

# An Approximate Dynamic Programming Approach to The Incremental Knapsack Problem

Ali Aouad

Department of Management Science and Operations, London Business School, Regent's Park,  
London, United Kingdom NW14SA, aaouad@london.edu

Danny Segev

Department of Statistics and Operations Research, School of Mathematical Sciences,  
Tel Aviv University, Tel Aviv 69978, Israel, segevdanny@tauex.tau.ac.il

We study the incremental knapsack problem, where one wishes to sequentially pack items into a knapsack whose capacity expands over a finite planning horizon, with the objective of maximizing time-averaged profits. While various approximation algorithms were developed under mitigating structural assumptions, obtaining non-trivial performance guarantees for this problem in its utmost generality has remained an open question thus far. In this paper, we devise a polynomial-time approximation scheme for general instances of the incremental knapsack problem, which is the strongest guarantee possible given existing hardness results. In contrast to earlier work, our algorithmic approach exploits an approximate dynamic programming formulation. Starting with a simple exponentially sized dynamic program, we prove that an appropriate composition of state pruning ideas yields a polynomially sized state space with negligible loss of optimality. The analysis of this formulation synthesizes various techniques, including new problem decompositions, parsimonious counting arguments, and efficient rounding methods, that may be of broader interest.

*Key words:* Incremental Knapsack, approximate dynamic programming, PTAS.

---

## 1. Introduction

Integer packing problems have traditionally been some of the most fundamental and well-studied computational questions in discrete optimization. Contributions along this line of research have ranged from advancing the theory of algorithm design to developing practical tools for tackling a wide range of real-life applications. One of the cornerstone resources in this context is the binary knapsack problem. Here, given a finite collection of items, each associated with profit and weight parameters, one wishes to identify the most profitable subset of items subject to a single packing constraint, in the form of a pre-specified upper bound on the total weight of selected items. Despite its seemingly simple formulation, binary knapsack turned out to incorporate just the right amount of combinatorial complexity to end up as one of Karp's 21 NP-complete problems (1972), and at the same time, to admit various fully polynomial-time approximation schemes (Ibarra and Kim 1975, Lawler 1979, Magazine and Oguz 1981, Kellerer and Pferschy 1999). Due to the large body of literature on the binary knapsack problem and its descendants, we refer the avid reader to excellent

books dedicated to this topic (Toth and Martello 1990, Kellerer et al. 2004) for an exhaustive overview.

*The incremental knapsack problem.* Knapsack-like problems were studied under various forms and shapes motivated by applications such as financial planning, energy management, and logistics. In particular, the thesis work of Sharp (2007) and Hartline (2008) introduced a novel multi-period generalization, dubbed the incremental knapsack problem, as part of their incremental optimization modeling framework; this approach extends classic combinatorial optimization problems into multi-period formulations with time-varying constraints. Deferring the finer details of this setting to Section 1.2, rather than having a single knapsack constraint, we are now considering a discrete planning horizon of  $T$  periods, which are associated with non-decreasing capacities  $W_1 \leq \dots \leq W_T$  as well as with non-negative coefficients  $\lambda_1, \dots, \lambda_T$ . An additional input ingredient is a collection of  $n$  items, where each item  $i \in [n]$  is characterized by a profit  $p_i$  and a weight  $w_i$ . These items are to be incrementally introduced along a nested sequence of subsets  $S_1 \subseteq \dots \subseteq S_T \subseteq [n]$ , subject to time-varying knapsack constraints, stipulating that  $w(S_t) \leq W_t$  for every  $t \in [T]$ . The goal is to maximize the  $\lambda$ -averaged profit,  $\sum_{t \in [T]} \lambda_t \cdot p(S_t)$ . Here,  $w(\cdot)$  and  $p(\cdot)$  are extensions of the weight and profit functions from individual items to item sets, meaning that  $w(S_t) = \sum_{i \in S_t} w_i$  and  $p(S_t) = \sum_{i \in S_t} p_i$ .

This modeling approach naturally captures resource allocation settings where the available resource (e.g., capital, production capacity, demand) grows in a predictable way over a finite planning horizon. A canonical example of such resource allocation problems is financial planning. For instance, well-informed households generally wish to optimize how their cumulative income is invested over time into low-volatility securities that procure regular payments; from that angle, time-varying discounts reflect specific households' investment profiles. Similarly, this approach can be utilized for modeling the selection of investments in bond markets and various government decision processes, as illustrated by Faenza and Malinovic (2018). Additionally, the incremental knapsack problem has certain structural similarities with sequential assortment optimization problems in revenue management (Davis et al. 2015, Gallego et al. 2020, Aouad and Segev 2020, Derakhshan et al. 2018) as well as with the generalized assignment problem (Shmoys and Tardos 1993, Chekuri and Khanna 2005, Fleischer et al. 2011, Feige and Vondrák 2010). Nevertheless, to our knowledge, connections to these lines of research do not lend themselves for the purpose of designing provably-good algorithms for the incremental knapsack problem.

*Previous work.* We proceed by demonstrating that this generalization of the standard binary knapsack problem creates intriguing computational challenges. To this end, we briefly summarize known hardness results and algorithmic findings related to the incremental knapsack problem.

- *Hardness:* Quite surprisingly, Bienstock et al. (2013) and Ye (2016) showed that the incremental knapsack problem is strongly NP-hard. Consequently, the existence of an FPTAS is ruled

out in this computational setting unless  $P = NP$ , thus creating a strong separation from its ancestor, the binary knapsack problem, which is known to admit various such approximation schemes (Ibarra and Kim 1975, Lawler 1979, Magazine and Oguz 1981, Kellerer and Pferschy 1999).

- *Special cases.* Positive results for special cases of the incremental knapsack problem have been appearing in the literature for quite some time. Hartline and Sharp (2006) devised a polynomial-time  $O(1/\log T)$ -approximation for the incremental subset sum problem, where all input parameters are integral and  $p_i = w_i$  for every  $i \in [n]$ . Hartline (2008, Chap. 3.3) subsequently proposed a pseudo-polynomial  $O(nW_T)$ -time dynamic program that yields a  $1/2$ -approximation for incremental subset sum. By extending the standard pseudo-polynomial dynamic program for the binary knapsack problem, Sharp (2007, Chap. 3.4) attained an approximation scheme in this setting. Her algorithm computes a  $(1 - \epsilon)$ -approximate solution in  $O(\frac{1}{\epsilon} \cdot (nW_T)^{O(T)})$  time, meaning that this approach is exponential in  $T$ . Subsequently, Della Croce et al. (2018) focused on the special case of incremental knapsack with  $T = 3$  time periods; they developed an LP-based algorithm with a tight approximation ratio of  $30/37$ .
- *Approximability under parametric assumptions.* More recently, research efforts have been put into algorithmic approaches that are provably good or computationally efficient in specific parametric regimes. For example, by combining LP rounding methods and disjunctive programming ideas, Bienstock et al. (2013) devised a polynomial-time constant-factor approximation; the specific constant attained grows as a function of the minimum parameter  $p$  satisfying  $W_t/W_1 \leq t^p$  for all  $t \in [T]$ . Later on, Della Croce et al. (2019) developed an enumeration-based approximation scheme that identifies an  $(1 - \epsilon)$ -approximate solution in  $O((nT)^{O(T/\epsilon)})$  time; due to its exponential dependency on  $T$ , this approach corresponds to a PTAS for a fixed number of time periods.
- *Time-invariant incremental knapsack.* The time-invariant formulation constitutes a special case of incremental knapsack for which stronger performance guarantees are known to exist. The latter heavily relies on a simplifying structural assumption, where all time periods are uniformly weighed within the objective function, meaning that  $\lambda_1 = \dots = \lambda_T$ . In this setting, Bienstock et al. (2013) attained a PTAS in the regime where the number of time periods is sublogarithmic in the number of items, i.e.,  $T = O(\sqrt{\log n})$ . Recently, Faenza and Malinovic (2018) developed a PTAS for the time-invariant incremental knapsack problem without additional assumptions. Their approach combines techniques such as problem sparsification, compact enumeration, and LP-rounding methods. As a byproduct of these ideas, the authors have also derived a PTAS for a special case of incremental knapsack with non-decreasing time period coefficients, namely, when  $\lambda_1 \leq \dots \leq \lambda_T$ .

*Research questions.* To summarize, while numerous research papers have shed light on various computational aspects, we are not aware of any way to leverage the above-mentioned advances to obtain non-trivial approximations for the incremental knapsack problem in its utmost generality. Quoting the work of Bienstock et al. (2013, Sec. 5): “*Our work leaves the following open questions . . . Is there a polynomial time algorithm for incremental knapsack with a constant-factor approximation ratio that makes no assumption on the growth rate of  $W_t$ ?*”. Similarly, this question has been restated by Faenza and Malinovic (2018, Sec. 3): “*Following Theorem 1, one could ask for a PTAS for the general incremental knapsack problem*”. Our paper seeks to resolve these fundamental open questions.

### 1.1. Our contributions

The main result of this paper resides in developing a PTAS for the incremental knapsack problem in its utmost generality, i.e., without mitigating assumptions on the underlying number of time periods, item weights, or any other parameter. Namely, we show that the optimal profit can be approached within any fixed degree of accuracy via an algorithmic approach whose running time is polynomial in the input size, as formally stated in Theorem 1 below; to avoid cumbersome expressions, we make use of  $\tilde{O}(\cdot)$  to hide polylogarithmic dependencies, i.e.,  $\tilde{O}(f(n)) = O(f(n) \cdot (\log f(n))^{O(1)})$ . Recalling that the incremental knapsack problem has previously been shown to be strongly NP-hard (Bienstock et al. 2013, Ye 2016), this result fully settles its complexity status.

**THEOREM 1.** *For any accuracy level  $\epsilon > 0$ , the incremental knapsack problem can be approximated within factor  $1 - \epsilon$  of optimal. The running time of our algorithm is  $O(n^{\tilde{O}(1/\epsilon^3)} \cdot |\mathcal{I}|^{O(1)})$ , where  $|\mathcal{I}|$  stands for the input size in binary representation.*

From an algorithmic perspective, while existing approximation schemes for special cases of the incremental knapsack problem mainly relied on LP-rounding methods and disjunctive programming (Bienstock et al. 2013, Faenza and Malinovic 2018), our approach is based on approximate dynamic programming. In a nutshell, the foremost ingredient of this formulation is an efficient state pruning method that takes advantage of how certain near-optimal solutions are structured. This approach combines various techniques, including newly developed problem decompositions, parsimonious counting arguments, and efficient rounding methods. For ease of exposition, our approximate dynamic programming ideas are presented in an incremental way throughout the technical parts of this paper, as we proceed to explain.

*Approximation scheme assuming  $\lambda^+$ -boundedness.* In Section 2, we begin by devising an approximation scheme whose running time includes a polynomial dependency on  $(\sum_{t \in [T]} \lambda_t)/\lambda_1$ . Clearly, the latter ratio by itself could be exponential in the input size, meaning that our proposed algorithm constitutes a PTAS only when  $(\sum_{t \in [T]} \lambda_t)/\lambda_1 = |\mathcal{I}|^{O(1)}$ , a property referred to as  $\lambda^+$ -boundedness. To derive this result, we establish a so-called prefix-like property of near-optimal solutions where, by slightly altering item profits, it is optimal to introduce items within each profit class by increasing weights. Given this property, we formulate an intermediate dynamic program whose states encapsulate, among other parameters, a “count vector” that specifies the number of items introduced thus far out of certain profit classes. Our main structural finding shows that this exponentially sized state space can be efficiently shrunk with a negligible loss of optimality. Specifically, we describe rounding operations that transform the original count vectors into highly structured vectors, of which there are polynomially many distinct occurrences.

*General approximation scheme.* Given these results, Section 3 is dedicated to devising a PTAS that applies to any instance of the incremental knapsack problem, with no assumptions whatsoever. This general purpose algorithm is developed via an efficient reduction to the  $\lambda^+$ -bounded setting. To this end, we prove that near-optimal solutions can be decomposed through the connected components of an appropriately defined bipartite graph, representing how items of different profit classes are assigned to “clusters” of time periods, each meeting the  $\lambda^+$ -boundedness property. Moreover, we argue that the best-possible such decomposition can be computed via further dynamic programming ideas.

## 1.2. Problem formulation

In an instance of the incremental knapsack problem, we are given a collection of  $n$  items, where each item  $i \in [n]$  has a profit of  $p_i$  and a weight of  $w_i$ . By reading the remainder of this formulation, it is easy to verify that both parameters can be assumed to be strictly positive. These items may be introduced across a discrete planning horizon that consists of  $T$  time periods, indexed  $1, \dots, T$ , whose respective knapsack capacities are designated by  $W_1 \leq \dots \leq W_T$ . Moreover, each period  $t \in [T]$  is associated with a non-negative coefficient  $\lambda_t$ .

To formally capture how items are inserted over time, we define a solution  $\mathcal{S} = \langle S_1, \dots, S_T \rangle$  to be a nested sequence of items sets  $S_1 \subseteq \dots \subseteq S_T$  falling within the above-mentioned capacities, i.e.,  $w(S_t) \leq W_t$  for every  $t \in [T]$ . We refer to the difference  $S_t \setminus S_{t-1}$  as the collection of items introduced in time period  $t$ , with the convention that  $S_0 = \emptyset$ , meaning that each set  $S_t$  is precisely the collection of items that were introduced up until and including period  $t$ , with a total profit of  $p(S_t)$ . The objective function  $\Psi(\mathcal{S})$  averages the profits of these item sets over all time periods through their  $\lambda$ -coefficients, namely,  $\Psi(\mathcal{S}) = \sum_{t \in [T]} \lambda_t \cdot p(S_t)$ . Our goal is to compute a solution that maximizes this function.

*Equivalent representation of  $\Psi$ .* Moving forward, it will be convenient to rewrite the objective function  $\Psi$  as a summation over “item contributions” rather than over time period profits. For this purpose, let  $\lambda^{t+} = \sum_{\tau=t}^T \lambda_\tau$ , noting that  $\lambda^{1+} \geq \dots \geq \lambda^{T+}$ . We assume without loss of generality that  $\lambda^{T+} > 0$ , since all time periods  $t \in [T]$  with  $\lambda_t = 0$  can be eliminated in advance. Additionally, with respect to a solution  $\mathcal{S} = \langle S_1, \dots, S_T \rangle$  and an item  $i \in S_T$ , we use  $\nu(i, \mathcal{S})$  to denote the unique time period in which this item was introduced. Consequently, by defining  $\Psi_i(\mathcal{S}) = p_i \cdot \lambda^{\nu(i, \mathcal{S})+}$  for any such item, it follows that the objective function can be decomposed into the sum of individual item contributions as follows:

$$\Psi(\mathcal{S}) = \sum_{t \in [T]} \lambda_t \cdot p(S_t) = \sum_{i \in S_T} \left( p_i \cdot \sum_{t=\nu(i, \mathcal{S})}^T \lambda_t \right) = \sum_{i \in S_T} p_i \cdot \lambda^{\nu(i, \mathcal{S})+} = \sum_{i \in S_T} \Psi_i(\mathcal{S}). \quad (1)$$

## 2. PTAS for $\lambda^+$ -Bounded Instances

In what follows, we devise an approximation scheme for the incremental knapsack problem whose running time incorporates a polynomial dependency on the ratio  $\rho = \lambda^{1+}/\lambda^{T+}$  between the largest and smallest  $\lambda^+$  values. The specifics of this result are formally stated in the next theorem.

**THEOREM 2.** *For any accuracy level  $\epsilon > 0$ , the incremental knapsack problem can be approximated within factor  $1 - \epsilon$  of optimal in  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$  time.*

### 2.1. The inverse problem

*Auxiliary formulation.* To provide an accessible account of this result, we introduce a closely related variant, dubbed inverse incremental knapsack. Here, given a profit requirement  $\phi \geq 0$ , suppose that our goal is to identify a solution  $\mathcal{S} = \langle S_1, \dots, S_T \rangle$  that minimizes the total weight of inserted items,  $w(S_T)$ , subject to obtaining a profit of at least  $\phi$ , namely,  $\Psi(\mathcal{S}) \geq \phi$ . Moving forward, we use  $\mathcal{S}^{*\phi}$  to designate a fixed optimal solution for the inverse problem, parameterized by  $\phi$ ; when the latter problem is infeasible, we indicate this situation by setting  $\mathcal{S}^{*\phi} = \perp$ . In the remainder of this section, we develop an approximation scheme for computing a super-optimal solution to the inverse incremental knapsack problem, at the expense of slightly violating the profit constraint. As formally stated in the next claim, our algorithm will incur a polynomial dependency on  $\rho$  as part of its running time.

**THEOREM 3.** *For any accuracy level  $\epsilon > 0$ , given a profit requirement  $\phi \geq 0$  for which  $\mathcal{S}^{*\phi} \neq \perp$ , we can compute in  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$  time a solution  $\mathcal{S}$  satisfying:*

- (a) *Super-optimality:*  $w(S_T) \leq w(S_T^{*\phi})$ .
- (b)  *$\epsilon$ -profit violation:*  $\Psi(\mathcal{S}) \geq (1 - \epsilon) \cdot \phi$ .

