Online Rack Placement in Large-Scale Data Centers Electronic Companion

EC.1. Multi-stage Stochastic Integer Programming (MSSIP)

This paper considers multi-stage stochastic (mixed-)integer programs with a separable objective function as the sum of per-period functions $f^t\left(\boldsymbol{x}^t, \boldsymbol{\xi}^{1:t}\right)$; additive constraints over the decision variables (Equation (18)), and uncertain parameters following an exogenous, time-independent and history-independent distribution $\mathcal{D}$.

EC.1.1. Benchmark algorithms

We compare the OSO algorithm to the following benchmarks to solve Equation (19):

– *Certainty-Equivalent (CE) resolving heuristic*. The CE algorithm proceeds as single-sample OSO, except that it replaces uncertain parameters by their mean. Since our uncertainty is i.i.d. over time, this results in a single sample path equal to the mean $\overline{\boldsymbol{\xi}}^{t+1:T} := (\overline{\boldsymbol{\xi}}, \ldots, \overline{\boldsymbol{\xi}})$. Again, the solution is re-optimized dynamically (Algorithm 2).

Algorithm 2 Certainty-Equivalent (CE) algorithm.

Repeat, for $t = 1, \ldots, T$:

Observe: Observe realization $\boldsymbol{\xi}^t$.

Compute mean: Compute the mean $\overline{\boldsymbol{\xi}}_{t+1:T} = (\overline{\boldsymbol{\xi}}, \ldots, \overline{\boldsymbol{\xi}})$ of the distribution $\mathcal{D}$.

Optimize: Solve the following problem; store optimal solution $(\widetilde{\boldsymbol{x}}^t, \widetilde{\boldsymbol{x}}^{t+1}, \ldots, \widetilde{\boldsymbol{x}}^T)$:

$$\min \quad f^t\left(\boldsymbol{x}^t, \boldsymbol{\xi}^{1:t}\right) + \sum_{\tau=t+1}^{T} f^\tau\left(\boldsymbol{x}^\tau, \left(\boldsymbol{\xi}^{1:t}, \overline{\boldsymbol{\xi}}^{t+1:\tau}\right)\right) + \Psi(\boldsymbol{x}^t) \tag{EC.1a}$$

$$\text{s.t.} \quad \boldsymbol{x}^t \in \mathcal{F}_t\left(\overline{\boldsymbol{x}}^{1:t-1}, \boldsymbol{\xi}^{1:t}\right) \tag{EC.1b}$$

$$\boldsymbol{x}^\tau \in \mathcal{F}_\tau\left(\left(\overline{\boldsymbol{x}}^{1:t-1}, \boldsymbol{x}^{t:\tau-1}\right), \left(\boldsymbol{\xi}^{1:t}, \overline{\boldsymbol{\xi}}^{t+1:\tau}\right)\right) \qquad \forall\, \tau \in \{t+1, \ldots, T\} \tag{EC.1c}$$

Implement: Implement $\overline{\boldsymbol{x}}^t = \widetilde{\boldsymbol{x}}^t$, discarding $(\widetilde{\boldsymbol{x}}^{t+1}, \ldots, \widetilde{\boldsymbol{x}}^T)$.

– *Myopic benchmark (Myo)*. This benchmark optimizes the immediate decision only without anticipating future uncertainty realizations, future decisions and future costs.

– *Hindsight-optimal benchmark (HO)*. This benchmark solves a deterministic optimization problem assuming full knowledge of uncertain parameters.

Algorithm 3 Myopic (Myo) benchmark.

Repeat, for $t = 1, \ldots, T$:

Observe: Observe realization $\boldsymbol{\xi}^t$.

Optimize: Solve the following problem; store optimal solution $\widetilde{\boldsymbol{x}}^t$:

$$\min \quad f^t\left(\boldsymbol{x}^t, \boldsymbol{\xi}^{1:t}\right) + \Psi(\boldsymbol{x}^t) \tag{EC.2a}$$

$$\text{s.t.} \quad \boldsymbol{x}^t \in \mathcal{F}_t\left(\overline{\boldsymbol{x}}^{1:t-1}, \boldsymbol{\xi}^{1:t}\right) \tag{EC.2b}$$

Implement: Implement $\overline{\boldsymbol{x}}^t = \widetilde{\boldsymbol{x}}^t$.

Algorithm 4 Hindsight-optimal (HO) benchmark.

Observe: Observe the true realizations $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^T$.

Optimize: Solve the following problem; store optimal solution $(\widetilde{\boldsymbol{x}}^1, \ldots, \widetilde{\boldsymbol{x}}^T)$:

$$\min \quad \sum_{t \in \mathcal{T}} f^t\left(\boldsymbol{x}^t, \boldsymbol{\xi}^{1:t}\right) \tag{EC.3a}$$

$$\text{s.t.} \quad \boldsymbol{x}^t \in \mathcal{F}_t\left(\boldsymbol{x}^{1:t-1}, \boldsymbol{\xi}^{1:t}\right) \quad \forall\, t \in \{1, \ldots, T\} \tag{EC.3b}$$

Repeat, for $t = 1, \ldots, T$:

Implement: Implement $\overline{\boldsymbol{x}}^t = \widetilde{\boldsymbol{x}}^t$.

EC.2. Online Resource Allocation

EC.2.1. Problem Statement and MSSIP Formulation

We first recall the definition of our online resource allocation problem:

Definition EC.1 (Online resource allocation). Items arrive one at a time, indexed by $t = \{1, \ldots, T\}$. There are m supply nodes and d resources, denoted by $\mathcal{J}$ and $\mathcal{K}$ respectively. Each resource k has capacity b_k. Each item is assigned to at most one supply node $j \in \mathcal{J}$; the assignment of item i to supply node j yields a reward r_j^t and consumes A_{jk}^t units of resource k. For item t, we can define the vector of actions $\boldsymbol{x}^t = \{x_j^t\}_{j \in \mathcal{J}}$, the vector of rewards $\boldsymbol{r}^t = \{r_j^t\}_{j \in \mathcal{J}}$ and the resource consumption matrix $\boldsymbol{A}^t = \{A_{jk}^t\}_{j \in \mathcal{J}, k \in \mathcal{K}}$. For each item t, the unknown parameters $\boldsymbol{\Xi}^t = (\boldsymbol{r}^t, \boldsymbol{A}^t)$ are i.i.d. according to distribution $\mathcal{D}$ (with realizations $\boldsymbol{\xi}^t$). Assignments are irrevocable, and the decision-maker wishes to maximize total assignment reward subject to resource constraints.

We denote by $\Delta_I = \left\{ \boldsymbol{x} \in \{0,1\}^m \,\middle|\, \sum_{j \in \mathcal{J}} x_j \leq 1 \right\}$ the action space at time t, which is a discrete simplex. Denoting the vector of resources by $\boldsymbol{b} \in \mathbb{R}^d$, the offline problem can be written as:

$$\text{OPT} \;=\; \max \quad \sum_{t=1}^{T} \boldsymbol{r}^t \cdot \boldsymbol{x}^t \tag{EC.4a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} (\boldsymbol{A}^t)^\top \boldsymbol{x}^t \leq \boldsymbol{b} \tag{EC.4b}$$

$$\boldsymbol{x}^t \in \Delta_I \qquad \forall\ t \in \{1,\dots,T\} \tag{EC.4c}$$

Correspondingly, we can write the multistage stochastic program as follows. The per-period objective functions and feasible sets are:

$$f^t(\boldsymbol{x}^t, \boldsymbol{\xi}^t) = \boldsymbol{r}^t \cdot \boldsymbol{x}^t \tag{EC.5}$$

$$\mathcal{F}_t(\boldsymbol{x}^{1:t-1}, \boldsymbol{\xi}^{1:t}) = \left\{ \boldsymbol{x}^t \in \Delta_I \;\middle|\; \sum_{\tau=1}^{t-1} (\boldsymbol{A}^\tau)^\top \boldsymbol{x}^\tau + (\boldsymbol{A}^t)^\top \boldsymbol{x}^t \le \boldsymbol{b} \right\} \tag{EC.6}$$

The multi-stage stochastic program can be expressed as follows:

$$\mathbb{E}_{\boldsymbol{\xi}^1}\left[\max_{\boldsymbol{x}^1 \in \mathcal{F}_1(\boldsymbol{\xi}^1)} \left\{ f^1(\boldsymbol{x}^1, \boldsymbol{\xi}^1) + \mathbb{E}_{\boldsymbol{\xi}^2}\left[\max_{\boldsymbol{x}^2 \in \mathcal{F}_2(\boldsymbol{x}^1, \boldsymbol{\xi}^{1:2})} \left\{ f^2(\boldsymbol{x}^2, \boldsymbol{\xi}^2) + \dots + \mathbb{E}_{\boldsymbol{\xi}^T}\left[\max_{\boldsymbol{x}^T \in \mathcal{F}_T(\boldsymbol{x}^{1:T-1}, \boldsymbol{\xi}^{1:T})} \left\{ f^T(\boldsymbol{x}^T, \boldsymbol{\xi}^T) \right\}\right] \dots \right\}\right]\right\}\right] \tag{EC.7}$$

The single-sample OSO algorithm ("OSO" henceforth) for the online resource allocation problem is given in Algorithm 5. The algorithm relies on one sample path at each iteration t, denoted by $\widetilde{\boldsymbol{\xi}}_t^{t+1:T} = (\widetilde{\boldsymbol{\xi}}_t^{t+1}, \dots, \widetilde{\boldsymbol{\xi}}_t^{T})$. By construction, it returns a feasible solution. We then proceed to prove optimality guarantees in Theorem 1.

Algorithm 5 Single-sample OSO algorithm for the online resource allocation.

Repeat, for $t = 1, \dots, T$:

Observe: Observe realization $\boldsymbol{\xi}^t = (\boldsymbol{r}^t, \boldsymbol{A}^t)$.

Sample: Choose *one* sample path $\widetilde{\boldsymbol{\xi}}_t^{t+1:T} = (\widetilde{\boldsymbol{\xi}}_t^{t+1}, \dots, \widetilde{\boldsymbol{\xi}}_t^{T})$ from distribution $\mathcal{D}$.

Optimize: Solve the following problem; store optimal solution $(\widetilde{\boldsymbol{x}}^t, \widetilde{\boldsymbol{x}}^{t+1}, \dots, \widetilde{\boldsymbol{x}}^T)$:

$$\max \quad \boldsymbol{r}^t \cdot \boldsymbol{x}^t + \sum_{\tau=t+1}^{T} \widetilde{\boldsymbol{r}}_t^\tau \cdot \boldsymbol{x}^\tau \tag{EC.8a}$$

$$\text{s.t.} \quad \sum_{\tau=1}^{t-1} (\boldsymbol{A}^\tau)^\top \overline{\boldsymbol{x}}^\tau + (\boldsymbol{A}^t)^\top \boldsymbol{x}^t + \sum_{\tau=t+1}^{T} (\widetilde{\boldsymbol{A}}_t^\tau)^\top \boldsymbol{x}^\tau \le \boldsymbol{b} \tag{EC.8b}$$

$$\boldsymbol{x}^\tau \in \Delta_I \qquad \forall\ t \in \{t, \dots, T\} \tag{EC.8c}$$

Implement: Implement $\overline{\boldsymbol{x}}^t = \widetilde{\boldsymbol{x}}^t$, discarding $(\widetilde{\boldsymbol{x}}^{t+1}, \dots, \widetilde{\boldsymbol{x}}^T)$.

EC.2.2. Proof of Theorem 1

EC.2.2.1. Preliminaries

To formally state our guarantees, we need the definition of an *equivariant* solver:

DEFINITION EC.2. An integer program solver is *equivariant* if, when we permute the items, the solution is permuted the same way: if it returns $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^T)$ as an optimal solution to

$$\max_{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^T \in \Delta_I} \left\{ \sum_{t=1}^{T} \boldsymbol{r}^t \cdot \boldsymbol{x}^t \;\middle|\; \sum_{t=1}^{T} (\boldsymbol{A}^t)^\top \boldsymbol{x}^t \leq \boldsymbol{b} \right\},$$

any permutation π of $\{1, \ldots, T\}$ gives $(\boldsymbol{x}^{\pi(1)}, \ldots, \boldsymbol{x}^{\pi(T)})$ as an optimal solution to:

$$\max_{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^T \in \Delta_I} \left\{ \sum_{t=1}^{T} \boldsymbol{r}^{\pi(t)} \cdot \boldsymbol{x}^{\pi(t)} \;\middle|\; \sum_{t=1}^{T} (\boldsymbol{A}^{\pi(t)})^\top \boldsymbol{x}^{\pi(t)} \leq \boldsymbol{b} \right\}.$$

Any solver can be made equivariant by sorting items according to a pre-specified order (e.g., such that the tuplets $(\boldsymbol{r}^t, \boldsymbol{A}^t)$ are in lexicographic order) and applying the inverse permutation.

We proceed to prove Theorem 1 as long as the algorithm uses an equivariant solver. We consider a discrete distribution $\mathcal{D}$; the general case follows from standard approximation arguments.

Without loss of generality, we assume that $A^t_{jk} \in [0,1]$ for all t, j, k and $b_k = B$ for all k. Otherwise, this can be obtained by rescaling the rows. Indeed, let B denotes the smallest resource capacity normalized to resource requirements, that is:

$$B = \min_{k \in \mathcal{K}} \left\{ \frac{b_k}{\max_{t \in \mathcal{T}, j \in \mathcal{J}} A^t_{jk}} \right\}$$

Then, we can rewrite the constraint

$$\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} A^t_{jk} x^t_j \leq b_k$$

as:

$$\sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{J}} \widetilde{A}^t_{jk} x^t_j \leq B,$$

with

$$\widetilde{A}^t_{jk} = \frac{B}{b_k} A^t_{jk} \leq \frac{A^t_{jk}}{\max_{t \in \mathcal{T}, j \in \mathcal{J}} A^t_{jk}} \leq 1, \qquad \forall\, t \in \mathcal{T},\; j \in \mathcal{J},\; k \in \mathcal{K}.$$

By assumption, $B \geq 1024 \cdot \log\left(\frac{2dT}{\varepsilon}\right) \cdot \frac{\log^3(1/\varepsilon)}{\varepsilon^2}$.

Let us perform the change of variables $\overline{\varepsilon} = \frac{\varepsilon}{8 \log^{1.5}(1/\varepsilon)}$. We note three properties:

– Since $8\varepsilon \log^{1.5}(1/\varepsilon) \leq 1 < 2dT$ for all $\varepsilon \leq 0.001$:

$$\begin{aligned} & 2dT \geq 8\varepsilon \log^{1.5}(1/\varepsilon) = \frac{\varepsilon^2}{\overline{\varepsilon}} \quad \text{(by definition of } \overline{\varepsilon}) \\ \iff \quad & \left(\frac{2dT}{\varepsilon}\right)^2 \geq \frac{2dT}{\overline{\varepsilon}} \\ \iff \quad & 2 \cdot \log\left(\frac{2dT}{\varepsilon}\right) \geq \log\left(\frac{2dT}{\overline{\varepsilon}}\right) \end{aligned} \tag{EC.9}$$

– For all $\varepsilon \leq 0.0001$:

$$\begin{aligned} & \log\left(\frac{16\log^{1.5}(1/\varepsilon)}{\varepsilon}\right) \leq \log^{1.5}(1/\varepsilon) & \\ \iff \quad & \log(2/\overline{\varepsilon}) \leq \log^{1.5}(1/\varepsilon) & \text{(by definition of } \overline{\varepsilon}) \\ \iff \quad & 8\overline{\varepsilon}\log(2/\overline{\varepsilon}) \leq 8\overline{\varepsilon}\log^{1.5}(1/\varepsilon) = \varepsilon & \text{(by definition of } \overline{\varepsilon}) \end{aligned} \tag{EC.10}$$

– By definition of $\overline{\varepsilon}$, $B \geq 16 \cdot \log\left(\frac{2dT}{\varepsilon}\right) \cdot \frac{1}{\overline{\varepsilon}^2}$, hence, per Equation (EC.9), $B \geq 8 \cdot \log\left(\frac{2dT}{\overline{\varepsilon}}\right) \cdot \frac{1}{\overline{\varepsilon}^2}$.

EC.2.2.2. Lemma EC.1: offline optimum scales with time and budget.

Definition EC.3. Let $\mathrm{OPT}(s, \boldsymbol{b}')$ be the optimum of the subproblem which considers the first s items and a resource capacity $\boldsymbol{b}' \in \mathbb{R}^d$:

$$\mathrm{OPT}(s, \boldsymbol{b}') \;=\; \max_{\boldsymbol{x}^1,\ldots,\boldsymbol{x}^s \in \Delta_I} \left\{ \sum_{t=1}^{s} \boldsymbol{r}^t \cdot \boldsymbol{x}^t \;\middle|\; \sum_{t=1}^{s} (\boldsymbol{A}^t)^\top \boldsymbol{x}^t \leq \boldsymbol{b}' \right\}, \tag{EC.11}$$

Note that $\mathrm{OPT}(s, \boldsymbol{b}')$ is a random variable and that $\mathrm{OPT}(T, B\mathbf{1})$ is equal to the original problem (where $\mathbf{1}$ denotes the d-dimensional vector of ones). Gupta and Molinaro (2016b) prove that $\mathrm{OPT}(s, \boldsymbol{b}')$ scales with the fraction of timesteps $\frac{s}{T}$ and with the smallest fraction of budget available $\min_k \frac{b'_k}{B}$ in the more restrictive case where $\boldsymbol{A}^t$ is a $1 \times d$ vector (i.e. there is only one supply node). We extend this result in Lemma EC.1 in the case where $\boldsymbol{A}^t$ is a $m \times d$ matrix, namely $\mathbb{E}\left[\mathrm{OPT}(s, \boldsymbol{b}')\right] \approx \min\left\{\frac{s}{T}, \min_k \frac{b'_k}{B}\right\} \cdot \mathbb{E}\left[\mathrm{OPT}\right]$.

Lemma EC.1. *Let $\zeta = \min\left\{\frac{s}{T}, \min_k \frac{b'_k}{B}\right\}$. If $s \geq \overline{\varepsilon}^2 T$ and $\min_k b'_k \geq \overline{\varepsilon}^2 B$, then:*

$$\mathbb{E}\left[\mathrm{OPT}(s, \boldsymbol{b}')\right] \geq \left[\zeta\left(1 - \overline{\varepsilon}\sqrt{\frac{1}{\zeta}}\right) - \frac{1+\overline{\varepsilon}}{T}\right] \cdot \mathbb{E}\left[\mathrm{OPT}\right] \tag{EC.12}$$

Proof of Lemma EC.1. Without loss of generality, all coordinates of $\boldsymbol{b}'$ are identical, so $\boldsymbol{b}' = B'\mathbf{1}$ and $B' = \min_k b'_k$, and $\frac{B'}{B} = \min_k \frac{b'_k}{B} \leq \frac{s}{T}$. This can be obtained by rescaling the rows.

