# Advanced Inventory Management

## Models and Algorithms

Hans Blanc

CentER Graduate Program in Business
Tilburg University
The Netherlands

Course 230261: Advanced Quantitative Logistics

January 7, 2011

# Preface

This syllabus contains the lecture notes for a part of the course Advanced Quantitative Logistics. This course was given as an elective in the OR track of the Graduate Program in Business at Tilburg University, The Netherlands, from 2003 until 2008. The subject of these lecture notes is Advanced Inventory Management. This concerns models and algorithms for inventory management with interaction between items or between stocking points. The prerequisites for this topic consist of models and algorithms for single item, single stocking point inventory systems with constant, time-varying and stochastic demand as discussed in most introductory textbooks on operations research and management science such as Winston [80, Ch. 16, 17, Sect. 20.7] and as summarized in the first chapter of these lecture notes.

# Contents

# Chapter 1

# Introduction to Inventory Management

Inventory management deals with ordering and stock keeping of goods for sale, production or distribution. Inventories are idle goods waiting for use or sale. Inventories are kept in many environments, for instance, in the mining-industry of minerals, in factories of raw materials, parts, work in progress and finished products, and in warehouses, depots and wholesale dealers of goods for distribution, and at shops and by retailers of goods for sale. The main reasons why inventories are held are that it is uneconomical to produce, to handle or to transport units one by one and that consumers often do not accept a delay in the delivery of goods or only want to buy goods that are on display or available in a shop, supermarket or department store. Inventory theory aims to develop models and algorithms as an aid to inventory management.

In Section 1.1 inventory policies will be classified, inventory costs and concepts will be described, and the main sources of uncertainty will be indicated. Section 1.2 contains a review of models and results for single item inventory management. Section 1.3 provides an overview of interactions that may occur in the control of inventories of multiple items and at multiple stocking points.

## 1.1 Definitions and concepts

In this section, general concepts in inventory management will be described and explained. The main sources of uncertainty where inventory management has to deal with are

**Demand:** the demand for items may fluctuate from day to day (due to stochastic behavior at retailers, due to variations in the production plan in a manufacturing environment), from month to month (due to a seasonal pattern) and during the lifetime of a product (an upward trend in the beginning, a downward trend towards the end);

**Lead time:** the total time that elapses between the reorder instant and the instant when goods are ready for use or sale. It consists of the handling time at the supplier (the time required for order picking, packing, and loading), the shipping time from the supplier to the stocking point and the handling time at the stocking point (the time required for unloading, unpacking, and placing on the shelf). When the goods still have to be produced after the reorder instant, it also includes the production time and possibly a set-up time for the production run.

In the practical situation of uncertain (stochastic) demand and nonnegligible lead times stockout occurrences cannot be completely avoided. For customers arriving when an item is out of stock, two cases are often distinguished:

- any demand is backordered and the backlog is filled as soon as a replenishment is delivered; customers are willing to wait if it is difficult to obtain the item elsewhere;
- any demand is lost; customers go elsewhere to buy the item or give up the intention of buying the item.

For some items, part of the demand may be backlogged and part may be lost. The distinction between the two extreme cases becomes less important when stockouts occur more rarely.

A prerequisite for applying inventory policies is a good forecasting method for future demand. A statistical analysis is required of historical demand and lead time data. Further, an ABC classification is often carried out. In most companies, a relatively small percentage (5%–20%) of all items account for a relatively large percentage (55%–65%) of all sales. These are called type A items. Most effort of forecasting and inventory control should be concentrated on this type of items. Further, there is a middle class of items consisting of 20%–30% of all items that account for 20%–40% of all sales. These type B items require less attention and their inventory can often be controlled by standard procedures. Finally, there are type C items consisting of 50%–75% of all items and accounting for only 5%–25% of all sales which are to be controlled by simple and safe procedures. In manufacturing environments, the classification of a raw material item may also be based on how critical the item is for the continuation of work, beside on its value.
The three most important questions to be answered by an inventory policy are

- When to review stocks?
  A distinction is made between
    - periodic review policies where stocks are reviewed at fixed time intervals, the review periods;
    - continuous review policies where stocks are reviewed after each transaction.
- When to order?
  A distinction is made between
    - periodic review policies where orders can only be placed at the periodic review instants;
    - continuous review policies which use reorder points in inventory positions.
- What to order?
  A distinction is made between
    - policies with a fixed order quantity;
    - policies with a fixed order-up-to level.

Next, we will discuss costs that may play a role when ordering and storing goods:

**Ordering cost:** the fixed cost of placing an order; this cost includes the cost of paperwork and accounting associated with an order which is independent of the size of an order; if the item is made internally rather than ordered from an external supplier, this cost is often called set-up cost and includes the cost of labor, material and idle time associated with setting up and shutting down a machine for a production run; if goods are ordered from another location within the same company, this cost may include internal shipping cost.

