}]$, and the definitions of F_a and F_c.

In addition,

$$x_\rho^*(W, V) \triangleq \mathbb{1}\{V > m_\rho W\}.$$

is an optimal primal solution to the fluid relaxation. This threshold control exhausts the expected capacity

$$\mathbb{E}[W x_\rho^*(W, V)] = \mathbb{E}[W \mathbb{1}\{V > m_\rho W\}] = F_a(m_\rho) = \rho, \tag{56}$$

and its expected reward is

$$\mathbb{E}[V x_\rho^*(W, V)] = F_c(m_\rho) = \phi^{\mathrm{F}}(\rho).$$

Finally, from the envelope theorem we have

$$\phi^{\mathrm{F}\prime}(\rho) = m_\rho = F_a^{-1}(\rho). \tag{57}$$

E.2.2 Policy Correspondence

Consider a state with k periods remaining and capacity c, where $c/k \in \mathcal{U}$. The dynamic fluid policy re-solves the fluid problem at the normalized remaining capacity c/k and uses the optimal dual price

$$\lambda^*(c/k) = m_{c/k} = F_a^{-1}(c/k)$$

according to (54). Therefore, our dynamic fluid policy and the OnLineGreedy policy of Lueker (1998) use the same optimal fluid dual price and hence the same acceptance threshold.

E.2.3 Verification of Assumptions

In this section, we verify that Conditions 1 and 2 of Lueker (1998) imply Assumption 2.1 and Assumptions 3.1 to 3.3. First, Condition 1 implies Assumption 2.1 with $\bar{r} = \bar{b} = d$. We then set $\mathsf{B} = \{1\}$ and $\phi_{\mathsf{B}}(\rho) = \phi^{\mathrm{F}}(\rho)$ and verify that Assumptions 3.1 to 3.3 are satisfied.

To verify Assumption 3.1, from (57) and the inverse function theorem we have

$$\phi''_{\mathsf{B}}(\rho) = \frac{1}{f_a(m_\rho)} = -\frac{1}{\alpha_\rho}. \tag{58}$$

Since f_a is Lipschitz and bounded away from zero on N_m, the map $\rho \mapsto 1/f_a(m_\rho)$ is also locally Lipschitz. Thus, $\phi_{\mathsf{B}} \in C^{2,1}(\mathcal{U})$, which verifies Assumption 3.1 with Hölder exponent $\alpha = 1$.

We now verify Assumption 3.2. The static fluid action in (4) is $a^{\mathrm{F}}(\rho) = x^*_\rho(W, V) = \mathbb{1}\{V > m_\rho W\}$. Equation (56) gives $\mathbb{E}\big[B_{a^{\mathrm{F}}(\rho)}\big] = \rho$. Hence, Assumption 3.2 holds.

Finally, we verify Assumption 3.3. With a binary action set, the consumption difference between the two actions is W, and the top-versus-competitor score gap is $|V - m_\rho W|$, as m_ρ is the optimal dual variable by (54). Fix a compact set $\mathcal{K} \Subset \mathcal{U}$ and choose $t_{\mathcal{K}} > 0$ so that $m_\rho \pm t \in N_m$ for every $\rho \in \mathcal{K}$ and $0 \le t \le t_{\mathcal{K}}$. We have

$$\mathbb{E}\big[W\mathbb{1}\{|V - m_\rho W| \le tW\}\big] = F_a(m_\rho - t) - F_a(m_\rho + t) = -\int_{m_\rho - t}^{m_\rho + t} f_a(s)\,\mathrm{d}s \le 2\overline{\alpha}t.$$

Hence, Assumption 3.3 holds with $\beta_{\mathcal{K}} = 1$ and $C_{\mathcal{K}} = 2\overline{\alpha}$.