Let $\boldsymbol{x}^\star = (\boldsymbol{x}^{\star,1}, \ldots, \boldsymbol{x}^{\star,T})$ be an optimal solution to the offline problem $\mathrm{OPT}(T, B\mathbf{1})$ by an equivariant solver. Also, let *Set* be the *set* of the realizations $\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^T\}$ of the problem. The definition of *Set* ignores the order, so conditioning on *Set* leaves the order of $\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^T\}$ uniformly random; in particular, conditioned on *Set*, the sequence of random variables $\boldsymbol{r}^1, \ldots, \boldsymbol{r}^T$ is exchangeable, so the distribution of $(\boldsymbol{r}^{\pi(1)}, \ldots, \boldsymbol{r}^{\pi(T)})$ is the same for every permutation π of $\{1, \ldots, T\}$. Due to the equivariance of the solution $\boldsymbol{x}^\star$, the distribution of $\boldsymbol{r}^t \cdot \boldsymbol{x}^{\star,t}$ is the same for all $t \in \{1, \ldots, T\}$, conditioned on *Set*; in particular, at time t, the contribution to the optimal solution is

$$\mathbb{E}\left[\boldsymbol{r}^t \cdot \boldsymbol{x}^{\star,t} \,\middle|\, Set\right] \;=\; \frac{1}{T} \cdot \mathbb{E}\left[\sum_{\tau=1}^{T} \boldsymbol{r}^\tau \cdot \boldsymbol{x}^{\star,\tau} \,\middle|\, Set\right] \;=\; \frac{\mathbb{E}\left[\mathrm{OPT} \mid Set\right]}{T}. \tag{EC.13}$$

Informally, the proof of the lemma relies on the observation that, for $\widetilde{s}$ slightly smaller than ζT, the solution truncated to its first $\widetilde{s}$ elements is a feasible solution with high probability to the problem given in Equation (EC.11), and yields an expected value $\widetilde{s}\frac{\mathbb{E}[\mathrm{OPT}]}{T} \approx \zeta\mathbb{E}[\mathrm{OPT}]$. This will give that $\mathrm{OPT}(s, \boldsymbol{b}')$ is larger than but close to $\zeta\mathbb{E}[\mathrm{OPT}]$. Let us formalize these arguments.

Let $\widetilde{B} = B'\left(1 - \bar{\varepsilon}\sqrt{B/B'}\right)$, and fix $\widetilde{s}$ to the integer in the interval $\left(\frac{T\widetilde{B}}{B} - 1, \frac{T\widetilde{B}}{B}\right]$. In particular, since $\widetilde{B} \le B' \le \frac{sB}{T}$, we have that $\widetilde{s} \le s$. We first show that the solution $\widehat{\boldsymbol{x}} = (\boldsymbol{x}^{\star,1}, \dots, \boldsymbol{x}^{\star,\widetilde{s}}, \boldsymbol{0}, \dots, \boldsymbol{0}) \in \Delta_I^m$ is feasible with high probability for Equation (EC.11). Again, conditioned on *Set*, the sequence of random variables $\boldsymbol{A}^1, \dots, \boldsymbol{A}^T$ is exchangeable.

Due to the equivariance of $\boldsymbol{x}^\star$, it follows that for each resource k, the sequence of random variables $\left\{ (\boldsymbol{A}^1)^\top \boldsymbol{x}^{\star,1}, \ \dots, \ (\boldsymbol{A}^T)^\top \boldsymbol{x}^{\star,T} \right\}$ is also exchangeable. From the feasibility of $\boldsymbol{x}^\star$ for $\mathrm{OPT}(T, B\boldsymbol{1})$, $\sum_{t=1}^T (\boldsymbol{A}^t)^\top \boldsymbol{x}^{\star,t} \le B\boldsymbol{1}$ in every scenario. From the concentration inequality for exchangeable sequences (Corollary EC.1 in EC.2.2.6), we obtain the following inequality for each resource k (using $\tau = \bar{\varepsilon}\sqrt{BB'}$ and $M = B$):

$$\begin{aligned}
\mathbb{P}\left(\sum_{t=1}^T \sum_{j\in\mathcal{J}} A_{jk}^t \widehat{x}_j^t \ge B' \,\middle|\, Set \right) &= \mathbb{P}\left(\sum_{t=1}^T \sum_{j\in\mathcal{J}} A_{jk}^t \widehat{x}_j^t \ge \widetilde{B} + \bar{\varepsilon}\sqrt{BB'} \,\middle|\, Set \right) \\
&\le \mathbb{P}\left(\sum_{t=1}^T \sum_{j\in\mathcal{J}} A_{jk}^t \widehat{x}_j^t \ge \frac{\widetilde{s}B}{T} + \bar{\varepsilon}\sqrt{BB'} \,\middle|\, Set \right) \\
&\le 2\exp\left(-\min\left\{ \frac{\bar{\varepsilon}^2 B' T}{8\widetilde{s}}, \ \frac{\bar{\varepsilon}\sqrt{BB'}}{2} \right\} \right).
\end{aligned} \tag{EC.14}$$

To upper bound the right-hand side, we use $B \ge \frac{8}{\bar{\varepsilon}^2}\log\left(\frac{2dT}{\bar{\varepsilon}}\right)$, $\widetilde{B} \le B'$ and $\widetilde{s} \le \frac{T\widetilde{B}}{B}$ to obtain:

$$\frac{\bar{\varepsilon}^2 B' T}{8\widetilde{s}} \ge \frac{\bar{\varepsilon}^2 B' B}{8\widetilde{B}} \ge \frac{\bar{\varepsilon}^2 B}{8} \ge \log\left(\frac{2dT}{\bar{\varepsilon}}\right). \tag{EC.15}$$

Moreover, the assumption that $B' \ge \bar{\varepsilon}^2 B$ implies:

$$\frac{\bar{\varepsilon}\sqrt{BB'}}{2} \ge \frac{\bar{\varepsilon}^2 B}{2} \ge 4\log\left(\frac{2dT}{\bar{\varepsilon}}\right). \tag{EC.16}$$

Thus, the solution violates the resource k constraint of $(\mathrm{OPT}(s, b'))$ with probability at most $\frac{\bar{\varepsilon}}{dT}$:

$$\forall\ k \in \mathcal{K}: \quad \mathbb{P}\left(\sum_{t=1}^T \sum_{j\in\mathcal{J}} A_{jk}^t \widehat{x}_j^t \ge B' \,\middle|\, Set \right) \le \frac{\bar{\varepsilon}}{dT}. \tag{EC.17}$$

Taking a union bound over all d constraints, the solution is feasible with high probability:

$$\mathbb{P}\left(\sum_{t=1}^T (\boldsymbol{A}^t)^\top \widehat{\boldsymbol{x}}^t \le \boldsymbol{b}' \,\middle|\, Set \right) \ge 1 - \frac{\bar{\varepsilon}}{T}. \tag{EC.18}$$

Let G be the good event that this feasibility condition holds (and G^c be its complement). Under this event, $\mathrm{OPT}(s, \boldsymbol{b}')$ is at least equal to $\sum_{t=1}^{\widetilde{s}} \boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t$. We obtain:

$$\mathbb{E}\left[\mathrm{OPT}(s, \boldsymbol{b}') \,|\, Set\right] \ge \mathbb{E}\left[\sum_{t=1}^{\widetilde{s}} \boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t \,\middle|\, G \text{ and } Set \right] \cdot \mathbb{P}\left(G \,|\, Set\right)$$

$$= \mathbb{E}\left[\sum_{t=1}^{\widetilde{s}} \boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t \,\middle|\, Set\right] - \mathbb{E}\left[\sum_{t=1}^{\widetilde{s}} \boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t \,\middle|\, G^c \text{ and } Set\right] \cdot \mathbb{P}\left(G^c \,|\, Set\right)$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{\widetilde{s}} \boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t \,\middle|\, Set\right] - \mathbb{E}\left[\mathrm{OPT} \,|\, G^c \text{ and } Set\right] \cdot \mathbb{P}\left(G^c \,|\, Set\right), \tag{EC.19}$$

where the last inequality stems from the feasibility of $\widehat{\boldsymbol{x}}$ in the full problem.

To bound the first term, recall that $\boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t = \boldsymbol{r}^t \cdot \boldsymbol{x}^{\star,t}$ for $t \leq \tilde{s}$, so Equation (EC.13) implies $\mathbb{E}\left[\boldsymbol{r}^t \cdot \widehat{\boldsymbol{x}}^t \,|\, Set\right] = \dfrac{\mathbb{E}\left[\mathrm{OPT} \,|\, Set\right]}{T}$. For the second term, notice that conditioning on Set fixes the items in the instance, hence the optimum OPT, so further conditioning on G^c has no effect. Thus:

$$\mathbb{E}\left[\mathrm{OPT}(s, \boldsymbol{b}') \,|\, Set\right] \geq \frac{\widetilde{s}}{T}\mathbb{E}\left[\mathrm{OPT} \,|\, Set\right] - \frac{\overline{\varepsilon}}{T}\mathbb{E}\left[\mathrm{OPT} \,|\, Set\right] \tag{EC.20}$$

$$\geq \left[\frac{\widetilde{B}}{B} - \frac{1}{T} - \frac{\overline{\varepsilon}}{T}\right] \cdot \mathbb{E}\left[\mathrm{OPT} \,|\, Set\right] \tag{EC.21}$$

$$= \left[\frac{B'}{B}\left(1 - \overline{\varepsilon}\sqrt{\frac{B}{B'}}\right) - \frac{1+\overline{\varepsilon}}{T}\right] \cdot \mathbb{E}\left[\mathrm{OPT} \,|\, Set\right] \tag{EC.22}$$

$$\geq \left[\zeta\left(1 - \overline{\varepsilon}\sqrt{\frac{1}{\zeta}}\right) - \frac{1+\overline{\varepsilon}}{T}\right] \cdot \mathbb{E}\left[\mathrm{OPT} \,|\, Set\right], \tag{EC.23}$$

where the last inequality uses that $\zeta \geq \overline{\varepsilon}^2$, and the function $x \mapsto x(1 - \overline{\varepsilon}\sqrt{1/x})$ is increasing for $x \geq \overline{\varepsilon}^2$. Taking expectation with respect to Set concludes the proof of the lemma. □

EC.2.2.3. Lemmas EC.2 and EC.3: Resource consumption scales with time

We next show that resources consumption scales with time and the resource capacity. This is, by time t, the OSO algorithm utilizes approximately a fraction $\frac{t}{T}$ of the resource budget for each resource. We formalize this via the following definitions:

Definition EC.4 (Occupancy vector). The occupancy vector of resources consumed by OSO at time t is defined as follows, where $\overline{\boldsymbol{x}}^t$ is the decision implemented by OSO at time t.

$$\boldsymbol{S}^t = \sum_{\tau=1}^{t} (\boldsymbol{A}^\tau)^\top \overline{\boldsymbol{x}}^\tau \in \mathbb{R}^d \tag{EC.24}$$

Definition EC.5. We denote by $\mathcal{H}_t$ the σ-algebra generated by the history of the OSO algorithm up to time t, i.e., the demand realization $\boldsymbol{\xi}^\tau = (\boldsymbol{r}^\tau, \boldsymbol{A}^\tau)$ and the sample path $\widetilde{\boldsymbol{\xi}}_\tau^{\tau+1:T}$ for $\tau \in \{1, \ldots, t\}$. We can condition on $\mathcal{H}_t$, giving the expectation operator $\mathbb{E}_{\mathcal{H}_t}[\cdot]$ (denoted by $\mathbb{E}_t[\cdot]$ for simplicity).

We show that by time t, the algorithm utilizes approximately a fraction $\frac{t}{T}$ of the overall budget, i.e., $\boldsymbol{S}^t$ is less than but close to $\frac{t}{T}B\mathbf{1}$ componentwise. In fact, we prove the following stronger result, which will be later used to show that $\boldsymbol{S}^t$ is concentrated around its mean:

Lemma EC.2. *For every $t \geq 1$, we have*

$$\mathbb{E}_{t-1}\left[\boldsymbol{S}^t\right] \leq \left(1 - \frac{1}{T-t+1}\right)\boldsymbol{S}^{t-1} + \frac{B\mathbf{1}}{T-t+1}. \tag{EC.25}$$

In particular, $\mathbb{E}\left[\boldsymbol{S}^t\right] \leq \frac{t}{T}B\mathbf{1}$.

Proof of Lemma EC.2. By construction, OSO yields a feasible solution :

$$\boldsymbol{S}^{t-1} + (\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t + (\widetilde{\boldsymbol{A}}_t^{t+1})^\top \widetilde{\boldsymbol{x}}^{t+1} + \cdots + (\widetilde{\boldsymbol{A}}_t^{T})^\top \widetilde{\boldsymbol{x}}^{T} \le B\mathbf{1}. \tag{EC.26}$$

Even conditioned on the history up to time $t-1$, the matrices $\boldsymbol{A}^t, \widetilde{\boldsymbol{A}}_t^{t+1}, \ldots, \widetilde{\boldsymbol{A}}_t^{T}$ are sampled i.i.d., and the solution $(\overline{\boldsymbol{x}}^t = \widetilde{\boldsymbol{x}}^t, \widetilde{\boldsymbol{x}}^{t+1}, \ldots, \widetilde{\boldsymbol{x}}^T)$ is by assumption equivariant. Therefore, conditioned on the history up to time $t-1$, the sequence of vectors $\left\{ (\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t,\ (\widetilde{\boldsymbol{A}}_t^{t+1})^\top \widetilde{\boldsymbol{x}}^{t+1},\ \ldots,\ (\widetilde{\boldsymbol{A}}_t^{T})^\top \widetilde{\boldsymbol{x}}^{T} \right\}$ is again an exchangeable sequence of random variables with equal expectation, conditioned on $\mathcal{H}_{t-1}$: $\mathbb{E}_{t-1}\left[(\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t\right] = \mathbb{E}_{t-1}\left[(\widetilde{\boldsymbol{A}}_t^{\tau})^\top \widetilde{\boldsymbol{x}}^{\tau}\right]$ for all $\tau > t$. From Equation (EC.26), we therefore have:

$$\boldsymbol{S}^{t-1} + (T-t+1)\cdot \mathbb{E}_{t-1}\left[(\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t\right] \le B\mathbf{1}. \tag{EC.27}$$

We obtain inequality in Equation (EC.25):

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\boldsymbol{S}^t\right] &= \boldsymbol{S}^{t-1} + \mathbb{E}_{t-1}\left[(\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t\right] && \text{(EC.28)}\\
&\le \boldsymbol{S}^{t-1} + \frac{B\mathbf{1} - \boldsymbol{S}^{t-1}}{T-t+1} && \text{(EC.29)}\\
&= \left(1 - \frac{1}{T-t+1}\right)\boldsymbol{S}^{t-1} + \frac{B\mathbf{1}}{T-t+1}. && \text{(EC.30)}
\end{aligned}$$

We now prove that $\mathbb{E}\left[\boldsymbol{S}^t\right] \le \frac{t}{T}B\mathbf{1}$ by induction on t. It clearly holds for $t=0$ (where we define $\boldsymbol{S}^0 = \mathbf{0}$ by convention). Assuming that it holds for $t-1$, we obtain:

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{S}^t\right] = \mathbb{E}\left[\mathbb{E}_{t-1}\left[\boldsymbol{S}^t\right]\right] &\le \left(1 - \frac{1}{T-t+1}\right)\mathbb{E}\left[\boldsymbol{S}^{t-1}\right] + \frac{B\mathbf{1}}{T-t+1} && \text{(EC.31)}\\
&\le \left(1 - \frac{1}{T-t+1}\right)\frac{t-1}{T}\cdot B\mathbf{1} + \frac{B\mathbf{1}}{T-t+1} && \text{(EC.32)}\\
&= \frac{t}{T}B\mathbf{1}, && \text{(EC.33)}
\end{aligned}$$

where the first inequality follows from inequality in Equation (EC.25) and the next inequality follows by the induction hypothesis. This concludes the proof. □

While this lemma guarantees that the occupation vector $\mathbb{E}\left[\boldsymbol{S}^t\right]$ is at most $\frac{t}{T}B\mathbf{1}$ in expectation, we need high-probability guarantees. We derive them in the next lemma.

Lemma EC.3. *For each $t \ge \frac{\overline{\varepsilon}^2 T}{4}$, we have:*

$$\mathbb{P}\left(\boldsymbol{S}^t \le \left(1 + \overline{\varepsilon}\sqrt{\frac{T}{t}}\right)\frac{t}{T}B\mathbf{1}\right) \ge 1 - \frac{\overline{\varepsilon}}{2T}. \tag{EC.34}$$

Proof of Lemma EC.3. We show that for every component $k \in \{1, \ldots, d\}$, S_k^t is concentrated around its expected value, namely that $S_k^t \le \left(1 + \overline{\varepsilon}\sqrt{\frac{T}{t}}\right)\frac{t}{T}B$ with probability at least $1 - \frac{\overline{\varepsilon}}{2dT}$. However, since the increments $(\boldsymbol{A}^1)^\top \overline{\boldsymbol{x}}^1, \ldots, (\boldsymbol{A}^t)^\top \overline{\boldsymbol{x}}^t$ are not independent (e.g., $\overline{\boldsymbol{x}}^\tau$ depends on $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^\tau$), we cannot use standard concentration inequalities. Instead, we rely on a concentration result for "self-centering" sequences, given in Theorem EC.2 (see EC.2.2.6).

Let us denote $\alpha_t = 1 - \frac{1}{T-t+1}$ and $\beta_t = \frac{B}{T-t+1}$, and let Y_t be the solution to the recurrence relation $y_t = \alpha_t y_{t-1} + \beta_t$ and $y_0 = 0$. From Theorem EC.2, we obtain, for any $\gamma \in (0,1]$:

$$\mathbb{P}\left(S_k^t \geq (1+2\gamma)Y_t\right) \leq \exp\left(-\gamma^2 Y_t\right). \tag{EC.35}$$

As in Lemma EC.2, we prove by induction on t that $Y_t = \frac{t}{T}B$: this holds for $t=0$, and then:

$$Y_t = \left(1 - \frac{1}{T-t+1}\right) Y_{t-1} + \frac{B}{T-t+1} = \left(1 - \frac{1}{T-t+1}\right)\frac{t-1}{T}B + \frac{B}{T-t+1} = \frac{t}{T}B. \tag{EC.36}$$

From Equation (EC.35) with $\gamma = \frac{\bar{\varepsilon}}{2}\sqrt{T/t}$ ($\gamma \leq 1$ by assumption), we derive:

$$\mathbb{P}\left(S_k^t \geq \left(1 + \bar{\varepsilon}\sqrt{\frac{T}{t}}\right)\frac{t}{T}B\right) \leq \exp\left(-\frac{\bar{\varepsilon}^2}{4}B\right). \tag{EC.37}$$

Recall that, by assumption, $B \geq \frac{8}{\bar{\varepsilon}^2}\log\left(\frac{2dT}{\bar{\varepsilon}}\right) \geq \frac{4}{\bar{\varepsilon}^2}\log\left(\frac{2dT}{\bar{\varepsilon}}\right)$. This yields:

$$\mathbb{P}\left(S_k^t \geq \left(1 + \bar{\varepsilon}\sqrt{\frac{T}{t}}\right)\frac{t}{T}B\right) \leq \exp\left(-\log(2dT/\bar{\varepsilon})\right) = \frac{\bar{\varepsilon}}{2dT}. \tag{EC.38}$$

Taking a union bound over all d coordinates, we obtain:

$$\mathbb{P}\left(\boldsymbol{S}^t \leq \left(1 + \bar{\varepsilon}\sqrt{\frac{T}{t}}\right)\frac{t}{T}B\mathbf{1}\right) \geq 1 - \frac{\bar{\varepsilon}}{2T}. \tag{EC.39}$$

This concludes the proof of the lemma. □

EC.2.2.4. Lemma EC.4: Reward of OSO algorithm bounded below.