**Purchasing cost:** the variable cost associated with purchasing a single unit of a good; this cost often includes variable labor cost, variable overhead cost and raw material cost associated with producing of handling a single unit; if goods are ordered from an external supplier, it also includes shipping cost; the external supplier may want to stimulate larger orders to save on shipping cost by offering quantity discounts; these cost only depend on the inventory policy in case of quantity discounts or lost sales.

**Holding cost:** the variable cost of holding a single unit of a good on stock during a unit time period; this cost often includes variable opportunity cost incurred by investing capital in inventory, storage cost, insurance cost, and cost due to possible theft, obsolescence, breakage and spoilage; the opportunity cost is often assumed to be a certain percentage, the so called carrying charge, of the purchasing cost; the carrying charge is strongly related to the interest rate.

**Handling cost:** the cost associated to the handling of goods in a warehouse; as far as this cost is proportional to the number of items handled it does not influence the minimization of the total inventory cost if all demand is satisfied; as far as this cost is proportional to the number of orders handled it can be incorporated in the ordering cost; this cost is important in the design and control of warehouses.

**Shipping cost:** the cost associated to the transport of goods from one stocking point to another; in case of an external supplier, the shipping cost is often included in the purchasing cost.

**Stockout cost:** in case of backlog of demand it is the extra cost associated to the administration and later delivery of goods; in case of lost sales it is the opportunity cost of lost profit on unsatisfied demand; in all cases, it may include a penalty cost for loss of future goodwill; it may also include extra cost for rush orders or overtime work; in many cases, stockout costs are difficult to assess and are therefore replaced by service level constraints (see below).

**Management cost:** the cost incurred by keeping track of inventory levels and by computing order quantities; this cost is usually not included in inventory models but should form an incentive to choose for inventory policies that are simple to implement.

In the stochastic demand models the following two service level constraints will be considered:

**Cycle service constraint:** the probability of no stockout in a reorder cycle must be at least a prescribed probability $\alpha$; the latter probability is called the cycle service level (this constraint is also called $P_1$-criterion);

**Fill rate constraint:** the fraction of the demand that is satisfied directly from stock must be at least a prescribed fraction $\beta$; the latter fraction is called the target fill rate (this constraint is also called $P_2$-criterion).

The following inventory concepts will be used in the various models:

**Inventory on hand:** the number of units actually present at the stocking point; it is also called the physical stock; this quantity plays a role in determining holding costs;

**Net inventory (net stock):** the inventory on hand minus the amount of backlog; this quantity can take positive and negative values;

**Inventory position:** the net stock plus the number of units on order but not yet delivered; this quantity is required for determining a reorder instant;

**Safety stock:** the average inventory position just before a delivery instant; this quantity is used as a protection against uncertainty in demand and against other irregularities like breakage and pilferage; it is related to the service level constraint or the cost of stockouts or losses.

Some general considerations in inventory management are:

- choose a model in agreement with the availability and the reliability of data (according to the general principle "garbage in → garbage out");
- the robustness of a model is important: the resulting replenishment policy should not depend too strongly on the assumptions (like the shape of the demand distribution): this requires sensitivity analysis;
- in practice, quantities are usually rounded off (reorder cycles are chosen in whole days or weeks; order quantities are chosen in packing units): a complicated, time-consuming algorithm for exact optimization then has little use;
- within companies, conflicts in interests or goals may exist between the purchasing department (which strives for quantity discounts and delivery of goods at the beginning of a season) and the logistics department (which has to cope with large quantities at the same time, and which may be saddled with superfluous stocks at the end of a season); such conflicts may be due to the remuneration system of a company;
- uncertainty in demand and in lead times, and desired service levels lead to safety stocks; on the other hand, risk of disappointing demand, technical obsolescence, going out of fashion or decay compel to reservedness toward large stocks.

## 1.2 Review of single-item policies

This section contains a review of inventory policies developed for individual items at a single stocking point. Most of this material can be found in introductory textbooks on operations research or management science, such as Winston [80], and in more specialized textbooks on inventory management, such as Silver et al. [64] and Axsäter [7]. For the case of stochastic demand, see also Tijms [65].

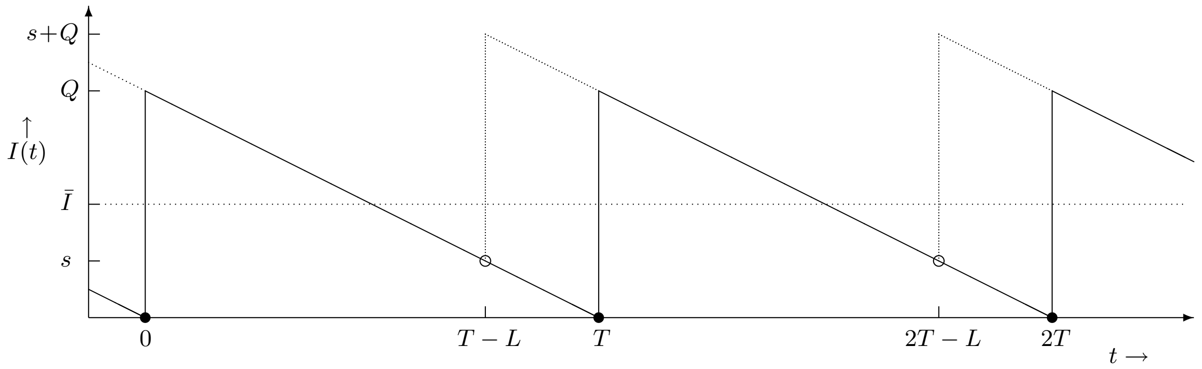

Figure 1.1: Saw-tooth pattern of inventory in deterministic fluid model with lead time $L < T$.