*Immediate consequence.* To derive Theorem 2, the important observation is that our algorithm for the inverse formulation can be leveraged to devise an approximation scheme for incremental knapsack in a straightforward way. To this end, letting  $\mathcal{S}^*$  be an optimal solution for the latter problem, suppose we were given in advance an estimate  $\phi \in [(1 - \epsilon) \cdot \Psi(\mathcal{S}^*), \Psi(\mathcal{S}^*)]$  for the optimal profit. It is easy to verify that, since  $\Psi(\mathcal{S}^*) \in [\lambda_{\min} p_{\min}, n \lambda_{\max} p_{\max}] \cup \{0\}$ , such an estimate can be obtained by enumerating over either 0 or quantities of the form  $\lambda_{\min} p_{\min} \cdot (1 + \epsilon)^k$  within this interval, for integer values of  $k$ . Clearly, for every  $\phi \leq \Psi(\mathcal{S}^*)$ , the  $\phi$ -parameterized inverse formulation is feasible, meaning that by applying Theorem 3 with an estimate  $\phi \in [(1 - \epsilon) \cdot \Psi(\mathcal{S}^*), \Psi(\mathcal{S}^*)]$ , we identify a solution  $\mathcal{S}$  satisfying  $\Psi(\mathcal{S}) \geq (1 - \epsilon) \cdot \phi \geq (1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)$ , which is a  $(1 - 2\epsilon)$ -approximate solution to the original incremental knapsack problem. As a side note, it is worth pointing out that the latter argument does not exploit the super-optimality of  $\mathcal{S}$ ; this property will play a crucial role in Section 3, once we consider arbitrarily structured instances.

## 2.2. Preliminaries: Prefix-like super-optimal solutions

Focusing on an inverse incremental knapsack instance parameterized by  $\phi$ , it would be convenient to fix the profit requirement  $\phi \geq 0$  from this point on. For simplicity of notation, we further drop the symbol  $\phi$  when referring to  $\mathcal{S}^{*\phi}$ , which results in having  $\mathcal{S}^*$  as an optimal solution to the inverse problem instance being considered.

*Profit classes.* We begin by scaling all profits such that  $\min_{i \in [n]} p_i = 1$ . In addition, we round down the profit of each item to the nearest integer power of  $1 + \epsilon$ . It is easy to verify that, as a result of this alteration, we are losing at most an  $\epsilon$ -fraction of the profit  $\Psi(\mathcal{S}^*)$ ; clearly, once the original item profits are restored, the profit of any solution can only increase. For simplicity, we keep denoting the modified profits by  $p_1, \dots, p_n$ . Following these preprocessing steps, for every  $\ell \in \mathbb{N}_0$ , we use  $\mathcal{P}_\ell = \{i \in [n] : p_i = (1 + \epsilon)^\ell\}$  to designate the collection of items associated with a profit of  $(1 + \epsilon)^\ell$ ; this set will be referred to as the profit class  $\mathcal{P}_\ell$ .

*Prefix-like solutions.* For any profit class  $\mathcal{P}_\ell$  and for any integer  $k \in [0, |\mathcal{P}_\ell|]$ , we use  $\mathcal{P}_\ell[k]$  to denote the set of  $k$  smallest weight items in  $\mathcal{P}_\ell$ , breaking ties arbitrarily. Also,  $\mathcal{P}_\ell[k_1, k_2]$  will be the set of  $k_1$ -th smallest,  $(k_1 + 1)$ -th smallest, up to  $k_2$ -th smallest weight items in  $\mathcal{P}_\ell$ , that is,  $\mathcal{P}_\ell[k_1, k_2] = \mathcal{P}_\ell[k_2] \setminus \mathcal{P}_\ell[k_1 - 1]$ . We say that a solution  $\mathcal{S} = \langle S_1, \dots, S_T \rangle$  has a prefix-like structure when, for every time period  $t \in [T]$  and class  $\mathcal{P}_\ell$ , the items picked by  $S_t$  from  $\mathcal{P}_\ell$  are those with smallest weight, i.e.,  $S_t \cap \mathcal{P}_\ell = \mathcal{P}_\ell[|S_t \cap \mathcal{P}_\ell|]$ .

Crucial to our algorithmic approach is the next lemma, which argues that the sequence of profit classes  $\mathcal{P}_0, \mathcal{P}_1, \dots$  contains an  $O(\frac{1}{\epsilon} \log \frac{n\rho}{\epsilon})$ -length interval such that the union of its classes contains a super-optimal prefix-like solution.

LEMMA 1. *There exist an  $O(\frac{1}{\epsilon} \log \frac{n\rho}{\epsilon})$ -length interval  $\mathcal{L} \subseteq \mathbb{N}_0$  and a prefix-like solution  $\tilde{\mathcal{S}} = \langle \tilde{S}_1, \dots, \tilde{S}_T \rangle$  such that:*

- (a) *The solution  $\tilde{\mathcal{S}}$  only introduces items from profit classes  $\{\mathcal{P}_\ell\}_{\ell \in \mathcal{L}}$ .*
- (b) *Super-optimality:  $w(\tilde{S}_T) \leq w(S_T^*)$ .*
- (c)  *$\epsilon$ -profit violation:  $\Psi(\tilde{\mathcal{S}}) \geq (1 - \epsilon) \cdot \Psi(\mathcal{S}^*)$ .*

*Proof.* To determine the interval  $\mathcal{L}$ , let  $\ell_{\max}$  be the maximal index of a profit class of which  $\mathcal{S}^*$  introduces at least one item, namely,  $\ell_{\max} = \max\{\ell \in \mathbb{N}_0 : S_T^* \cap \mathcal{P}_\ell \neq \emptyset\}$ . Our interval is defined as  $\mathcal{L} = [[\ell_{\max} - L + 1]^+, \ell_{\max}]$ , where  $L = \lceil \log_{1+\epsilon}(\frac{n\rho}{\epsilon}) \rceil$ , meaning that it is indeed of the required length. Having specified the interval  $\mathcal{L}$ , we construct a corresponding prefix-like solution  $\tilde{\mathcal{S}} = \langle \tilde{S}_1, \dots, \tilde{S}_T \rangle$  by setting  $\tilde{S}_t = \bigcup_{\ell \in \mathcal{L}} \mathcal{P}_\ell[|S_t^* \cap \mathcal{P}_\ell|]$  for every  $t \in [T]$ . In other words, for every  $\ell \in \mathcal{L}$ , the set  $\tilde{S}_t$  picks out of the profit class  $\mathcal{P}_\ell$  exactly the same number of items as  $S_t^*$  does; however, these items are picked by order of increasing weight.

One straightforward observation is that, since no items are introduced out of profit classes  $\{\mathcal{P}_\ell\}_{\ell \notin \mathcal{L}}$ , property (a) is clearly satisfied. In addition, this definition guarantees that  $\tilde{\mathcal{S}}$  is indeed a feasible solution by noting that, for every time period  $t \in [T]$ ,

$$w(\tilde{S}_t) = \sum_{\ell \in \mathcal{L}} w(\mathcal{P}_\ell[|S_t^* \cap \mathcal{P}_\ell|]) \leq \sum_{\ell \in \mathcal{L}} w(S_t^* \cap \mathcal{P}_\ell) \leq \sum_{\ell \in \mathbb{N}_0} w(S_t^* \cap \mathcal{P}_\ell) = w(S_t^*),$$

where the first inequality is derived by observing that  $S_t^* \cap \mathcal{P}_\ell$  is a subset of  $|S_t^* \cap \mathcal{P}_\ell|$  items in profit class  $\mathcal{P}_\ell$ , while  $\mathcal{P}_\ell[|S_t^* \cap \mathcal{P}_\ell|]$  is the lightest subset of  $\mathcal{P}_\ell$  consisting of that many items. Beyond feasibility, the inequality we have just proven also shows that property (b) is satisfied.

To prove property (c), we first argue that for every  $\ell \in \mathcal{L}$ ,

$$\begin{aligned} \sum_{i \in \tilde{S}_T \cap \mathcal{P}_\ell} \Psi_i(\tilde{\mathcal{S}}) &= \sum_{t \in [T]} \sum_{i \in (\tilde{S}_t \setminus \tilde{S}_{t-1}) \cap \mathcal{P}_\ell} \Psi_i(\tilde{\mathcal{S}}) \\ &= (1 + \epsilon)^\ell \cdot \sum_{t \in [T]} \lambda^{t+} \cdot \left| (\tilde{S}_t \setminus \tilde{S}_{t-1}) \cap \mathcal{P}_\ell \right| \\ &= (1 + \epsilon)^\ell \cdot \sum_{t \in [T]} \lambda^{t+} \cdot \left| (S_t^* \setminus S_{t-1}^*) \cap \mathcal{P}_\ell \right| \\ &= \sum_{t \in [T]} \sum_{i \in (S_t^* \setminus S_{t-1}^*) \cap \mathcal{P}_\ell} \Psi_i(\mathcal{S}^*) \\ &= \sum_{i \in S_T^* \cap \mathcal{P}_\ell} \Psi_i(\mathcal{S}^*). \end{aligned}$$

Here, the second equality holds since  $\Psi_i(\tilde{\mathcal{S}}) = p_i \cdot \lambda^{\nu(i, \tilde{\mathcal{S}})+} = (1 + \epsilon)^\ell \cdot \lambda^{t+}$  for all items  $i \in (\tilde{S}_t \setminus \tilde{S}_{t-1}) \cap \mathcal{P}_\ell$ , whereas the fourth equality follows from an analogous argument with respect to  $\mathcal{S}^*$ . As a result,

based on representation (1) of the objective function, to show that  $\Psi(\tilde{\mathcal{S}}) \geq (1 - \epsilon) \cdot \Psi(\mathcal{S}^*)$ , it suffices to argue that  $\sum_{i \in \mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T} \Psi_i(\mathcal{S}^*) \leq \epsilon \cdot \Psi(\mathcal{S}^*)$ . For this purpose, note that

$$\begin{aligned} \sum_{i \in \mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T} \Psi_i(\mathcal{S}^*) &= \sum_{i \in \mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T} p_i \cdot \lambda^{\nu(i, \mathcal{S}^*)+} \\ &\leq \left| \mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T \right| \cdot \lambda^{1+} \cdot (1 + \epsilon)^{\ell_{\max} - L} \\ &\leq n \cdot \lambda^{1+} \cdot (1 + \epsilon)^{\ell_{\max}} \cdot \frac{\epsilon}{n\rho} \\ &= \epsilon \cdot \lambda^{T+} \cdot (1 + \epsilon)^{\ell_{\max}} \\ &\leq \epsilon \cdot \Psi(\mathcal{S}^*) . \end{aligned}$$

Here, the first inequality holds since  $\lambda^{1+} \geq \dots \geq \lambda^{T+}$  and since  $\mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T$  is necessarily a subset of  $\bigcup_{\ell \in [0, [\ell_{\max} - L] +]} \mathcal{P}_\ell$ , by definition of  $\ell_{\max}$ , meaning that every item  $i \in \mathcal{S}_T^* \setminus \tilde{\mathcal{S}}_T$  has a profit of at most  $(1 + \epsilon)^{\ell_{\max} - L}$ . The second inequality is obtained by observing that  $(1 + \epsilon)^{-L} \leq \frac{\epsilon}{n\rho}$  in view of our choice for  $L = \lceil \log_{1+\epsilon}(\frac{n\rho}{\epsilon}) \rceil$ . The last inequality holds since, by definition of  $\ell_{\max}$ , at least one item  $i \in \mathcal{P}_{\ell_{\max}}$  is introduced by  $\mathcal{S}^*$ , implying that its individual profit contribution  $\Psi_i(\mathcal{S}^*)$  is at least  $\lambda^{T+} \cdot (1 + \epsilon)^{\ell_{\max}}$ .  $\square$

### 2.3. Dynamic programming formulation

Taking advantage of Lemma 1, we restrict our attention to prefix-like solutions with respect to the profit classes  $\{\mathcal{P}_\ell\}_{\ell \in \mathcal{L}}$ . For convenience, rather than representing such solutions as a sequence of items sets  $\mathcal{S} = \langle \mathcal{S}_1, \dots, \mathcal{S}_T \rangle$ , we use instead an equivalent sequence of utilization vectors  $\mathcal{N} = \langle N_1, \dots, N_T \rangle$ . Here, each  $N_t$  is an integer-valued vector of dimension  $|\mathcal{L}|$ , in which each coordinate  $N_{t,\ell}$  specifies the number of items picked thus far from the profit class  $\mathcal{P}_\ell$ , i.e.,  $N_{t,\ell} = |\mathcal{S}_t \cap \mathcal{P}_\ell|$ .

*States.* Each state of our dynamic program is described by two parameters:

- The index of the current period,  $t$ . This parameter takes one of the values  $0, 1, \dots, T$ .
- The collection of items utilized thus far, specified by the utilization vector  $N_t$ . This parameter belongs to the family of vectors  $\mathcal{U}_{\text{all}} = [|\mathcal{P}_{\min \mathcal{L}}|]_0 \times \dots \times [|\mathcal{P}_{\max \mathcal{L}}|]_0$ .

It is important to emphasize that the family of vectors we are currently working with is of cardinality  $|\mathcal{U}_{\text{all}}| = \Omega(n^{\Omega(|\mathcal{L}|)}) = \Omega(n^{\Omega(\frac{1}{\epsilon} \log \frac{n\rho}{\epsilon})})$ , meaning that the state space described so far is exponentially sized. Our goal in subsequent sections will be to develop a more compact, and yet still super-optimal state space representation.

*Value function.* The value function  $F(t, N_t)$  stands for the maximum-possible accumulated profit contribution  $\Psi(N_{[1,t]}) = \sum_{\tau \in [t]} (\lambda^{\tau+} \cdot \sum_{\ell \in \mathcal{L}} (1 + \epsilon)^\ell \cdot (N_{\tau,\ell} - N_{\tau-1,\ell}))$  of a  $t$ -period solution  $N_{[1,t]} = \langle N_1, \dots, N_t \rangle$  that satisfies the following conditions:

1. *Utilization vector  $N_t$  at period  $t$ :* The set of items introduced across time periods  $1, \dots, t$  is described by the utilization vector  $N_t$ .

2. *Feasibility*: All capacities are satisfied, i.e.,  $w(N_\tau) \leq W_\tau$  for every  $\tau \in [t]$ , with the convention that  $w(N_\tau) = \sum_{\ell \in \mathcal{L}} w(\mathcal{P}_\ell[N_\tau, \ell])$ .

When there is no solution satisfying conditions 1 and 2, we define  $F(t, N_t) = -\infty$ . Consequently, we are interested in finding a utilization vector  $N_T \in \mathcal{U}_{\text{all}}$  that minimizes  $w(N_T)$  out of those that meet the criterion  $F(T, N_T) \geq (1 - \epsilon) \cdot \phi$ . At this stage, it is worth pointing out that, by considering the prefix-like solution  $\tilde{\mathcal{S}}$  constructed in Lemma 1, it follows that the utilization vector  $\tilde{N}_T$  associated with  $\tilde{S}_T$  satisfies  $F(T, \tilde{N}_T) \geq (1 - \epsilon) \cdot \Psi(\mathcal{S}^*) \geq (1 - \epsilon) \cdot \phi$  and  $w(\tilde{N}_T) \leq w(S_T^*)$ . Therefore, the previously mentioned criterion is met by some utilization vector, and concurrently, since we return the one whose total weight is minimized, the items introduced have a total weight of at most  $w(S_T^*)$ . Hence, by following the actions taken by our dynamic program, we obtain an inverse incremental knapsack solution that satisfies the properties required in Theorem 3.

*Recursive equations.* Next, we argue that the value function  $F$  satisfies a collection of recursive equations, leading to a dynamic programming formulation of the inverse incremental knapsack problem. The important observation is that the feasibility of any  $t$ -period solution  $N_{[1,t]} = \langle N_1, \dots, N_t \rangle$  is equivalent to jointly ensuring that:

- The induced  $(t - 1)$ -period solution  $N_{[1,t-1]}$  is feasible by itself.
- Our extension to period  $t$  preserves feasibility, meaning that  $N_{t-1} \leq N_t$  and  $w(N_t) \leq W_t$ .

Consequently, for every  $t \in [T]$  and  $N_t \in \mathcal{U}_{\text{all}}$ , the value function  $F$  can be recursively expressed as

$$F(t, N_t) = \max_{N_{t-1} \in \mathcal{U}_{\text{all}}} \left\{ F(t-1, N_{t-1}) + \lambda^{t+} \cdot \sum_{\ell \in \mathcal{L}} (1 + \epsilon)^\ell \cdot (N_{t,\ell} - N_{t-1,\ell}) \right\}$$

s.t.  $N_{t-1} \leq N_t$ ,

$w(N_t) \leq W_t$ .

To better understand the recursion above, note that the expression  $\lambda^{t+} \cdot \sum_{\ell \in \mathcal{L}} (1 + \epsilon)^\ell \cdot (N_{t,\ell} - N_{t-1,\ell})$  is precisely the marginal profit contribution of the items introduced in period  $t$ , bearing in mind that we have already accumulated profit contributions amounting to  $F(t-1, N_{t-1})$  up to and including period  $t-1$ . Specifically, on top of the latter quantity, in period  $t$  we are introducing  $N_{t,\ell} - N_{t-1,\ell}$  items from each profit class  $\mathcal{P}_\ell$ , whose profits are all uniformly equal to  $(1 + \epsilon)^\ell$ . When there is no extendible  $(t-1)$ -period solution, i.e., one satisfying  $N_{t-1} \leq N_t$  and  $w(N_t) \leq W_t$ , we have  $F(t, N_t) = -\infty$ . Finally, the terminal states of this dynamic program are reached at  $t = 0$ , in which case  $F(0, 0) = 0$  and  $F(0, N_0) = -\infty$  for all other states.