We now bound the reward of the OSO algorithm at time t, i.e. $\boldsymbol{r}^t \cdot \bar{\boldsymbol{x}}^t$. From Lemma EC.3, there is about $\left(1 - \frac{t-1}{T}\right)B$ of the budget left in each of the constraints, and so the remaining value should be $\left(1 - \frac{t-1}{T}\right)\mathbb{E}[\mathrm{OPT}]$. Moreover, since there are $T-t+1$ variables in the remaining problem, we expect that $\bar{\boldsymbol{x}}^t$ accrues a value of $\frac{1}{T-t+1}\left(1 - \frac{t-1}{T}\right)\mathbb{E}[\mathrm{OPT}]$, or $\frac{1}{T}\mathbb{E}[\mathrm{OPT}]$. This is formalized below.

Lemma EC.4. *For every t satisfying $\bar{\varepsilon}^2 T \leq t \leq (1-2\bar{\varepsilon})T$ we have:*

$$\mathbb{E}\left[\boldsymbol{r}^t \cdot \bar{\boldsymbol{x}}^t\right] \geq \left[1 - \bar{\varepsilon}\sqrt{\frac{T}{(1-\bar{\varepsilon})T - t}} - 2\bar{\varepsilon}\frac{\sqrt{Tt}}{T-t} - \frac{1+\bar{\varepsilon}}{T-t+1}\right]\frac{\mathbb{E}[\mathrm{OPT}]}{T}. \tag{EC.40}$$

Proof of Lemma EC.4. Fix t such that $\bar{\varepsilon}^2 T \leq t \leq (1-2\bar{\varepsilon})T$, and consider the solution $(\bar{\boldsymbol{x}}^t, \tilde{\boldsymbol{x}}^{t+1}, \ldots, \tilde{\boldsymbol{x}}_T)$ obtained by the OSO algorithm at this time. By definition, we have:

$$\boldsymbol{r}^t \cdot \bar{\boldsymbol{x}}^t + \sum_{\tau=t+1}^{T} \tilde{\boldsymbol{r}}_t^\tau \cdot \tilde{\boldsymbol{x}}^\tau = \max_{\boldsymbol{x}^t,\ldots,\boldsymbol{x}^T \in \Delta_I}\left\{\boldsymbol{r}^t \cdot \boldsymbol{x}^t + \sum_{\tau=t+1}^{T}\tilde{\boldsymbol{r}}_t^\tau \cdot \boldsymbol{x}^\tau \,\middle|\, (\boldsymbol{A}^t)^\top \boldsymbol{x}^t + \sum_{\tau=t+1}^{T}(\tilde{\boldsymbol{A}}_t^\tau)^\top \boldsymbol{x}^\tau \leq B\mathbf{1} - \boldsymbol{S}^{t-1}\right\} \tag{EC.41}$$

Conditioning on the history $\mathcal{H}_{t-1}$ fixes the occupation vector $\boldsymbol{S}^{t-1}$, hence the right-hand side of the resource constraint. The expected value of the stochastic program is equal to $\mathbb{E}\left[\mathrm{OPT}(T-t+1, B\mathbf{1}-\boldsymbol{S}^{t-1})\right]$. It comes:

$$\mathbb{E}_{t-1}\left[\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t+\sum_{\tau=t+1}^{T}\widetilde{\boldsymbol{r}}_t^{\tau}\cdot\widetilde{\boldsymbol{x}}^{\tau}\right]=\mathbb{E}\left[\mathrm{OPT}(T-t+1,B\mathbf{1}-\boldsymbol{S}^{t-1})\right]. \tag{EC.42}$$

As earlier, conditioned on $\mathcal{H}_{t-1}$ the random variables $\left\{\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t, \widetilde{\boldsymbol{r}}_t^{t+1}\cdot\widetilde{\boldsymbol{x}}^{t+1},\ldots,\widetilde{\boldsymbol{r}}_t^{T}\cdot\widetilde{\boldsymbol{x}}^{T}\right\}$ form an exchangeable sequence and thus have the same conditional expectations. Thus, all the terms on the left-hand side of Equation (EC.42) have the same expectation. In particular,

$$\mathbb{E}_{t-1}\left[\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t\right]=\frac{1}{T-t+1}\cdot\mathbb{E}\left[\mathrm{OPT}(T-t+1,B\mathbf{1}-\boldsymbol{S}^{t-1})\right]. \tag{EC.43}$$

Now, let $\gamma_t=\overline{\varepsilon}\sqrt{\frac{T}{t}}$. From Lemma EC.3, we know that, with probability at least $1-\frac{\overline{\varepsilon}}{2T}$, we have:

$$\boldsymbol{S}^{t-1}\le(1+\gamma_{t-1})\frac{t-1}{T}B\mathbf{1}\implies B\mathbf{1}-\boldsymbol{S}^{t-1}\ge\left(1-(1+\gamma_{t-1})\frac{t-1}{T}\right)B\mathbf{1}. \tag{EC.44}$$

Denote the above good event happening by G. When G transpires, letting $\zeta_t=\min\left\{\frac{T-t+1}{T},1-(1+\gamma_{t-1})\frac{t-1}{T}\right\}=1-(1+\gamma_{t-1})\frac{t-1}{T}$, we have from Lemma EC.1 that:

$$\mathbb{E}\left[\mathrm{OPT}(T-t+1,B\mathbf{1}-\boldsymbol{S}^{t-1})\right]\ge\left[\zeta_t\left(1-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}\right)-\frac{1+\overline{\varepsilon}}{T}\right]\cdot\mathbb{E}\left[\mathrm{OPT}\right]. \tag{EC.45}$$

Note that the assumptions in Lemma EC.1 are met because $t\le(1-2\overline{\varepsilon})T$ and $\overline{\varepsilon}\in(0,1]$. Then:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t\right]&\ge\mathbb{E}\left[\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t\,\middle|\,G\right]\cdot\mathbb{P}(G)\\&\ge\left(1-\frac{\overline{\varepsilon}}{2T}\right)\left[\zeta_t\left(1-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}\right)-\frac{1+\overline{\varepsilon}}{T}\right]\frac{\mathbb{E}\left[\mathrm{OPT}\right]}{T-t+1}\\&\ge\left[\left(1-\frac{\overline{\varepsilon}}{T}\right)\left(1-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}\right)\frac{\zeta_t}{T-t+1}-\frac{1+\overline{\varepsilon}}{T(T-t+1)}\right]\mathbb{E}\left[\mathrm{OPT}\right].\end{aligned} \tag{EC.46}$$

To further lower bound the right-hand side, using the definitions of ζ_t and γ_{t-1} we have

$$\frac{\zeta_t}{T-t+1}=\frac{(T-t+1)-\gamma_{t-1}(t-1)}{T(T-t+1)}=\frac{1}{T}\left(1-\overline{\varepsilon}\frac{\sqrt{T}\sqrt{t-1}}{T-t+1}\right)\ge\frac{1}{T}\left(1-\overline{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t}\right) \tag{EC.47}$$

Substituting into Equation (EC.46) and using $(1-a)(1-b)\ge1-a-b$ for $a,b\ge0$, we get:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{r}^t\cdot\overline{\boldsymbol{x}}^t\right]&\ge\left[\left(1-\frac{\overline{\varepsilon}}{T}\right)\left(1-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}\right)\left(1-\overline{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t}\right)-\frac{1+\overline{\varepsilon}}{T-t+1}\right]\frac{\mathbb{E}\left[\mathrm{OPT}\right]}{T}\\&\ge\left[1-\frac{\overline{\varepsilon}}{T}-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}-\overline{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t}-\frac{1+\overline{\varepsilon}}{T-t+1}\right]\frac{\mathbb{E}\left[\mathrm{OPT}\right]}{T}\\&\ge\left[1-\overline{\varepsilon}\sqrt{\frac{1}{\zeta_t}}-2\overline{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t}-\frac{1+\overline{\varepsilon}}{T-t+1}\right]\frac{\mathbb{E}\left[\mathrm{OPT}\right]}{T}.\end{aligned} \tag{EC.48}$$

We can complete the lower bound of this right-hand side using

$$\zeta_t = \frac{(T-t+1)}{T} - \bar{\varepsilon}\sqrt{\frac{t-1}{T}} \geq \frac{(1-\bar{\varepsilon})T-t}{T} \tag{EC.49}$$

We obtain:

$$\mathbb{E}\left[\boldsymbol{r}^t \cdot \bar{\boldsymbol{x}}^t\right] \geq \left[1 - \bar{\varepsilon}\sqrt{\frac{T}{(1-\bar{\varepsilon})T-t}} - 2\bar{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t} - \frac{1+\bar{\varepsilon}}{T-t+1}\right]\frac{\mathbb{E}[\mathrm{OPT}]}{T}. \tag{EC.50}$$

This concludes the proof of the lemma. □

EC.2.2.5. Proof of Theorem 1.

Let ALG be the value achieved by the OSO algorithm. From Lemma EC.4, we have:

$$\mathbb{E}[\mathrm{ALG}] \geq \sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\left(1 - \bar{\varepsilon}\sqrt{\frac{T}{(1-\bar{\varepsilon})T-t}} - 2\bar{\varepsilon}\frac{\sqrt{T}\sqrt{t}}{T-t} - \frac{1+\bar{\varepsilon}}{T-t+1}\right)\frac{\mathbb{E}[\mathrm{OPT}]}{T}. \tag{EC.51}$$

Since the function $\sqrt{\frac{T}{(1-\bar{\varepsilon})T-t}}$ is increasing in t, we can use an integral to upper bound the sum:

$$\begin{aligned}
\sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\sqrt{\frac{T}{(1-\bar{\varepsilon})T-t}} &\leq \int_0^{(1-2\bar{\varepsilon})T}\sqrt{\frac{T}{(1-\bar{\varepsilon})T-t}}\,\mathrm{d}t \\
&= \sqrt{T}\int_{\bar{\varepsilon}T}^{(1-\bar{\varepsilon})T}\frac{1}{\sqrt{y}}\,\mathrm{d}y \\
&= 2\sqrt{T}\left(\sqrt{(1-\bar{\varepsilon})T} - \sqrt{\bar{\varepsilon}T}\right) \\
&\leq 2T.
\end{aligned} \tag{EC.52}$$

Similarly, since the function $\frac{\sqrt{t}}{T-t}$ is increasing in t, we have:

$$\sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\frac{\sqrt{T}\sqrt{t}}{T-t} = \sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\frac{\sqrt{t/T}}{1-(t/T)} \leq \int_0^{(1-2\bar{\varepsilon})T}\frac{\sqrt{t/T}}{1-t/T}\,\mathrm{d}t = T\int_0^{1-2\bar{\varepsilon}}\frac{\sqrt{y}}{1-y}\,\mathrm{d}y. \tag{EC.53}$$

Therefore,

$$\begin{aligned}
\sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\frac{\sqrt{T}\sqrt{t}}{T-t} &\leq T\cdot\left[-2\sqrt{y} + \log\left(\frac{1+\sqrt{y}}{1-\sqrt{y}}\right)\right]\Bigg|_0^{1-2\bar{\varepsilon}} \\
&\leq T\log\left(\frac{1+\sqrt{1-2\bar{\varepsilon}}}{1-\sqrt{1-2\bar{\varepsilon}}}\right) \leq T\log\left(\frac{2}{1-\sqrt{1-2\bar{\varepsilon}}}\right) \\
&\leq T\log(2/\bar{\varepsilon}),
\end{aligned} \tag{EC.54}$$

where the last inequality uses the fact that $\sqrt{1+x} \leq 1 + \frac{x}{2}$ for all $x \in [-1, \infty)$.

Finally, the third negative term can be bounded as

$$\begin{aligned}
\sum_{t=\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T-1}\frac{1+\bar{\varepsilon}}{T-t+1} &\leq \int_{\bar{\varepsilon}^2 T}^{(1-2\bar{\varepsilon})T}\frac{1+\bar{\varepsilon}}{T-t+1}\,\mathrm{d}t = \int_{1+2\bar{\varepsilon}T}^{1+(1-\bar{\varepsilon}^2)T}\frac{1+\bar{\varepsilon}}{y}\,\mathrm{d}y \\
&= (1+\bar{\varepsilon})\log(y)\Bigg|_{1+2\bar{\varepsilon}T}^{1+(1-\bar{\varepsilon}^2)T} \leq (1+\bar{\varepsilon})\log(1/\bar{\varepsilon})
\end{aligned} \tag{EC.55}$$

Combining these bounds into Equation (EC.51), we conclude:

$$\begin{aligned}\mathbb{E}[\mathrm{ALG}] &\geq \left(1-2\bar{\varepsilon}-2\bar{\varepsilon}\log(2/\bar{\varepsilon})-\frac{1+\bar{\varepsilon}}{T}\log(1/\bar{\varepsilon})\right)\mathbb{E}[\mathrm{OPT}] \\ &\geq (1-8\bar{\varepsilon}\log(2/\bar{\varepsilon}))\mathbb{E}[\mathrm{OPT}] \\ &\geq (1-\varepsilon)\mathbb{E}[\mathrm{OPT}]. \end{aligned} \tag{EC.56}$$

The second inequality uses the assumption that $T \geq \frac{1}{\bar{\varepsilon}}$ (otherwise, the facts that $B \geq \frac{1}{\varepsilon}$ and the matrices $\boldsymbol{A}^t$ have entries in $[0,1]$ make all constraints redundant and the problem becomes trivial). The third inequality comes from Equation (EC.10). This concludes the proof of Theorem 1. □

EC.2.2.6. Concentration Inequalities

In this section, we collect and show concentration inequalities that are used in the proof of Theorem 1. These include a concentration inequality for exchangeable sequences (Corollary EC.1), and a concentration inequality for affine stochastic processes (Theorem EC.2).

We make use of the following result from Van Der Vaart and Wellner (1996):

Theorem EC.1 **(Theorem 2.14.19 in Van Der Vaart and Wellner (1996)).** *Let $A = \{a_1,\dots,a_n\}$ be a set of real numbers in $[0,1]$. Let S be a random subset of A of size s and let $A_S=\sum_{i\in S}a_i$. Setting $\bar{a}=\frac{1}{n}\sum_{i=1}^n a_i$ and $\sigma^2=\frac{1}{n}\sum_{i=1}^n(a_i-\bar{a})^2$, we have, for every $\tau>0$:*

$$\mathbb{P}(|A_S - s\bar{a}| \geq \tau) \leq 2\exp\left(-\frac{\tau^2}{2s\sigma^2+\tau}\right). \tag{EC.57}$$

A sequence $X_1,\dots,X_n$ of random variables is *exchangeable* if its distribution is permutation invariant, i.e., the distribution of the vector $(X_{\pi(1)},X_{\pi(2)},\dots,X_{\pi(n)})$ is the same for all permutations π of $\{1,\dots,n\}$. The following result is the main result of this section.

Corollary EC.1. *Let $X_1,\dots,X_n$ be an exchangeable sequence of random variables, i.i.d. according to distribution $\mathcal{D}$, in the interval $[0,1]$. Assume that $\sum_{i=1}^n X_i \leq M$ with probability 1. Then for every $s\in\{1,\dots,n\}$ and $\tau>0$, we have:*

$$\mathbb{P}\left(X_1+\cdots+X_s \geq \frac{sM}{n}+\tau\right) \leq 2\exp\left(-\min\left\{\frac{\tau^2 n}{8sM},\frac{\tau}{2}\right\}\right). \tag{EC.58}$$

Proof of Corollary EC.1. We prove the result in the case where the distribution $\mathcal{D}$ is discrete. The general case following from standard approximation arguments.

Consider a set $A=\{a_1,\dots,a_n\}$ of n values in $[0,1]$, and define $\bar{a}=\frac{1}{n}\sum_{i=1}^n a_i$ and $\sigma^2=\frac{1}{n}\sum_{i=1}^n(a_i-\bar{a})^2$. Condition on the *set* $\{X_1,\dots,X_n\}$ being equal to A, leaving their *order* free. Under this conditioning, $X_1,\dots,X_s$ is just a random subset of size s from A, and $\bar{a}=\frac{1}{n}\sum_{i=1}^n X_i \leq \frac{M}{n}$. From Theorem EC.1, we get:

$$\mathbb{P}\left(X_1+\cdots+X_s \geq \frac{sM}{n}+\tau \,\middle|\, \{X_1,\dots,X_n\}=A\right) \leq 2\exp\left(-\frac{\tau^2}{2s\sigma^2+\tau}\right). \tag{EC.59}$$

Since the a_i's belong to the interval $[0,1]$, the variance term can be bounded as

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i-\overline{a})^2 \le \frac{1}{n}\sum_{i=1}^{n}|a_i-\overline{a}| \le \frac{1}{n}\left(\sum_{i=1}^{n}|a_i|+\sum_{i=1}^{n}|\overline{a}|\right) = 2\overline{a} \le \frac{2M}{n}. \tag{EC.60}$$

Further using the inequality $\frac{1}{a+b} \ge \frac{1}{2}\min\{\frac{1}{a},\frac{1}{b}\}$ for non-negative a,b, we obtain:

$$\exp\left(-\frac{\tau^2}{2s\sigma^2+\tau}\right) \le \exp\left(-\frac{\tau^2}{4sM/n+\tau}\right) \le \exp\left(-\frac{1}{2}\min\left\{\frac{\tau^2 n}{4sM},\tau\right\}\right). \tag{EC.61}$$

Taking the expectation of Equation (EC.59) over all possible sets A completes the proof. □

Theorem EC.2. *Consider a sequence $X_1,\dots,X_T$ of (possibly dependent) random variables in $[0,1]$ adapted to a filtration $\mathcal{H}_1,\dots,\mathcal{H}_T$. Define the partial sums $S_t = X_1+\cdots+X_t$ for $t\in\{0,1,\dots,T\}$ (with $S_0=0$). Furthermore, suppose that there are sequences $\alpha_1,\dots,\alpha_T\in[0,1]$ and $\beta_1,\dots,\beta_T\ge 0$ such that $\mathbb{E}[S_t\,|\,\mathcal{H}_{t-1}] \le \alpha_t S_{t-1}+\beta_t$ for all t. Then for every $\gamma\in(0,1]$, we have:*

$$\mathbb{P}(S_T \ge (1+2\gamma)Y_T) \le \exp\left(-\gamma^2 Y_T\right), \tag{EC.62}$$

where $Y_1,\dots,Y_T$ is the solution to the recursion $y_t=\alpha_t y_{t-1}+\beta_t$ (with $y_0=0$).

We make use of a lemma on the moment-generating function of affine transformations of S_t.