### 1.2.1 The basic EOQ model

The basic inventory model assumes that demand occurs at a constant, known rate of $D$ units per unit of time. The cost of placing an order is $a$ dollars per order. The holding cost is $h$ dollars per unit per unit of time. All demand must be satisfied from stock. As a consequence, the average purchasing cost per unit of time is a policy-independent constant amount of $D$ times the purchasing price of a unit of the good, and can be omitted in the optimization procedure. In the absence of uncertainty in the demand and the lead time, it is optimal to have a zero inventory level at delivery instants. Based on a fluid approximation in which goods flow out of stock at a constant rate $D$, and the stock is instantaneously replenished with order quantities $Q$ at delivery instants, cf. Figure 1.1, the following results can be derived. The relevant inventory cost for a given fixed order quantity $Q$ consists of ordering costs and holding costs. Since the number of orders per unit of time is $D/Q$ and the average inventory level is $\bar{I} = \frac{1}{2}Q$, it follows that the average inventory cost per unit of time, $C(Q)$, is

$$C(Q) = a\,\frac{D}{Q} + h\bar{I} = a\,\frac{D}{Q} + \tfrac{1}{2}hQ. \tag{1.1}$$

The optimal order quantity is found by differentiation of the cost function, cf. Appendix B.1:

$$Q^* = \sqrt{\frac{2aD}{h}}. \tag{1.2}$$

The quantity $Q^*$ is usually referred to as the economic order quantity (EOQ). The reorder cycle (time) $T$, the time between two successive replenishments, for a given order quantity $Q$ follows as

$$T = Q/D. \tag{1.3}$$

The average inventory cost per unit of time can also be formulated as function of the reorder cycle $T$:

$$C(T) = a\,\frac{1}{T} + \tfrac{1}{2}hDT. \tag{1.4}$$

The optimal reorder cycle is

$$T^* = \frac{Q^*}{D} = \sqrt{\frac{2a}{hD}}. \tag{1.5}$$

**Example 1.1** Figure 1.2 shows the average inventory cost per unit of time both as function of the order quantity ($C(Q)$) and as function of the reorder cycle ($C(T)$) for the case $a = \$5$, $D = 500$ and $h = \$2$; here, $Q^* = 50$ and $T^* = 0.1$, while the minimum average inventory cost per unit of time is $C^* = 100$. □

In general, the minimum average inventory cost per unit of time follows by substitution of the optimal order quantity into the cost function:

$$C^* = aD\sqrt{\frac{h}{2aD}} + \tfrac{1}{2}h\sqrt{\frac{2aD}{h}} = \tfrac{1}{2}\sqrt{2ahD} + \tfrac{1}{2}\sqrt{2ahD} = \sqrt{2ahD}. \tag{1.6}$$