E.2.4 Coefficient Correspondence

Define

$$\kappa(\rho) \triangleq F_c(m_\rho) \qquad \text{and} \qquad \sigma^2_\rho \triangleq \mathbb{V}\mathrm{ar}\big(W\mathbb{1}\{V \ge m_\rho W\}\big)$$

following Lueker (1998). By (55), we have $\phi^{\mathrm{F}}(\rho) = F_c(m_\rho) = \kappa(\rho)$. Moreover, by definition,

$$\Sigma(\rho) = \mathbb{V}\mathrm{ar}\big(W\mathbb{1}\{V \ge m_\rho W\}\big) = \sigma^2_\rho.$$

Substituting this equation and (58) into (5) yields

$$\Gamma(\rho) = -\frac{1}{2}\phi''_{\mathsf{B}}(\rho)\Sigma(\rho) = \frac{\sigma^2_\rho}{2\alpha_\rho}.$$

In Lueker's notation, $\alpha_0 = \alpha_{\beta_0}$ and $\sigma_0^2 = \sigma_{\beta_0}^2$. Hence, at $\rho = \beta_0$, $\kappa(\beta_0) = \phi^{\mathrm{F}}(\beta_0)$ and $\sigma_0^2/(2\alpha_0) = \Gamma(\beta_0)$. Therefore, the first-order and logarithmic coefficients in Lueker's Theorem 2 coincide with ours in Theorem 4.1.

E.2.5 The Case with Uniformly Distributed Weights and Values

In this section, we consider the special case in which V and W are independent standard uniform random variables. We restrict attention to $\rho = c/n \in \mathcal{U} \triangleq (0, 1/2)$, on which the capacity constraint is binding in the fluid relaxation.

The Fluid Value Function and Dynamic Fluid Policy For any dual variable $\lambda \geq 0$, we have:

$$\mathcal{V}(\lambda) = \mathbb{E}[(V - \lambda W)^+] = \begin{cases} \frac{1}{2} - \frac{\lambda}{2} + \frac{\lambda^2}{6}, & 0 \leq \lambda \leq 1, \\ \frac{1}{6\lambda}, & \lambda \geq 1. \end{cases}$$

Therefore,

$$\phi^{\mathrm{F}}(\rho) = \min_{\lambda \geq 0} \{\lambda\rho + \mathcal{V}(\lambda)\} = \begin{cases} \sqrt{\frac{2}{3}\rho}, & 0 < \rho \leq 1/6, \\ \frac{1}{2} - \frac{3}{2}(1/2 - \rho)^2, & 1/6 \leq \rho < 1/2, \end{cases} \tag{59}$$

and the optimal dual variable is

$$\lambda^*(\rho) = \begin{cases} \frac{1}{\sqrt{6\rho}}, & 0 < \rho \leq 1/6, \\ 3(1/2 - \rho), & 1/6 \leq \rho < 1/2. \end{cases} \tag{60}$$

To obtain the above, note that the objective $\lambda\rho + \mathcal{V}(\lambda)$ is convex in λ and has an interior minimizer. As a result, the optimal solution $\lambda^*(\rho)$ is characterized by the first-order condition; substituting it into the objective gives $\phi^{\mathrm{F}}(\rho)$.

At a state with c units of remaining capacity and n periods remaining, the dynamic fluid policy accepts the item if and only if $W \leq c$ and $V > \lambda^*(c/n)\, W$, with $\lambda^*(\rho)$ given above.

Verification of Assumptions We next verify that Assumption 2.1 and Assumptions 3.1 – 3.3 hold for the uniform case, with binding-resource set $\mathsf{B} = \{1\}$ and fluid function $\phi_{\mathsf{B}}(\rho) = \phi^{\mathrm{F}}(\rho)$.

Assumption 2.1 holds with $\bar{r} = \bar{b} = 1$.

To verify Assumption 3.1, we differentiate (59) to obtain the following derivatives:

$$\phi^{\mathrm{F}\prime}(\rho) = \begin{cases} \dfrac{1}{\sqrt{6\rho}}, & 0 < \rho < 1/6, \\ 3(1/2-\rho), & 1/6 < \rho < 1/2, \end{cases}$$

$$\phi^{\mathrm{F}\prime\prime}(\rho) = \begin{cases} -\dfrac{1}{2\sqrt{6}\,\rho^{3/2}}, & 0 < \rho < 1/6, \\ -3, & 1/6 < \rho < 1/2. \end{cases}$$