Lemma EC.5. *Consider a function $f(x)=ax+b$ with $a\in[0,1]$. Under the assumptions from Theorem EC.2, for all $\gamma\in(0,1]$ we have*

$$\mathbb{E}[\exp(\gamma f(S_t))] \le \mathbb{E}[\exp(\gamma f(\alpha_t S_{t-1}+(1+\gamma)\beta_t))]. \tag{EC.63}$$

Proof of Lemma EC.5. Since conditioning on $\mathcal{H}_{t-1}$ fixes the sum S_{t-1}, we observe that:

$$\mathbb{E}[\exp(\gamma f(S_t))] = \mathbb{E}[\exp(\gamma(aS_t+b))] = \mathbb{E}[\exp(\gamma(aS_{t-1}+b))\cdot\mathbb{E}[\exp(\gamma a X_t)\,|\,\mathcal{H}_{t-1}]]. \tag{EC.64}$$

We get:

$$\begin{aligned}
\mathbb{E}[\exp(\gamma a X_t)\,|\,\mathcal{H}_{t-1}] &\le \mathbb{E}\left[1+\gamma a X_t+\gamma^2 a^2 X_t^2\,\middle|\,\mathcal{H}_{t-1}\right] \quad (\text{since } e^x \le 1+x+x^2 \text{ for } x\in[0,1])\\
&\le \mathbb{E}\left[1+(\gamma+\gamma^2)aX_t\,\middle|\,\mathcal{H}_{t-1}\right]\\
&= 1+(\gamma+\gamma^2)a\cdot\mathbb{E}[X_t\,|\,\mathcal{H}_{t-1}]\\
&\le \exp\left((\gamma+\gamma^2)a\cdot\mathbb{E}[X_t\,|\,\mathcal{H}_{t-1}]\right). \quad (\text{since } 1+x\le e^x)
\end{aligned} \tag{EC.65}$$

Since $S_t = S_{t-1}+X_t$, the assumption $\mathbb{E}[S_t\,|\,\mathcal{H}_{t-1}] \le \alpha_t S_{t-1}+\beta_t$ implies $\mathbb{E}[X_t\,|\,\mathcal{H}_{t-1}] \le (\alpha_t - 1)S_{t-1}+\beta_t$. Applying this to Equation (EC.65) and $\gamma^2 a(\alpha_t-1)S_t\le 0$, we get:

$$\begin{aligned}
\mathbb{E}[\exp(\gamma a X_t)\,|\,\mathcal{H}_{t-1}] &\le \exp\left((\gamma+\gamma^2)a((\alpha_t-1)S_{t-1}+\beta_t)\right)\\
&\le \exp\left(\gamma a((\alpha_t-1)S_{t-1}+\beta_t)+\gamma^2 a\beta_t\right).
\end{aligned} \tag{EC.66}$$

From Equation (EC.64), it comes:

$$\begin{aligned}\mathbb{E}\left[\exp\left(\gamma f(S_t)\right)\right] &\leq \mathbb{E}\left[\exp\left(\gamma(a(\alpha_t S_{t-1}+(1+\gamma)\beta_t)+b)\right)\right]\\ &= \mathbb{E}\left[\exp\left(\gamma f(\alpha_t S_{t-1}+(1+\gamma)\beta_t)\right)\right]. \qquad \text{(EC.67)}\end{aligned}$$

This concludes the proof of the lemma. □

Proof of Theorem EC.2. Define the affine function $f_t(x)=\alpha_t x+(1+\gamma)\beta_t$, so that Lemma EC.5 can be expressed as $\mathbb{E}\left[\exp\left(\gamma f(S_t)\right)\right] \leq \mathbb{E}\left[\exp\left(\gamma f(f_t(S_{t-1}))\right)\right]$. Applying it repeatedly gives:

$$\mathbb{E}\left[\exp\left(\gamma S_T\right)\right] \leq \mathbb{E}\left[\exp\left(\gamma f_T(S_{T-1})\right)\right] \leq \mathbb{E}\left[\exp\left(\gamma f_T(f_{T-1}(S_{T-2}))\right)\right] \qquad \text{(EC.68)}$$

$$\leq \ldots \leq \mathbb{E}\left[\exp\left(\gamma f_T(f_{T-1}(\ldots f_2(f_1(0))))\right)\right]. \qquad \text{(EC.69)}$$

We can still apply Lemma EC.5 because the composed function $f_T \circ f_{T-1} \circ \cdots \circ f_t$ is still affine of the form $ax+b$ with $a\in[0,1]$ and $b\geq 0$ (indeed, $a=\alpha_T\ldots\alpha_t\in[0,1]$ and b is obtained by taking products and sums of the α_t's and β_t's, which are all non-negative).

Moreover, we prove by induction on $t=1,\ldots,T$ that the composition of these functions satisfies:

$$f_t(f_{t-1}(\ldots f_2(f_1(0))))=(1+\gamma)Y_t. \qquad \text{(EC.70)}$$

For $t=0$, we have $f_1(0)=(1+\gamma)\beta_1=(1+\gamma)Y_1$. For $t\geq 1$, we have:

$$f_t(f_{t-1}(\ldots f_2(f_1(0))))=f_t((1+\gamma)Y_{t-1})=(1+\gamma)\left[\alpha_t Y_{t-1}+\beta_t\right]=(1+\gamma)Y_t. \qquad \text{(EC.71)}$$

Thus, we get the moment-generating function upper bound $\mathbb{E}\left[\exp\left(\gamma S_T\right)\right] \leq \exp\left(\gamma(1+\gamma)Y_T\right)$. Finally, applying Markov's inequality we get:

$$\begin{aligned}\mathbb{P}\left(S_T \geq (1+2\gamma)Y_T\right) &= \mathbb{P}\left(\exp\left(\gamma S_T\right) \geq \exp\left(\gamma(1+2\gamma)Y_T\right)\right)\\ &\leq \frac{\mathbb{E}\left[\exp\left(\gamma S_T\right)\right]}{\exp\left(\gamma(1+2\gamma)Y_T\right)}\\ &\leq \frac{\exp\left(\gamma(1+\gamma)Y_T\right)}{\exp\left(\gamma(1+2\gamma)Y_T\right)}\\ &= \exp\left(-\gamma^2 Y_T\right). \qquad \text{(EC.72)}\end{aligned}$$

This concludes the proof of the theorem. □

EC.2.2.7. Proof of Proposition 1.

Consider the online resource allocation problem with T items of size 1, one supply, and one resource in quantity $\sqrt{T}$, i.e., $|\mathcal{J}|=1$, $|\mathcal{K}|=1$, $A_{11}^t=1$ for all $t\in\{1,\ldots,T\}$, and $b_1=\sqrt{T}$. Assume that item values r_j^t are equal to 1 with probability $1/\sqrt{T}$ and to a small value $\varphi>0$ with probability $1-1/\sqrt{T}$. Since there are on average $\sqrt{T}$ items of value 1, it can be shown that $\mathbb{E}\left[\mathrm{OPT}\right]\approx\sqrt{T}$.

Per Theorem 1, for any $\varepsilon > 0$, the OSO algorithm achieves a value within a multiplicative factor of $1-\varepsilon$ of the $\sqrt{T}$ optimum for large enough T. However, a myopic decision-making rule would always assign items to the supply until all $\sqrt{T}$ resources have been consumed, leading to an expected value of $\sqrt{T}\left(1 \cdot \frac{1}{\sqrt{T}} + \varphi \cdot \left(1 - \frac{1}{\sqrt{T}}\right)\right) = \varphi\sqrt{T} + 1 - \varphi \approx \varphi\sqrt{T}$. Therefore, a myopic approach achieves a fraction φ of the optimal value, which can be made arbitrarily small. □

EC.2.2.8. Proof of Proposition 2.

Consider the online resource allocation problem with one supply bin ($m = 1$), $d > 2$ resources, and $T > d^2$ time periods. All rewards are known and equal to $r_1^t = 1$. Each resource $k \in \mathcal{K}$ has capacity $b_k = \sqrt{T}$. The assignment of an incoming item into the supply bin consumes one unit of one resource with probability $1/\sqrt{T}$ and one unit of each resource with probability $1 - d/\sqrt{T}$. Formally, we define the following probability distribution with support over $d+1$ possible realizations; for simplicity, we encode it via a random variable ξ_t in $\{0, \ldots, d\}$.

$$(A_{1k}^t)_{k\in\mathcal{K}} = \begin{cases} \boldsymbol{e}_1 & \text{with probability } 1/\sqrt{T} \text{ (encoded via } \xi_t = 1) \\ \vdots & \\ \boldsymbol{e}_d & \text{with probability } 1/\sqrt{T} \text{ (encoded via } \xi_t = d) \\ \mathbf{1} & \text{with probability } 1 - d/\sqrt{T} \text{ (encoded via } \xi_t = 0) \end{cases} \tag{EC.73}$$

The mean-based certainty-equivalent (CE) algorithm for the multi-dimensional knapsack problem solves the following problem at each iteration, and implements the current-period solution $\overline{x}^t := x^t$.

$$\max \quad \sum_{\tau=1}^{t-1} r^\tau \overline{x}^\tau + r^t x^t + \sum_{\tau=t+1}^{T} \mathbb{E}\left[r^\tau\right] x^\tau \tag{EC.74}$$

$$\text{s.t.} \quad \sum_{\tau=1}^{t-1} A_{1k}^\tau \overline{x}^\tau + A_{1k}^t x^t + \sum_{\tau=t+1}^{T} \mathbb{E}\left[A_{1k}^\tau\right] x^\tau \le b_k, \quad \forall\, k \in \mathcal{K} \tag{EC.75}$$

$$x^\tau \in \{0,1\} \qquad \forall\, \tau \in \{t, \ldots, T\} \tag{EC.76}$$

Hindsight-optimal solution. Let $N_\ell = \sum_{t=1}^{T} \mathbb{1}(\xi_t = \ell)$ characterize the number of time periods with realization $\ell \in \{0, \ldots, d\}$. Then, $(N_1, \ldots, N_d, N_0)$ follows a multinomial distribution with sum T and parameters $\left(\frac{1}{\sqrt{T}}, \ldots, \frac{1}{\sqrt{T}}, 1 - \frac{d}{\sqrt{T}}\right)$. The hindsight-optimal policy can be formulated via decision variable G_ℓ characterizing the number of times an item of type ℓ is chosen. Then, the hindsight-optimal solution, denoted by $\mathrm{OPT}(\boldsymbol{\xi})$, is obtained from the following optimization problem:

$$\begin{aligned} \mathrm{OPT}(\boldsymbol{\xi}) = \max \quad & \sum_{j=0}^{d} G_\ell \\ \text{s.t.} \quad & G_\ell \le N_\ell & \forall\, \ell \in \{0, \ldots, d\} \\ & G_\ell + G_0 \le \sqrt{T} & \forall\, \ell \in \{1, \ldots, d\} \end{aligned}$$

Since items of type 0 involve higher resource consumption, they get de-prioritized in favor of all other item types. The optimal solution is therefore given by:

$$G_\ell = \min\{N_\ell, \sqrt{T}\}, \ \forall\ \ell \in \{1,\ldots,d\}$$
$$G_0 = \sqrt{T} - \max_{\ell\in\{1,\ldots,d\}} G_\ell$$

We can bound $\mathbb{E}[\mathrm{OPT}(\boldsymbol{\xi})]$ as follows, for a small enough $\varepsilon > 0$ and a large enough T:

$$\begin{aligned}
\mathbb{E}[\mathrm{OPT}(\boldsymbol{\xi})] &= \sum_{j=0}^{d} \mathbb{E}[G_\ell] \\
&\geq \sum_{j=1}^{d} \mathbb{E}[G_\ell] \\
&= d\cdot\left(\mathbb{E}\left[N_1 \,\middle|\, N_1 \leq \sqrt{T}\right]\cdot \mathbb{P}\left(N_1 \leq \sqrt{T}\right) + \sqrt{T}\cdot\mathbb{P}\left(N_1 \geq \sqrt{T}\right)\right) \quad \text{(by symmetry)} \\
&\geq d\cdot\left(\frac{1}{2}\left(\sqrt{T} - (\sqrt{T}-1)^{1/2}\frac{1/\sqrt{2\pi}}{1/2}\right) + \frac{1}{2}\sqrt{T}\right) - \varepsilon \\
&= d\sqrt{T} - d\frac{1}{\sqrt{2\pi}}(\sqrt{T}-1)^{1/2} - \varepsilon \\
&\geq d\sqrt{T} - dT^{1/4}
\end{aligned}$$

The critical step lies in the second inequality. Since $(N_1,\ldots,N_d,N_0)$ follows a multinomial distribution with sum T and parameters $\left(\frac{1}{\sqrt{T}},\ldots,\frac{1}{\sqrt{T}},1-\frac{d}{\sqrt{T}}\right)$, N_1 follows a binomial distribution with T trials and success probability $\frac{1}{\sqrt{T}}$. Therefore, its mean is $\sqrt{T}$ and its variance is $\sqrt{T}-1$. By the central limit theorem, we therefore know that N_1 converges to a normal distribution with mean $\sqrt{T}$ and variance $\sqrt{T}-1$ as T grow large. Thus, $\mathbb{P}\left(N_1 \leq \sqrt{T}\right) \approx 0.5$ and $\mathbb{E}\left[N_1 \,\middle|\, N_1 \leq \sqrt{T}\right] \approx \sqrt{T} - (\sqrt{T}-1)^{1/2}\frac{1/\sqrt{2\pi}}{1/2}$. This proves that $\mathbb{E}[\mathrm{OPT}(\boldsymbol{\xi})] \geq d\sqrt{T} - dT^{1/4}$.

<u>OSO solution.</u> Per Theorem 1, the single-sample OSO algorithm achieves a value close to $\mathbb{E}[\mathrm{OPT}(\boldsymbol{\xi})]$, the hindsight optimum, for large enough T. Specifically, the single-sample OSO algorithm obtains a value of at least $(1-\varepsilon_{d,T,B})\cdot\mathbb{E}[\mathrm{OPT}(\boldsymbol{\xi})]$, with ε scaling approximately as $\mathcal{O}\left(\sqrt{\frac{\log(dT)}{B}}\right)$.

<u>CE solution.</u> The expected resource utilization in each period, denoted by $\overline{A}$, is equal to:

$$\overline{A} = \mathbb{E}[A^t_{1k}] = \frac{1}{\sqrt{T}} + \left(1-\frac{d}{\sqrt{T}}\right) = 1 - \frac{d-1}{\sqrt{T}} < 1, \quad \forall k \in \mathcal{K} \tag{EC.77}$$

Consider a decision epoch $t \in \{1,\ldots,T\}$ where the capacity of all resources is equal to B. Assume that $\xi_t = 0$, i.e., that the incoming item consumes one unit of each resource. Since the mean item consumes $\overline{A} < 1$ unit of each resource while yielding the same reward, the CE solution will reject the incoming item as long as the remaining time horizon is long enough. Similarly, if $\xi_t = j \in \{1,\ldots,d\}$, the incoming item consumes one unit of resource j whereas the mean item consumes $\overline{A} < 1$ unit of each resource while yielding the same reward. Again, the CE solution will serve as many copies of the mean item as possible rather than the incoming item, and it will therefore reject the incoming item as long as the remaining time horizon is long enough.

Therefore, the CE solution will not serve any incoming item starting from $t=1$ for the first $T-\sqrt{T}/A$ (since the capacities of all resources remain equal). In turn, the CE algorithm will accept the remaining $\sqrt{T}/A$ items up to capacity. The CE objective value, referred to as CE, therefore achieves an expected value of $\mathbb{E}\left[\mathrm{CE}(\boldsymbol{\xi})\right] \leq \sqrt{T}/A$.

Next, we prove that the CE solution achieves a competitive ratio of at most $1/d$. Suppose for the sake of contradiction that its competitive ratio is $\frac{1}{d}+\varepsilon$ for some $\varepsilon>0$. Then the following holds:

$$
\begin{aligned}
& \mathbb{E}\left[\mathrm{CE}(\boldsymbol{\xi})\right] \geq \left(\frac{1}{d}+\varepsilon\right)\mathbb{E}\left[\mathrm{OPT}(\boldsymbol{\xi})\right]-\alpha \\
\Longrightarrow \quad & \frac{\sqrt{T}}{1-\frac{d-1}{\sqrt{T}}} \geq \left(\frac{1}{d}+\varepsilon\right)\cdot\left(d\sqrt{T}-dT^{1/4}\right)-\alpha \\
\Longrightarrow \quad & T \geq \left(\sqrt{T}-(d-1)\right)\cdot\left((1+\varepsilon d)\sqrt{T}-(1+\varepsilon d)T^{1/4}-\alpha\right) \\
\Longrightarrow \quad & 0 \geq \varepsilon dT-(1+\varepsilon d)T^{3/4}-(\alpha+(d-1)(1+\varepsilon d))\sqrt{T}+(d-1)(1+\varepsilon d)T^{1/4}+(d-1)\alpha
\end{aligned}
$$

which is a contradiction for large enough T.

Therefore, CE achieves a competitive ratio of at most $1/d$, and OSO can achieve a competitive ratio of $1-\varepsilon_{d,T,B}$ with ε scaling as $\mathcal{O}\left(\sqrt{\frac{\log(dt)}{B}}\right)$, This proves that OSO can achieve unbounded (multiplicative) benefits over CE as the number of resources d grows infinitely. $\square$

EC.2.3. Computational Results

EC.2.3.1. Need for Resolving Heuristics

We first show that the online resource allocation problem (Equation (19)) remains intractable with off-the-shelf stochastic programming and dynamic programming methods even in moderately-sized instances, thus motivating the need for efficient resolving heuristics such as the OSO algorithm studied in this paper. We consider an online multidimensional knapsack problem over T periods with $|\mathcal{J}|=1$ supply node, and $|\mathcal{K}|=d$ resources of capacity B. Items arrive one at a time, and the resource consumption variables A^t_{1k} are binary (equal to 0 or 1 with probability 0.5) and independent. The complexity of the problem is driven by the number of items T, the number of resources d, and the capacity B.

Table EC.1 compares the objective values and computational times of the OSO algorithm against (i) a multi-stage stochastic programming formulation based on a scenario-tree representation (SP), and (ii) the policy computed via dynamic programming (DP). Note that the scenario-tree representation leads to highly intractable integer optimization instances even for small problems. The full scenario tree involves 2^d scenarios at each time period, hence $\mathcal{O}(2^{dT})$ overall nodes over the

Metric	T	B	d	SP	DP	OSO-1	OSO-5	OSO-10	OSO-20
Obj.	4	2	2	3.34	3.34	3.34	3.33	3.33	3.34
	6	3	2	5.1	5.1	5.11	5.09	5.12	5.09
	8	4	2	6.94	6.94	6.96	6.9	6.96	6.94
	10	5	2	8.91	8.91	8.89	8.84	8.91	8.9
	12	6	2	—	10.74	10.73	10.62	10.72	10.74
	14	7	2	—	12.64	12.62	12.55	12.63	12.64
	16	8	2	—	14.56	14.52	14.41	14.55	14.55
	18	9	2	—	16.45	16.37	16.2	16.44	16.39
	20	10	2	—	18.41	18.36	18.24	18.38	18.37
	20	10	3	—	17.96	17.92	17.75	17.9	17.92
	20	10	4	—	17.64	17.5	17.43	17.57	17.59
	20	10	5	—	17.34	17.16	17.11	17.3	17.35
	20	10	6	—	—	16.84	16.81	17.09	17.09
Time (s)	4	2	2	2.65×10^{-3}	1.93×10^{-4}	3.13×10^{-3}	3.00×10^{-3}	3.35×10^{-3}	3.84×10^{-3}
	6	3	2	4.87×10^{-2}	5.46×10^{-4}	2.96×10^{-3}	4.39×10^{-3}	4.36×10^{-3}	4.64×10^{-3}
	8	4	2	1.04	1.16×10^{-3}	4.27×10^{-3}	4.90×10^{-3}	5.34×10^{-3}	6.86×10^{-3}
	10	5	2	3.59×10^{1}	2.08×10^{-3}	5.35×10^{-3}	6.55×10^{-3}	6.80×10^{-3}	9.27×10^{-3}
	12	6	2	—	3.47×10^{-3}	6.24×10^{-3}	7.26×10^{-3}	8.75×10^{-3}	1.16×10^{-2}
	14	7	2	—	5.35×10^{-3}	9.88×10^{-3}	9.54×10^{-3}	1.14×10^{-2}	1.47×10^{-2}
	16	8	2	—	7.89×10^{-3}	8.72×10^{-3}	1.37×10^{-2}	1.39×10^{-2}	1.79×10^{-2}
	18	9	2	—	1.10×10^{-2}	1.05×10^{-2}	1.70×10^{-2}	1.69×10^{-2}	2.23×10^{-2}
	20	10	2	—	1.50×10^{-2}	1.11×10^{-2}	1.71×10^{-2}	1.87×10^{-2}	2.66×10^{-2}
	20	10	3	—	5.19×10^{-1}	1.17×10^{-2}	2.36×10^{-2}	2.15×10^{-2}	3.25×10^{-2}
	20	10	4	—	1.30×10^{1}	1.16×10^{-2}	1.59×10^{-2}	2.19×10^{-2}	3.74×10^{-2}
	20	10	5	—	3.35×10^{2}	1.43×10^{-2}	2.66×10^{-2}	4.27×10^{-2}	4.47×10^{-2}
	20	10	6	—	—	1.26×10^{-2}	2.03×10^{-2}	2.84×10^{-2}	4.34×10^{-2}

—: Instances which did not complete within a 1-hour time limit.