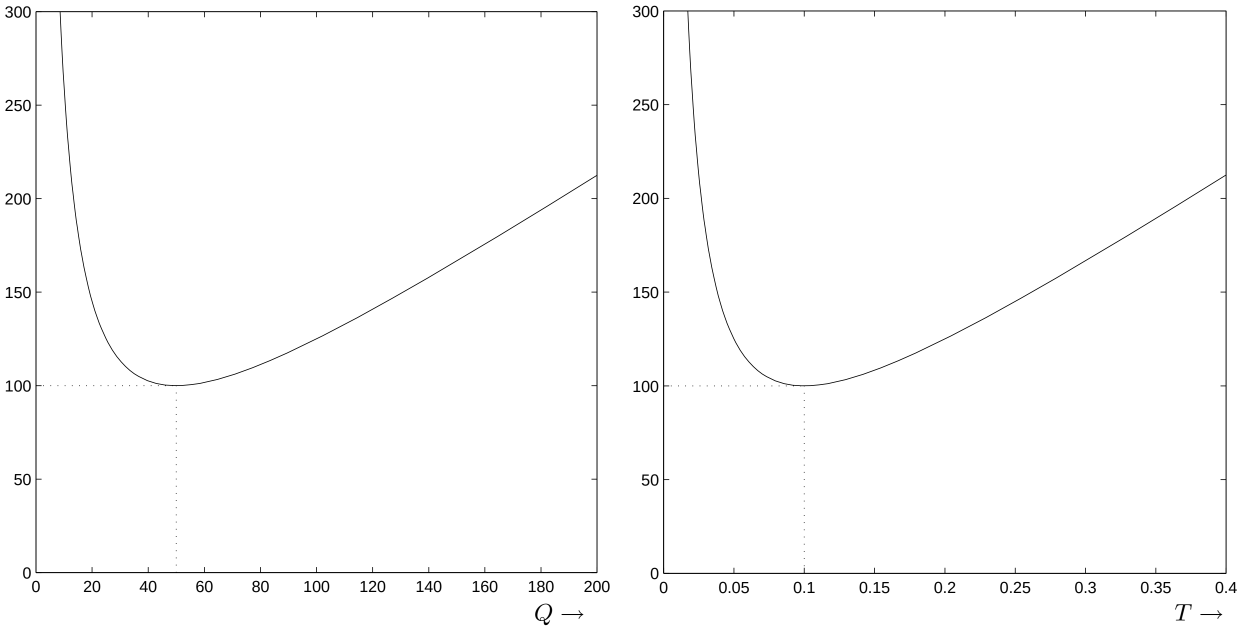

Figure 1.2: Average inventory cost as function of order quantity (left) and of reorder cycle (right).

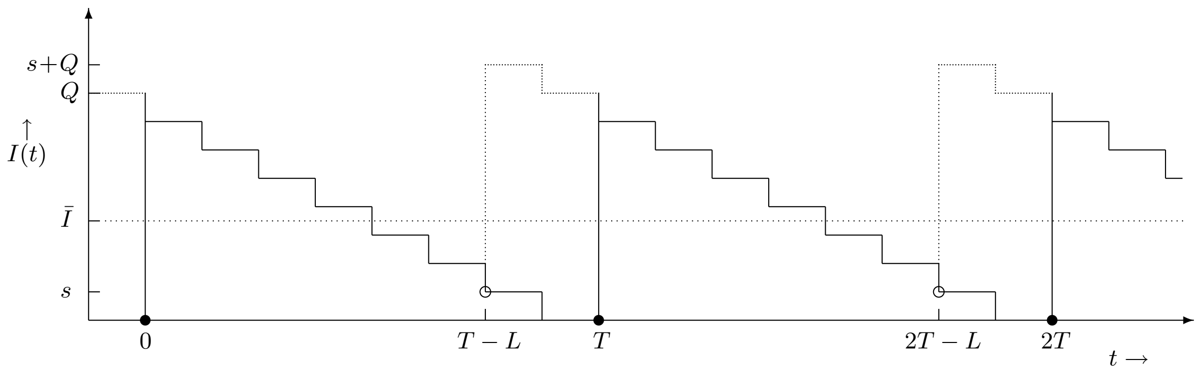

Figure 1.3: Step pattern of inventory in deterministic unit-by-unit model with lead time $L < T$.

Observe that for the optimal order quantity (and the optimal reorder cycle) the average ordering cost is equal to the average holding cost. The optimal number of orders per unit of time is

$$M^* = \frac{D}{Q^*} = \frac{1}{T^*} = \sqrt{\frac{hD}{2a}}. \tag{1.7}$$

In general, $Q^*$, $T^*$, and $M^*$ are not whole numbers. Since they are interrelated, it is usually not possible to round them all to integer values. A way to find the optimal integer order quantity is by determining the value $Q$ for which the cost $C(Q)$ is equal to the cost $C(Q+1)$, cf. (1.1). The equation

$$a\,\frac{D}{Q} + \tfrac{1}{2}hQ = a\,\frac{D}{Q+1} + \tfrac{1}{2}h(Q+1),$$

reduces to

$$Q(Q+1) = 2aD/h.$$

Similarly, the equation $C(Q) = C(Q-1)$ implies $Q(Q-1) = 2aD/h$. This means that the integer order quantity $\tilde{Q}$ is optimal if

$$\tilde{Q}(\tilde{Q}-1) < 2aD/h < \tilde{Q}(\tilde{Q}+1) \quad \text{or} \quad \sqrt{\tilde{Q}(\tilde{Q}-1)} < Q^* < \sqrt{\tilde{Q}(\tilde{Q}+1)}. \tag{1.8}$$

In this way, the interval $0 < Q^* < \sqrt{2}$ is rounded to $\tilde{Q} = 1$, the interval $\sqrt{2} < Q^* < \sqrt{6}$ to $\tilde{Q} = 2$, the interval $\sqrt{6} < Q^* < \sqrt{12}$ to $\tilde{Q} = 3$, etcetera. This way of rounding corrects for the skewness of the cost function near