## 2.4. Approximate dynamic program: High-level overview

*Restriction to  $\mathcal{U}_{\text{small}}$ .* In what follows, we explain how our original dynamic programming formulation can be altered, when considering an appropriately restricted subset of the utilization vectors

$\mathcal{U}_{\text{all}}$ . To this end, let  $\mathcal{U}_{\text{small}} \subseteq \mathcal{U}_{\text{all}}$  be any family of utilization vectors with  $0 \in \mathcal{U}_{\text{small}}$ . Given such a set, one can define an approximate dynamic program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$  that would mimic the exact program  $F$ , when restricted to the family  $\mathcal{U}_{\text{small}}$ . That is, while each state is still of the form  $(t, N_t)$ , we now require that  $N_t \in \mathcal{U}_{\text{small}}$  instead of  $N_t \in \mathcal{U}_{\text{all}}$ . In addition, rather than searching over  $N_{t-1} \in \mathcal{U}_{\text{all}}$ , we substitute this constraint with  $N_{t-1} \in \mathcal{U}_{\text{small}}$ . As a result, the recursive equations we obtain for  $\tilde{F}_{\mathcal{U}_{\text{small}}}(t, N_t)$  are given by:

$$\begin{aligned} \tilde{F}_{\mathcal{U}_{\text{small}}}(t, N_t) = \max_{N_{t-1} \in \mathcal{U}_{\text{small}}} & \left\{ \tilde{F}_{\mathcal{U}_{\text{small}}}(t-1, N_{t-1}) + \lambda^{t+} \cdot \sum_{\ell \in \mathcal{L}} (1+\epsilon)^\ell \cdot (N_{t,\ell} - N_{t-1,\ell}) \right\} \\ \text{s.t.} & \quad N_{t-1} \leq N_t, \\ & \quad w(N_t) \leq W_t. \end{aligned} \quad (2)$$

It is easy to verify that every utilization vector  $N_T \in \mathcal{U}_{\text{small}}$  with  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T) \neq -\infty$  can be converted in a straightforward way into a solution  $\mathcal{S}$  with a profit of  $\Psi(\mathcal{S}) = \tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T)$  and a total weight of  $w(\mathcal{S}_T) = w(N_T)$ .

*A small-sized family  $\mathcal{U}_{\text{small}}$  with near-optimal profit.* Let OPT be the maximum value of  $F(T, N_T)$  over all utilization vectors  $N_T \in \mathcal{U}_{\text{all}}$  with  $w(N_T) \leq w(S_T^*)$ ; we remind the reader that  $S^*$  is a fixed optimal solution to the inverse incremental knapsack instance being considered. As noted in Section 2.3, the utilization vector  $\tilde{N}_T$ , corresponding to the prefix-like solution  $\tilde{\mathcal{S}}$  given by Lemma 1, satisfies  $F(T, \tilde{N}_T) \geq (1-\epsilon) \cdot \phi$  and  $w(\tilde{N}_T) \leq w(S_T^*)$ , immediately showing that  $\text{OPT} \geq (1-\epsilon) \cdot \phi$ . Similarly, let us define  $\text{OPT}_{\mathcal{U}_{\text{small}}}$  as the maximum value of  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T)$  over all utilization vectors  $N_T \in \mathcal{U}_{\text{small}}$  with  $w(N_T) \leq w(S_T^*)$ . In the remainder of this section, we present an explicit construction of a family  $\mathcal{U}_{\text{small}}$  formed by  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot \log(\frac{w_{\text{max}}}{w_{\text{min}}}))$  utilization vectors, for which  $\text{OPT}_{\mathcal{U}_{\text{small}}} \geq (1-2\epsilon) \cdot \text{OPT} \geq (1-3\epsilon) \cdot \phi$ . Consequently, our algorithm returns the solution  $\mathcal{S}$  associated with the utilization vector  $N_T \in \mathcal{U}_{\text{small}}$  that minimizes  $w(N_T)$  over those meeting the criterion  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T) \geq (1-3\epsilon) \cdot \phi$ . By the preceding discussion, this solution is guaranteed to simultaneously satisfy  $w(\mathcal{S}_T) \leq w(S_T^*)$  and  $\Psi(\mathcal{S}) \geq (1-3\epsilon) \cdot \phi$ , thereby completing the proof of Theorem 3.

At a high level, given an arbitrary utilization vector  $N \in \mathcal{U}_{\text{all}}$ , we describe two alterations for transforming the latter into a highly structured vector,  $N^{(2)}$ . Informally speaking, this vector preserves the main features of  $N$ , in the sense of generating a nearly-matching total profit while still being feasible with respect to the time period capacities. However, the additional structural properties entailed by our alterations would ensure that the family of vectors  $\mathcal{U}_{\text{small}} = \{N^{(2)} : N \in \mathcal{U}_{\text{all}}\}$  is sufficiently small. Specifically, in Section 2.5, we begin by defining the up-rounding operation, where  $N$  is mapped to a vector  $N^{(1)}$  by appropriately rounding up some of its coordinates; unfortunately, up-rounding by itself may lead to capacity violations at various time periods. Next, in Section 2.6, we define the truncation operation, where  $N^{(1)}$  is mapped to the final utilization vector

$N^{(2)}$ . Intuitively, this operation discards certain items within each profit class in order to offset the excess capacity within the up-rounding  $N^{(1)}$ , and thus, to restore feasibility. Finally, in Section 2.7, we bound the optimality loss entailed by the approximate formulation  $\tilde{F}_{\mathcal{U}_{\text{small}}}$  and explain how to efficiently construct the family of utilization vectors  $\mathcal{U}_{\text{small}}$ .

## 2.5. The up-rounding operation: Definitions and properties

In what follows, we define an explicit mapping from each utilization vector  $N \in \mathcal{U}_{\text{all}}$  to its up-rounding  $N^{(1)}$ , which is a new utilization vector with very specific structure.

*Light and heavy classes.* To avoid cumbersome notation, we assume without loss of generality that  $1/\epsilon$  is an integer and that  $\epsilon < 1/4$ . We begin by defining a vector-specific dichotomy of the profit classes  $\{\mathcal{P}_\ell\}_{\ell \in \mathcal{L}}$ , depending on the number of items picked by the utilization vector  $N$  out of each one. For this purpose, the profit class  $\mathcal{P}_\ell$  is said to be light for  $N$  when  $N_\ell \leq 1/\epsilon$ . The collection of all light classes is denoted by  $\text{light}(N)$ . Every other class, with strictly more than  $1/\epsilon$  items, is called heavy for  $N$ , with the analogous notation  $\text{heavy}(N)$ .

*Constructing  $N^{(1)}$  via up-rounding.* Each coordinate of the vector  $N^{(1)}$  will be defined depending on whether its corresponding profit class is light or heavy. Specifically, out of each light class,  $N^{(1)}$  picks exactly the same number of items as  $N$  does, meaning that  $N_\ell^{(1)} = N_\ell$  for every  $\ell \in \text{light}(N)$ . In contrast, for each heavy class  $\ell \in \text{heavy}(N)$ , the way we define  $N_\ell^{(1)}$  is more elaborate:

- We begin by introducing a shorthand notation  $W_{\text{heavy}}(N)$  for the total weight of all items picked by  $N$  out of heavy classes, beyond the  $1/\epsilon$  lightest ones in each class, i.e.,

$$W_{\text{heavy}}(N) = \sum_{\ell \in \text{heavy}(N)} w\left(\mathcal{P}_\ell \left[\frac{1}{\epsilon} + 1, N_\ell\right]\right).$$

- For every  $\ell \in \text{heavy}(N)$ , let  $\mathcal{E}_\ell$  be an over-estimate for  $\mathcal{P}_\ell$ 's contribution  $w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell])$  toward the sum  $W_{\text{heavy}}(N)$ , up to an additive factor of  $\text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$ . Here, the operator  $\text{Power}_2[\cdot]$  rounds its non-negative argument up to the nearest integer power of 2, with the convention that  $\text{Power}_2[0] = 0$ . Specifically, our estimate is of the form  $\mathcal{E}_\ell = \mu_\ell \cdot \text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$ , where  $\mu_\ell \geq 1$  is the unique integer for which

$$(\mu_\ell - 1) \cdot \text{Power}_2\left[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)\right] < w\left(\mathcal{P}_\ell \left[\frac{1}{\epsilon} + 1, N_\ell\right]\right) \leq \underbrace{\mu_\ell \cdot \text{Power}_2\left[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)\right]}_{\mathcal{E}_\ell}. \quad (3)$$

- Finally, out of the profit class  $\mathcal{P}_\ell$  in order of increasing weights, the vector  $N^{(1)}$  picks the maximum number of items beyond the first  $1/\epsilon$  ones such that their total weight is still within the estimate  $\mathcal{E}_\ell$ . That is,

$$N_\ell^{(1)} = \max\left\{k \in \left[\frac{1}{\epsilon} + 1, |\mathcal{P}_\ell|\right] : w\left(\mathcal{P}_\ell \left[\frac{1}{\epsilon} + 1, k\right]\right) \leq \mathcal{E}_\ell\right\}. \quad (4)$$

As a side note, while one has to delve deep into our analysis to fully digest why the exact form we have chosen for  $\mathcal{E}_\ell$  is required, its main purpose is to eventually end up with a small-sized family  $\mathcal{U}_{\text{small}} \subseteq \mathcal{U}_{\text{all}}$  of utilization vectors, still allowing the approximate dynamic program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$  to collect a near-optimal profit. However, an equally important purpose, which is absolutely necessary for our construction, is to ensure certain monotonicity properties (see Lemma 2), achievable by rounding  $\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)$  up to the nearest integer power of 2; we do not know how to establish such properties by, for instance, rounding up to a power of  $1 + \epsilon$ , although this operation seems more intuitive.

*Is “up-rounding” really rounding up?* Based on the above definitions, it is not entirely clear whether we are indeed rounding up each coordinate of  $N$  to obtain  $N^{(1)}$ , or whether some coordinates can actually decrease. To show that  $N \leq N^{(1)}$ , we first observe that for  $\ell \in \text{light}(N)$ , one clearly has  $N_\ell^{(1)} = N_\ell$  by definition. For  $\ell \in \text{heavy}(N)$ , due to our method for choosing  $N_\ell^{(1)}$  in (4), we necessarily have  $N_\ell^{(1)} \geq N_\ell$ , since  $w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell]) \leq \mathcal{E}_\ell$  by definition of  $\mathcal{E}_\ell$ . These observations also imply that  $\text{light}(N^{(1)}) = \text{light}(N)$  and  $\text{heavy}(N^{(1)}) = \text{heavy}(N)$ , meaning that the up-rounding operation preserves the light and heavy classes.

*Key properties.* The next lemma summarizes the main properties of the up-rounding operation that will be utilized in our subsequent analysis.

LEMMA 2. *The up-rounding operation satisfies the following properties:*

- (a) *Monotonicity:*  $N^{(1)} \leq M^{(1)}$  for every pair of vectors  $N \leq M$ .
- (b) *Weight beyond first  $1/\epsilon$  items:*  $W_{\text{heavy}}(N^{(1)}) \leq (1 + 2\epsilon) \cdot W_{\text{heavy}}(N)$ .

For ease of presentation, since the proof is rather technical, its specifics are deferred to Appendix A.1. At a high level, the above-mentioned properties stem from appropriately choosing the multiplier  $\text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$ . By selecting a sufficiently small multiplier, that explicitly depends on the total weight quantity  $W_{\text{heavy}}(N)$  and the number of classes  $|\mathcal{L}|$ , we ensure that the excessive items due to our up-rounding operation only account for an  $O(\epsilon)$  fraction of the total weight, as stated in property (b). In contrast, arguing for the monotonicity of this operation, corresponding to property (a), is more challenging. Here, choosing the multiplier  $\text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$  as an integer power of 2 will play an important role, as further explained in Appendix A.1.

## 2.6. The truncation operation: Definitions and properties

In what follows, given a utilization vector  $N$  and its up-rounding  $N^{(1)}$ , we explain how to define the truncation  $N^{(2)}$  of the latter vector. Before presenting this construction, we remind the reader that the up-rounding operation was shown in Section 2.5 to be class-preserving, meaning that  $\text{light}(N^{(1)}) = \text{light}(N)$  and  $\text{heavy}(N^{(1)}) = \text{heavy}(N)$ .

*Constructing  $N^{(2)}$  via truncation.* Similarly to the up-rounding operation, each coordinate of the vector  $N^{(2)}$  will be defined depending on whether its corresponding profit class is light or heavy. Specifically, out of each light class,  $N^{(2)}$  picks exactly the same number of items as  $N^{(1)}$  and  $N$  do, that is,  $N_\ell^{(2)} = N_\ell^{(1)} = N_\ell$  for every  $\ell \in \text{light}(N)$ . However, for each heavy class  $\ell \in \text{heavy}(N)$ , we will define  $N_\ell^{(2)}$  by eliminating a certain number of items from  $N_\ell^{(1)}$ . To formalize this notion, let  $\Delta_\ell = N_\ell^{(1)} - \frac{1}{\epsilon}$  be the number of items picked by  $N^{(1)}$  from the profit class  $\mathcal{P}_\ell$  beyond the first  $1/\epsilon$  items. Then,  $N_\ell^{(2)}$  is obtained from  $N_\ell^{(1)}$  by eliminating its last  $\lceil 2\epsilon \cdot \Delta_\ell \rceil$  items, that is,  $N_\ell^{(2)} = N_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell \rceil$ .

*Is truncation class-preserving?* Before turning our attention to specific structural properties of the newly obtained vector  $N^{(2)}$ , it is worth pointing out that, similarly to up-rounding, the truncation operation is class-preserving. To verify this claim, we first observe that for  $\ell \in \text{light}(N^{(1)})$ , one clearly has  $N_\ell^{(2)} = N_\ell^{(1)}$  by definition, implying that  $\ell \in \text{light}(N^{(2)})$ . In addition, to argue that each class  $\ell \in \text{heavy}(N^{(1)})$  remains heavy for  $N^{(2)}$ , we should prove that none of the first  $1/\epsilon$  items in  $N_\ell^{(1)}$  will be eliminated. Put differently, the number of eliminated items,  $\lceil 2\epsilon \cdot \Delta_\ell \rceil$ , does not exceed  $N_\ell^{(1)} - \frac{1}{\epsilon} = \Delta_\ell$ . For this purpose, we know that  $\Delta_\ell \geq 1$ , as class  $\ell$  is heavy. Now, when  $\Delta_\ell = 1$ , we have  $\lceil 2\epsilon \cdot \Delta_\ell \rceil = \lceil 2\epsilon \rceil = 1$  since  $\epsilon < 1/4$ , and the claim holds. When  $\Delta_\ell \geq 2$ , we have  $\lceil 2\epsilon \cdot \Delta_\ell \rceil \leq \frac{\Delta_\ell}{2} + 1 \leq \Delta_\ell$  for the same reason. To summarize, we have  $\text{light}(N^{(2)}) = \text{light}(N^{(1)}) = \text{light}(N)$  and  $\text{heavy}(N^{(2)}) = \text{heavy}(N^{(1)}) = \text{heavy}(N)$ .

*Key properties.* The next lemma, whose proof is provided in Appendix A.2, summarizes the main properties of our construction.

LEMMA 3. *The truncation operation satisfies the following properties:*

- (a) *Monotonicity:*  $N^{(2)} \leq M^{(2)}$  for every pair of vectors  $N \leq M$ .
- (b) *Weight restoration:*  $w(N^{(2)}) \leq w(N)$ .
- (c) *Items picked out of heavy classes:*  $N_\ell^{(2)} \geq (1 - 2\epsilon) \cdot N_\ell$  for every  $\ell \in \text{heavy}(N)$ .

## 2.7. Analysis: State space size and performance guarantee

*Main analytical questions.* Following the discussion in Sections 2.4-2.6, the family of utilization vectors  $\mathcal{U}_{\text{small}} = \{N^{(2)} : N \in \mathcal{U}_{\text{all}}\}$  is now well-defined. Nevertheless, a number of basic questions still have to be addressed:

1. What is the cardinality of  $\mathcal{U}_{\text{small}}$  and how do we efficiently construct this set?
2. What is the overall running time of the dynamic program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$ ?
3. What is the optimality loss of  $\text{OPT}_{\mathcal{U}_{\text{small}}}$  relative to  $\text{OPT}$ ?

The remainder of this section provides a detailed treatment for each of these questions.

*Answer 1: Bounding and constructing  $\mathcal{U}_{\text{small}}$ .* Even though we have shown in Section 2.3 that the original collection  $\mathcal{U}_{\text{all}}$  of utilization vectors may generally be of size  $\Omega(n^{\Omega(|\mathcal{L}|)}) = \Omega(n^{\Omega(\frac{1}{\epsilon} \log \frac{n\rho}{\epsilon})})$ , it turns out that  $\mathcal{U}_{\text{small}}$  is considerably smaller, as stated in the next claim.