Table EC.1 Performance comparison in the online multi-dimensional knapsack problem. "SP": Stochastic programming. "DP": Dynamic programming. "OSO-S": OSO algorithm with S sample paths per iteration. Solutions evaluated on 100 instances.

multi-stage horizon, $\mathcal{O}(2^{dT})$ integer decision variables and $\mathcal{O}(2^{d(T-1)})$ constraints. The stochastic programming model becomes intractable very quickly, with as few as 12 time periods, 1 supply node, 2 resources, and binary uncertainty; in comparison, in the next section, we will solve instances with up to 100 time periods, 10 supply nodes, 20 resources, and continuous uncertainty. Whereas these results rely on an exhaustive scenario tree, they also suggest that stochastic programming remains intractable even with small-sample representations of uncertainty. For example, with $d=2$, the scenario tree involves 4 possible realizations at each time period, and leads to an intractable formulation with $T=12$ and $|\mathcal{J}|=1$. Even small-sample scenario-tree approximations based, for example, on sample average approximation or scenario reduction, would result in similarly intractable stochastic programming formulations.

The dynamic programming algorithm is more scalable but still terminates up to 4-5 orders of magnitude slower than OSO, and times out in comparatively small instances (e.g., 20 time periods, 1 supply node, and 6 resources). This again stems from the exponential growth in problem size, with $\mathcal{O}(T(2B)^d)$ possible states. In comparison, the OSO algorithm is very computationally efficient, terminating in fractions of a second on these simple examples. Furthermore, the OSO solutions are very close to optimal. In low-dimensional problem instances, even the single-sample variant of OSO leads to virtually identical solutions as the DP algorithm (variations are due to the randomness associated with the 100 out-of-sample scenarios). When the number of resources d grows larger, the OSO algorithm benefits from more sample paths.

Next, we compare the OSO algorithm to the other resolving heuristics and the perfect-information benchmark.

EC.2.3.2. Comparison on online resource allocation problems

We provide additional computational results to complement the results from Section 4.3 for the multi-dimensional knapsack problem, the online generalized assignment problem, and the general online resource allocation problem. Recall that these problems involve very high-dimensional discrete allocation problems with up to 50 time periods, 10 supply nodes, and 50 resources; or up to 100 time periods, 10 supply nodes, and 20 resources. All uncertain resource consumption parameters A^t_{jk} follow a bimodal distribution parametrized by ψ, described in the main text and shown in Figure EC.1. At each iteration, the integer program is solved with Gurobi 11 and Julia 1.10, with 2 threads. For the multidimensional knapsack and online generalized assignment problems, each IP was solved with termination criteria of either 60 seconds or a relative gap of 0.01%. For the more challenging online resource allocation instances, we used a termination criteria of either 100 seconds per iteration or a relative gap of 1%. The capacity parameter B was chosen to vary over a range which is not too small (such that all algorithms cannot serve demand) nor too large (such that all algorithms can trivially serve all items).

Tables EC.2 to EC.4 compare the algorithms' solutions for each problem; Figure EC.2 to Figure EC.4 show them visually in absolute terms, without normalization. Throughout, the myopic policy induces a significant loss as compared to the hindsight-optimal solution, of up to 30% for multidimensional knapsack, 37% for online generalized assignment, and 50% for online resource allocation. The CE benchmark improves upon the myopic solution in all settings except the multidimensional knapsack with bimodal resource consumption. Then, single-sample OSO yields significant improvements in solution quality over the CE benchmark, (up to 39% for online resource allocation). These improvements are stronger when the distribution is more bimodal (and hence the mean is less representative of a sample) and when capacity is smaller. Finally, multi-sample OSO can yield additional improvements but these become marginally smaller as the number of sample paths increases. Although the single-sample and small-sample OSO methods involve longer computational times, these methods still yield high-quality solutions within the time limit.

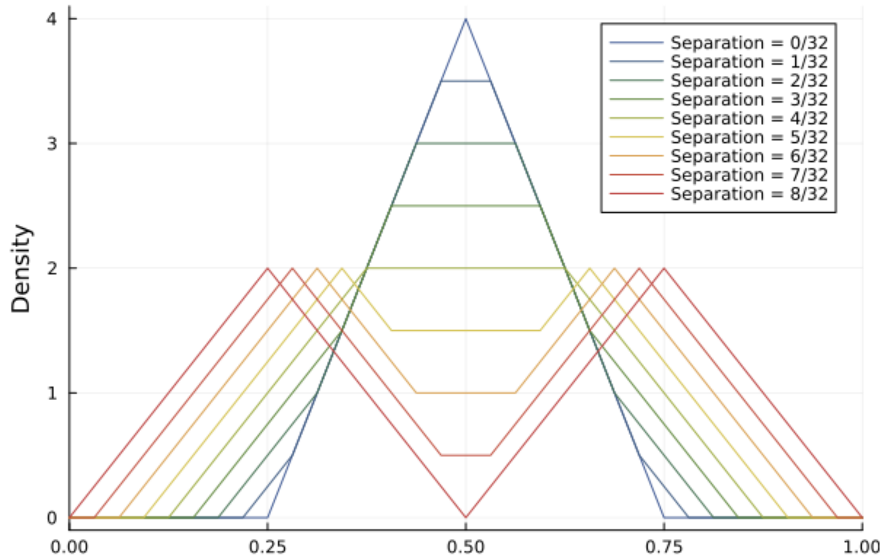


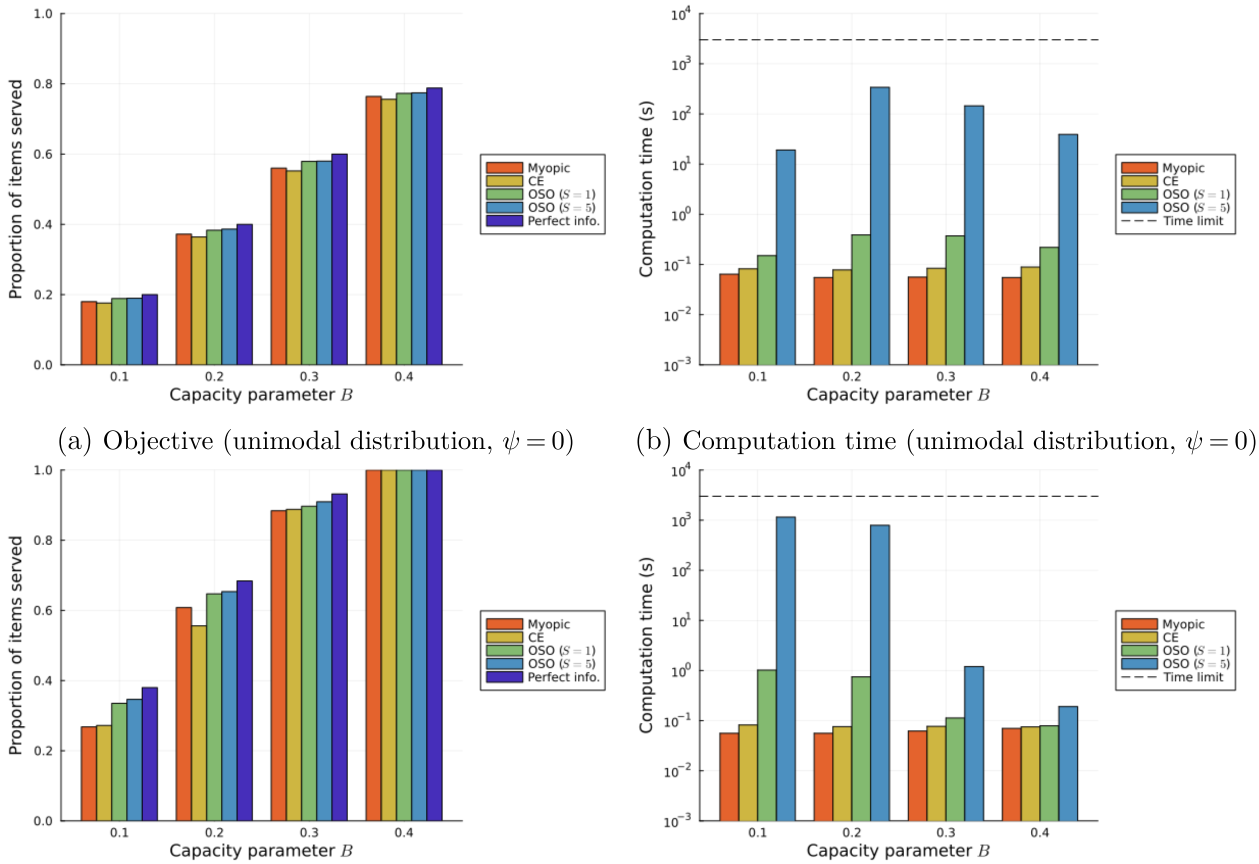
Figure EC.1 Bimodal distribution with separation parameter $\psi \in [0, 0.25]$.

		Objective (% of perfect-info)					Computation time (s)			
ψ	B	Myopic	CE	OSO-1	(over CE)	OSO-5	Myopic	CE	OSO-1	OSO-5
0.0	0.1	90.00	87.9	94.36	+7.3%	94.74	0.064	0.082	0.150	19.1
0.0	0.2	92.97	90.98	95.80	+5.3%	96.59	0.055	0.078	0.388	339
0.0	0.3	93.31	91.99	96.53	+4.9%	96.67	0.056	0.084	0.372	145
0.0	0.4	96.94	95.94	98.08	+2.2%	98.28	0.055	0.089	0.219	39.0
0.25	0.1	70.32	71.45	88.20	+23.4%	91.15	0.056	0.082	1.02	1150
0.25	0.2	88.81	81.22	94.60	+16.5%	95.55	0.056	0.076	0.748	790
0.25	0.3	94.83	95.25	96.21	+1.0%	97.60	0.062	0.077	0.113	1.20
0.25	0.4	100.0	100.0	100.0	+0.0%	100.0	0.070	0.075	0.079	0.191

Table EC.2 Objective relative to the hindsight optimum (in percent) and computation time (in seconds) for the online multi-dimensional knapsack problem. "OSO-1", "OSO-5": OSO algorithm with 1, 5 sample paths per iteration.

		Objective (% of perfect-info)						Computation time (s)					
ψ	B	Myopic	CE	OSO-1	OSO-5	OSO-10	OSO-20	Myopic	CE	OSO-1	OSO-5	OSO-10	OSO-20
0.0	0.2	74.38	95.93	93.58	94.17	95.98	95.37	0.077	0.221	0.575	6.44	14.8	35.8
0.0	0.3	77.84	96.64	96.12	95.56	96.38	96.09	0.076	0.228	4.22	53.0	164	502
0.0	0.4	80.39	95.98	95.49	96.69	97.69	97.8	0.059	0.238	17.3	188	682	1200
0.0	0.5	84.35	91.57	94.86	97.84	98.24	98.48	0.069	0.229	21.5	45.3	118	264
0.25	0.2	68.22	78.63	84.94	87.37	88.12	91.41	0.080	0.230	7.03	76.7	200	535
0.25	0.3	63.24	68.52	81.49	86.23	86.76	87.27	0.071	0.252	213	1030	1310	1560
0.25	0.4	69.11	72.3	84.79	88.47	90.07	89.99	0.059	0.226	841	1620	1850	1980
0.25	0.5	85.97	88.78	96.87	98.07	98.24	99.60	0.060	0.228	0.562	10.1	11.7	23.1

Table EC.3 Objective relative to the hindsight optimum (in percent) and computation time (in seconds) for the online generalized assignment problem. "OSO-S": OSO algorithm with S sample paths per iteration.



(a) Objective (unimodal distribution, $\psi = 0$)

(b) Computation time (unimodal distribution, $\psi = 0$)

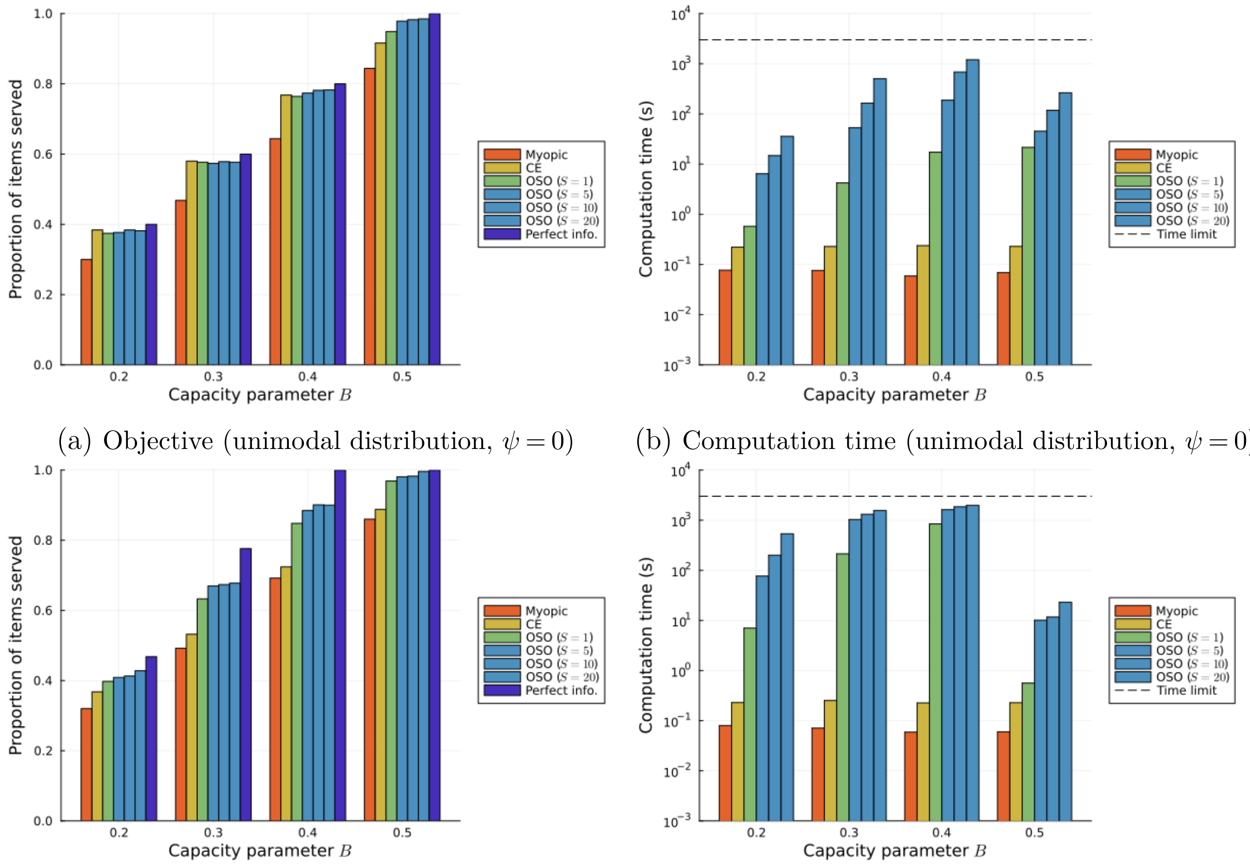
(c) Objective (bimodal distribution, $\psi = 1/4$)

(d) Computation time (bimodal distribution, $\psi = 1/4$)

Figure EC.2 Normalized objectives and computation times for the online multidimensional knapsack problem.

		Objective (% of perfect-info)						Computation time (s)			
ψ	B	Myopic	CE	(over Myo.)	OSO-1	(over CE)	OSO-5	Myopic	CE	OSO-1	OSO-5
0.0	0.1	78.80	84.76	+7.6%	87.48	+3.2%	90.68	0.039	0.600	87.7	5770
0.0	0.2	77.32	81.89	+5.9%	91.77	+12.1%	93.66	0.040	0.701	442	7890
0.0	0.3	80.27	83.33	+3.8%	93.78	+12.5%	95.11	0.042	0.869	686	8310
0.25	0.1	50.72	60.71	+19.7%	84.33	+38.9%	88.44	0.040	0.637	150	7440
0.25	0.2	54.11	66.00	+22.0%	89.61	+35.7%	93.24	0.039	1.20	670	8490
0.25	0.3	60.87	73.75	+21.2%	92.71	+25.7%	95.50	0.042	1.42	1890	6590

Table EC.4 Objective relative to the hindsight optimum (in percent) and computation time (in seconds) for the online resource allocation problem. "OSO-1", "OSO-5": OSO algorithm with 1, 5 sample paths per iteration.



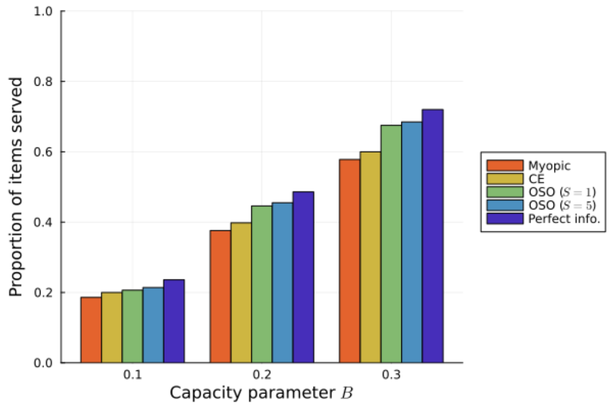
(a) Objective (unimodal distribution, $\psi = 0$)

(b) Computation time (unimodal distribution, $\psi = 0$)

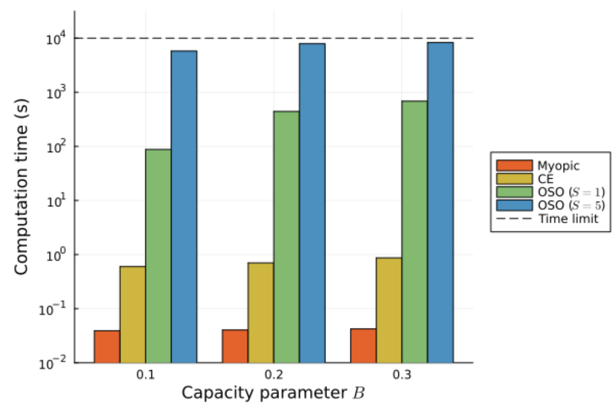
(c) Objective (bimodal distribution, $\psi = 1/4$)

(d) Computation time (bimodal distribution, $\psi = 1/4$)

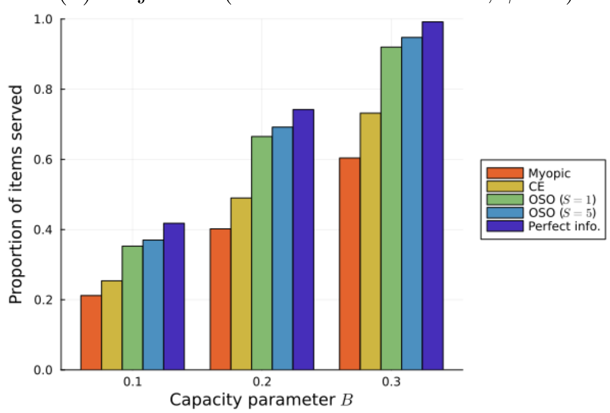
Figure EC.3 Normalized objectives and computation times for the online generalized assignment problem.