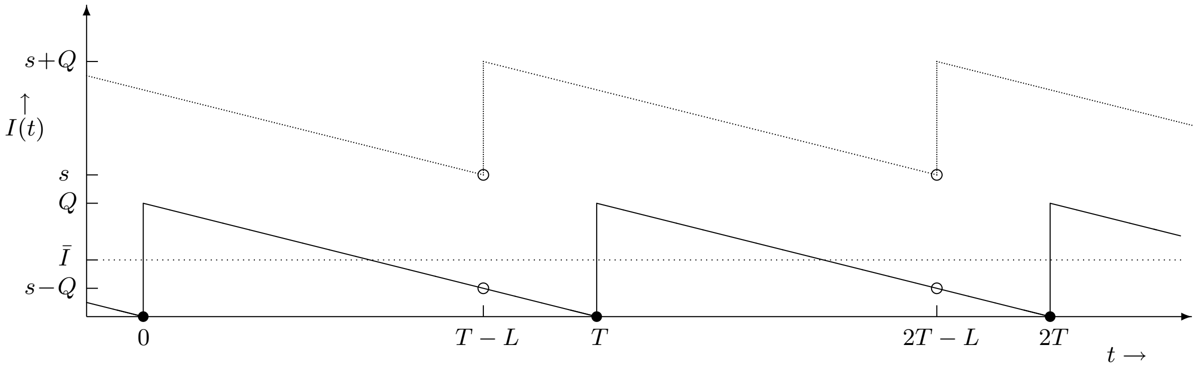

Figure 1.4: Saw-tooth pattern of inventory in deterministic fluid model with lead time $T < L < 2T$.

its minimum. It should be noted that the smooth inventory pattern in Figure 1.1 is in fact an approximation for the step pattern for discrete units as shown in Figure 1.3. The fluid approximation is better according as larger quantities are involved (fast movers). For items that are sold or used in small quantities (slow movers) the discrete nature of the demand and the inventory pattern should be taken into consideration.

A useful property of the EOQ formula is that the cost function (1.1) is rather flat near its minimum in many cases, cf. Figure 1.2. This implies that the optimal order quantity $Q^*$ is rather insensitive with respect to the cost parameters.

The above derivations do not involve a possible lead time $L$ between the instant at which an order is placed and the instant at which the order is delivered. If the lead time $L$ is constant, the optimal values $Q^*$, $T^*$ and $M^*$ are not affected by this lead time. The lead time $L$ only determines the instant when an order has to be placed, namely, $L$ units of time before the order is needed. The demand during the lead time is $D \cdot L$. Since the order has to be delivered at the instant when the inventory on hand reaches the level 0 in deterministic systems, the inventory position at reorder instants should be equal to $D \cdot L$. The inventory position at the reorder instant is called the reorder point, and will be denoted by $s$. In deterministic systems, it is seen that

$$s = D \cdot L. \tag{1.9}$$

Figure 1.1 includes the reorder instants and the reorder point for the case that the lead time $L$ is smaller than the reorder cycle $T$. In this case, the inventory position (dotted line) only differs from the inventory on hand on the time intervals $(mT - L, mT)$, $m = 1, 2, \ldots$. Figure 1.4 shows the reorder instants and the reorder point for the case $T < L < 2T$. In this case, there is always at least one order in transit so that the inventory position is higher than the inventory on hand at every instant. In these and later figures, ○ marks the inventory level at a reorder instant and • marks the inventory level just before a delivery instant.

### 1.2.2 The production lot size model

Consider the following variant of the basic EOQ model in which units are not purchased from an external supplier but internally manufactured. Let $a$ denote the set-up cost of starting a production run, $h$ the holding cost per unit per unit of time, $D$ the demand rate and $p$ the production rate, that is, the number of units that can be produced per unit of time. It is assumed that $p > D$. In a production environment the way in which items or substances become available to satisfy demand may affect the optimal lot size. If the produced items become available as a batch at the end of a production run, the optimal lot size is determined by the optimal order quantity (1.2). As lead time one has to take $L = u + Q^*/p$ with $u$ a possible set-up time for a production run. The reorder point then becomes $s = DL = Du + DQ^*/p$, cf. (1.9).

However, if the produced units become available for satisfying demand at a rate $p$ during the production run, the start of the production can be delayed until the instant at which the inventory reaches the level $s = Du$. In this case, the maximum inventory is not $Q$ but the inventory level at the end of the production run. To produce $Q$ units takes a time $\tau = Q/p$, and during this time the inventory increases at a net rate of $p - D$ up to, cf. Figure 1.5,

$$I_{\max} = (p - D)\tau = (p - D)Q/p. \tag{1.10}$$

Hence, the average inventory level is $\bar{I} = \frac{1}{2}(p - D)Q/p$, and the cost function becomes