LEMMA 4.  $|\mathcal{U}_{\text{small}}| = O((n\rho)^{\bar{O}(1/\epsilon^2)} \cdot \log(\frac{w_{\text{max}}}{w_{\text{min}}}))$ .

*Proof.* To establish the desired claim, the crucial observation is that, for any  $N \in \mathcal{U}_{\text{all}}$ , its truncated up-rounded image  $N^{(2)} \in \mathcal{U}_{\text{small}}$  is uniquely determined, given the following pieces of information:

1. *The identity of  $N^{(2)}$ 's light and heavy classes.* There are only  $2^{|\mathcal{L}|}$  possible ways of assigning a light/heavy label to each of the classes  $\{\mathcal{P}_\ell\}_{\ell \in \mathcal{L}}$ .
2. *The number of items picked by  $N^{(2)}$  out of each light class.* In regard to these features, since  $N_\ell^{(2)} \leq \frac{1}{\epsilon}$  for all  $\ell \in \text{light}(N^{(2)})$ , there are only  $O((\frac{1}{\epsilon})^{|\text{light}(N^{(2)})|}) = O((\frac{1}{\epsilon})^{O(|\mathcal{L}|)})$  possible values that can jointly be taken by  $(N_\ell^{(2)})_{\ell \in \text{light}(N^{(2)})}$ .
3. *The number of items picked by  $N^{(2)}$  out of each heavy class.* Deriving an upper bound on the number of possible values jointly taken by  $(N_\ell^{(2)})_{\ell \in \text{heavy}(N^{(2)})}$  is somewhat more involved. For readability purposes, we separately show in Lemma 5 that this quantity is only  $O(2^{O(|\mathcal{L}|/\epsilon)} \cdot \log(\frac{nw_{\text{max}}}{w_{\text{min}}}))$ .

By combining the above remarks, we have just shown that

$$|\mathcal{U}_{\text{small}}| = O\left(2^{|\mathcal{L}|} \cdot \left(\frac{1}{\epsilon}\right)^{O(|\mathcal{L}|)} \cdot 2^{O(|\mathcal{L}|/\epsilon)} \cdot \log\left(\frac{nw_{\text{max}}}{w_{\text{min}}}\right)\right) = O\left((n\rho)^{\bar{O}(1/\epsilon^2)} \cdot \log\left(\frac{w_{\text{max}}}{w_{\text{min}}}\right)\right),$$

where the second equality follows from plugging in  $|\mathcal{L}| = O(\frac{1}{\epsilon} \log \frac{n\rho}{\epsilon})$ .  $\square$

LEMMA 5. *The number of values jointly taken by  $(N_\ell^{(2)})_{\ell \in \text{heavy}(N^{(2)})}$  is  $O(2^{O(|\mathcal{L}|/\epsilon)} \cdot \log(\frac{nw_{\text{max}}}{w_{\text{min}}}))$ .*

*Proof.* A close inspection of how truncations are defined in Section 2.6 reveals that  $N_\ell^{(2)}$  is uniquely determined by  $N_\ell^{(1)}$ . Therefore, since  $\text{heavy}(N^{(1)}) = \text{heavy}(N^{(2)})$ , we can instead focus on the possible values for  $(N_\ell^{(1)})_{\ell \in \text{heavy}(N^{(1)})}$ . In turn, the way up-roundings are defined in Section 2.5 guarantees these values to be uniquely determined by  $\text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$  and by  $(\mu_\ell)_{\ell \in \text{heavy}(N)}$ .

With respect to the first parameter, the quantity  $W_{\text{heavy}}(N) = \sum_{\ell \in \text{heavy}(N)} w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell])$  clearly resides within the interval  $[w_{\text{min}}, nw_{\text{max}}]$ , where  $w_{\text{min}} = \min_{i \in [n]} w_i$  and  $w_{\text{max}} = \max_{i \in [n]} w_i$  stand for the minimum and maximum weights of any item, respectively. Consequently,  $\text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N)]$  is an integer power of 2 contained in the interval  $[\frac{\epsilon}{|\mathcal{L}|} \cdot w_{\text{min}}, \frac{2\epsilon}{|\mathcal{L}|} \cdot nw_{\text{max}}]$ , and there are only  $O(\log(\frac{nw_{\text{max}}}{w_{\text{min}}}))$  such powers. Now, as far as the values  $(\mu_\ell)_{\ell \in \text{heavy}(N)}$  are concerned, we exploit the next claim, which places an upper bound on their summation. To avoid deviating from the broader argument, we provide the proof in Appendix A.3.

CLAIM 1.  $\sum_{\ell \in \text{heavy}(N)} \mu_\ell \leq \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon}$ .

Now, basic counting ideas show that the number of different solutions in non-negative integers to the inequality  $\sum_{\ell \in \text{heavy}(N)} \mu_\ell \leq \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon}$  is precisely

$$\sum_{k=0}^{\lfloor \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon} \rfloor} \binom{|\text{heavy}(N)| + k - 1}{k} \leq \sum_{k=0}^{\lfloor \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon} \rfloor} \binom{|\mathcal{L}| + \lfloor \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon} \rfloor}{k} \leq 2^{|\mathcal{L}| + \lfloor \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon} \rfloor} = 2^{O(|\mathcal{L}|/\epsilon)}. \quad \square$$

In regard to efficiently generating  $\mathcal{U}_{\text{small}}$ , the important observation is that the proof of Lemma 4 actually provides an explicit construction. Specifically, following the arguments employed to bound the cardinality of this set, one can simply enumerate over all possible combinations of the next three features:

1. The identity of light and heavy classes.
2. The number of items picked out of each light class.
3. The number of items picked out of each heavy class.

As previously shown, there are only  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot \log(\frac{w_{\max}}{w_{\min}}))$  such combinations, which can naively be enumerated in  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$  time.

*Answer 2: Running time.* We now turn our attention to evaluate the running time incurred by the approximate dynamic program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$ . For this purpose, we have just observed that the reduced family of utilization vectors  $\mathcal{U}_{\text{small}}$  can be constructed in  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$  time. Now, for the purpose of solving the recursive equations (2) through which  $\tilde{F}_{\mathcal{U}_{\text{small}}}$  is defined, at each state  $(t, N_t)$  of this dynamic program, the optimal action to be taken can straightforwardly be identified in  $O(|\mathcal{L}| \cdot |\mathcal{U}_{\text{small}}|)$  time. All in all, recalling that  $|\mathcal{U}_{\text{small}}| = O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot \log(\frac{w_{\max}}{w_{\min}}))$  by Lemma 4, the running time stated in Theorem 3 is obtained by noting that the overall number of elementary operations performed is  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$ .

*Answer 3: Performance guarantee.* We proceed by analyzing the approximate dynamic program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$  in terms of its profit-wise performance. In the next claim, we bound the optimality loss incurred by replacing our original state space with the one induced by utilization vectors within  $\mathcal{U}_{\text{small}}$ , thereby completing the proof of Theorem 3.

LEMMA 6.  $\text{OPT}_{\mathcal{U}_{\text{small}}} \geq (1 - 2\epsilon) \cdot \text{OPT}$ .

*Proof.* Recall that  $\text{OPT}$  is given by the maximal value of  $F(T, N_T)$  attained by a utilization vector  $N_T \in \mathcal{U}_{\text{all}}$  for which  $w(N_T) \leq w(S_T^*)$ . We denote the sequence of states traversed by the original dynamic program  $F$  to attain this value by  $(0, N_0^*), (1, N_1^*), \dots, (T, N_T^*)$ , noting that  $N_0^* = 0$ .

The crucial claim we exploit is that, when each of the above-mentioned utilization vectors is replaced by its truncated up-rounding, the sequence of states  $(0, 0), (1, N_1^{*(2)}), \dots, (T, N_T^{*(2)})$  we obtain is necessarily feasible along the approximate program  $\tilde{F}_{\mathcal{U}_{\text{small}}}$ . To understand why this claim holds, we have by definition  $N_t^{*(2)} \in \mathcal{U}_{\text{small}}$  for every  $t \in [T]_0$ , and therefore, it remains to explain

why  $N_{t-1}^{*(2)}$  is a feasible solution to the structural constraints of the recursive equations (2), through which  $\tilde{F}_{\mathcal{U}_{\text{small}}}(t, N_t^{*(2)})$  is defined. Specifically, these constraints stipulate that  $N_{t-1}^{*(2)} \leq N_t^{*(2)}$  and  $w(N_t^{*(2)}) \leq W_t$ . For this purpose, note that  $w(N_t^{*(2)}) \leq w(N_t^*) \leq W_t$ , where the first inequality follows from Lemma 3(b) and the second inequality is a constraint of the dynamic program  $F$ . Along the same lines, since  $N_{t-1}^* \leq N_t^*$  is a concurrent constraint of  $F$ , we have  $N_{t-1}^{*(2)} \leq N_t^{*(2)}$  as well, by Lemma 3(a).

Now, recalling that  $\text{OPT}_{\mathcal{U}_{\text{small}}}$  is the maximal value of  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T)$  attained by a utilization vector  $N_T \in \mathcal{U}_{\text{small}}$  for which  $w(N_T) \leq w(S_T^*)$ , we have just shown that  $\text{OPT}_{\mathcal{U}_{\text{small}}} \geq \tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T^{*(2)})$ , since  $w(N_T^{*(2)}) \leq w(N_T^*) \leq w(S_T^*)$ , where the first inequality follows from Lemma 3(b). Thus, to conclude the proof, it suffices to argue that  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T^{*(2)}) \geq (1 - 2\epsilon) \cdot \text{OPT}$ . For this purpose, note that

$$\begin{aligned} \tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T^{*(2)}) &= \sum_{t \in [T]} \left( \tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_t^{*(2)}) - \tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_{t-1}^{*(2)}) \right) \\ &\geq \sum_{t \in [T]} \left( \lambda^{t+} \cdot \sum_{\ell \in \mathcal{L}} (1 + \epsilon)^\ell \cdot (N_{t,\ell}^{*(2)} - N_{t-1,\ell}^{*(2)}) \right) \\ &= \sum_{\ell \in \mathcal{L}} \left( (1 + \epsilon)^\ell \cdot \sum_{t \in [T]} \lambda^{t+} (N_{t,\ell}^{*(2)} - N_{t-1,\ell}^{*(2)}) \right) \\ &= \sum_{\ell \in \mathcal{L}} \left( (1 + \epsilon)^\ell \cdot \sum_{t \in [T]} \lambda_t \cdot N_{t,\ell}^{*(2)} \right), \end{aligned}$$

where the inequality above holds since  $N_{t-1}^{*(2)}$  is one of the feasible actions within the dynamic programming equation (2) through which  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_t^{*(2)})$  is defined. On the other hand, using nearly identical transitions, we have

$$\begin{aligned} \text{OPT} &= F(T, N_T^*) \\ &= \sum_{t \in [T]} (F(t, N_t^*) - F(t-1, N_{t-1}^*)) \\ &= \sum_{t \in [T]} \left( \lambda^{t+} \cdot \sum_{\ell \in \mathcal{L}} (1 + \epsilon)^\ell \cdot (N_{t,\ell}^* - N_{t-1,\ell}^*) \right) \\ &= \sum_{\ell \in \mathcal{L}} \left( (1 + \epsilon)^\ell \cdot \sum_{t \in [T]} \lambda_t \cdot N_{t,\ell}^* \right). \end{aligned}$$

Consequently, we show that  $\tilde{F}_{\mathcal{U}_{\text{small}}}(T, N_T^{*(2)}) \geq (1 - 2\epsilon) \cdot \text{OPT}$  by comparing between these quantities on a term-by-term basis. Specifically, as mentioned in Section 2.6, we have  $N_{t,\ell}^{*(2)} = N_{t,\ell}^*$  for every  $\ell \in \text{light}(N_t^*)$ . In addition, by Lemma 3(c), we have  $N_{t,\ell}^{*(2)} \geq (1 - 2\epsilon) \cdot N_{t,\ell}^*$  for every  $\ell \in \text{heavy}(N_t^*)$ .

□

### 3. PTAS for General Instances

In this section, we devise a polynomial-time approximation scheme for arbitrarily structured instances of the incremental knapsack problem by exploiting our algorithm for the  $\lambda^+$ -bounded setting as a subroutine. The main result we establish is summarized in Theorem 1, which is restated here for readability purposes.

**THEOREM 1.** *For any accuracy level  $\epsilon > 0$ , the incremental knapsack problem can be approximated within factor  $1 - \epsilon$  of optimal. The running time of our algorithm is  $O(n^{\tilde{O}(1/\epsilon^3)} \cdot |\mathcal{I}|^{O(1)})$ , where  $|\mathcal{I}|$  stands for the input size in binary representation.*

#### 3.1. Technical overview

To establish our main result, we present an approximate reduction to the  $\lambda^+$ -bounded setting, previously examined in Section 2. For this purpose, we develop a dynamic programming approach that decomposes arbitrarily structured instances of the incremental knapsack problem into mutually disjoint sub-problems, each focusing on a restricted subset of time periods, collectively referred to as a cluster. These sub-problems are instances of the inverse incremental knapsack problem, shown to satisfy the  $\lambda^+$ -bounded property with an  $O((n/\epsilon)^{O(1/\epsilon)})$  ratio between their extremal  $\lambda^+$  values. Consequently, we leverage our algorithm for the  $\lambda^+$ -bounded setting as a subroutine to efficiently approximate each sub-problem. In what follows, we outline the main technical ingredients of our approach.

*Clustering time periods (Section 3.2).* Given a general instance of the incremental knapsack problem, we start-off by defining clusters of time periods. Each such cluster is formed by an interval of successive time periods, such that within clusters,  $\lambda^+$  values may differ by a factor of  $O((n/\epsilon)^{O(1/\epsilon)})$ . Moreover, these clusters are “well-separated”, in the sense that there is a large gap between their underlying  $\lambda^+$  values, a notion that will be made formal later on.

*Near-optimal uncrossing-stars solutions (Section 3.3).* Building on this construction, we inspect the relationship between clusters and profit classes within near-optimal solutions. Specifically, we examine the bipartite graph induced by such solutions, having cluster vertices on one side and profit class vertices on the other, whose edges represent whether or not at least one item in a given profit class is introduced within a given cluster of time periods. Our main structural result proves the existence of a near-optimal solution that induces a highly structured bipartite graph, taking the form of disjoint uncrossing-stars. Informally, each connected component of this graph is a star centered at a time period cluster, which is joined by edges to profit classes that follow a very specific ordering property.

*Dynamic programming decomposition (Sections 3.4-3.6).* Finally, we leverage the above-mentioned existence result to obtain a polynomial-time approximation scheme by showing that uncrossing-stars solutions can be efficiently computed. Specifically, we argue that one can identify a near-optimal uncrossing-stars solution via a dynamic programming based decomposition. Interestingly, the recursive equations involved in detecting each star-shaped solution boil down to handling a  $\lambda^+$ -bounded sub-problem, allowing us to utilize the approximation scheme of Section 2 as a subroutine.

### 3.2. Clustering time periods

*Defining clusters.* We define the sequence of time period intervals  $I_1 = \{t \in [T] : \lambda^{t+} \in (\frac{\epsilon}{n} \cdot \lambda^{1+}, \lambda^{1+}]\}$ ,  $I_2 = \{t \in [T] : \lambda^{t+} \in ((\frac{\epsilon}{n})^2 \cdot \lambda^{1+}, \frac{\epsilon}{n} \cdot \lambda^{1+}]\}$ , so on and so forth. Assuming without loss of generality that  $1/\epsilon$  is an integer, we randomly pick  $\xi \sim U\{0, \dots, \frac{1}{\epsilon} - 1\}$ . With respect to this choice, any interval  $I_m$  with  $(m \bmod \frac{1}{\epsilon}) = \xi$  is said to be marked, while all other intervals are unmarked. We use  $\mathcal{C}_1$  to denote the set of time periods belonging to the (unmarked) intervals appearing prior to the first marked interval (i.e.,  $I_1, \dots, I_{\xi-1}$ ). Then,  $\mathcal{C}_2$  will denote the set of time periods in the (unmarked) intervals appearing between the first and second marked intervals (i.e.,  $I_{\xi+1}, \dots, I_{\xi+\frac{1}{\epsilon}-1}$ ), so on and so forth, up until the last non-empty such set,  $\mathcal{C}_M$ . We refer to  $\mathcal{C}_1, \dots, \mathcal{C}_M$  as time period clusters, noting that within a single cluster,  $\lambda^{t+}$  values differ by a factor of at most  $(n/\epsilon)^{(1/\epsilon)-1}$  and that the overall number of clusters is  $M = O(\log_{(n/\epsilon)^{1/\epsilon}}(\frac{\lambda^{1+}}{\lambda^{T+}})) = O(|\mathcal{I}|^{O(1)})$ .