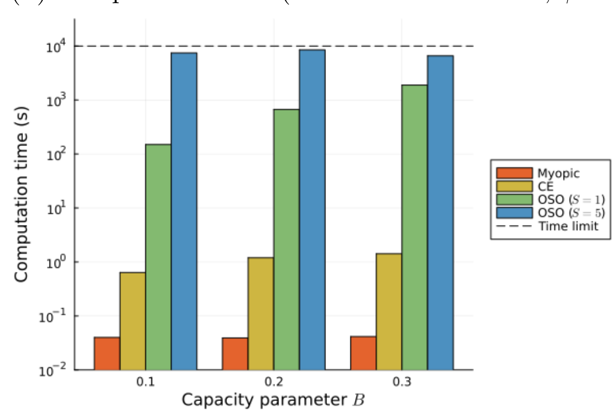
(a) Objective (unimodal distribution, $\psi = 0$)



(b) Computation time (unimodal distribution, $\psi = 0$)



(c) Objective (bimodal distribution, $\psi = 1/4$)



(d) Computation time (bimodal distribution, $\psi = 1/4$)

Figure EC.4 Normalized objectives and computation times for the online resource allocation problem.

EC.3. Online Batched Bin-packing

EC.3.1. Problem Statement and MSSIP Formulation

Definition EC.6 (Online batched bin packing). Items arrive over T time periods. All bins have capacity B. At each time period t, a batch $\mathcal{I}^t$ of q items are revealed. Each item $i \in \mathcal{I}^t$ has size $V_i^t \in \{0,1,\ldots,B\}$. The objective is to pack items in as few bins as possible.

We use the flow-based formulation from Valério de Carvalho (1999); this formulation exhibits much stronger scalability than other formulation with a looser linear relaxation, in our instances. This formulation relies on a network representation with node set: $\mathcal{N} = \{0,\ldots,B\}$, packing arcs: $\{(i,j) : 0 \le i < j \le B\}$, and loss arcs corresponding to wasted capacity $\{(i,i+1) : 0 \le i < B\}$. A packing is characterized by a path from node 0 to node B. Figure EC.5 shows an example.

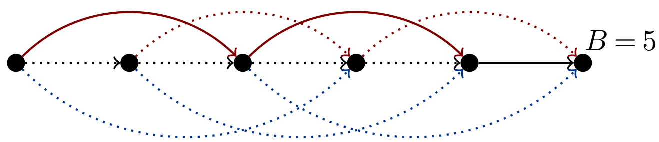


Figure EC.5 Flow-based bin packing representation. Red (resp. blue) arcs denote items of size 2 (resp. 3). The bin capacity is 5. The solution in solid lines packs two items of size 2.

We define the following decision variables:

$$\begin{aligned} x_{ij} &= \text{number of items of size } j-i \text{ placed in any bin starting in "position" } i \\ w_j &= \text{number of loss arcs used across all bins starting in "position" } j \\ z &= \text{number of bins opened} \end{aligned}$$

The offline bin-packing formulation for a set of items $\mathcal{I}$ is given as follows. Equation (EC.78a) minimizes the number of bins. Equation (EC.78b) defines packing solutions as a path from the source to the sink, and Equation (EC.78c) ensures that all items are placed in a bin (denoting $c_s(\mathcal{I})$ as the number of size-s items in $\mathcal{I}$).

$$\min \quad z \tag{EC.78a}$$

$$\text{s.t.} \quad \sum_{i=0}^{j-1} x_{ij} - \sum_{k=j+1}^{B} x_{jk} = \begin{cases} -z + w_j & j = 0 \\ -w_{j-1} + w_j & j \neq 0, B \\ -w_{j-1} + z & j = B \end{cases} \tag{EC.78b}$$

$$\sum_{j=0}^{B-s} x_{j,j+s} = c_s(\mathcal{I}) \quad \forall\ s = 1,\ldots,B \tag{EC.78c}$$

$$z, \boldsymbol{x}, \boldsymbol{w} \text{ nonnegative integer} \tag{EC.78d}$$

Definition EC.7. Let $\mathcal{I}$ be a set of items. We denote by $\mathcal{F}(\mathcal{I})$ the feasible set of Equation (EC.78), projected on the $\boldsymbol{x}$ and z variables:

$$\mathcal{F}(\mathcal{I}) := \left\{ \boldsymbol{x} \in \mathbb{Z}_+^{B(B+1)/2}, z \in \mathbb{Z}_+ \,\middle|\, \exists\, \boldsymbol{w} \in \mathbb{Z}_+^B : \text{ Equations (EC.78b) and (EC.78c)} \right\} \tag{EC.79}$$

In the online problem, items come in batch $\mathcal{I}^t$ at time $t \in \mathcal{T}$. We denote the uncertainty in batch t by $\boldsymbol{\Xi}^t = \boldsymbol{V}^t$, with realized batches $\mathcal{I}^t$ and sampled batches $\widetilde{\mathcal{I}}^t$; we denote by $\boldsymbol{x}^t, z^t$ the decision variables at time $t \in \mathcal{T}$. Let also $\mathcal{I}^{1:t-1} := \mathcal{I}^1 \cup \cdots \cup \mathcal{I}^{t-1}$ denote the set of all past items, and by $\boldsymbol{x}^{1:t-1} := \sum_{\tau=1}^{t-1} \boldsymbol{x}^\tau$ and $z^{1:t-1} := \sum_{\tau=1}^{t-1} z^\tau$ the cumulative past decisions (vector of utilization and number of bins). We can express the per-period objective function and feasible set for the current decisions $(\boldsymbol{x}^t, z^t)$ as:

$$f^t(\boldsymbol{x}^t, z^t, \mathcal{I}^{1:t}) := z^t \tag{EC.80}$$

$$\mathcal{F}_t(\boldsymbol{x}^{1:t-1}, z^{1:t-1}, \mathcal{I}^{1:t}) := \left\{ \boldsymbol{x}^t \in \mathbb{Z}_+^{B(B+1)/2},\ z^t \in \mathbb{Z}_+ \,\middle|\, (\boldsymbol{x}^{1:t-1} + \boldsymbol{x}^t, z^{1:t-1} + z^t) \in \mathcal{F}(\mathcal{I}^{1:t}) \right\} \tag{EC.81}$$

Therefore, the multi-stage stochastic programming formulation can be expressed as follows:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Xi}^{1:T}}\Bigg[\max_{(\boldsymbol{x}^1, z^1) \in \mathcal{F}_1(\mathcal{I}^1)} \Bigg\{ f^1(\boldsymbol{x}^1, z^1, \mathcal{I}^1) + \mathbb{E}_{\boldsymbol{\Xi}^{2:T}}\Bigg[\max_{(\boldsymbol{x}^2, z^2) \in \mathcal{F}_2(\boldsymbol{x}^1, z^1, \mathcal{I}^{1:2})} \Bigg\{ f^2(\boldsymbol{x}^2, z^2, \mathcal{I}^{1:2}) + \ldots \\ + \mathbb{E}_{\boldsymbol{\Xi}^T}\Bigg[\max_{(\boldsymbol{x}^T, z^T) \in \mathcal{F}_T(\boldsymbol{x}^{1:T-1}, z^{1:T-1}, \mathcal{I}^{1:T})} \Big\{ f^T(\boldsymbol{x}^T, z^T, \mathcal{I}^{1:T}) \Big\} \Bigg] \ldots \Bigg\} \Bigg] \Bigg\} \Bigg]\end{aligned} \tag{EC.82}$$

The single-sample OSO algorithm for the online batched bin packing problem is given in Algorithm 6. We consider a variant where all uncertainty realizations are sampled at the beginning of the horizon, for ease of theoretical analysis. We define the following problem at time t:

Definition EC.8. We denote by $\mathrm{IP}^t\left(\mathcal{I}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right)$ the following integer program which is solved at time t of the OSO algorithm, giving optimal solution $(\widetilde{\boldsymbol{x}}^t, \widetilde{\boldsymbol{x}}^{t+1:T}, \widetilde{z}^t, \widetilde{z}^{t+1:T})$:

$$\min \quad \overline{z}^{1:t-1} + z^t + z^{t+1:T} \tag{EC.83a}$$

$$\text{s.t.} \quad (\boldsymbol{x}^t, z^t) \in \mathcal{F}_t(\overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}, \mathcal{I}^{1:t}) \tag{EC.83b}$$

$$(\boldsymbol{x}^\tau, z^\tau) \in \mathcal{F}_t(\overline{\boldsymbol{x}}^{1:t-1} + \boldsymbol{x}^{t:\tau-1}, \overline{z}^{1:t-1} + z^{t:\tau-1}, \mathcal{I}^{1:t} \cup \widetilde{\mathcal{I}}^{t+1:\tau}) \quad \forall\, \tau \in \{t+1, \ldots, T\} \tag{EC.83c}$$

We call $(\widetilde{\boldsymbol{x}}^t, \ldots, \widetilde{\boldsymbol{x}}^T)$ an optimal extension of $\overline{\boldsymbol{x}}^{1:t-1}$ given $(\mathcal{I}^t, \widetilde{\mathcal{I}}^{t+1:T})$. We also denote by $\mathrm{OPT}^t\left(\mathcal{I}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right)$ the corresponding optimal objective value, which is the minimum number of bins that can hold the items $\mathcal{I}^{1:t} \cup \widetilde{\mathcal{I}}^{t+1:T}$.

EC.3.2. Proof of Theorem 2.

The proof proceeds by tracking the evolution of $\mathrm{OPT}^t\left(\widetilde{\mathcal{I}}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right)$, that is, of the cost estimate given that the algorithm has already made decisions $\overline{\boldsymbol{x}}^{1:t-1}$ using $\overline{z}^{1:t-1}$ bins. When $t = 1$, this cost is $\mathrm{OPT}^1\left(\widetilde{\mathcal{I}}^1, \widetilde{\mathcal{I}}^{2:T} \,\middle|\, \mathbf{0}, 0\right)$, which is equal to the true optimum in expectation (since $\mathcal{I}^t$ and $\widetilde{\mathcal{I}}^t$

Algorithm 6 Single-sample OSO for the online batched bin packing problem.

Sample: Sample batches $\widetilde{\mathcal{I}}^1,\ldots,\widetilde{\mathcal{I}}^T$ from the common distribution $\mathcal{D}$.

Repeat, for $t\in\{1,\ldots,T\}$:

Observe: Observe true batch of items $\mathcal{I}^t$.

Optimize: Solve $\mathrm{IP}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$, with optimal solution $(\widetilde{\boldsymbol{x}}^t,\widetilde{\boldsymbol{x}}^{t+1:T},\widetilde{z}^t,\widetilde{z}^{t+1:T})$.

Implement: Implement $\overline{\boldsymbol{x}}^t=\widetilde{\boldsymbol{x}}^t$ and $\overline{z}^t=\widetilde{z}^t$, discarding $\widetilde{\boldsymbol{x}}^{t+1:T}$ and $\widetilde{z}^{t+1:T}$.

are sampled from the same distribution). At time $t=T+1$, this cost is $\mathrm{OPT}^{T+1}\left(\varnothing,\varnothing\,\middle|\,\overline{\boldsymbol{x}}^{1:T},\overline{z}^{1:T}\right)$, which is the total cost of the OSO algorithm over the true instance. Therefore, to prove Theorem 2, it suffices to bound:

$$\mathrm{OPT}^{t+1}\left(\widetilde{\mathcal{I}}^{t+1},\widetilde{\mathcal{I}}^{t+2:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t},\overline{z}^{1:t}\right)-\mathrm{OPT}^{t}\left(\widetilde{\mathcal{I}}^{t},\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right). \tag{EC.84}$$

Consider a fixed round t. Since $(\widetilde{\boldsymbol{x}}^t,\widetilde{\boldsymbol{x}}^{t+1:T},\widetilde{z}^t,\widetilde{z}^{t+1:T})$ is optimal for $\mathrm{IP}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$, and $\overline{\boldsymbol{x}}^t:=\widetilde{\boldsymbol{x}}^t$, $(\overline{\boldsymbol{x}}^t,\widetilde{\boldsymbol{x}}^{t+1},\ldots,\widetilde{\boldsymbol{x}}^T)$ is an optimal extension of $\overline{\boldsymbol{x}}^{1:t-1}$ given items $(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T})$. Therefore, $(\widetilde{\boldsymbol{x}}^{t+1},\ldots,\widetilde{\boldsymbol{x}}^T)$ is an optimal extension of $\overline{\boldsymbol{x}}^{1:t}$ given items $(\widetilde{\mathcal{I}}^{t+1},\widetilde{\mathcal{I}}^{t+2:T})$. That is:

$$\mathrm{OPT}^{t}\left(\mathcal{I}^{t},\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)=\mathrm{OPT}^{t+1}\left(\widetilde{\mathcal{I}}^{t+1},\widetilde{\mathcal{I}}^{t+2:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t},\overline{z}^{1:t}\right) \tag{EC.85}$$

Thus, upper bounding the difference in Equation (EC.84) is equivalent to upper bounding

$$\mathrm{OPT}^{t}\left(\mathcal{I}^{t},\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)-\mathrm{OPT}^{t}\left(\widetilde{\mathcal{I}}^{t},\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right). \tag{EC.86}$$

That is, given $\overline{\boldsymbol{x}}^{1:t-1}$, we need to show that the total cost is not significantly impacted whether the next batch is $\mathcal{I}^t$ (the actual one) or $\widetilde{\mathcal{I}}^t$ (the sampled one). We leverage "coupling" between the item sizes in $\mathcal{I}^t$ and $\widetilde{\mathcal{I}}^t$ to design an assignment for $\mathcal{I}^t$ from an assignment for $\widetilde{\mathcal{I}}^t$. We make use of *monotone matchings*, which match two values if the latter is at least as large as the former.

Definition EC.9 (Monotone matching). Given two sequences $a_1,\ldots,a_n\in\mathbb{R}$ and $b_1,\ldots,b_n\in\mathbb{R}$, a *monotone matching* π from the a_ℓ's to the b_ℓ's is an injective function from a subset $L\in\{1,\ldots,n\}$ to $\{1,\ldots,n\}$ such that $a_\ell\leq b_{\pi(\ell)}$ for all $\ell\in L$. We say that a_ℓ is *matched* to $b_{\pi(\ell)}$ if $\ell\in L$, and *unmatched* otherwise.

Intuitively, thinking of $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ as sequences of item sizes, a monotone matching indicates that, if item $b_{\pi(\ell)}$ is assigned to some bin, then we can replace it with item a_ℓ without violating the bin's capacity. In other words, we can use an assignment of the items $b_1,\ldots,b_n$ to come up with an assignment of the matched items in $a_1,\ldots,a_n$ using the same bins. Rhee and Talagrand (1993a) showed that if the two sequences are i.i.d. from the same distribution, then almost all items can be matched using a monotone matching.

Theorem EC.3 (**Monotone Matching Theorem (Rhee and Talagrand 1993a)**). *Define independent random variables $A_1,\dots,A_n$ and $B_1,\dots,B_n$ from distribution $\mathcal{D}$ over $[0,1]$. There is a constant c such that with probability at least $1-\exp\left(-c\log^{3/2} n\right)$ there is a monotone matching π of the A_ℓ's to the B_ℓ's where at most $c\sqrt{n}\log^{3/4} n$ of the A_ℓ's are unmatched.*

Using this result we can upper bound the difference given in Equation (EC.86) as follows.

Lemma EC.6. *There is a constant c such that with probability at least $1-\exp\left(-c\log^{3/2} q\right)$,*

$$\mathrm{OPT}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)-\mathrm{OPT}^t\left(\widetilde{\mathcal{I}}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)\le c\sqrt{q}\log^{3/4} q. \tag{EC.87}$$

Proof of Lemma EC.6. The strategy will be to start with the packing solution for $\mathrm{IP}^t\left(\widetilde{\mathcal{I}}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$ and construct a "good enough" packing solution for $\mathrm{IP}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$ i.e., one that uses at most $c\sqrt{q}\log^{3/4} q$ more bins. Then the optimal solution $\mathrm{OPT}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$ will also use at most $c\sqrt{q}\log^{3/4} q$ more bins than $\mathrm{OPT}^t\left(\widetilde{\mathcal{I}}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$, completing the proof.

Let $V_1^t,\dots,V_q^t$ be the realized item sizes in the t-th batch, and $\widetilde{V}_1^t,\dots,\widetilde{V}_q^t$ be the sampled item sizes in the t-th batch. Let $(\widetilde{\boldsymbol{x}}^t,\widetilde{\boldsymbol{x}}^{t+1:T},\widetilde{z}^t,\widetilde{z}^{t+1:T})$ be the optimal solution for $\mathrm{IP}^t\left(\widetilde{\mathcal{I}}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$ given the previous assignments $\overline{\boldsymbol{x}}^{1:t-1}$, with auxillary variables $\boldsymbol{w}^{\text{now}}$ and $\boldsymbol{w}^{\text{next}}$ corresponding to Equations (EC.83b) and (EC.83c). We use a monotone matching to construct a feasible solution for $\mathrm{IP}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$.

Let π be a monotone matching from $\{V_1^t/B,\dots,V_q^t/B\}$, the normalized true item sizes in batch t, to $\{\widetilde{V}_1^t/B,\dots,\widetilde{V}_q^t/B\}$, the normalized sampled item sizes in batch t, given by Theorem EC.3. This implies that for all $\ell\in\{1,\dots,q\}$ matched by π, we have $V_\ell^t\le\widetilde{V}_{\pi(\ell)}^t$. We construct a new solution $(\widehat{\boldsymbol{x}}^t,\widehat{z}^t,\widehat{\boldsymbol{w}}^{\text{now}})$ from $(\widetilde{\boldsymbol{x}}^t,\widetilde{z}^t,\boldsymbol{w}^{\text{now}})$ according to Algorithm 7. The overall idea is that for matched elements in $\{V_1^t,\dots,V_q^t\}$, we can replace sampled items with true items that are not larger than those sampled items. True items that are unmatched are assigned to a new bin each.

We first verify that $(\widehat{\boldsymbol{x}}^t,\widetilde{\boldsymbol{x}}^{t+1:T},\widehat{z}^t,\widetilde{z}^{t+1:T})$ is a solution that packs batches $\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}$ given the history, i.e. is a feasible solution to $\mathrm{IP}^t\left(\mathcal{I}^t,\widetilde{\mathcal{I}}^{t+1:T}\,\middle|\,\overline{\boldsymbol{x}}^{1:t-1},\overline{z}^{1:t-1}\right)$. We show this in two parts:

1. Equation (EC.83b). We claim that $\widehat{\boldsymbol{w}}^{\text{now}}$ would certify that $(\overline{\boldsymbol{x}}^{1:t-1}+\widehat{\boldsymbol{x}}^t,\overline{z}^{1:t-1}+\widehat{z}^t)$ belongs in $\mathcal{F}(\mathcal{I}^{1:t-1}\cup\mathcal{I}^t)$. First note that $\boldsymbol{w}^{\text{now}}$ would certify that $(\overline{\boldsymbol{x}}^{1:t-1}+\widetilde{\boldsymbol{x}}^t,\overline{z}^{1:t-1}+\widetilde{z}^t)$ belongs in $\mathcal{F}(\mathcal{I}^{1:t-1}\cup\widetilde{\mathcal{I}}^t)$. Next, at each iteration of either for loop in Algorithm 7, flow conservation (Equation (EC.78b)) is maintained at all nodes $j\ne 0,B$; flow balance is also maintained at $j=0,B$ regardless of whether a new bin is opened for unmatched items in $\mathcal{I}^t$. Hence flow conservation remains satisfied for $(\overline{\boldsymbol{x}}^{1:t-1}+\widehat{\boldsymbol{x}}^t,\overline{z}^{1:t-1}+\widehat{z}^t,\widehat{\boldsymbol{w}}^{\text{now}})$.