At $\rho = 1/6$, the two branches of the fluid function and their first two derivatives agree:

$$\phi^{\mathrm{F}}(1/6-) = \phi^{\mathrm{F}}(1/6+) = 1/3, \quad \phi^{\mathrm{F}\prime}(1/6-) = \phi^{\mathrm{F}\prime}(1/6+) = 1, \quad \phi^{\mathrm{F}\prime\prime}(1/6-) = \phi^{\mathrm{F}\prime\prime}(1/6+) = -3.$$

Consequently, $\phi^{\mathrm{F}}(\rho)$ is twice continuously differentiable on $(0, 1/2)$. In addition, the second-order derivative $\phi^{\mathrm{F}\prime\prime}(\rho)$ is Lipschitz (i.e., $\alpha = 1$) on every compact subset of $(0, 1/2)$. To see this, note that on a compact interval to the left of the junction $\rho = 1/6$, the derivative of $\phi^{\mathrm{F}\prime\prime}$ is bounded; on the right-hand branch, $\phi^{\mathrm{F}\prime\prime}$ is constant. If $x < 1/6 < y$, continuity at the junction and a left-branch derivative bound L give $|\phi^{\mathrm{F}\prime\prime}(x) - \phi^{\mathrm{F}\prime\prime}(y)| \le L(1/6 - x) \le L|y - x|$. Therefore, $\phi_{\mathsf{B}} = \phi^{\mathrm{F}} \in C^{2,1}(0, 1/2)$, and this verifies Assumption 3.1.

We now verify that Assumption 3.2 holds. For every dual variable $\lambda \ge 0$,

$$\mathbb{E}\big[W\,\mathbb{1}\{V \ge \lambda W\}\big] = \begin{cases} \frac{1}{2} - \frac{\lambda}{3}, & 0 \le \lambda \le 1, \\ \frac{1}{6\lambda^2}, & \lambda \ge 1. \end{cases} \tag{61}$$

The static fluid action in (4) is

$$a^{\mathrm{F}}(\rho) \triangleq \mathbb{1}\{V > \lambda^*(\rho) W\}.$$

Therefore,

$$\mathbb{E}\left[B_{a^{\mathrm{F}}(\rho)}\right] = \mathbb{E}\big[W a^{\mathrm{F}}(\rho)\big] = \mathbb{E}\Big[W\mathbb{1}\big\{V > \lambda^*(\rho)\,W\big\}\Big] = \rho\,,$$

where the last equality follows from (60) and (61). Consequently, Assumption 3.2 holds. Because $\mathsf{B}^c = \varnothing$, the slack-resource conditions are vacuous.

Finally, we verify Assumption 3.3. The only top-versus-competitor gap is $S_{a^{\mathrm{F}}}(\rho) - S_{1-a^{\mathrm{F}}}(\rho) =$

$|V - \lambda^*(\rho)W|$, and its weight is $d_{01} = W$. Therefore, uniformly over $\rho \in \mathcal{K}$,

$$\begin{aligned}\mathbb{E}\left[d_{a^{\mathrm{F}}(\rho),1-a^{\mathrm{F}}(\rho)}\mathbb{1}\left\{S_{a^{\mathrm{F}}(\rho)}(\rho) - S_{1-a^{\mathrm{F}}(\rho)}(\rho) \le t\, d_{a^{\mathrm{F}}(\rho),1-a^{\mathrm{F}}(\rho)}\right\}\right] &= \mathbb{E}\Big[W \cdot \mathbb{1}\Big\{|V - \lambda^*(\rho)W| \le tW\Big\}\Big] \\ &\le 2t \cdot \mathbb{E}\left[W^2\right] = \frac{2t}{3}\,.\end{aligned}$$

The inequality follows because, conditional on W, the interval of V-values satisfying the event has length at most $2tW$. Therefore, Assumption 3.3 holds with $\beta_{\mathcal{K}} = 1$ and $C_{\mathcal{K}} = \frac{2}{3}$.