*The cost of ignoring marked intervals.* Let  $\mathcal{S}^* = \langle S_1^*, \dots, S_T^* \rangle$  be a fixed optimal solution, with a profit of  $\Psi(\mathcal{S}^*) = \sum_{i \in S_T^*} \Psi_i(\mathcal{S}^*)$ , written according to representation (1). Suppose we create a new feasible solution  $\mathcal{S}_\xi$  by modifying  $\mathcal{S}^*$  as follows: In each period  $t$  belonging to a marked interval, we do not introduce new items. In other words, all items in  $S_t^* \setminus S_{t-1}^*$  are deleted from  $S_t^*, \dots, S_T^*$ . Equivalently,  $\mathcal{S}_\xi$  is constructed by leaving only items that were introduced in unmarked time periods. Clearly, due to the randomness in choosing  $\xi$ , this is a random solution. The next lemma shows that, in expectation over the choice of  $\xi$ , the profit of  $\mathcal{S}_\xi$  is near optimal. As a result, at least one choice deterministically guarantees the bound stated below.

LEMMA 7.  $\mathbb{E}_\xi[\Psi(\mathcal{S}_\xi)] = (1 - \epsilon) \cdot \Psi(\mathcal{S}^*)$ .

*Proof.* To prove the desired claim, we first observe that each item  $i \in S_T^*$  contributes  $\Psi_i(\mathcal{S}^*)$  toward  $\Psi(\mathcal{S}_\xi)$  when the time period  $\nu(i, \mathcal{S}^*)$  in which it is introduced resides in an unmarked interval, and does not contribute at all otherwise. Therefore,

$$\mathbb{E}_\xi[\Psi(\mathcal{S}_\xi)] = \mathbb{E}_\xi \left[ \sum_{i \in S_T^*} \Psi_i(\mathcal{S}^*) \cdot \mathbb{I}[\nu(i, \mathcal{S}^*) \text{ in unmarked interval}] \right]$$

$$\begin{aligned}
&= \sum_{i \in \mathcal{S}_T^*} \Psi_i(\mathcal{S}^*) \cdot \mathbb{P}_\xi(\nu(i, \mathcal{S}^*) \text{ in unmarked interval}) \\
&= (1 - \epsilon) \cdot \sum_{i \in \mathcal{S}_T^*} \Psi_i(\mathcal{S}^*) \\
&= (1 - \epsilon) \cdot \Psi(\mathcal{S}^*) ,
\end{aligned}$$

where the third equality holds since any interval is unmarked with probability  $1 - \epsilon$ .  $\square$

### 3.3. Uncrossing-stars solutions

Before our formal construction, it is worth instilling some intuition regarding the purpose of near-optimal solutions that are supported on well-separated clusters. In essence, by creating  $\frac{n}{\epsilon}$ -sized gaps between clusters, we show that items in any given profit class have negligible profit contributions in all but a single cluster of time periods; this correspondence between profit classes and clusters follows an index-based ordering. On this account, we generate a highly-structured near-optimal solution by “filtering” all negligible profit contributions. Taking a graph theoretic perspective, the latter solution can be represented as a collection of uncrossing stars, as illustrated in Figure 1.

To formalize these notions, by enumerating over the collection of values  $0, \dots, \frac{1}{\epsilon} - 1$ , we may assume without loss of generality that a particular choice of  $\xi$  for which  $\Psi(\mathcal{S}_\xi) \geq (1 - \epsilon) \cdot \Psi(\mathcal{S}^*)$  is known in advance. Now let  $\hat{\mathcal{S}} = \langle \hat{S}_1, \dots, \hat{S}_T \rangle$  be a maximum-profit solution that only introduces items in time periods belonging to unmarked intervals, where “unmarked” and “marked” are defined with respect to  $\xi$ . Clearly,  $\Psi(\hat{\mathcal{S}}) \geq \Psi(\mathcal{S}_\xi) \geq (1 - \epsilon) \cdot \Psi(\mathcal{S}^*)$ .

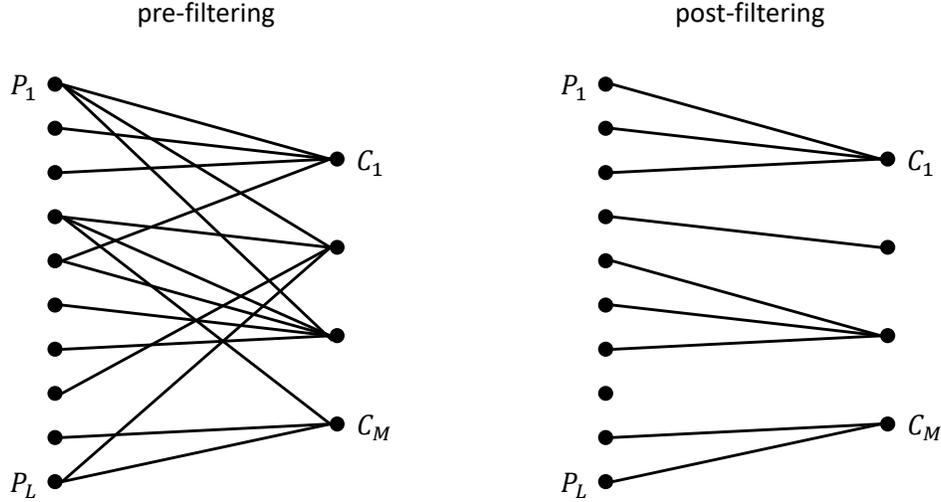
*The solution graph.* From this point on, we focus our attention on solutions  $\mathcal{S} = \langle S_1, \dots, S_T \rangle$  that only make use of time periods within unmarked intervals. For each such solution, we define an undirected bipartite graph  $G(\mathcal{S})$  as follows:

- Vertices: On one side, we have the time period clusters  $\mathcal{C}_1, \dots, \mathcal{C}_M$ . On the other side, we have the profit classes  $\mathcal{P}_0, \dots, \mathcal{P}_L$ , where  $L$  stands for the maximal index of a non-empty profit class.
- Edges: There is an edge  $(\mathcal{C}_m, \mathcal{P}_\ell)$  if and only if at least one item belonging to the profit class  $\mathcal{P}_\ell$  is introduced by  $\mathcal{S}$  in a time period belonging to  $\mathcal{C}_m$ .

To better understand the upcoming discussion, we advise the reader to consult Figure 1.

*Post-filtering graph properties.* We proceed by filtering items whose profit contribution is negligible. Specifically, let  $\tilde{\mathcal{S}} = \langle \tilde{S}_1, \dots, \tilde{S}_T \rangle$  be the solution obtained from  $\hat{\mathcal{S}} = \langle \hat{S}_1, \dots, \hat{S}_T \rangle$  by eliminating all items  $i \in \hat{S}_T$  for which  $\Psi_i(\hat{\mathcal{S}}) \leq \frac{\epsilon}{n} \cdot \Psi(\hat{\mathcal{S}})$ ; it is easy to verify that  $\Psi(\tilde{\mathcal{S}}) \geq (1 - \epsilon) \cdot \Psi(\hat{\mathcal{S}})$ . The crucial claim is that this filtering procedure results in a greatly simplified graph  $G(\tilde{\mathcal{S}})$ . Lemma 8 below shows that this graph is a collection of uncrossing stars, formally defined as follows:

1.  $\mathcal{P}$ -degrees: Each of the vertices  $\mathcal{P}_0, \dots, \mathcal{P}_L$  has a degree of at most 1.



**Figure 1** The pre-filtering graph  $G(\mathcal{S})$  appears on the left, whereas its post-filtering counterpart  $G(\tilde{\mathcal{S}})$  appears on the right, noting that the latter is a collection of uncrossing stars.

2. *Uncrossing*: There are no crossing edges, namely, a pair of edges  $(\mathcal{C}_{m_1}, \mathcal{P}_{\ell_1})$  and  $(\mathcal{C}_{m_2}, \mathcal{P}_{\ell_2})$  with  $\ell_1 < \ell_2$  and  $m_1 > m_2$ .

Put differently, property 1 implies that every connected component of  $G(\tilde{\mathcal{S}})$  is star-shaped, centered at a time period cluster. Property 2 imposes that, for any two such stars, say centered at  $\mathcal{C}_{m_1}$  and  $\mathcal{C}_{m_2}$  with  $m_1 < m_2$ , all profit classes within  $\mathcal{C}_{m_1}$ 's star have lower indices than those within  $\mathcal{C}_{m_2}$ 's star, which will be instrumental for algorithmic purposes, as shown in Sections 3.4-3.6.

LEMMA 8.  $G(\tilde{\mathcal{S}})$  is a collection of uncrossing stars.

*Proof.* In what follows, we prove the  $\mathcal{P}$ -degrees property, noting that the uncrossing property can be established via nearly identical arguments and is omitted to avoid duplicity. To arrive at a contradiction, suppose that  $G(\tilde{\mathcal{S}})$  has some vertex  $\mathcal{P}_\ell$  with degree at least 2, and let  $(\mathcal{C}_{m_1}, \mathcal{P}_\ell)$  and  $(\mathcal{C}_{m_2}, \mathcal{P}_\ell)$  be two of its adjacent edges; without loss of generality,  $m_1 < m_2$ . By construction of  $G(\tilde{\mathcal{S}})$ , at least one item  $i_1 \in \mathcal{P}_\ell$  is introduced by  $\tilde{\mathcal{S}}$  in a time period belonging to  $\mathcal{C}_{m_1}$ , implying that the pre-filtered solution  $\hat{\mathcal{S}}$  is associated with a profit of

$$\Psi(\hat{\mathcal{S}}) \geq \Psi_{i_1}(\hat{\mathcal{S}}) = \Psi_{i_1}(\tilde{\mathcal{S}}) = p_{i_1} \cdot \lambda^{\nu(i_1, \tilde{\mathcal{S}})+} \geq (1 + \epsilon)^\ell \cdot \lambda^{\max \mathcal{C}_{m_1}+}, \quad (5)$$

where the last inequality holds since items in  $\mathcal{P}_\ell$  have a uniform profit of  $(1 + \epsilon)^\ell$  and since  $\nu(i_1, \tilde{\mathcal{S}}) \in \mathcal{C}_{m_1}$ , over which the smallest  $\lambda^+$  value is attained at  $\lambda^{\max \mathcal{C}_{m_1}+}$ , as  $\lambda^{1+} \geq \dots \geq \lambda^{T+}$ . Similarly, at least one item  $i_2 \in \mathcal{P}_\ell$  is introduced in a time period belonging to  $\mathcal{C}_{m_2}$ , meaning that its contribution to the pre-filtered profit satisfies

$$\Psi_{i_2}(\hat{\mathcal{S}}) = \Psi_{i_2}(\tilde{\mathcal{S}}) = p_{i_2} \cdot \lambda^{\nu(i_2, \tilde{\mathcal{S}})+} \leq (1 + \epsilon)^\ell \cdot \lambda^{\min \mathcal{C}_{m_2}+} \leq (1 + \epsilon)^\ell \cdot \frac{\epsilon}{n} \cdot \lambda^{\max \mathcal{C}_{m_1}+} \leq \frac{\epsilon}{n} \cdot \Psi(\hat{\mathcal{S}}),$$

where the second inequality holds since  $\lambda^{\min \mathcal{C}_{m_2+}} \leq (\frac{\epsilon}{n})^{m_2-m_1} \cdot \lambda^{\max \mathcal{C}_{m_1+}} \leq \frac{\epsilon}{n} \cdot \lambda^{\max \mathcal{C}_{m_1+}}$ , by construction of the time period clusters, and the third inequality is obtained by plugging-in the bound in (5). We have just arrived at a contradiction: Since  $\Psi_{i_2}(\hat{\mathcal{S}}) \leq \frac{\epsilon}{n} \cdot \Psi(\hat{\mathcal{S}})$ , the item  $i_2$  should have been eliminated as part of our filtering procedure, while constructing the solution  $\tilde{\mathcal{S}}$ .  $\square$

### 3.4. The exact dynamic program

In the remainder of this section, we develop a dynamic programming approach for computing a near-optimal uncrossing-stars solution by leveraging our algorithm for the  $\lambda^+$ -bounded setting.

*State space.* Each state of our dynamic program is described by three parameters:

- The index  $m$  of the current cluster, taking one of the values  $0, 1, \dots, M$ .
- The index  $\ell_m$  of the current profit class, whose possible values are  $-1, 0, \dots, L$ .
- The total profit  $\varphi_m$  accumulated thus far. For the time being,  $\varphi_m$  will be treated as a continuous parameter, restricted to the interval  $[0, \lambda^{1+} \cdot \sum_{i \in [n]} p_i]$ .

We mention in passing that  $m = 0$  does not correspond to a concrete cluster, and we therefore define  $\mathcal{C}_0 = \emptyset$  by convention. Similarly,  $\ell_m = -1$  is an auxiliary value, and we set  $\mathcal{P}_{-1} = \emptyset$  for completeness. These edge cases will be useful in prescribing the terminal conditions of our dynamic program. In addition, due to making use of a continuous profit parameter  $\varphi_m$ , the initial dynamic programming formulation we propose should not be viewed as a discrete algorithm, but rather as a characterization of optimal uncrossing-stars solutions. We will explain in Section 3.5 how to efficiently approximate this dynamic program through an appropriate discretization of  $\varphi_m$ .

*Value function.* Letting  $\tau_m = \max \mathcal{C}_m$ , the value function  $F(m, \ell_m, \varphi_m)$  stands for the minimum-possible total weight  $w(S_{\tau_m})$  incurred by a  $\tau_m$ -period uncrossing-stars solution  $\mathcal{S} = \langle S_1, \dots, S_{\tau_m} \rangle$  that satisfies the following conditions:

1. *[1, m]-clustered periods:* Items may be introduced only within time periods belonging to one of the clusters  $\mathcal{C}_1, \dots, \mathcal{C}_m$ . In other words,  $S_t \setminus S_{t-1} = \emptyset$  for every  $t \notin \biguplus_{\mu \in [m]} \mathcal{C}_\mu$ .
2. *[0,  $\ell_m$ ]-profit classes:* We are allowed to introduce only items that reside within the profit classes  $\mathcal{P}_0, \dots, \mathcal{P}_{\ell_m}$ , meaning that  $S_{\tau_m} \subseteq \biguplus_{\ell \in [\ell_m]_0} \mathcal{P}_\ell$ .
3. *Total profit:* We are restricted to attain a profit of at least  $\varphi_m$ , or put differently,  $\Psi(\mathcal{S}) \geq \varphi_m$ .

A close inspection of this definition reveals that we have actually formulated a multi-cluster instance of the inverse incremental knapsack problem, whose feasibility set is further limited to uncrossing-stars solutions. We designate the latter instance by  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ , noting that  $F(m, \ell_m, \varphi_m)$  corresponds to its optimum value. When  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$  is infeasible, we define  $F(m, \ell_m, \varphi_m) = \infty$ . Clearly, the optimal profit attained by an uncrossing-stars solution is precisely the maximal value  $\varphi_M$  for which  $F(M, L, \varphi_M) < \infty$ . As an immediate conclusion of Section 3.3, the latter parameter can be related to the optimal profit as follows.

**COROLLARY 1.**  $F(M, L, \varphi_M) < \infty$  for any  $\varphi_M \leq (1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)$ .

*Optimal substructure.* In order to derive recursive equations for this value function, for every  $\phi \geq 0$  and  $\omega \geq 0$ , let us define a single-cluster instance  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell^-, \ell^+]$  of the inverse incremental knapsack problem as follows:

- The underlying collection of items is given by those residing in the profit classes  $\mathcal{P}_{\ell^-}, \dots, \mathcal{P}_{\ell^+}$ .
- The time periods are only those within the cluster  $\mathcal{C}_m$ . Each such period  $t$  is given an adjusted capacity of  $[W_t - \omega]^+$ .
- The objective is to compute a solution  $\mathcal{S} = \langle S_{\min \mathcal{C}_m}, \dots, S_{\max \mathcal{C}_m} \rangle$  that minimizes  $w(S_{\max \mathcal{C}_m})$  subject to the profit requirement  $\Psi(\mathcal{S}) \geq \phi$ . Here, the profit function  $\Psi$  is specialized for the time periods in question, namely,  $\Psi(\mathcal{S}) = \sum_{t \in \mathcal{C}_m} \lambda_t \cdot p(S_t)$ .

Moving forward, we will use  $\text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell^-, \ell^+])$  to designate the optimum value of  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell^-, \ell^+]$ , with the convention that  $\text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell^-, \ell^+]) = \infty$  when this instance is infeasible.

Now, to understand where such instances fit in, suppose we are given an optimal  $\tau_m$ -period uncrossing-stars solution  $\mathcal{S} = \langle S_1, \dots, S_{\tau_m} \rangle$  for the multi-cluster instance  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ , satisfying in particular  $w(S_{\tau_m}) = F(m, \ell_m, \varphi_m)$ . Our recursive approach proceeds by decomposing  $\mathcal{S}$  into two separate solutions:

- The restriction of  $\mathcal{S}$  to time periods in lower-indexed clusters,  $\mathcal{S}^{< m} = \langle S_1, \dots, S_{\tau_{m-1}} \rangle$ .
- The marginal sequence of items introduced within cluster  $\mathcal{C}_m$ , given by  $\mathcal{S}^m = \langle S_{\min \mathcal{C}_m} \setminus S_{\tau_{m-1}}, \dots, S_{\max \mathcal{C}_m} \setminus S_{\tau_{m-1}} \rangle$ .