Also, Equation (EC.78c) is satisfied. This is because $\sum_{j=0}^{B-s}\widetilde{x}_{j,j+s}^t=c_s(\widetilde{\mathcal{I}}^t)$ (since $\sum_{j=0}^{B-s}\overline{x}_{j,j+s}^{1:t-1}+\widetilde{x}_{j,j+s}^t=c_s(\mathcal{I}^{1:t-1}\cup\widetilde{\mathcal{I}}^t)$ and $\sum_{j=0}^{B-s}\overline{x}_{j,j+s}^{1:t-1}=c_s(\mathcal{I}^{1:t-1})$), and at each iteration of either for loop in Algorithm 7 one item in $\widetilde{\mathcal{I}}^t$ is swapped out for one item in $\mathcal{I}^t$. Therefore, at termination, $\sum_{j=0}^{B-s}\widehat{x}_{j,j+s}^t=$

Algorithm 7 Algorithm for constructing a solution $(\widehat{\boldsymbol{x}}^t, \widehat{z}^t, \widehat{\boldsymbol{w}}^{\text{now}})$ from $(\widetilde{\boldsymbol{x}}^t, \widetilde{z}^t, \boldsymbol{w}^{\text{now}})$.

Initialization: Define π as a matching between item sizes in $\mathcal{I}^t$ and item sizes in $\widetilde{\mathcal{I}}^t$.

Define $\widehat{\boldsymbol{x}}^t \leftarrow \widetilde{\boldsymbol{x}}^t$, $\widehat{z}^t \leftarrow \widetilde{z}^t$ and $\widehat{\boldsymbol{w}}^{\text{now}} \leftarrow \boldsymbol{w}^{\text{now}}$.

Define $L, M \subset \{1, \ldots, q\}$ as domain, range of π.

Define $L^c \leftarrow \{1, \ldots, q\} \setminus L$ and $M^c \leftarrow \{1, \ldots, q\} \setminus M$.

for $\ell \in L$ **do** $\triangleright$ Replace item $\widetilde{V}^t_{\pi(\ell)}$ with item V^t_ℓ in the same bin

Define $m := \pi(\ell)$.

Define $s := V^t_\ell$ and $\widetilde{s} := \widetilde{V}^t_m$; $V^t_\ell \le \widetilde{V}^t_m$, so $s \le \tilde{s}$.

Define j as the minimum such that $\widehat{x}^t_{j,j+\widetilde{s}} \ge 1$.

Replace packing arc $(j, j+\widetilde{s})$ with a packing arc $(j, j+s)$ and empty arcs from s to $\widetilde{s}$:

$$\widehat{x}^t_{j,j+\widetilde{s}} \leftarrow \widehat{x}^t_{j,j+\widetilde{s}} - 1 \tag{EC.88a}$$

$$\widehat{x}^t_{j,j+s} \leftarrow \widehat{x}^t_{j,j+s} + 1 \tag{EC.88b}$$

$$\widehat{w}^{\text{now}}_k \leftarrow \widehat{w}^{\text{now}}_k + 1 \quad \forall\, k \in \{s, \ldots, \widetilde{s} - 1\} \tag{EC.88c}$$

end for

for $\ell \in L^c$ **do** $\triangleright$ Replace any remaining $\widetilde{V}^t_m$ with item V^t_ℓ in its own bin

Pick any $m \in M^c$; update $M^c \leftarrow M^c \setminus \{m\}$.

Define $s := V^t_\ell$ and $\widetilde{s} := \widetilde{V}^t_m$.

Define j as the minimum such that $\widehat{x}^t_{j,j+\widetilde{s}} \ge 1$.

Replace packing arc $(j, j+\widetilde{s})$ with empty arcs from j to $j+\widetilde{s}$;

$$\widehat{x}^t_{j,j+\widetilde{s}} \leftarrow \widehat{x}^t_{j,j+\widetilde{s}} - 1 \tag{EC.89a}$$

$$\widehat{w}^{\text{now}}_k \leftarrow \widehat{w}^{\text{now}}_k + 1 \quad \forall\, k \in \{j, \ldots, j+\widetilde{s} - 1\} \tag{EC.89b}$$

Add flow on packing arc $(0, s)$ and empty arcs for the rest of the box:

$$\widehat{x}^t_{0,s} \leftarrow \widehat{x}^t_{0,s} + 1 \tag{EC.90a}$$

$$\widehat{z}^t \leftarrow \widehat{z}^t + 1 \tag{EC.90b}$$

$$\widehat{w}^{\text{now}}_k \leftarrow \widehat{w}^{\text{now}}_k + 1 \quad \forall\, k \in \{s, \ldots, B - 1\} \tag{EC.90c}$$

end for

return $(\widehat{\boldsymbol{x}}^t, \widehat{\boldsymbol{w}}^{\text{now}})$

$c_s(\mathcal{I}^t)$, and $\sum_{j=0}^{B-s} \overline{x}^{1:t-1}_{j,j+s} + \widehat{x}^t_{j,j+s} = c_s(\mathcal{I}^{1:t-1} \cup \mathcal{I}^t)$. Hence, $(\widehat{\boldsymbol{x}}^t, \widehat{z}^t)$ satisfies Equation (EC.83b): $(\overline{\boldsymbol{x}}^{1:t-1} + \widehat{\boldsymbol{x}}^t, \overline{z}^{1:t-1} + \widehat{z}^t) \in \mathcal{F}(\mathcal{I}^{1:t-1} \cup \mathcal{I}^t)$.

2. Equation (EC.83c). We claim that $\widehat{\boldsymbol{w}}^{\text{now}} + \boldsymbol{w}^{\text{next}} - \boldsymbol{w}^{\text{now}}$ would certify that $(\overline{\boldsymbol{x}}^{1:t-1} + \widehat{\boldsymbol{x}}^t + \widetilde{\boldsymbol{x}}^{t+1:T}, \overline{z}^{1:t-1} + \widehat{z}^t + \widetilde{z}^{t+1:T})$ belongs in $\mathcal{F}(\mathcal{I}^{1:t-1} \cup \mathcal{I}^t \cup \widetilde{\mathcal{I}}^{t+1:T})$. Firstly it is positive, since $\widehat{\boldsymbol{w}}^{\text{now}}$ starts from $\boldsymbol{w}^{\text{now}}$ and is never decremented in Algorithm 7. Next, Equation (EC.78b) is satisfied for:

- $\left(\overline{\boldsymbol{x}}^{1:t-1} + \widetilde{\boldsymbol{x}}^t, \overline{z}^{1:t-1} + \widetilde{z}^t, \boldsymbol{w}^{\text{now}}\right)$ and $\left(\overline{\boldsymbol{x}}^{1:t-1} + \widetilde{\boldsymbol{x}}^t, +\widetilde{\boldsymbol{x}}^{t+1:T}, \overline{z}^{1:t-1} + \widetilde{z}^t, +\widetilde{z}^{t+1:T}, \boldsymbol{w}^{\text{next}}\right)$, and therefore the difference $(\widetilde{\boldsymbol{x}}^{t+1:T}, \widetilde{z}^{t+1:T}, \boldsymbol{w}^{\text{next}} - \boldsymbol{w}^{\text{now}})$;
- $\left(\overline{\boldsymbol{x}}^{1:t-1} + \widehat{\boldsymbol{x}}^t, \overline{z}^{1:t-1} + \widehat{z}^t, \widehat{\boldsymbol{w}}^{\text{now}}\right)$ from above;
- and therefore the sum $\left(\overline{\boldsymbol{x}}^{1:t-1} + \widehat{\boldsymbol{x}}^t, +\widetilde{\boldsymbol{x}}^{t+1:T}, \overline{z}^{1:t-1} + \widehat{z}^t, +\widetilde{z}^{t+1:T}, \widehat{\boldsymbol{w}}^{\text{now}} + \boldsymbol{w}^{\text{next}} - \boldsymbol{w}^{\text{now}}\right)$.

Finally, Equation (EC.78c) is satisfied because $\widetilde{\boldsymbol{x}}^{t+1:T}$ satisfies the counts of $\widetilde{\mathcal{I}}^{t+1:T}$ for each item size.

We next evaluate the quality of the constructed solution $(\widehat{\boldsymbol{x}}^t, \widetilde{\boldsymbol{x}}^{t+1:T}, \widehat{z}^t, \widetilde{z}^{t+1:T})$. This is by definition equal to $\overline{z}^{1:t-1} + \widehat{z}^t + \widetilde{z}^{t+1:T}$. $\widehat{z}^t$ is equal to $\widetilde{z}^t$ plus the number of unmatched elements in π, which from Theorem EC.3 is at most $c\sqrt{q}\log^{3/4} q$ with probability at least $1 - \exp\left(-c\log^{3/2} q\right)$. Therefore, with probability at least $1 - \exp\left(-c\log^{3/2} q\right)$, we have:

$$\mathrm{OPT}^t\left(\mathcal{I}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right) \le \overline{z}^{1:t-1} + \widehat{z}^t + \widetilde{z}^{t+1:T} \tag{EC.91}$$

$$\le \overline{z}^{1:t-1} + \widetilde{z}^t + \widetilde{z}^{t+1:T} + c\sqrt{q}\log^{3/4} q \tag{EC.92}$$

$$= \mathrm{OPT}^t\left(\widetilde{\mathcal{I}}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right) + c\sqrt{q}\log^{3/4} q \tag{EC.93}$$

which concludes the lemma. □

Proof of Theorem 2. By taking a union bound, Lemma EC.6 holds for all $t \in \{1, \dots, T\}$ with probability at least $1 - T\exp\left(-c\log^{3/2} q\right)$. Under such event, using Equation (EC.85) we have

$$\mathrm{OPT}^{t+1}\left(\widetilde{\mathcal{I}}^{t+1}, \widetilde{\mathcal{I}}^{t+2:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t}, \overline{z}^{1:t}\right) = \mathrm{OPT}^t\left(\mathcal{I}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right) \tag{EC.94}$$

$$\le \mathrm{OPT}^t\left(\widetilde{\mathcal{I}}^t, \widetilde{\mathcal{I}}^{t+1:T} \,\middle|\, \overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}\right) + c\sqrt{q}\log^{3/4} q \tag{EC.95}$$

Telescoping this inequality over $t \in \{1, \dots, T\}$, we obtain:

$$\mathrm{OPT}^{T+1}\left(\varnothing, \varnothing \,\middle|\, \overline{\boldsymbol{x}}^{1:T}, \overline{z}^{1:T}\right) \le \mathrm{OPT}^1\left(\widetilde{\mathcal{I}}^1, \widetilde{\mathcal{I}}^{2:T} \,\middle|\, \mathbf{0}, 0\right) + cT\sqrt{q}\log^{3/4} q \tag{EC.96}$$

Recall that $\mathrm{OPT}^{T+1}\left(\varnothing, \varnothing \,\middle|\, \overline{\boldsymbol{x}}^{1:T}, \overline{z}^{1:T}\right)$ is the cost of our OSO algorithm on the true items, denoted by $\mathrm{ALG}(\mathcal{I}^{1:T})$, and that $\mathrm{OPT}^1\left(\widetilde{\mathcal{I}}^1, \widetilde{\mathcal{I}}^{2:T} \,\middle|\, \mathbf{0}, 0\right)$ is the offline optimum for the sampled items, denoted by $\mathrm{OPT}(\widetilde{\mathcal{I}}^{1:T})$. Thus, with probability at least $1 - T\exp\left(-c\log^{3/2} q\right)$, we have:

$$\mathrm{ALG}(\mathcal{I}^{1:T}) \le \mathrm{OPT}(\widetilde{\mathcal{I}}^{1:T}) + cT\sqrt{q}\log^{3/4} q. \tag{EC.97}$$

Let G be the event that this inequality holds, and let G^c be its complement. Note that G is a random event that depends on both $\mathcal{I}^{1:T}$ (the problem instance) and $\widetilde{\mathcal{I}}^{1:T}$ (the sampling procedure). Taking expectations over the true item sizes, we obtain:

$$\mathbb{E}_{\mathcal{I}^{1:T}}\left[\mathrm{ALG}(\mathcal{I}^{1:T}) \,\middle|\, G\right] \le \mathrm{OPT}(\widetilde{\mathcal{I}}^{1:T}) + cT\sqrt{q}\log^{3/4} q. \tag{EC.98}$$

Therefore, since all items can fit in n bins, we have:

$$\mathbb{E}_{\mathcal{I}^{1:T}}\left[\mathrm{ALG}(\mathcal{I}^{1:T})\right] = \mathbb{E}_{\mathcal{I}^{1:T}}\left[\mathrm{ALG}(\mathcal{I}^{1:T})\,\middle|\, G\right]\mathbb{P}(G) + \mathbb{E}_{\mathcal{I}^{1:T}}\left[\mathrm{ALG}(\mathcal{I}^{1:T})\,\middle|\, G^c\right](1-\mathbb{P}(G)) \tag{EC.99}$$

$$\leq \mathbb{E}_{\mathcal{I}^{1:T}}\left[\mathrm{ALG}(\mathcal{I}^{1:T})\,\middle|\, G\right]\cdot 1 + nT\exp\left(-c\log^{3/2} q\right) \tag{EC.100}$$

$$\leq \mathrm{OPT}(\widetilde{\mathcal{I}}^{1:T}) + cT\sqrt{q}\log^{3/4} q + nT\exp\left(-c\log^{3/2} q\right) \tag{EC.101}$$

and taking a further expectation over the sampled item sizes gives:

$$\mathbb{E}\left[\mathrm{ALG}(\mathcal{I}^{1:T})\right] \leq \mathbb{E}\left[\mathrm{OPT}(\widetilde{\mathcal{I}}^{1:T})\right] + cT\sqrt{q}\log^{3/4} q + nT\exp\left(-c\log^{3/2} q\right) \tag{EC.102}$$

$$= \mathbb{E}\left[\mathrm{OPT}(\mathcal{I}^{1:T})\right] + cT\sqrt{q}\log^{3/4} q + nT\exp\left(-c\log^{3/2} q\right) \tag{EC.103}$$

Finally, the assumption on q guarantees that $n\exp\left(-c\log^{3/2} q\right) \leq \sqrt{q}\log^{3/4} q$, so the expected cost is at most $\mathbb{E}\left[\mathrm{OPT}(\mathcal{I}^{1:T})\right] + \mathcal{O}(T\sqrt{q}\,\log^{3/4} q)$. We conclude by leveraging the fact that $T = \frac{n}{q}$. □

EC.3.3. Computational Results

We compare the OSO algorithm to the resolving heuristics benchmarks for the online batched bin packing problem. The OSO algorithm provisions for a problem-specific regularizer; we add the following regularizer to the OSO problem solved at time t to encourage packing items into fuller bins:

$$\Psi(\boldsymbol{x}^t) = \frac{1}{|\mathcal{I}^t|}\sum_{i=0}^{B-1}\sum_{j=i+1}^{B}\left(1 - \frac{j^2}{B^2}\right)x_{ij}^t \tag{EC.104}$$

This regularizer is similar to other approaches for online bin-packing. Gupta and Radovanović (2020) use a penalty of the form $\exp(-\varepsilon_t N_t(h))$, where $N_t(h)$ denotes the number of bins filled to h at time t to discourage actions which deplete bins of levels with small $N(h)$. Instead, since we investigate bin packing instances with relatively large bins and fewer items, $N(h)$ is often 1 or 0 in our context; due to our batched setting, our regularizer prioritizes packing items into fuller bins.

The problem that OSO solves at each iteration $t \in \mathcal{T}$ is therefore given by:

$$\min \quad \overline{z}^{1:t-1} + z^t + z^{t+1:T} + \Psi(\boldsymbol{x}^t) \tag{EC.105a}$$

$$\text{s.t.} \quad (\boldsymbol{x}^t, z^t) \in \mathcal{F}_t(\overline{\boldsymbol{x}}^{1:t-1}, \overline{z}^{1:t-1}, \mathcal{I}^{1:t}) \tag{EC.105b}$$

$$(\boldsymbol{x}^\tau, z^\tau) \in \mathcal{F}_t(\overline{\boldsymbol{x}}^{1:t-1} + \boldsymbol{x}^{t:\tau-1}, \overline{z}^{1:t-1} + z^{t:\tau-1}, \mathcal{I}^{1:t} \cup \widetilde{\mathcal{I}}^{t+1:\tau}) \quad \forall\ \tau \in \{t+1,\dots,T\} \tag{EC.105c}$$

We construct online batched bin packing instances with T time periods, each with a batch of $|\mathcal{I}^t| = q$ items. Item sizes are drawn from a uniform distribution over $\{0,1,\dots,100\}$ with $B = 100$. For each combination of parameters, we generate 10 random instances and, for each one, we run OSO 5 times. We impose a time limit of 100 seconds for the integer program solved at each iteration. Table EC.5 reports the objective values and computation times.

			Objective increase				Computation time (s)			
With $\Psi(\cdot)$?	T	q	Myopic	CE	OSO-1	OSO-5	Myopic	CE	OSO-1	OSO-5
✗	16	64	+4.794%	+4.490%	+3.634%	+0.933%	2.811	4.780	11.93	393.7
✗	32	32	+6.693%	+6.440%	+4.942%	+1.153%	3.405	6.956	14.46	572.5
✗	64	16	+6.627%	+8.198%	+5.536%	+1.506%	4.434	8.428	20.25	398.3
✓	16	64	+1.439%	+1.363%	+1.318%	+0.715%	52.56	74.63	188.0	437.5
✓	32	32	+1.811%	+2.007%	+1.915%	+1.212%	73.60	108.4	211.8	1026
✓	64	16	+2.096%	+2.154%	+2.141%	+1.420%	5.898	69.80	159.8	850.6

Table EC.5 Geometric mean of percentage increase in bins opened over hindsight-optimal benchmark, and mean computational time in seconds. Averages taken over 10 random instances, and 5 random samples per instance for OSO. "OSO-1", "OSO-5": OSO algorithm with 1, 5 sample paths per iteration.

These results confirm our insights from the online resource allocation problem, showing that OSO improves upon the myopic and CE solutions, and that performance improves with more samples per iteration at the cost of longer computation times. Without the regularizer, the CE solution barely improves upon the myopic solution with large and medium batch sizes, and actually leads to deteriorated solutions with small batch sizes; in comparison, OSO-1 consistently returns higher-quality solutions than both benchmarks, with reductions in wasted capacity around 1–2 percentage points. Increasing the number of sample paths can achieve further cost improvements, albeit with one to two orders of magnitude increases in computational times. Moreover, these results show the impact of the regularizer in the online batched bin packing problem. Adding the regularizer can result in solution improvements across the board, at the cost of longer computational times. Still, the OSO-5 solution without the regularizer outperforms all benchmarks with the regularizer, further demonstrating the benefits of the sampling approach at the core of the OSO algorithm. Altogether, these results highlight the role of sampling and re-optimization to manage online arrivals in batched bin packing.