Implication of Theorem 4.1 and Derivation of the Logarithmic Coefficient By Theorem 4.1, for every compact set $\mathcal{K} \Subset (0, 1/2)$, uniformly over all states satisfying $\rho = c/n \in \mathcal{K}$, both $V_n^{\mathrm{OPT}}(c)$ and $V_n^{\mathrm{F}}(c)$ admit the expansion

$$n\phi^{\mathrm{F}}(\rho) - \Gamma(\rho)\log n + O_{\mathcal{K}}(1).$$

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

We now derive the logarithmic coefficient. By direct integration, we have

$$\mathbb{E}\Big[\big(Wa^{\mathrm{F}}(\rho)\big)^2\Big] = \mathbb{E}\big[W^2\,\mathbb{1}\big\{V > \lambda^*(\rho)\,W\big\}\big] = \begin{cases}\big(12\lambda^*(\rho)^3\big)^{-1}, & 0 < \rho \le 1/6,\\ \frac{1}{3} - \frac{1}{4}\lambda^*(\rho), & 1/6 \le \rho < 1/2.\end{cases}$$

Consequently, $\Sigma(\rho) = \mathbb{E}[(Wa^{\mathrm{F}}(\rho))^2] - \rho^2$, and

$$\Gamma(\rho) = -\frac{1}{2}\Sigma(\rho)\phi^{\mathrm{F}\prime\prime}(\rho) = \begin{cases}\dfrac{3 - \sqrt{6\rho}}{24}, & 0 < \rho \le 1/6,\\ \dfrac{-24\rho^2 + 18\rho - 1}{16}, & 1/6 \le \rho < 1/2.\end{cases}$$

E.3 Further Details on the Online Linear Programming Problem in Section 5.3

In this section, we show that Assumptions 1–6 of Bray (2025), with Assumption 6 equivalent to their Lemma 1, imply our Assumption 2.1 and Assumptions 3.1–3.3, except for the local α-Hölder condition on $\nabla^2\phi_{\mathsf{B}}$ required in part 2 of Assumption 3.1.

Bray's Assumptions 1–4 imply our Assumption 2.1. In the remainder of this section, we take $\mathsf{B} = [m]$ and $\phi_{\mathsf{B}} = \phi^{\mathrm{F}}$ and verify the other assumptions. We first characterize the fluid value function ϕ^{F} and its derivatives in Section E.3.1. We then verify Assumptions 3.1–3.3 in the remaining subsections.

E.3.1 Characterization of the Fluid Value Function ϕ^{F} and Its Derivatives

We follow Bray's notation, so b, β, and y denote vectors. Under the binary-action representation $(R_0, \mathbf{B}_0) = (0, \mathbf{0})$ and $(R_1, \mathbf{B}_1) = (u, \mathbf{a})$ and for a normalized inventory vector b, Bray's limiting dual objective $\Lambda_\infty^b(y) = b^{\mathrm{T}} y + \mathbb{E}[(u - \mathbf{a}^{\mathrm{T}} y)^+]$ is exactly the objective of our fluid dual problem (2). For notational convenience, define $g(y) \triangleq \mathbb{E}[(u - \mathbf{a}^{\mathrm{T}} y)^+]$, so that $\Lambda_\infty^b(y) = b^{\mathrm{T}} y + g(y)$.

Given a dual price y, define the expected resource consumption by

$$q(y) \triangleq \mathbb{E}\big[\mathbf{a}\mathbb{1}\{u > \mathbf{a}^{\mathrm{T}} y\}\big] \in \mathbb{R}_+^m.$$

For y in the neighborhood of y_β specified by Bray's Assumption 6, we have

$$\nabla_y \Lambda_\infty^b(y) = b - q(y) \qquad \text{and} \qquad \nabla_y^2 \Lambda_\infty^b(y) = H(y) \triangleq -\mathrm{D}q(y)\,, \tag{62}$$

where $\mathrm{D}q(y)$ denotes the Jacobian matrix of $q(y)$. We prove (62) at the end of this subsection.