Focusing on the restriction  $\mathcal{S}^{< m}$ , let  $\varphi_{m-1} = \Psi(\mathcal{S}^{< m})$  be the profit obtained by this solution, and let  $\ell_{m-1} \in [\ell_m]_0$  be the maximum index of a profit class out of which at least one item is introduced, i.e.,  $\ell_{m-1} = \max\{\ell \in [\ell_m]_0 : S_{\tau_{m-1}} \cap \mathcal{P}_\ell \neq \emptyset\}$ . The next claim, whose proof is presented in Appendix A.4, reveals the optimal substructure pertaining to our decomposition of  $\mathcal{S}$  into  $\mathcal{S}^{< m}$  and  $\mathcal{S}^m$ .

LEMMA 9. *The solution  $\mathcal{S}^{< m}$  is optimal for the multi-cluster instance  $\mathcal{I}_{\text{multi}}^{\varphi_{m-1}}[m-1, \ell_{m-1}]$ , and  $\mathcal{S}^m$  is optimal for the single-cluster instance  $\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, w(S_{\tau_{m-1}}^{< m})}[m, \ell_{m-1} + 1, \ell_m]$ .*

In light of the preceding discussion, we have just shown that

$$\begin{aligned}
F(m, \ell_m, \varphi_m) &= w(S_{\tau_m}) \\
&= w(S_{\tau_{m-1}}^{< m}) + w(S_{\tau_m}^m) \\
&= F(m-1, \ell_{m-1}, \varphi_{m-1}) + \text{OPT}\left(\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, w(S_{\tau_{m-1}}^{< m})}[m, \ell_{m-1} + 1, \ell_m]\right) \\
&= F(m-1, \ell_{m-1}, \varphi_{m-1}) + \text{OPT}\left(\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, F(m-1, \ell_{m-1}, \varphi_{m-1})}[m, \ell_{m-1} + 1, \ell_m]\right). \quad (6)
\end{aligned}$$

Here, the first equality holds since  $\mathcal{S}$  is an optimal solution to  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ . The second equality is obtained by observing that  $S_{\tau_m}$  is a disjoint union of  $S_{\tau_{m-1}}^{< m}$  and  $S_{\tau_m}^m$ , by definition. The third and fourth equalities are direct implications of Lemma 9.

*Recursive equations.* By taking advantage of this structural insight, we can now argue that the value function  $F$  satisfies the following recursive equation:

$$\begin{aligned} F(m, \ell_m, \varphi_m) &= \min_{\ell_{m-1}, \varphi_{m-1}} \left\{ F(m-1, \ell_{m-1}, \varphi_{m-1}) + \text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) \right\} \\ \text{s.t.} \quad &\ell_{m-1} \in [\ell_m]_0, \quad \varphi_{m-1} \in [0, \varphi_m] \\ &\phi = \varphi_m - \varphi_{m-1}, \\ &\omega = F(m-1, \ell_{m-1}, \varphi_{m-1}). \end{aligned}$$

It is worth mentioning that  $\phi$  and  $\omega$  are not decision variables in this formulation; they simply serve as shorthand notation, to avoid having  $\text{OPT}(\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, F(m-1, \ell_{m-1}, \varphi_{m-1})}[\dots])$  in the objective function. To better understand these recursive equations, note that decomposition (6) proves the existence of at least one solution  $(\ell_{m-1}, \varphi_{m-1})$  to the maximization problem on the right hand side with an objective value of  $F(m, \ell_m, \varphi_m)$ . In parallel, for any choice of  $\ell_{m-1} \in [\ell_m]_0$  and  $\varphi_{m-1} \in [0, \varphi_m]$ , their corresponding objective value provides an upper bound on  $F(m, \ell_m, \varphi_m)$ . Indeed, we can simply glue together the optimal solutions to  $\mathcal{I}_{\text{multi}}^{\varphi_{m-1}}[m-1, \ell_{m-1}]$  and  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]$ , thus forming a feasible solution to the multi-cluster instance  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ , implying that

$$\begin{aligned} &F(m-1, \ell_{m-1}, \varphi_{m-1}) + \text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) \\ &= \text{OPT}(\mathcal{I}_{\text{multi}}^{\varphi_{m-1}}[m-1, \ell_{m-1}]) + \text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) \\ &\geq \text{OPT}(\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]) \\ &= F(m, \ell_m, \varphi_m), \end{aligned}$$

The terminal states of the dynamic program  $F$  correspond to  $m = 0$  or  $\ell_m = -1$ , in which case  $F(m, \ell_m, 0) = 0$  by definition and  $F(m, \ell_m, \varphi_m) = \infty$  for  $\varphi_m > 0$ . Finally, from a computational standpoint, beyond the continuity of the profit parameters, our objective function involves an exact solution to single-cluster instance  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]$ . To address this computational challenge, we utilize the approximation scheme developed in Section 2, as explained in the next section.

### 3.5. The approximate program

With the above-mentioned obstacles in mind, we turn our attention to devising a polynomial-time approximation for the exact formulation  $F$ . To formalize this notion, we introduce two alterations of the original dynamic program:

- *Profit discretization.* In order to discretize the required profit  $\varphi_m$ , we restrict this parameter to taking values within the set

$$\Phi = \{0\} \cup \left\{ \left(1 + \frac{\epsilon}{M}\right)^k \cdot \delta : 0 \leq k \leq \left\lceil \frac{M}{\epsilon} \cdot \log \left( \frac{M}{\epsilon} \cdot \frac{\lambda^{1+}}{\lambda_T} \right) \right\rceil \right\}, \quad (7)$$

where  $\delta = \frac{\epsilon}{M} \cdot \lambda_T \cdot p_{\max}$ . One can easily verify that this set is of cardinality  $|\Phi| = O(\frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot |\mathcal{I}|^{O(1)})$ , by recalling that  $M = O(|\mathcal{I}|^{O(1)})$ , as shown in Section 3.2. From this point on, we use  $\mathcal{D}$  to designate the discrete state space, in which  $\varphi_m$  is restricted to  $\Phi$ , meaning that

$$\mathcal{D} = \{(m, \ell_m, \varphi_m) : m \in \{0, 1, \dots, M\}, \ell_m \in \{-1, 0, \dots, L\}, \varphi_m \in \Phi\} .$$

- *Approximating single-cluster problems.* In addition, rather than shooting for an optimal solution to single-cluster instances of the form  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]$ , we utilize the approximation scheme proposed in Theorem 3 as a subroutine. Consequently, we are guaranteed to obtain in  $O((n\rho)^{\tilde{O}(1/\epsilon^2)} \cdot |\mathcal{I}|^{O(1)})$  time a super-optimal solution, i.e., one having a total weight of

$$\widetilde{\text{OPT}}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) \leq \text{OPT}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) , \quad (8)$$

while ensuring that its profit is at least  $(1 - \epsilon) \cdot \phi$ . In this context,  $\rho$  stands for the ratio between the largest and smallest  $\lambda^{t+}$  values within the single-cluster instance  $\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]$ . As noted in Section 3.2, these values differ by a factor of at most  $(n/\epsilon)^{(1/\epsilon)-1}$ . Hence, by plugging in  $\rho = (n/\epsilon)^{1/\epsilon}$ , we infer that the time required to approximate each such instance is  $O(n^{\tilde{O}(1/\epsilon^3)} \cdot |\mathcal{I}|^{O(1)})$ .

*The approximate formulation.* These two alterations allow us to formulate an approximate dynamic program  $\tilde{F}$ , defined by mimicking our exact program  $F$ . Specifically, for every discretized state  $(m, \ell_m, \varphi_m) \in \mathcal{D}$ , we have

$$\begin{aligned} \tilde{F}(m, \ell_m, \varphi_m) &= \min_{\ell_{m-1}, \varphi_{m-1}} \left\{ \tilde{F}(m-1, \ell_{m-1}, \varphi_{m-1}) + \widetilde{\text{OPT}}(\mathcal{I}_{\text{single}}^{\phi, \omega}[m, \ell_{m-1} + 1, \ell_m]) \right\} \\ \text{s.t.} \quad &\ell_{m-1} \in [\ell_m]_0 , \quad \varphi_{m-1} \in [0, \varphi_m] \cap \Phi \\ &\phi = [(\varphi_m - \varphi_{m-1}) - \zeta(\varphi_{m-1})]^+ , \\ &\omega = \tilde{F}(m-1, \ell_{m-1}, \varphi_{m-1}) . \end{aligned}$$

To shed light on the definition of  $\phi$ , we mention that due to discretizing the profit requirement  $\varphi_m$ , a certain degree of inaccuracy may potentially accumulate throughout our recursive calls. For this reason, naively following the exact formulation  $F$  and setting our marginal profit requirement as  $\phi = \varphi_m - \varphi_{m-1}$  may not be an appropriate choice. Instead, we settle for a slightly smaller profit of  $\phi = [(\varphi_m - \varphi_{m-1}) - \zeta(\varphi_{m-1})]^+$ , where  $\zeta(\cdot)$  is a correction function whose exact form will be determined later on in a way that offsets the above-mentioned inaccuracies. As before, terminal states of our approximate dynamic program correspond to  $m = 0$  or  $\ell_m = -1$ , in which case  $\tilde{F}(m, \ell_m, 0) = 0$  by definition and  $\tilde{F}(m, \ell_m, \varphi_m) = \infty$  for  $\varphi_m > 0$ .

*Overall running time.* Summarizing our earlier discussion, the number of states over which we compute the value function  $\tilde{F}$  is only  $|\mathcal{D}| = O(ML \cdot |\Phi|) = O(\frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot |\mathcal{I}|^{O(1)})$ . At each state, solving its recursive equation involves  $O(M \cdot |\mathcal{D}|)$  calls to the subroutine given by Theorem 3, which approximates single-cluster instances of the inverse incremental knapsack problem. As previously noted, each such call can be executed in  $O(n^{\tilde{O}(1/\epsilon^3)} \cdot |\mathcal{I}|^{O(1)})$  time. All in all, our approximate dynamic programming formulation  $\tilde{F}$  requires  $O(n^{\tilde{O}(1/\epsilon^3)} \cdot |\mathcal{I}|^{O(1)})$  time to be solved.

*Final solution.* Lastly, it remains to explain how the various  $\tilde{F}(\cdot, \cdot, \cdot)$  values we obtain can ultimately be translated back into an uncrossing-stars solution. To this end, let  $\tilde{\varphi}_M$  be the maximum-possible profit  $\varphi_M \in \Phi$  for which  $\tilde{F}(M, L, \varphi_M) < \infty$ . Given this quantity, consider the sequence  $(0, \tilde{\ell}_0, \tilde{\varphi}_0), (1, \tilde{\ell}_1, \tilde{\varphi}_1), \dots, (M, \tilde{\ell}_M, \tilde{\varphi}_M)$  of states traversed by the dynamic program  $\tilde{F}$  in reaching state  $(M, \tilde{\ell}_M, \tilde{\varphi}_M)$ , with the convention that  $\tilde{\ell}_0 = -1$ ,  $\tilde{\varphi}_0 = 0$ , and  $\tilde{\ell}_M = L$ . Our solution  $\bar{\mathcal{S}} = \langle \bar{\mathcal{S}}_1, \dots, \bar{\mathcal{S}}_T \rangle$  is formed by gluing together the single-cluster solutions generated along this sequence. In other words, for each state  $(m, \tilde{\ell}_m, \tilde{\varphi}_m)$ , let  $\tilde{\mathcal{S}}^m$  be the solution obtained by approximating the single-cluster instance  $\mathcal{I}_{\text{single}}^{\tilde{\varphi}_m, \tilde{\omega}_m}[m, \tilde{\ell}_{m-1} + 1, \tilde{\ell}_m]$ , where  $\tilde{\varphi}_m = [(\tilde{\varphi}_m - \tilde{\varphi}_{m-1}) - \zeta(\tilde{\varphi}_{m-1})]^+$  and  $\tilde{\omega}_m = \tilde{F}(m-1, \tilde{\ell}_{m-1}, \tilde{\varphi}_{m-1})$ . Then, at any time period  $t \in \mathcal{C}_m$ , the items introduced by  $\bar{\mathcal{S}}$  are simply those introduced by  $\tilde{\mathcal{S}}^m$ , i.e.,  $\bar{\mathcal{S}}_t \setminus \bar{\mathcal{S}}_{t-1} = \tilde{\mathcal{S}}_t^m \setminus \tilde{\mathcal{S}}_{t-1}^m$ .

### 3.6. Analysis

*Feasibility.* Prior to examining how profitable  $\bar{\mathcal{S}}$  is, we address a more basic question, asking why this solution is indeed feasible. The next lemma argues that our construction is guaranteed to satisfy the capacity constraints of all time periods.

LEMMA 10.  $w(\bar{\mathcal{S}}_t) \leq W_t$  for every  $t \in [T]$ .

*Proof.* We establish the desired claim by induction over the cluster indices  $m \in [M]$ . For the base case of  $m = 1$ , note that  $\tilde{\mathcal{S}}_1$  is a solution to the single-cluster instance  $\mathcal{I}_{\text{single}}^{\tilde{\varphi}_1, \tilde{\omega}_1}[1, \tilde{\ell}_0 + 1, \tilde{\ell}_1]$ , where  $\tilde{\omega}_1 = \tilde{F}(0, \tilde{\ell}_0, \tilde{\varphi}_0) = \tilde{F}(0, -1, 0) = 0$  in view of our terminal conditions. Consequently, for every time period  $t \in \mathcal{C}_1$ , it follows that  $w(\bar{\mathcal{S}}_t) = w(\tilde{\mathcal{S}}_t^1) \leq [W_t - \tilde{\omega}_1]^+ = W_t$ , where the latter inequality holds since this period was given an adjusted capacity of  $[W_t - \tilde{\omega}_1]^+$  within  $\mathcal{I}_{\text{single}}^{\tilde{\varphi}_1, \tilde{\omega}_1}[1, \tilde{\ell}_0 + 1, \tilde{\ell}_1]$ , as explained in Section 3.4.

To prove the inductive step, with respect to a cluster index  $m \geq 2$ , the construction of  $\bar{\mathcal{S}}$  implies that, for every time period  $t \in \mathcal{C}_m$ ,

$$\bar{\mathcal{S}}_t \setminus \bar{\mathcal{S}}_{\tau_{m-1}} = \biguplus_{q=\tau_{m-1}+1}^t (\bar{\mathcal{S}}_q^m \setminus \bar{\mathcal{S}}_{q-1}^m) = \biguplus_{q=\tau_{m-1}+1}^t (\tilde{\mathcal{S}}_q^m \setminus \tilde{\mathcal{S}}_{q-1}^m) = \tilde{\mathcal{S}}_t^m,$$

with the convention that  $\tilde{S}_{\tau_{m-1}}^m = \emptyset$ . Consequently,

$$\begin{aligned}
w(\bar{S}_t) &= w(\bar{S}_{\tau_{m-1}}) + w(\bar{S}_t \setminus \bar{S}_{\tau_{m-1}}) \\
&= \sum_{\mu \in [m-1]} w(\tilde{S}_{\tau_\mu}^\mu) + w(\tilde{S}_t^m) \\
&= \sum_{\mu \in [m-1]} \widetilde{\text{OPT}} \left( \mathcal{I}_{\text{single}}^{\tilde{\varphi}_\mu, \tilde{\omega}_\mu} [\mu, \tilde{\ell}_{\mu-1} + 1, \tilde{\ell}_\mu] \right) + w(\tilde{S}_t^m) \\
&= \tilde{F}(m-1, \tilde{\ell}_{m-1}, \tilde{\varphi}_{m-1}) + w(\tilde{S}_t^m), \tag{9}
\end{aligned}$$

where the last equality is obtained by expanding the recursive equation of our approximate dynamic program  $\tilde{F}$ . However,  $\tilde{\omega}_m = \tilde{F}(m-1, \tilde{\ell}_{m-1}, \tilde{\varphi}_{m-1})$  by definition, while  $w(\tilde{S}_t^m) \leq [W_t - \tilde{\omega}_m]^+ = W_t - \tilde{\omega}_m$ , since  $\tilde{S}^m$  is a solution to the single-cluster instance  $\mathcal{I}_{\text{single}}^{\tilde{\varphi}_m, \tilde{\omega}_m} [m, \tilde{\ell}_{m-1} + 1, \tilde{\ell}_m]$ . By plugging these observations into (9), we have just shown that  $w(\bar{S}_t) \leq W_t$ .  $\square$

*Optimality gap.* We now turn our attention to proving that  $\bar{S}$  is profit-wise comparable to the optimal incremental knapsack solution  $\mathcal{S}^*$ , thereby concluding the proof of Theorem 1. A crucial part of our overall argument shows that, over the discretized state space  $\mathcal{D}$ , the value function  $\tilde{F}$  provides a lower bound on the original value function  $F$ . To this end, it is important to mention that we still have to prescribe the exact form of the correction function  $\zeta(\cdot)$  that appears within the recursive equations of  $\tilde{F}$ . For this purpose, we went through the analysis below with  $\zeta(\cdot)$  as a parameter and determined the properties it has to satisfy; simply choosing  $\zeta(\varphi) = \frac{\epsilon}{M} \cdot \varphi + \delta$  turns out to be sufficient.

LEMMA 11.  $\Psi(\bar{S}) \geq (1 - 7\epsilon) \cdot \Psi(\mathcal{S}^*)$ .

*Proof.* Our analysis proceeds by establishing two claims, one relating the profit  $\Psi(\bar{S})$  to the profit parameter  $\tilde{\varphi}_M$ , as defined in Section 3.5, and the other relating between the value functions  $F$  and  $\tilde{F}$ . Since the arguments involved are rather technical, we provide both proofs in Appendices A.5 and A.6.