$$C(Q) = a\,\frac{D}{Q} + \tfrac{1}{2}hQ(1 - D/p). \tag{1.11}$$

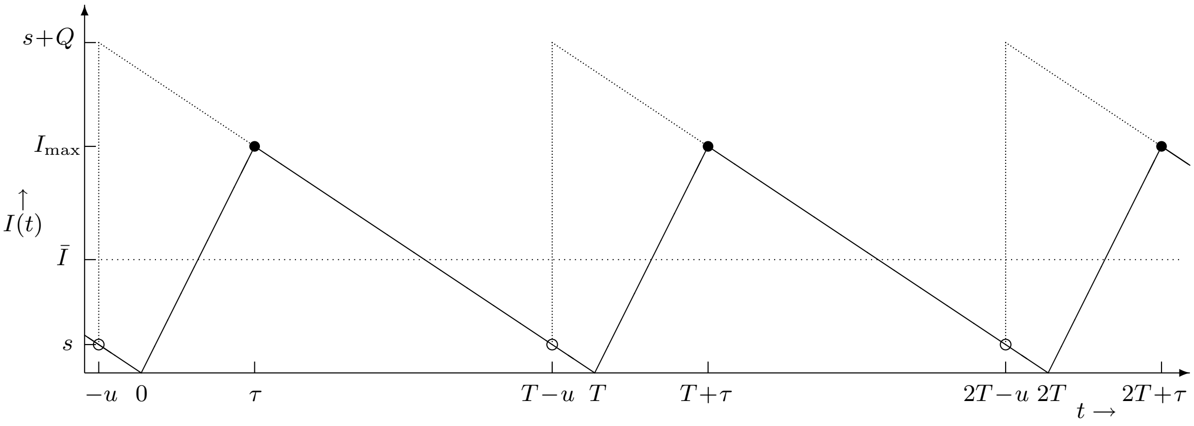

Figure 1.5: Triangular pattern of inventory on hand in deterministic production lot-size model.

The optimal production lot size becomes

$$Q^* = \sqrt{\frac{2aD}{h(1-D/p)}}, \tag{1.12}$$

the optimal production cycle length is

$$T^* = \frac{Q^*}{D} = \sqrt{\frac{2a}{h(1-D/p)D}}, \tag{1.13}$$

and the minimum average cost per unit of time is

$$C^* = \sqrt{2ah(1-D/p)D}. \tag{1.14}$$

In fact, all results of the model with batch delivery can be translated to the model with gradual delivery by replacing $h$ by $h(1-D/p)$. Figure 1.5 shows the triangular pattern of the inventory on hand and the saw-tooth pattern of the inventory position for a system with gradual delivery.

### 1.2.3 The EOQ model with quantity discounts

Some suppliers offer quantity discounts on purchasing prices to their customers to prevent the handling, packing and shipment of small quantities or to stimulate larger sales. In this situation, it is assumed that the holding cost is proportional to the purchasing price $v$, that is, $h = rv$, with $r$ a carrying charge. Further, suppose that there exists one or more ($\ell$) price break points $V_j$, $V_1 \geq V_2 \geq \cdots \geq V_\ell$, and discount factors $d_j$, $d_1 \geq d_2 \geq \cdots \geq d_\ell$, such that a discount factor $d_j$ is awarded if the value of an order without discount exceeds the price break point $V_j$. Let $v_0$ denote the basic price per unit. Then, the price per unit is

$$\begin{aligned} &v(Q) = v_0, && \text{if } v_0 Q < V_\ell, \\ &v(Q) = v_0(1-d_j), && \text{if } V_j \leq v_0 Q < V_{j-1}, \quad j = 2, \ldots, \ell, \\ &v(Q) = v_0(1-d_1), && \text{if } v_0 Q \geq V_1. \end{aligned} \tag{1.15}$$

The cost function becomes, including the variable purchasing cost,

$$C(Q) = a\,\frac{D}{Q} + \tfrac{1}{2} r v(Q) Q + v(Q) D. \tag{1.16}$$

**Algorithm 1.1** [EOQ with quantity discount]

Step 1: Start with the highest discount factor: $m = 1$.

Step 2: Determine the optimal order quantity $Q^*_m$ according to (1.2) with $h = rv_0(1-d_m)$. If $Q^*_m$ is feasible ($Q^*_m \geq V_m/v_0$), compute $C(Q^*_m)$ and goto Step 3; otherwise, compute $C(V_m/v_0)$, decrease $m$ by 1 and repeat Step 2.