Bray's Assumption 5 guarantees a minimizer $y_\beta > \mathbf{0}$ of Λ_∞^β. Because y_β lies in the interior of $\mathbb{R}_+^m$, the gradient of $\Lambda_\infty^\beta(y)$ at y_β must be zero. Substituting $b = \beta$ and $y = y_\beta$ into (62) yields

$$\mathbf{0} = \nabla_y \Lambda_\infty^\beta(y_\beta) = \beta - q(y_\beta), \qquad \text{so} \qquad q(y_\beta) = \beta. \tag{63}$$

We remark that β is componentwise positive (i.e., $\beta > \mathbf{0}$). Suppose, to the contrary, that $\beta_j = 0$ for some j. The identity $q(y_\beta) = \beta$ gives $q_j(y_\beta) = \beta_j = 0$. Since $q_j(y) \geq 0$ for every y, the differentiable function q_j attains its global minimum at the interior point y_β. Its first-order condition is therefore $\frac{\partial}{\partial y_k} q_j(y_\beta) = 0$ for every $k \in [m]$. Equivalently, row j of $\mathrm{D}q(y_\beta)$ is zero, which contradicts the nonsingularity of $\mathrm{D}q(y_\beta)$. Hence, $\beta > \mathbf{0}$.

By Lemma 1 of Bray (2025), the matrix $H(y)$ is continuous and full rank in a neighborhood of y_β. The implicit function theorem then yields open neighborhoods $\mathcal{U}$ of β and $\mathcal{V}$ of y_β and a continuously differentiable map $b \mapsto y_b \in \mathcal{V}$ such that, for every $b \in \mathcal{U}$, $q(y_b) = b$ and y_b is the unique $y \in \mathcal{V}$ satisfying $q(y) = b$. We state this result in Proposition E.3 and defer its proof to the end of the subsection.

Proposition E.3. *There exist open, convex neighborhoods $\mathcal{U} \subseteq (0,\infty)^m$ of β and $\mathcal{V} \subseteq (0,\infty)^m$ of y_β and a continuously differentiable map $y : \mathcal{U} \to \mathcal{V}$ such that, for every $b \in \mathcal{U}$, $q(y(b)) = b$ and $y(b)$ is the unique $y \in \mathcal{V}$ satisfying $q(y) = b$. For ease of notation, let $y_b \triangleq y(b)$. Then $\mathrm{D}_b y_b = -H(y_b)^{-1}$.*

We note that y_b is the unique global minimizer of Λ_∞^b. The identity $q(y_b) = b$ implies $\nabla_y \Lambda_\infty^b(y_b) =$

$\mathbf{0}$ by (62). Moreover, the map $y \mapsto (u - \mathbf{a}^{\mathrm{T}} y)^{+}$ is convex for every realization of $(u, \mathbf{a})$, so Λ_{∞}^{b} is convex. Hence, its zero-gradient point y_b is a global minimizer. Finally, since H is positive definite on $\mathcal{V}$, Λ_{∞}^{b} is strictly convex on $\mathcal{V}$; therefore, y_b is the unique global minimizer.

We now characterize the fluid value function ϕ^{F} and its derivatives. Consider the threshold control

$$x_b(u, \mathbf{a}) \triangleq \mathbb{1}\{u > \mathbf{a}^{\mathrm{T}} y_b\}.$$

Its expected resource consumption is $\mathbb{E}[\mathbf{a} x_b] = q(y_b) = b$, so it is feasible for the fluid relaxation at b and exhausts every resource. Its expected reward satisfies

$$\begin{aligned}
\mathbb{E}[u x_b] &= \mathbb{E}\big[(u - \mathbf{a}^{\mathrm{T}} y_b)^{+}\big] + \mathbb{E}[\mathbf{a}^{\mathrm{T}} y_b \mathbb{1}\{u > \mathbf{a}^{\mathrm{T}} y_b\}] \\
&= g(y_b) + y_b^{\mathrm{T}} q(y_b) \\
&= g(y_b) + y_b^{\mathrm{T}} b = \Lambda_{\infty}^{b}(y_b).
\end{aligned}$$

Therefore, weak duality implies that the threshold control x_b and the dual price y_b are optimal for the primal and dual problems of the fluid relaxation, respectively, and

$$\phi^{\mathrm{F}}(b) = \Lambda_{\infty}^{b}(y_b) = b^{\mathrm{T}} y_b + g(y_b).$$

For each coordinate $i \in [m]$, since y_b is the unique minimizer of $\phi^{\mathrm{F}}(b)$, the envelope theorem gives $\frac{\partial \phi^{\mathrm{F}}(b)}{\partial b_i} = (y_b)_i$. Consequently,

$$\nabla \phi^{\mathrm{F}}(b) = y_b, \qquad \nabla^2 \phi^{\mathrm{F}}(b) = \mathrm{D} y_b = -H(y_b)^{-1}, \tag{64}$$

where the last equality follows from Proposition E.3.