CLAIM 2.  $\Psi(\bar{S}) \geq (1 - 2\epsilon) \cdot \tilde{\varphi}_M - M\delta$ .

CLAIM 3.  $\tilde{F}(m, \ell_m, \varphi_m) \leq F(m, \ell_m, \varphi_m)$ , for every state  $(m, \ell_m, \varphi_m) \in \mathcal{D}$ .

We remind the reader that, as an intermediate step, we have shown in Corollary 1 that  $F(M, L, \varphi_M) < \infty$  for any  $\varphi_M \leq (1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)$ . In turn, Claim 3 implies that  $\tilde{F}(M, L, \varphi_M) < \infty$  for any  $\varphi_M \in \Phi \cap [0, (1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)]$ . Therefore, since  $\tilde{\varphi}_M$  is the maximum-possible profit  $\varphi_M \in \Phi$  for which  $\tilde{F}(M, L, \varphi_M) < \infty$ , we clearly have

$$\tilde{\varphi}_M \geq \max \{ \Phi \cap [0, (1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)] \} \geq \frac{(1 - 2\epsilon) \cdot \Psi(\mathcal{S}^*)}{1 + \frac{\epsilon}{M}} - \delta \geq (1 - 3\epsilon) \cdot \Psi(\mathcal{S}^*) - \delta,$$

where the middle inequality follows from the definition of  $\Phi$  in (7). Plugging this relation back into Claim 2, we derive the desired bound on  $\Psi(\bar{\mathcal{S}})$  by observing that

$$\begin{aligned} \Psi(\bar{\mathcal{S}}) &\geq (1 - 2\epsilon) \cdot \tilde{\varphi}_M - M\delta \\ &\geq (1 - 5\epsilon) \cdot \Psi(\mathcal{S}^*) - (M + 1) \cdot \delta \\ &\geq (1 - 5\epsilon) \cdot \Psi(\mathcal{S}^*) - 2\epsilon \cdot \lambda_T p_{\max} \\ &\geq (1 - 7\epsilon) \cdot \Psi(\mathcal{S}^*) . \end{aligned}$$

Here, the third inequality holds since  $\delta = \frac{\epsilon}{M} \cdot \lambda_T \cdot p_{\max}$ , and the fourth inequality is obtained by noting that  $\lambda_T p_{\max} \leq \Psi(\mathcal{S}^*)$ , since an objective value of  $\lambda_T p_{\max}$  can trivially be attained for the incremental knapsack problem and since  $\mathcal{S}^*$  is an optimal solution.  $\square$

## 4. Concluding Remarks

*Additional applications.* Interestingly, the applicability of our technical ideas beyond the present setting has been demonstrated by recent papers, focusing primarily on revenue management problems. Specifically, well-separated clusters along the lines of Section 3.2 were exploited by Bai et al. (2021) for computing near-optimal solutions to the assortment optimization problem under the multi-purchase Multinomial Logit model. In addition, Jiang and Feldman (2021) found the state-space reduction methods we proposed in Sections 2.5-2.7 to be useful in devising a PTAS for a sequential display optimization problem.

*Subsequent work.* Following a preliminary version of the current paper, Faenza et al. (2020) have investigated the generalized incremental knapsack problem, in which item profits have a general dependency on their insertion time. In other words, letting  $p_{it}$  be the marginal profit contribution due to introducing item  $i$  at time period  $t$ , the quantities  $\{p_{it}\}_{t \in [T]}$  are allowed to be completely unrelated. In our setting, these contributions take a simpler product form,  $p_{it} = p_i \cdot \lambda^{t+}$ . Faenza et al. (2020) proved that generalized incremental knapsack can be approximated within factor  $\frac{1}{2} - \epsilon$  in polynomial time. Additionally, the authors have proposed a quasi-PTAS, computing  $(1 - \epsilon)$ -approximate solutions in slightly super-polynomial time; specifically, this approach includes an exponential dependency on  $O(\log^{O(1)} |\mathcal{I}|)$ , where  $|\mathcal{I}|$  stands for the input size in binary representation. Even though the latter bounds are weaker in absolute terms than the polynomial-time approximation scheme we have developed here, it is worth keeping in mind that these results are incomparable since, as the name implies, the generalized incremental knapsack problem is a more general computational setting. From a technical standpoint, the methods involved are very different in nature, and we are not aware of any way to leverage our findings for the purpose of devising a PTAS for generalized incremental knapsack or vice versa.

## Acknowledgments

We thank Jacob Feldman (WUSTL) for his feedback on an earlier version of this paper.

---

## References

- Aouad A, Segev D (2020) Display optimization for vertically differentiated locations under Multinomial Logit preferences. *Management Science* (forthcoming).
- Bai Y, Feldman J, Segev D, Topaloglu H, Wagner L (2021) Assortment optimization under the multi-purchase Multinomial Logit choice model. Available as SSRN report #3866734.
- Bienstock D, Sethuraman J, Ye C (2013) Approximation algorithms for the incremental knapsack problem via disjunctive programming. ArXiv preprint arXiv:1311.4563.
- Chekuri C, Khanna S (2005) A polynomial time approximation scheme for the multiple knapsack problem. *SIAM Journal on Computing* 35(3):713–728.
- Davis JM, Topaloglu H, Williamson DP (2015) Assortment optimization over time. *Operations Research Letters* 43(6):608–611.
- Della Croce F, Pferschy U, Scatamacchia R (2018) Approximating the 3-period incremental knapsack problem. *Journal of Discrete Algorithms* 52-53:55–69.
- Della Croce F, Pferschy U, Scatamacchia R (2019) On approximating the incremental knapsack problem. *Discrete Applied Mathematics* 264:26–42.
- Derakhshan M, Golrezaei N, Manshadi V, Mirrokni V (2018) Product ranking on online platforms. *Management Science* (forthcoming).
- Faenza Y, Malinovic I (2018) A PTAS for the time-invariant incremental knapsack problem. *Proceedings of the 5th International Symposium on Combinatorial Optimization*, 157–169.
- Faenza Y, Segev D, Zhang L (2020) Approximation algorithms for the generalized incremental knapsack problem. ArXiv preprint arXiv:2009.07248.
- Feige U, Vondrák J (2010) The submodular welfare problem with demand queries. *Theory of Computing* 6(1):247–290.
- Fleischer L, Goemans MX, Mirrokni VS, Sviridenko M (2011) Tight approximation algorithms for maximum separable assignment problems. *Mathematics of Operations Research* 36(3):416–431.
- Gallego G, Li A, Truong V, Wang X (2020) Approximation algorithms for product framing and pricing. *Operations Research* 68(1):134–160.
- Hartline J, Sharp A (2006) An incremental model for combinatorial maximization problems. *Proceedings of the 5th International Workshop on Experimental Algorithms*, 36–48.
- Hartline JRK (2008) *Incremental Optimization*. Ph.D. thesis, Department of Computer Science, Cornell University, adviser: Dexter Kozen.
- Ibarra OH, Kim CE (1975) Fast approximation algorithms for the knapsack and sum of subset problems. *Journal of the ACM* 22(4):463–468.

- Jiang P, Feldman J (2021) Display optimization under the Multinomial Logit choice model: Balancing revenue and customer satisfaction. Working paper.
- Karp RM (1972) Reducibility among combinatorial problems. *Complexity of computer computations*, 85–103 (Springer).
- Kellerer H, Pferschy U (1999) A new fully polynomial time approximation scheme for the knapsack problem. *Journal of Combinatorial Optimization* 3(1):59–71.
- Kellerer H, Pferschy U, Pisinger D (2004) *Knapsack Problems* (Springer-Verlag Berlin Heidelberg).
- Lawler EL (1979) Fast approximation algorithms for knapsack problems. *Mathematics of Operations Research* 4(4):339–356.
- Magazine MJ, Oguz O (1981) A fully polynomial approximation algorithm for the 0-1 knapsack problem. *European Journal of Operational Research* 8(3):270–273.
- Sharp AM (2007) *Incremental Algorithms: Solving Problems in a Changing World*. Ph.D. thesis, Department of Computer Science, Cornell University, adviser: Dexter Kozen.
- Shmoys DB, Tardos É (1993) An approximation algorithm for the generalized assignment problem. *Mathematical Programming* 62:461–474.
- Toth P, Martello S (1990) *Knapsack Problems: Algorithms and Computer Implementations* (Wiley).
- Ye C (2016) *On the Trade-offs between Modeling Power and Algorithmic Complexity*. Ph.D. thesis, Department of Industrial Engineering and Operations Research, Columbia University, adviser: Jay Sethuraman.

## Appendix A: Additional Proofs

### A.1. Proof of Lemma 2

*Proof of item (a).* Given a pair of utilization vectors,  $N \leq M$ , we show that  $N^{(1)} \leq M^{(1)}$  by considering two cases, depending on whether a given profit class is light or heavy for  $N$ .

*Case 1:*  $\ell \in \text{light}(N)$ . This is the easy scenario, since

$$N_\ell^{(1)} = N_\ell \leq M_\ell \leq M_\ell^{(1)},$$

where the first equality is simply the definition of  $N_\ell^{(1)}$  for light classes and the last inequality is due to having  $M \leq M^{(1)}$ , following the up-rounding operation.

*Case 2:*  $\ell \in \text{heavy}(N)$ . We begin by noting that, since  $N \leq M$ , it follows that  $\ell \in \text{heavy}(M)$  as well. Consequently, by definition of  $N_\ell^{(1)}$  and  $M_\ell^{(1)}$ , we know that these values are precisely:

$$\begin{aligned} N_\ell^{(1)} &= \max \left\{ k \in \left[ \frac{1}{\epsilon} + 1, |\mathcal{P}_\ell| \right] : w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, k \right] \right) \leq \mathcal{E}_\ell^{(N)} \right\}, \\ M_\ell^{(1)} &= \max \left\{ k \in \left[ \frac{1}{\epsilon} + 1, |\mathcal{P}_\ell| \right] : w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, k \right] \right) \leq \mathcal{E}_\ell^{(M)} \right\}. \end{aligned}$$

While these expressions are nearly identical, the estimate we make use of in the  $N_\ell^{(1)}$ -related expression is  $\mathcal{E}_\ell^{(N)} = \mu_\ell^{(N)} \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right]$ , where  $\mu_\ell^{(N)} \geq 1$  is the unique integer for which

$$\left( \mu_\ell^{(N)} - 1 \right) \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] < w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell \right] \right) \leq \mu_\ell^{(N)} \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right].$$

On the other hand, in the  $M_\ell^{(1)}$ -related expression, our estimate is  $\mathcal{E}_\ell^{(M)} = \mu_\ell^{(M)} \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right]$ , where  $\mu_\ell^{(M)} \geq 1$  is the unique integer for which

$$\left( \mu_\ell^{(M)} - 1 \right) \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right] < w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, M_\ell \right] \right) \leq \mu_\ell^{(M)} \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right].$$

Therefore, to show that  $N_\ell^{(1)} \leq M_\ell^{(1)}$ , it remains to prove that  $\mathcal{E}_\ell^{(N)} \leq \mathcal{E}_\ell^{(M)}$ . For this purpose, note that

$$\begin{aligned} \left( \mu_\ell^{(N)} - 1 \right) \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] &< w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell \right] \right) \\ &\leq w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, M_\ell \right] \right) \\ &\leq \mu_\ell^{(M)} \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right], \end{aligned}$$

where the second inequality holds since  $N \leq M$ . By rearranging the bound we have just attained, it follows that  $\mu_\ell^{(N)} - 1 < \mu_\ell^{(M)} \cdot \frac{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right]}{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right]}$ ; for this transition, note that  $\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] > 0$ , as  $W_{\text{heavy}}(N) \geq w(\mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell \right]) \geq \min_{i \in \mathcal{P}_\ell} w_i > 0$ , by our initial assumption that all item weights are strictly positive (see Section 1.2). However, since  $W_{\text{heavy}}(M) \geq W_{\text{heavy}}(N)$ , we know that  $\frac{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right]}{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right]}$  is an integer. Moreover, since  $\mu_\ell^{(N)}$  and  $\mu_\ell^{(M)}$  are integers as well, we necessarily have  $\mu_\ell^{(N)} \leq \mu_\ell^{(M)} \cdot \frac{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(M) \right]}{\text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right]}$ . This inequality is precisely  $\mathcal{E}_\ell^{(N)} \leq \mathcal{E}_\ell^{(M)}$ , by definition of these two estimates.

*Proof of item (b).* To obtain the desired upper bound on  $W_{\text{heavy}}(N^{(1)})$ , note that

$$\begin{aligned} W_{\text{heavy}}(N^{(1)}) &= \sum_{\ell \in \text{heavy}(N^{(1)})} w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell^{(1)} \right] \right) \\ &\leq \sum_{\ell \in \text{heavy}(N)} \mathcal{E}_\ell \\ &\leq \sum_{\ell \in \text{heavy}(N)} \left( w \left( \mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell \right] \right) + \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] \right) \\ &\leq W_{\text{heavy}}(N) + 2\epsilon \cdot \frac{|\text{heavy}(N)|}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \\ &\leq (1 + 2\epsilon) \cdot W_{\text{heavy}}(N). \end{aligned}$$

Here, the first inequality holds since  $\text{heavy}(N^{(1)}) = \text{heavy}(N)$  and since  $w(\mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell^{(1)} \right]) \leq \mathcal{E}_\ell$ , by definition of  $N_\ell^{(1)}$ . The second inequality is obtained by recalling that  $\mathcal{E}_\ell - \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] < w(\mathcal{P}_\ell \left[ \frac{1}{\epsilon} + 1, N_\ell \right])$ .

## A.2. Proof of Lemma 3

*Proof of item (a).* Given a pair of utilization vectors,  $N \leq M$ , we have already shown in Lemma 2(a) that  $N^{(1)} \leq M^{(1)}$ . In order to further argue that  $N^{(2)} \leq M^{(2)}$ , our proof proceeds by considering four cases, depending on whether a given profit class is light or heavy for  $N^{(1)}$  and  $M^{(1)}$ .

1.  $\ell \in \text{light}(N^{(1)})$  and  $\ell \in \text{light}(M^{(1)})$ . In this scenario,

$$N_\ell^{(2)} = N_\ell^{(1)} \leq M_\ell^{(1)} = M_\ell^{(2)},$$

where the first and last equalities follow from how the truncation operation is defined for light profit classes, and the middle inequality holds since  $N^{(1)} \leq M^{(1)}$ .

2.  $\ell \in \text{light}(N^{(1)})$  and  $\ell \in \text{heavy}(M^{(1)})$ . As explained in Section 2.6, the truncation operation is class-preserving, meaning in particular that  $\ell \in \text{light}(N^{(2)})$  and  $\ell \in \text{heavy}(M^{(2)})$ . Therefore,  $N_\ell^{(2)} \leq \frac{1}{\epsilon} < M_\ell^{(2)}$ .
3.  $\ell \in \text{heavy}(N^{(1)})$  and  $\ell \in \text{light}(M^{(1)})$ . This scenario is not possible, since  $N^{(1)} \leq M^{(1)}$ .
4.  $\ell \in \text{heavy}(N^{(1)})$  and  $\ell \in \text{heavy}(M^{(1)})$ . In this case, we begin by recalling that  $N_\ell^{(2)} = N_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell^{(N)} \rceil$  and  $M_\ell^{(2)} = M_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell^{(M)} \rceil$ , where  $\Delta_\ell^{(N)} = N_\ell^{(1)} - \frac{1}{\epsilon}$  and  $\Delta_\ell^{(M)} = M_\ell^{(1)} - \frac{1}{\epsilon}$ , noting that  $\Delta_\ell^{(N)} \leq \Delta_\ell^{(M)}$  as  $N^{(1)} \leq M^{(1)}$ . Now, when  $\Delta_\ell^{(N)} = \Delta_\ell^{(M)}$ , we immediately get  $N_\ell^{(2)} \leq M_\ell^{(2)}$ . In the complementary case, where  $\Delta_\ell^{(N)} \leq \Delta_\ell^{(M)} - 1$ , note that

$$\begin{aligned} M_\ell^{(2)} &= M_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell^{(M)} \rceil \\ &= \frac{1}{\epsilon} + \Delta_\ell^{(M)} - \lceil 2\epsilon \cdot \Delta_\ell^{(M)} \rceil \\ &\geq \frac{1}{\epsilon} + (1 - 2\epsilon) \cdot \Delta_\ell^{(M)} - 1 \\ &\geq \frac{1}{\epsilon} + (1 - 2\epsilon) \cdot (\Delta_\ell^{(N)} + 1) - 1 \\ &\geq \frac{1}{\epsilon} + \Delta_\ell^{(N)} - \lceil 2\epsilon \cdot \Delta_\ell^{(N)} \rceil - 2\epsilon \\ &= N_\ell^{(2)} - 2\epsilon. \end{aligned}$$

Thus, since  $\epsilon < 1/2$ , we indeed have  $N_\ell^{(2)} \leq M_\ell^{(2)}$ .