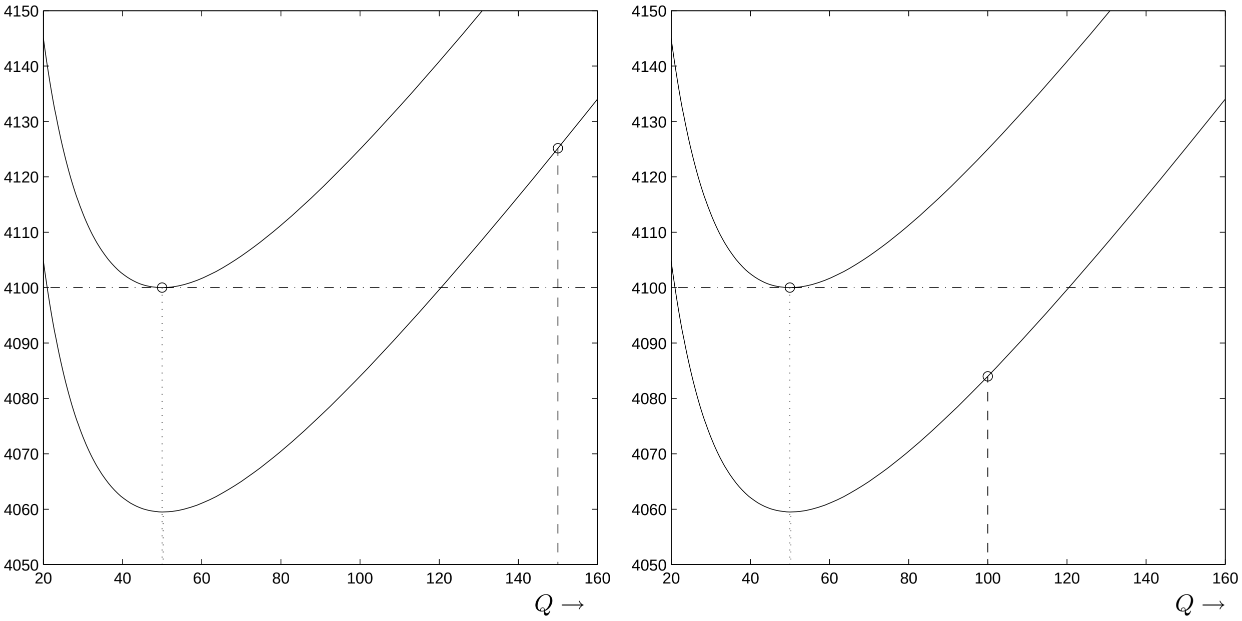

Figure 1.6: Average inventory cost as function of order quantity with and without discount.

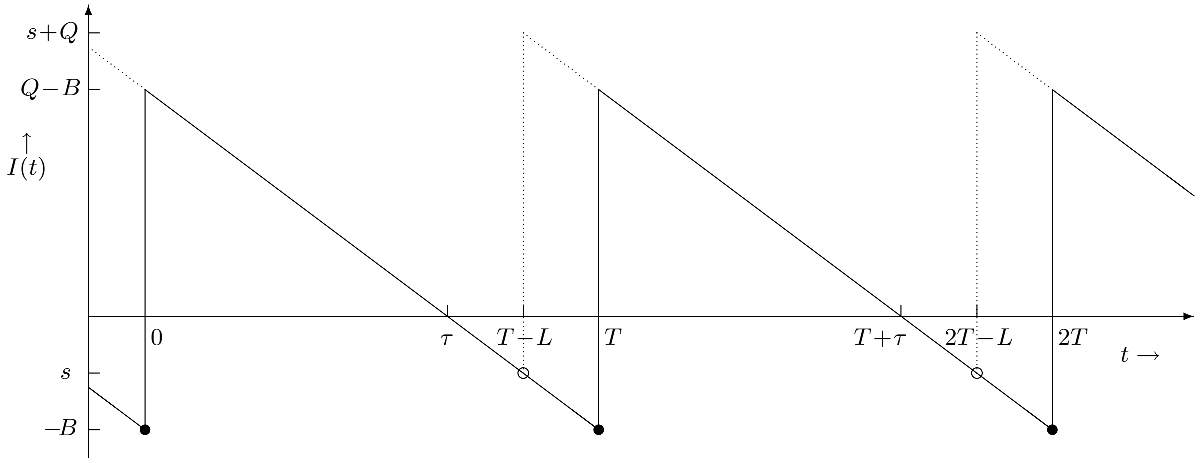

Figure 1.7: Saw-tooth pattern of net inventory in deterministic fluid model with planned stockouts.

Step 3: The optimal order quantity is that quantity from the set $\{Q_m^*, V_{m+1}/v_0, \ldots, V_1/v_0\}$ with the lowest cost.

Here, take $d_{\ell+1} \doteq 0$ and note that $Q_{\ell+1}^*$ is feasible in any case. □

**Example 1.2** Consider an item with the following data: $a = \$5$, $D = 500$, $v_0 = \$8$ and $r = \$0.25$ so that $h = \$2$. Suppose that the supplier offers a discount price of $v_1 = \$7.92$ for order quantities of 150 units or more. This means in the above notations: $\ell = 1$, $d_1 = 0.01$ and $V_1 = 1200$. Figure 1.6 (left) shows the average inventory cost per unit of time with (lower graph) and without (upper graph) discount. According to Algorithm 1.1 first the optimal order quantity with the discount price is computed: $Q_1^* = 50.25$. This order quantity is not feasible. Hence, the average cost at $V_1/v_0$ is computed: $C(150) = \$4{,}125.17$. Then, the optimal order quantity without discount is computed: $Q_2^* = 50.00$, with corresponding average cost $C(50) = \$4{,}100.00$. Hence, it is optimal to order 50 units and ignore the discount offer. However, if the discount price would be offered for order quantities of 100 units or more, the quantity $Q_1^* = 50.25$ is still not feasible, but since $C(100) = \$4{,}084.00$ it is optimal to order 100 units and accept the discount offer, cf. Figure 1.6 (right). □