Proof of Equation (62). Fix y in the neighborhood on which Bray's Assumption 6 holds, and let e_j be the jth standard basis vector. For every $h > 0$,

$$0 \le \mathbb{E}[a_j \mathbb{1}\{u = \mathbf{a}^{\mathrm{T}} y\}] \le q_j(y - h e_j) - q_j(y + h e_j).$$

Since $q(y)$ is continuous at y by Bray's Assumption 6, letting $h \downarrow 0$ gives

$$\mathbb{E}[a_j \mathbb{1}\{u = \mathbf{a}^{\mathrm{T}} y\}] = 0.$$

Because $a_j \mathbb{1}\{u = \mathbf{a}^{\mathrm{T}} y\}$ is nonnegative and has expectation zero, $a_j \mathbb{1}\{u = \mathbf{a}^{\mathrm{T}} y\} = 0$ almost surely.

Next, for any $h \neq 0$, we have

$$\left| \left[u - \mathbf{a}^{\mathrm{T}} \left(y + h e_j \right) \right]^+ - \left(u - \mathbf{a}^{\mathrm{T}} y \right)^+ \right| \leq |h| a_j.$$

Dividing by $|h|$ gives

$$\left| \frac{\left[u - \mathbf{a}^{\mathrm{T}} \left(y + h e_j \right) \right]^+ - \left(u - \mathbf{a}^{\mathrm{T}} y \right)^+}{h} \right| \leq a_j.$$

Since $0 \leq a_j \leq \alpha_j$ by Bray's Assumption 4, the dominated convergence theorem yields

$$\begin{aligned}
\frac{\partial g(y)}{\partial y_j} &= \lim_{h \to 0} \mathbb{E} \left\{ \frac{\left[u - \mathbf{a}^{\mathrm{T}} \left(y + h e_j \right) \right]^+ - \left(u - \mathbf{a}^{\mathrm{T}} y \right)^+}{h} \right\} \\
&= \mathbb{E} \left\{ \lim_{h \to 0} \frac{\left[u - \mathbf{a}^{\mathrm{T}} \left(y + h e_j \right) \right]^+ - \left(u - \mathbf{a}^{\mathrm{T}} y \right)^+}{h} \right\} \\
&= \mathbb{E} \left[\frac{\partial}{\partial y_j} \left(u - \mathbf{a}^{\mathrm{T}} y \right)^+ \right] = -\mathbb{E}\big[a_j \mathbb{1}\{ u > \mathbf{a}^{\mathrm{T}} y \} \big] = -q_j(y) \,,
\end{aligned}$$

The third and fourth equalities use

$$\frac{\partial}{\partial y_j} \big(u - \mathbf{a}^{\mathrm{T}} y \big)^+ = -a_j \mathbb{1}\{ u > \mathbf{a}^{\mathrm{T}} y \}, \;\; a.s.$$

which holds because $a_j \mathbb{1}\{ u = \mathbf{a}^{\mathrm{T}} y \} = 0$ almost surely.

Computing this derivative coordinatewise gives the first equality in (62), and differentiating once more gives the second equality in (62). $\square$

Proof of Proposition E.3. Let $\mathcal{V}_0$ denote the neighborhood of y_β on which Bray's Lemma 1 holds. Define function $F : \mathbb{R}^m \times \mathcal{V}_0 \to \mathbb{R}^m$ by $F(b, y) \triangleq b - q(y)$. (63) gives $F(\beta, y_\beta) = \mathbf{0}$. The Jacobians of F with respect to b and y are

$$\mathrm{D}_b F(b, y) = I_m, \qquad \mathrm{D}_y F(b, y) = -\mathrm{D} q(y) = H(y).$$