*Proof of item (b).* To prove that  $w(N^{(2)}) \leq w(N)$ , we first express these quantities as:

$$\begin{aligned} w(N^{(2)}) &= \sum_{\ell \in \text{light}(N^{(2)})} w\left(\mathcal{P}_\ell \left[1, N_\ell^{(2)}\right]\right) + \sum_{\ell \in \text{heavy}(N^{(2)})} w\left(\mathcal{P}_\ell \left[1, \frac{1}{\epsilon}\right]\right) \\ &\quad + \underbrace{\sum_{\ell \in \text{heavy}(N^{(2)})} w\left(\mathcal{P}_\ell \left[\frac{1}{\epsilon} + 1, N_\ell^{(2)}\right]\right)}_{W_{\text{heavy}}(N^{(2)})}, \end{aligned}$$

$$\begin{aligned}
w(N) &= \sum_{\ell \in \text{light}(N)} w(\mathcal{P}_\ell[1, N_\ell]) + \sum_{\ell \in \text{heavy}(N)} w\left(\mathcal{P}_\ell\left[1, \frac{1}{\epsilon}\right]\right) \\
&\quad + \underbrace{\sum_{\ell \in \text{heavy}(N)} w\left(\mathcal{P}_\ell\left[\frac{1}{\epsilon} + 1, N_\ell\right]\right)}_{W_{\text{heavy}}(N)}.
\end{aligned}$$

As explained in Section 2.6, we know that the light and heavy classes of  $N^{(2)}$  and  $N$  are identical, and therefore, the first two summands in each of the above expressions are identical, noting that  $N_\ell^{(2)} = N_\ell$  for every  $\ell \in \text{light}(N^{(2)}) = \text{light}(N)$ . Consequently, it suffices to show that  $W_{\text{heavy}}(N^{(2)}) \leq W_{\text{heavy}}(N)$ .

To this end, let us consider some class  $\ell \in \text{heavy}(N)$ . Recall that  $N_\ell^{(2)} = N_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell \rceil$  and that the internal order between the items in  $\mathcal{P}_\ell$  is increasing by weight. Therefore, by eliminating the  $\lceil 2\epsilon \cdot \Delta_\ell \rceil$  heaviest items picked by  $N_\ell^{(1)}$  from this class, we have deleted a fraction of at least  $\frac{\lceil 2\epsilon \cdot \Delta_\ell \rceil}{\Delta_\ell} \geq 2\epsilon$  of the total weight  $w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell^{(1)}])$ . In other words,  $w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell^{(2)}]) \leq (1 - 2\epsilon) \cdot w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell^{(1)}])$ . Summing this inequality over all heavy classes, it follows that

$$W_{\text{heavy}}(N^{(2)}) \leq (1 - 2\epsilon) \cdot W_{\text{heavy}}(N^{(1)}) \leq (1 - 2\epsilon) \cdot (1 + 2\epsilon) \cdot W_{\text{heavy}}(N) \leq W_{\text{heavy}}(N),$$

where the second inequality follows from Lemma 2(b).

*Proof of item (c).* To prove that  $N_\ell^{(2)} \geq (1 - 2\epsilon) \cdot N_\ell$  for every  $\ell \in \text{heavy}(N)$ , we initially recall that  $N \leq N^{(1)}$ , as shown in Section 2.5. Hence, it suffices to argue that  $N_\ell^{(2)} \geq (1 - 2\epsilon) \cdot N_\ell^{(1)}$ . To this end, since  $N_\ell^{(2)} = N_\ell^{(1)} - \lceil 2\epsilon \cdot \Delta_\ell \rceil$  and  $\Delta_\ell = N_\ell^{(1)} - \frac{1}{\epsilon}$ , the latter inequality follows by observing that

$$\lceil 2\epsilon \cdot \Delta_\ell \rceil \leq 2\epsilon \cdot \Delta_\ell + 1 = 2\epsilon \cdot \left(N_\ell^{(1)} - \frac{1}{\epsilon}\right) + 1 \leq 2\epsilon \cdot N_\ell^{(1)}.$$

### A.3. Proof of Claim 1

We prove the desired bound by contradiction. To this end, suppose that we actually have

$\sum_{\ell \in \text{heavy}(N)} \mu_\ell > \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon}$ . Then,

$$\begin{aligned}
&\sum_{\ell \in \text{heavy}(N)} (\mu_\ell - 1) \cdot \text{Power}_2 \left[ \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \right] \\
&\geq \frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \cdot \sum_{\ell \in \text{heavy}(N)} \mu_\ell - 2\epsilon \cdot \frac{|\text{heavy}(N)|}{|\mathcal{L}|} \cdot W_{\text{heavy}}(N) \\
&\geq \left( \frac{\epsilon}{|\mathcal{L}|} \cdot \sum_{\ell \in \text{heavy}(N)} \mu_\ell - \frac{1}{2} \right) \cdot W_{\text{heavy}}(N) \\
&> W_{\text{heavy}}(N) \\
&= \sum_{\ell \in \text{heavy}(N)} w\left(\mathcal{P}_\ell\left[\frac{1}{\epsilon} + 1, N_\ell\right]\right),
\end{aligned}$$

where the second and third inequalities hold since  $\epsilon \leq 1/4$  and  $\sum_{\ell \in \text{heavy}(N)} \mu_\ell > \frac{3}{2} \cdot \frac{|\mathcal{L}|}{\epsilon}$ , respectively. Therefore, there is at least one index  $\ell \in \text{heavy}(N)$  for which  $(\mu_\ell - 1) \cdot \text{Power}_2[\frac{\epsilon}{|\mathcal{L}|} \cdot W_{\text{heavy}(N)}] > w(\mathcal{P}_\ell[\frac{1}{\epsilon} + 1, N_\ell])$ , contradicting the definition of  $\mu_\ell$  in equation (3).

#### A.4. Proof of Lemma 9

We first note that, by definition,  $\mathcal{S}^{< m}$  is a feasible  $\tau_{m-1}$ -period uncrossing-stars solution to the multi-cluster instance  $\mathcal{I}_{\text{multi}}^{\varphi_{m-1}}[m-1, \ell_{m-1}]$ . In addition,  $\mathcal{S}^m$  is feasible for the single-cluster instance  $\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, w(S_{\tau_{m-1}}^{< m})}[m, \ell_{m-1} + 1, \ell_m]$ , again by definition. Now, to arrive at a contradiction, suppose that at least one of these solutions is sub-optimal for its corresponding instance. Then, there are feasible solutions  $\check{\mathcal{S}}^{< m}$  and  $\check{\mathcal{S}}^m$  to  $\mathcal{I}_{\text{multi}}^{\varphi_{m-1}}[m-1, \ell_{m-1}]$  and  $\mathcal{I}_{\text{single}}^{\varphi_m - \varphi_{m-1}, w(S_{\tau_{m-1}}^{< m})}[m, \ell_{m-1} + 1, \ell_m]$ , respectively, for which

$$w(\check{\mathcal{S}}_{\tau_{m-1}}^{< m}) + w(\check{\mathcal{S}}_{\tau_m}^m) < w(S_{\tau_{m-1}}^{< m}) + w(S_{\tau_m}^m).$$

Therefore, to conclude the proof, it suffices to show that the  $\tau_m$ -period uncrossing-stars solution  $\check{\mathcal{S}}$  obtained by gluing  $\check{\mathcal{S}}^{< m}$  and  $\check{\mathcal{S}}^m$  together is feasible for the multi-cluster instance  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ . This finding contradicts the optimality of  $\mathcal{S}$ , since the inequality above shows that the total weight  $w(\check{\mathcal{S}}_{\tau_m}) = w(\check{\mathcal{S}}_{\tau_{m-1}}^{< m}) + w(\check{\mathcal{S}}_{\tau_m}^m)$  is strictly smaller than the total weight  $w(S_{\tau_m}) = w(S_{\tau_{m-1}}^{< m}) + w(S_{\tau_m}^m)$ . To establish the feasibility of  $\check{\mathcal{S}}$  for  $\mathcal{I}_{\text{multi}}^{\varphi_m}[m, \ell_m]$ , we verify each of the conditions involved:

1. *[1, m]-clustered periods:* By definition, any item introduced by  $\check{\mathcal{S}}$  is either introduced by  $\check{\mathcal{S}}^{< m}$ , implying that this event occurs at a time period belonging to one of the clusters  $\mathcal{C}_1, \dots, \mathcal{C}_{m-1}$ , or by  $\check{\mathcal{S}}^m$ , in which case, this event occurs at a time period in  $\mathcal{C}_m$ .
2. *[0,  $\ell_m$ ]-profit classes:* Along the same lines, all items introduced by  $\check{\mathcal{S}}^{< m}$  reside within the profit classes  $\mathcal{P}_0, \dots, \mathcal{P}_{\ell_{m-1}}$ , whereas those introduced by  $\check{\mathcal{S}}^m$  reside within  $\mathcal{P}_{\ell_{m-1}+1}, \dots, \mathcal{P}_{\ell_m}$ .
3. *Total profit:* Here, the important observation is that, since  $\check{\mathcal{S}}$  introduces precisely the same items as  $\check{\mathcal{S}}^{< m}$  and  $\check{\mathcal{S}}^m$  do, at precisely the same time periods, its profit can be bounded as follows:

$$\Psi(\check{\mathcal{S}}) = \Psi(\check{\mathcal{S}}^{< m}) + \Psi(\check{\mathcal{S}}^m) \geq \varphi_{m-1} + (\varphi_m - \varphi_{m-1}) = \varphi_m.$$

#### A.5. Proof of Claim 2

To obtain the desired bound, we observe that the total profit attained by  $\bar{\mathcal{S}}$  can be decomposed as follows:

$$\Psi(\bar{\mathcal{S}}) = \sum_{m \in [M]} \Psi(\check{\mathcal{S}}^m) \tag{10}$$

$$\geq (1 - \epsilon) \cdot \sum_{m \in [M]} \tilde{\phi}_m \tag{11}$$

$$\geq (1 - \epsilon) \cdot \sum_{m \in [M]} \left( \tilde{\varphi}_m - \left(1 + \frac{\epsilon}{M}\right) \cdot \tilde{\varphi}_{m-1} - \delta \right) \quad (12)$$

$$\geq (1 - \epsilon) \cdot \tilde{\varphi}_M - \frac{\epsilon}{M} \cdot \sum_{m \in [M-1]} \tilde{\varphi}_m - M\delta \quad (13)$$

$$\geq (1 - 2\epsilon) \cdot \tilde{\varphi}_M - M\delta. \quad (14)$$

Equality (10) holds since, by definition of  $\bar{\mathcal{S}}$ , each item introduced by  $\bar{\mathcal{S}}$  appears in exactly one of the single-cluster solutions  $\tilde{\mathcal{S}}^1, \dots, \tilde{\mathcal{S}}^M$ , where it is introduced at precisely the same time period. To derive inequality (11), note that since  $\tilde{\mathcal{S}}^m$  is obtained by applying the approximation scheme proposed in Theorem 3 with respect to the single-cluster instance  $\mathcal{I}_{\text{single}}^{\tilde{\varphi}_m, \tilde{\omega}^m} [m, \tilde{\ell}_{m-1} + 1, \tilde{\ell}_m]$ , we are guaranteed to collect a profit of  $\Psi(\tilde{\mathcal{S}}^m) \geq (1 - \epsilon) \cdot \tilde{\varphi}_m$ . Inequality (12) holds since  $\tilde{\varphi}_m = [(\tilde{\varphi}_m - \tilde{\varphi}_{m-1}) - \zeta(\tilde{\varphi}_{m-1})]^+ = [\tilde{\varphi}_m - (1 + \frac{\epsilon}{M}) \cdot \tilde{\varphi}_{m-1} - \delta]^+$ . Inequality (13) proceeds by noting that  $\tilde{\varphi}_0 = 0$ . Inequality (14) is obtained by observing that  $\tilde{\varphi}_1 \leq \dots \leq \tilde{\varphi}_M$ , as the dynamic program  $\tilde{F}$  is forced to pick each  $\tilde{\varphi}_{m-1}$  out of  $[0, \tilde{\varphi}_m] \cap \Phi$ .

### A.6. Proof of Claim 3

We prove the claim by induction over the discretized state space  $\mathcal{D}$ , following the order by which states are considered in the approximate program  $\tilde{F}$ . The base case corresponds to terminal states  $(m, \ell_m, \varphi_m)$ , for which  $m = 0$  or  $\ell_m = -1$ . Here, we have  $\tilde{F}(m, \ell_m, \varphi_m) = F(m, \ell_m, \varphi_m)$ , according to the terminal conditions of  $F$  and  $\tilde{F}$ . To establish the induction step, consider a state  $(m, \ell_m, \varphi_m) \in \mathcal{D}$  with  $m \geq 1$  and  $\ell_m \geq 0$ . Let  $(\ell_{m-1}^*, \varphi_{m-1}^*)$  be an optimal solution to the minimization problem given by our recursive equation for  $F(m, \ell_m, \varphi_m)$ . Next, we define  $\varphi_{m-1}^\Phi$  as the result of rounding  $\varphi_{m-1}^*$  down to the nearest value in  $\Phi$ , which implies that

$$\varphi_{m-1}^\Phi \leq \varphi_{m-1}^* \leq \left(1 + \frac{\epsilon}{M}\right) \cdot \varphi_{m-1}^\Phi + \delta = \varphi_{m-1}^\Phi + \zeta(\varphi_{m-1}^\Phi), \quad (15)$$

given the definition of  $\Phi$  in (7). Lastly, we define the auxiliary notation  $\phi^\Phi = [(\varphi_m - \varphi_{m-1}^\Phi) - \zeta(\varphi_{m-1}^\Phi)]^+$  and  $\omega^\Phi = \tilde{F}(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi)$ . With these definitions, our recursive equation for  $\tilde{F}$  ensures that

$$\begin{aligned} \tilde{F}(m, \ell_m, \varphi_m) &\leq \tilde{F}(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) + \widetilde{\text{OPT}} \left( \mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi} [m, \ell_{m-1}^* + 1, \ell_m] \right) \\ &\leq F(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) + \widetilde{\text{OPT}} \left( \mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi} [m, \ell_{m-1}^* + 1, \ell_m] \right) \\ &\leq F(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) + \text{OPT} \left( \mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi} [m, \ell_{m-1}^* + 1, \ell_m] \right). \end{aligned} \quad (16)$$

Here, the second inequality proceeds from the induction hypothesis, and the third inequality employs the super-optimality property of the single-cluster solution we obtain for  $\mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi} [m, \ell_{m-1}^* + 1, \ell_m]$ , as accounted for in (8).

On the other hand, letting  $\phi^* = \varphi_m - \varphi_{m-1}^*$  and  $\omega^* = F(m-1, \ell_{m-1}^*, \varphi_{m-1}^*)$ , the optimality of  $(\ell_{m-1}^*, \varphi_{m-1}^*)$  for the minimization problem that defines  $F(m, \ell_m, \varphi_m)$  implies

$$\begin{aligned} F(m, \ell_m, \varphi_m) &= F(m-1, \ell_{m-1}^*, \varphi_{m-1}^*) + \text{OPT}\left(\mathcal{I}_{\text{single}}^{\phi^*, \omega^*}[m, \ell_{m-1}^* + 1, \ell_m]\right) \\ &\geq F(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) + \text{OPT}\left(\mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi}[m, \ell_{m-1}^* + 1, \ell_m]\right). \end{aligned} \quad (17)$$

The inequality above holds since, as we explain below,  $\omega^\Phi \leq \omega^*$  and  $\phi^\Phi \leq \phi^*$ . Therefore, any feasible solution to  $\mathcal{I}_{\text{single}}^{\phi^*, \omega^*}[m, \ell_{m-1}^* + 1, \ell_m]$  is also feasible to  $\mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi}[m, \ell_{m-1}^* + 1, \ell_m]$ , and hence,  $\text{OPT}(\mathcal{I}_{\text{single}}^{\phi^\Phi, \omega^\Phi}[m, \ell_{m-1}^* + 1, \ell_m]) \leq \text{OPT}(\mathcal{I}_{\text{single}}^{\phi^*, \omega^*}[m, \ell_{m-1}^* + 1, \ell_m])$ . Noting that the right hand sides of (16) and (17) are identical, we conclude that  $\tilde{F}(m, \ell_m, \varphi_m) \leq F(m, \ell_m, \varphi_m)$ , as required.

Finally, to see why  $\omega^\Phi \leq \omega^*$  as argued above, note that

$$\omega^\Phi = \tilde{F}(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) \leq F(m-1, \ell_{m-1}^*, \varphi_{m-1}^\Phi) \leq F(m-1, \ell_{m-1}^*, \varphi_{m-1}^*) = \omega^*,$$

where the first inequality follows from the induction hypothesis, and the second inequality holds since  $\varphi_{m-1}^\Phi \leq \varphi_{m-1}^*$  by inequality (15). Similarly, in order to show that  $\phi^\Phi \leq \phi^*$ , the important observation is that

$$\phi^\Phi = [(\varphi_m - \varphi_{m-1}^\Phi) - \zeta(\varphi_{m-1}^\Phi)]^+ \leq [\varphi_m - \varphi_{m-1}^*]^+ = \varphi_m - \varphi_{m-1}^* = \phi^*,$$

where the inequality above holds since  $\varphi_{m-1}^* \leq \varphi_{m-1}^\Phi + \zeta(\varphi_{m-1}^\Phi)$ , by inequality (15).