EC.4. Rack Placement

Here, we provide details on our experimental setup for the rack placement problem, and the modeling modifications made during the deployment process.

EC.4.1. Experimental Setup

We build a simulated datacenter with two identical rooms. Each room has 4 top-level UPS devices; each UPS device is connected to 6 mid-level PDU devices, and each PDU device is connected to 3 leaf-level PSU devices. The regular capacity of each mid-level PDU and leaf-level PSU is respectively 20% and 60% of their parent's capacity. The top-level UPS devices have regular capacities equal to 75% of their failover capacities, while the regular capacities of the PDU and PSU devices is 50% of their failover capacities. Each room has 36 rows, each with 20 tiles. Each row is connected to two PSU devices in the same room with different parent PDU and UPS devices.

The reward is identical across demand requests, set to $r_i = 200$, while the number of racks n_i, power requirement per rack ρ_i, and cooling requirement per rack γ_i for demand request i are constructed from empirical distributions.

In computational experiments, the perfect-information benchmark was solved with a 1-hour time limit. The resolving heuristics (OSO and the myopic and CE benchmarks) were solved with a 300-second limit per iteration and a 1% optimality gap. We considered instances with a total of 150 items, thus 150, 30, or 15 time periods (corresponding to batch sizes of 1, 5, 10 respectively).

EC.4.2. Modeling Modifications in Production

We detail the modifications we made to our rack placement model discussed in Section 5 to closely align recommendations with real-world considerations and preferences from data center managers. These modifications come in the form of the following secondary objectives. Throughout, we applied a small weight to these objectives to retain the primary goal of maximizing data center utilization.

– *Room minimization.* This objective incentivizes compact data center configurations to reduce operational overhead. The room of row $r \in \mathcal{R}$ is denoted by $\text{room}(r) \in \mathcal{M}$. We add a new binary variable w^t_{im} denoting if room m contains demand $i \in \mathcal{I}_t$. We add a term $-\sum_{i\in\mathcal{I}_t}\sum_{m\in\mathcal{M}}\lambda_m w^t_{im}$ to the objective to penalize placements in emptier rooms, where λ_m is larger for emptier rooms. We link this variable to the placement decisions $\boldsymbol{y}^t$:

$$y^t_{ir} \le w^t_{i,\text{room}(r)} \qquad \forall\, i \in \mathcal{I}_t,\ \forall\, r \in \mathcal{R} \tag{EC.106}$$

– *Row minimization.* This objective also incentivizes compact configurations to reduce operational overhead. We add a new binary variable z^t_r indicating whether row $r \in \mathcal{R}$ contains racks from the current demand batch $\mathcal{I}_t$. We add a term $-\sum_{r\in\mathcal{R}}\theta_r z^t_r$ to the objective where θ_r is larger for rows $r \in \mathcal{R}$ with fewer placed racks. We add the following linking constraints:

$$y^t_{ir} \le z^t_r \qquad \forall\, i \in \mathcal{I}_t,\ \forall\, r \in \mathcal{R} \tag{EC.107}$$

– *Tile group minimization.* This objective incentivizes placing multi-rack reservations on identical tile groups to facilitate customer service down the road and, again, to reduce overhead—in practice, tiles belonging to the same tile groups are located in the same part of the row, so this objective promotes contiguity. We add a binary variable v^t_{ij} denoting if tile group j is used by demand $i \in \mathcal{I}_t$, penalizing it by a parameter τ. We add the following linking constraint:

$$x^t_{ij} \le n_i \cdot v^t_{ij} \qquad \forall\, i \in \mathcal{I}_t,\ \forall\, j \in \mathcal{J} \tag{EC.108}$$

– *Power balance.* This objective encourages balanced power loads to avoid overloads and mitigate maintenance operations. Let $\mathcal{P}_m^{UPS}$ store top-level power devices in room $m \in \mathcal{M}$. The power devices in $\mathcal{P}_m^{UPS}$ share a load of $\sum_{p \in \mathcal{P}_m^{UPS}} P_p$; all capacities P_p are identical in practice; and there are $\binom{|\mathcal{P}_m^{UPS}|}{2}$ pairs of distinct power devices in room m. Accordingly, if all pairs of distinct power devices share the same load, each pair will power a load of $\frac{1}{\binom{|\mathcal{P}_m^{UPS}|}{2}} \sum_{p' \in \mathcal{P}_m^{UPS}} P_{p'}$. We refer to this quantity as the pair-wise balanced load. We first minimize the surplus load $\Phi^t \in \mathbb{R}_+$ for any pair of top-level UPS devices (p,q) as the difference between the total load allocated to power devices p and q and the pair-wise balanced load:

$$\Phi^t \geq \sum_{\tau=1}^{t-1} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \sum_{i \in \mathcal{I}_\tau} \rho_i \overline{x}_{ij}^\tau + \sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \rho_i x_{ij}^t - \frac{1}{\binom{|\mathcal{P}_m^{UPS}|}{2}} \sum_{q \in \mathcal{P}_m^{UPS}} P_q, \quad \forall\, m \in \mathcal{M},\ \forall\, p,p' \in \mathcal{P}_m^{UPS} \tag{EC.109}$$

Second, we minimize the largest power load difference across all pairs of top-level UPS devices. This is written as $\Gamma_{\mathrm{U}}^t - \Gamma_{\mathrm{L}}^t$, where $\Gamma_{\mathrm{U}}^t, \Gamma_{\mathrm{L}}^t \in \mathbb{R}_+$ are defined as follows:

$$\Gamma_{\mathrm{U}}^t \geq \sum_{\tau=1}^{t-1} \sum_{i \in \mathcal{I}_\tau} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \rho_i \overline{x}_{ij}^\tau + \sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \rho_i x_{ij}^t, \qquad \forall\, m \in \mathcal{M},\ \forall\, p,p' \in \mathcal{P}_m^{UPS} \tag{EC.110}$$

$$\Gamma_{\mathrm{L}}^t \leq \sum_{\tau=1}^{t-1} \sum_{i \in \mathcal{I}_\tau} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \rho_i \overline{x}_{ij}^\tau + \sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{J}_p \cap \mathcal{J}_{p'}} \rho_i x_{ij}^t, \qquad \forall\, m \in \mathcal{M},\ \forall\, p,p' \in \mathcal{P}_m^{UPS} \tag{EC.111}$$

The modified objective function at time t is then given by

$$f_t(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{\xi}^{1:t}) - \sum_{i \in \mathcal{I}_t} \sum_{m \in \mathcal{M}} \lambda_m w_{im}^t - \tau \sum_{i \in \mathcal{I}_t} \sum_{j \in \mathcal{J}} v_{ij}^t - \sum_{r \in \mathcal{R}} \theta_r z_r^t - \alpha \Phi^t - \beta(\Gamma_{\mathrm{U}}^t - \Gamma_{\mathrm{L}}^t) \tag{EC.112}$$

EC.4.3. Production Setting

In our rack placement algorithm deployed in production, the room minimization parameter λ_m is equal to 40, 3 and 0, for rooms up to 0%, 20%, 100% full respectively at the start of batch $\mathcal{I}_t$. The row minimization parameter θ_r is equal to 2, 1, and 0 for rows up to 0%, 50%, and 100% full respectively at the start of batch $\mathcal{I}_t$. We have objective weights of $\tau = 1$, $\alpha = 10^{-3}$, $\beta = 10^{-5}$ for the tile group minimization parameter, the power surplus parameter, and the power load difference parameter. Since reward parameters are set to $r_i = 200$, these objectives remain secondary objectives in the model.

One difference between the generic multi-stage stochastic optimization framework considered in this paper and the rack placement problem is that the latter does not evolve in a well-specified finite horizon T; rather, the horizon terminates when the data center can no longer accommodate incoming requests. Accordingly, we define a moving horizon: at each time period t, we sample requests for k_t periods, where k_t is determined so that future requests fill all non-empty rooms. We also prioritize incoming demands over sampled demands via a corresponding weight in the objective function—in particular, we ensure that the current requests are placed if they can be placed.

EC.5. Propensity Score Matching Analysis.

We use propensity score matching (PSM) to corroborate our OLS regression estimates reported in Section 6. Specifically, we partition data centers into high- and low-adoption categories; we use a logistic regression model with the seven control variables to obtain each data center's propensity score; and we then match each high-adoption data center to its neighbors within the low-adoption population, allowing for replacement (Rosenbaum and Rubin 1983). For robustness, we repeat the procedure with two thresholds between high- and low-adoption data centers (60% and 45%) and with 1, 2 and 4 neighbors within the low-adoption population for each high-adoption data center. The propensity score model has an area under the curve of 0.75 with a 60% cutoff and of 0.71 with a 45% cutoff, which satisfy the target threshold of 0.70 (DeFond et al. 2017).

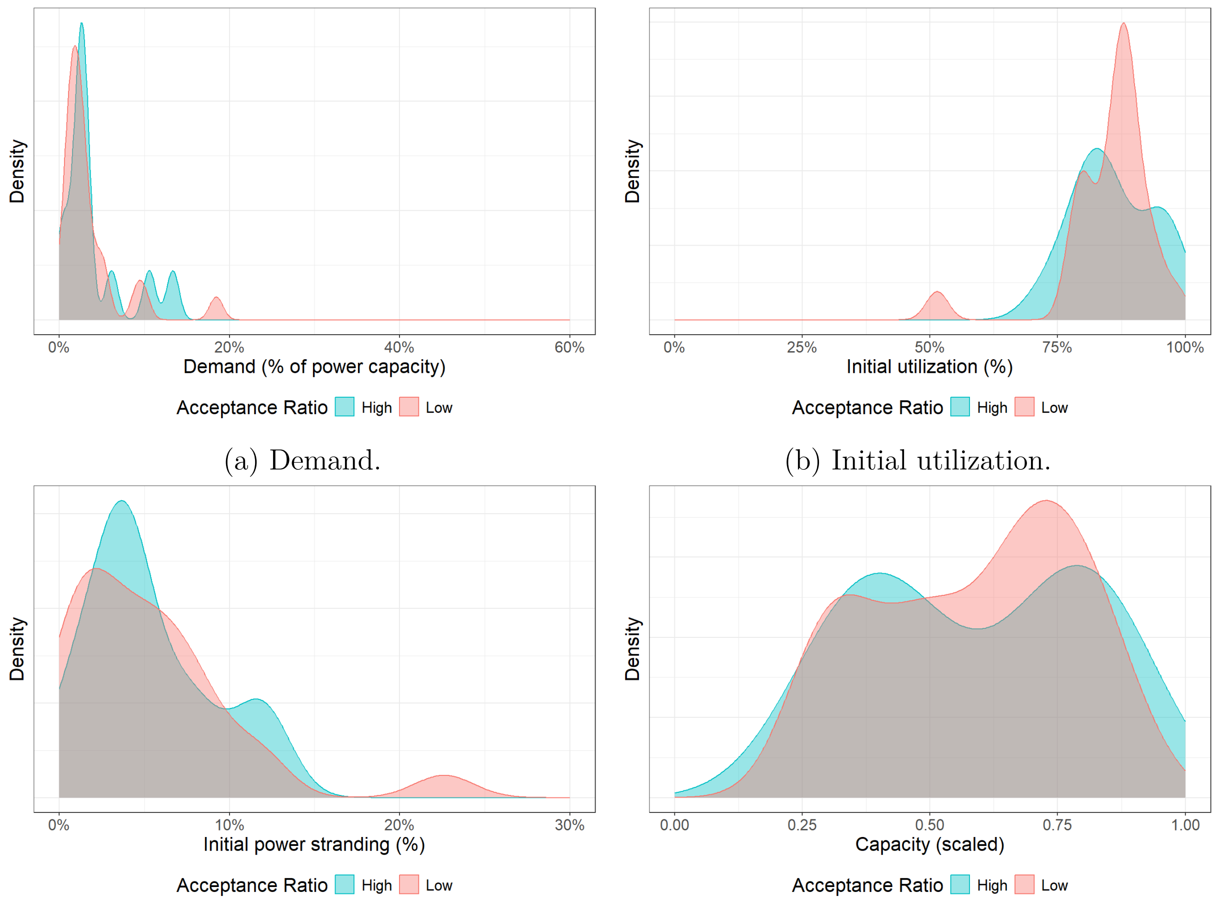
Figure EC.6 reports the distribution of the four continuous control variables across the high-adoption population and the matched low-adoption population. The visualization suggests that the two matched groups are highly balanced. To corroborate this observation, Table EC.6 shows that PSM mitigates the differences between the high- and low-adoption data centers across control variables. Specifically, the table indicates a standardized bias around 0.1, a variance ratio below 2, and a Kolmogorov-Smirnov statistic around 0.1–0.3, all of which are reflective of balanced distributions (Stuart et al. 2013, Rubin 2001). In turn, despite slight remaining disparities due to the small samples and the inherent variability in the control variables, the PSM method matches high-adoption data centers to "similar" low-adoption data centers per the control variables.

Table EC.6 Distributional balance after PSM (using 60% and 45% thresholds and four neighbors).

	Threshold: 60%				Threshold: 45%			
	SB	VR	KS	(p-value)	SB	VR	KS	(p-value)
Demand	0.1321	1.0353	0.2292	(0.5745)	−0.0392	1.2156	0.1563	(0.6274)
Initial utilization	0.0877	0.8033	0.2708	(0.3745)	0.0828	1.9774	0.2292	(0.1962)
Initial power stranding	−0.0084	0.5659	0.2083	(0.6821)	−0.1315	0.4573	0.1771	(0.4661)
IT capacity	0.0221	1.2828	0.1667	(0.8572)	−0.0859	1.1487	0.1771	(0.4321)
Rooms	0.1522	1.4485	0.1458	(0.6246)	−0.1010	1.4169	0.1667	(0.2466)
"Flex" architecture	0.0000	1.0682	0.0000	(1.0000)	0.1251	1.0490	0.0625	(0.6496)
Location (US = 1, Europe = 0)	−0.1549	1.3147	0.0625	(0.6915)	−0.1877	1.2238	0.0833	(0.4435)

Threshold: boundary used to differentiate high-adoption data centers vs. low-adoption data centers.
Standardized bias (SB): difference between means, divided by the pooled standard deviation.
Variance ratio (VR): ratio between the sample variance of the high- and low-adoption data centers.
Kolmogorov-Smirnov (KS) statistic: measure of the distance between cumulative distribution functions.

Finally, we use the PSM dataset to corroborate our previous findings. Figure EC.7 shows the distribution of the outcome variable partitioned between the high-adoption data centers and the matched population, using 60% and 45% thresholds. The qualitative observations echo those from Figure 12, in that the distribution shifts to the left, leading to a lower average increase in power



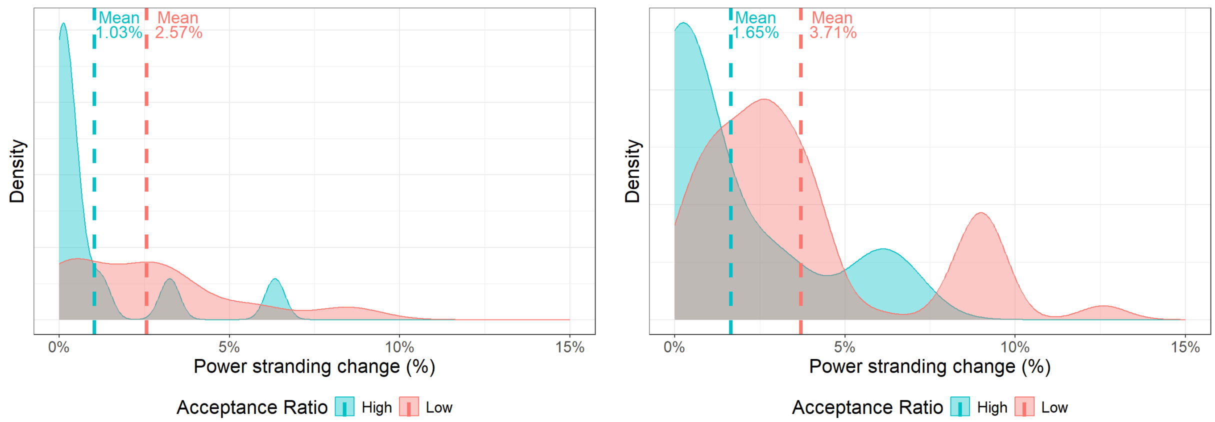
(a) Demand. (b) Initial utilization.

(c) Initial power stranding. (d) Power capacity.

Figure EC.6 Distribution of continuous control variables across data centers after PSM (using a 60% threshold and four neighbors).

stranding among high-adoption data centers than low-adoption ones. Quantitative evidence confirms the impact of adoption on power stranding after controlling for covariates via PSM (+1.03% vs. +2.57% with a 60% threshold, and 1.65% vs. 3.71% with a 45% threshold). The differences are of the same order of magnitude to the one found in the raw data (Figure 12) and remain statistically significant—at the 5% level with a 60% threshold and at the 1% level with a 45% threshold. These results confirm that differences on power stranding are not merely due to the confounding effect of third variables, but can be attributed to differences in adoption of our algorithmic tool.

To conclude, Table EC.7 reports the regression estimates, with and without controls, using the eight PSM samples corresponding to the two adoption thresholds and the three matching neighborhoods along with a no-PSM baseline. By design, the no-PSM estimates without controls are identical to those from Figure 12, and the PSM estimates with four estimates and without controls are identical to those from Figure EC.7; the others provide robustness tests with one and two neighbors, and by adding the control variables in a PSM regression specification. However, the no-PSM estimates do not exactly coincide with those from Table 2 because of the different treatment variable—namely, we



(a) Threshold: 60%. (b) Threshold: 45%.

Figure EC.7 Distribution of the outcome variable across data centers after PSM (using 45% and 60% thresholds between high- and low-adoption data centers, and four low-adoption neighbors for each high-adoption data center), smoothed via Kernel Density Estimation from the `geom_density` function in the `ggplot2` package in R.

used a continuous measure of adoption between 0 and 1 in Table 2 versus a binary treatment variable separating high-adoption from low-adoption data centers in Table EC.7. These results confirm that the impact of adoption on power stranding is negative across all specifications and statistically significant in the majority of cases—in 13 out of the 16 specifications. The magnitude of the coefficients ranges from -1.1% to -2.2%; this suggests that moving a data center from the low-adoption to the high-adoption category could reduce power stranding by 1 to 2 percentage points, which is also consistent with our baseline analysis.

Table EC.7 Regression estimates of the treatment effect across PSM specifications.

		Threshold: 60%				Threshold: 45%			
Controls?		No PSM	1N	2N	4N	No PSM	1N	2N	4N
No	Effect	-0.01988	-0.01547	-0.01213	-0.01542	-0.01734	-0.01243	-0.01725	-0.02064
	p-value	0.01264**	0.09962*	0.1121	0.02887**	0.03854**	0.06167*	0.00702***	0.00069***
Yes	Effect	-0.02259	-0.01196	-0.01319	-0.01601	-0.02037	-0.01108	-0.01802	-0.02178
	p-value	0.0304**	0.216	0.138	0.0441**	0.0220**	0.0880*	0.00572***	0.00032***

1N, 2N, 4N: Results after PSM using 1, 2, and 4 neighbors, respectively.

*, **, and *** indicate significance levels of 10%, 5%, and 1%.