By Bray's Lemma 1, F is continuously differentiable on $\mathbb{R}^m \times \mathcal{V}_0$. Moreover, $\mathrm{D}_y F(\beta, y_\beta) = H(y_\beta)$ is invertible. The implicit function theorem therefore yields open neighborhoods $\mathcal{U}$ of β and $\mathcal{V} \subseteq \mathcal{V}_0$ of y_β, together with a continuously differentiable map $b \mapsto y_b \in \mathcal{V}$ such that, for each $b \in \mathcal{U}$, $F(b, y_b) = \mathbf{0}$ and y_b is the unique $y \in \mathcal{V}$ satisfying $F(b, y) = \mathbf{0}$. Equivalently,

$$F(b, y_b) = \mathbf{0} \iff q(y_b) = b, \qquad b \in \mathcal{U}.$$

By shrinking both neighborhoods if necessary, we may take $\mathcal{U}$ and $\mathcal{V}$ to be convex. In addition, since $\beta > \mathbf{0}$ and $y_\beta > \mathbf{0}$, we may shrink them further so that $\mathcal{U}, \mathcal{V} \subseteq (0,\infty)^m$.

To compute the derivative of y_b, differentiate $F(b, y_b) = \mathbf{0}$ with respect to b. By the chain rule,

$$\mathbf{0} = \mathrm{D}_b F(b, y_b) + \mathrm{D}_y F(b, y_b)\, \mathrm{D}_b y_b = I_m + H(y_b)\, \mathrm{D}_b y_b.$$

Hence, $\mathrm{D}_b y_b = -H(y_b)^{-1}$. □

E.3.2 Verification of Assumption 3.1

Define the set of binding resources $\mathsf{B} = [m]$ and the fluid function $\phi_{\mathsf{B}} = \phi^{\mathrm{F}}$. For every $b \in \mathcal{U}$, we have $\nabla^2 \phi_{\mathsf{B}}(b) = -H(y_b)^{-1}$ by (64). Since (i) the map $y(b) = y_b$ is continuous on $\mathcal{U}$ by Proposition E.3, (ii) $H(y)$ is continuous and invertible on $\mathcal{V}$ by Bray's Lemma 1, and (iii) matrix inversion is continuous on the set of nonsingular matrices, $\nabla^2 \phi_{\mathsf{B}}$ is continuous on $\mathcal{U}$, and hence $\phi_{\mathsf{B}} \in C^2(\mathcal{U})$.

Assumption 3.1 part 2 additionally requires that $\nabla^2 \phi_{\mathsf{B}}$ is locally α-Hölder continuous for some $\alpha \in (0, 1]$. This is the only requirement in Assumption 3.1 not implied by Bray's assumptions.

E.3.3 Verification of Assumption 3.2

The static fluid policy in (4) is

$$a^{\mathrm{F}}(b) = \mathbb{1}\{u > \mathbf{a}^{\mathrm{T}} y_b\} = x_b(u, \mathbf{a}).$$

By Proposition E.3, $q(y_b) = b$ for all $b \in \mathcal{U}$. Hence, for any resource $j \in [m]$,

$$\mu_j(b) = \mathbb{E}[a_j \mathbb{1}\{u > \mathbf{a}^{\mathrm{T}} y_b\}] = q_j(y_b) = b_j.$$

As a result, all resource constraints bind for every $b \in \mathcal{U}$. This verifies Assumption 3.2.

E.3.4 Verification of Assumption 3.3

Because the action set is binary, there is only one competitor to the selected action. For either selected action, the consumption difference is $d_{01}^{\mathsf{B}} = \|\mathbf{a}\|_1$, and the score gap between the selected action and its competitor is $|u - \mathbf{a}^{\mathrm{T}} y_b|$. Hence, the expectation inside the supremum in Assumption 3.3 equals

$$\mathbb{E}\big[\, \|\mathbf{a}\|_1\, \mathbb{1}\{|u - \mathbf{a}^{\mathrm{T}} y_b| \le t\, \|\mathbf{a}\|_1\}\, \big].$$

Fix a compact set $\mathcal{K}_{\mathsf{B}} \Subset \mathcal{U}$. Since the map $b \mapsto y_b \in \mathcal{V}$ is continuous on $\mathcal{U}$, $A \triangleq \{y_b : b \in \mathcal{K}_{\mathsf{B}}\} \subseteq \mathcal{V}$ is compact. Since $\mathcal{V}$ is an open and convex set, we can choose $t_{\mathcal{K}} > 0$ sufficiently small so that the compact set

$$B \triangleq \Big\{y_b + s\mathbf{1} : b \in \mathcal{K}_{\mathsf{B}},\ |s| \leq 2t_{\mathcal{K}}\Big\} \subseteq \mathcal{V}.$$

Define

$$M_{\mathcal{K}} \triangleq \sup\Big\{ \|\mathrm{D}q(y)\| : y \in B\Big\} < \infty.$$

The upper bound $M_{\mathcal{K}}$ is finite because the Jacobian $\mathrm{D}q(y)$ is continuous by Bray's Assumption 6 and B is compact.

Fix a point $b \in \mathcal{K}_{\mathsf{B}}$ and set $y = y_b$. Take $0 \leq t \leq t_{\mathcal{K}}$, and let $\delta = t + \eta$ with $0 < \eta \leq t_{\mathcal{K}}$. Note that $\delta \leq 2t_{\mathcal{K}}$, hence $y + s\mathbf{1} \in B$ whenever $|s| \leq \delta$. When $\|\mathbf{a}\|_1 > 0$, the event $|u - y^{\mathrm{T}}\mathbf{a}| \leq t\,\|\mathbf{a}\|_1$ implies

$$u - (y - \delta\mathbf{1})^{\mathrm{T}}\mathbf{a} \geq (\delta - t)\,\|\mathbf{a}\|_1 = \eta\,\|\mathbf{a}\|_1 > 0$$

and

$$u - (y + \delta\mathbf{1})^{\mathrm{T}}\mathbf{a} \leq (t - \delta)\,\|\mathbf{a}\|_1 = -\eta\,\|\mathbf{a}\|_1 < 0.$$

Therefore, on $\{\|\mathbf{a}\|_1 > 0\}$, we have

$$\mathbb{1}\{|u - y^{\mathrm{T}}\mathbf{a}| \leq t\|\mathbf{a}\|_1\} \leq \mathbb{1}\{u > (y - \delta\mathbf{1})^{\mathrm{T}}\mathbf{a}\} - \mathbb{1}\{u > (y + \delta\mathbf{1})^{\mathrm{T}}\mathbf{a}\}.$$

Multiplying the preceding inequality by $\|\mathbf{a}\|_1 = \mathbf{1}^{\mathrm{T}}\mathbf{a}$ and taking expectations yields:

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{a}\|_1\,\mathbb{1}\{|u - y^{\mathrm{T}}\mathbf{a}| \leq t\,\|\mathbf{a}\|_1\}\right] &\leq \mathbf{1}^{\mathrm{T}}\Big\{q\big(y - (t+\eta)\mathbf{1}\big) - q\big(y + (t+\eta)\mathbf{1}\big)\Big\} \\ &= -\int_{-(t+\eta)}^{t+\eta} \mathbf{1}^{\mathrm{T}}\,\mathrm{D}q(y + s\mathbf{1})\,\mathbf{1}\,\mathrm{d}s \leq 2mM_{\mathcal{K}}(t+\eta)\,.\end{aligned}$$

The equality follows from the fundamental theorem of calculus along the line $s \mapsto y + s\mathbf{1}$, and the second inequality follows because $|\mathbf{1}^{\mathrm{T}}\,\mathrm{D}q(y + s\mathbf{1})\,\mathbf{1}| \leq \|\mathbf{1}\|_2^2\,\|\mathrm{D}q(y + s\mathbf{1})\|_2 \leq mM_{\mathcal{K}}$. Letting $\eta \to 0$ implies that, for any $b \in \mathcal{K}_{\mathsf{B}}$ and $0 \leq t \leq t_{\mathcal{K}}$,

$$\mathbb{E}[\|\mathbf{a}\|_1\,\mathbb{1}\{|u - \mathbf{a}^{\mathrm{T}}y_b| \leq t\,\|\mathbf{a}\|_1\}] \leq 2mM_{\mathcal{K}}t.$$

Consequently, Assumption 3.3 holds with $\beta_{\mathcal{K}} = 1$ and $C_{\mathcal{K}} = 2mM_{\mathcal{K}}$.