### 1.2.4 The EOQ model with planned stockouts

In this section we consider the variant of the basic EOQ model in which stockouts are allowed. It is assumed that all stockouts are backlogged instantaneously at a delivery instant. The accumulated amount of backlog at a delivery instant will be denoted by $B$, and will be treated as a decision variable beside the order quantity $Q$. It will be assumed that $B \leq Q$ (it is readily verified that this is optimal). The parameters $a$, $h$ and $D$ have the same meaning as in Section 1.2.1. Further, $b$ will represent the stockout cost per unit short per unit of time. Figure 1.7 shows the saw-tooth pattern of the net inventory as function of time. When the net inventory is positive it represents the inventory on hand. The inventory on hand is zero whenever stockout occurs; then, the net inventory is negative and its absolute value represents the accumulated amount of backlog. In this model, the reorder point $s$ can be positive or negative. It is given by, cf. (1.9),

$$s = D \cdot L - B. \tag{1.17}$$

Next, the average cost per unit of time will be determined for this model. The number of orders per unit of time is $D/Q$ and the reorder cycle is $T = Q/D$, as in the basic EOQ model. During a time $\tau$ the net inventory is positive. This time is determined by $D\tau = Q - B$. Hence, a fraction $\tau/T$ of each cycle the net inventory is positive, and the average level is $\frac{1}{2}(Q - B)$. A fraction $\tau/T$ of each cycle the net inventory is negative, and the average backlog is $\frac{1}{2}B$. Hence, the average ordering, holding and stockout cost per unit of time is

$$C(Q, B) = a\,\frac{D}{Q} + \tfrac{1}{2}h(Q - B)\,\frac{\tau}{T} + \tfrac{1}{2}bB\,\frac{T - \tau}{T}.$$

Since $\tau/T = (Q - B)/Q$, the average cost per unit of time can be expressed in the decision variables $Q$ and $B$:

$$C(Q, B) = a\,\frac{D}{Q} + \tfrac{1}{2}h\,\frac{(Q - B)^2}{Q} + \tfrac{1}{2}b\,\frac{B^2}{Q}. \tag{1.18}$$

The partial derivative of this cost function with respect to $B$ vanishes if $h(Q - B) = bB$. This means that the optimal order quantity for a fixed total backlog per cycle is

$$\hat{Q}(B) = \frac{h + b}{h}\,B, \tag{1.19}$$

with corresponding average cost per unit of time

$$\hat{C}(B) = \frac{ah}{h + b}\,\frac{D}{B} + \tfrac{1}{2}bB. \tag{1.20}$$

From this function it readily follows that the optimal backlog per cycle, the optimal order quantity and the minimum average cost per unit of time are:

$$B^* = \sqrt{\frac{2ahD}{b(h + b)}}, \qquad Q^* = \sqrt{\frac{2aD(h + b)}{hb}}, \qquad C^* = \sqrt{\frac{2ahbD}{(h + b)}}. \tag{1.21}$$

The actual fill rate, cf. Section 1.1, for the optimal policy is

$$\Psi = 1 - \frac{B^*}{Q^*} = 1 - \frac{h}{h + b} = \frac{b}{h + b}. \tag{1.22}$$

Alternatively, this implies that a target fill rate $\beta$ corresponds to a stockout cost of

$$b = \frac{\beta h}{1 - \beta}. \tag{1.23}$$

In the limit as $b \to \infty$, that is, if the stockout cost becomes very high, it follows from the above that $B^* \to 0$, $Q^* \to \sqrt{2aD/h}$ and $C^* \to \sqrt{2ahD}$, which represents the basic EOQ policy, cf. Section 1.2.1.

### 1.2.5 Time-varying demand

In this section we consider situations in which demand fluctuates over time, but is known in advance for a number of future periods (days, weeks, months). Time-varying demand may be due to trends or seasonal patterns, and may also arise as a consequence of material requirement planning in a manufacturing environment.

In the context of time-varying demand a discrete time model is considered with a finite planning horizon $H$ (a number of periods). Here, a single item is treated in isolation. The demand in period $t$ is denoted as $D_t$, $t = 1, \ldots, H$. The cost of placing an order is $a$ dollar per order. Orders are delivered at the beginning of a period. The demand for that period is set aside, and holding costs of $h$ dollar per unit per period are only charged over the excess inventory. All demand must be fulfilled and no back orders are allowed. The decision variables are $Q_t$, the order quantity for the beginning of period $t$ ($t = 1, \ldots, H$). The inventory level at the end of period $t$ is denoted by $I_t$ ($t = 0, \ldots, H$). It is assumed that the initial inventory level is zero: $I_0 = 0$. The inventory levels can be described by the recursive equations

$$I_t = I_{t-1} + Q_t - D_t, \quad t = 1, \ldots, H. \tag{1.24}$$

The objective is to minimize the total ordering and holding costs over $H$ periods:

$$C(Q_1, \ldots, Q_H) = \sum_{t=1}^{H} \left[ a\delta(Q_t) + hI_t \right], \tag{1.25}$$

here, $\delta(.)$ is a dummy function: $\delta(0) = 0$, $\delta(q) = 1$ if $q > 0$. The order quantities are restricted by the requirements that $I_t \geq 0$, $t = 1, \ldots, H$. The minimization of the cost function (1.25) is facilitated by the observation that the optimal order quantities are zero or equal to the demand of a whole number of consecutive periods:

$$Q_t = 0 \quad \text{or} \quad Q_t = D_t + \cdots + D_{t+\ell}, \quad \ell = 0, 1, \ldots, H - t, \quad \text{for} \quad t = 1, \ldots, H. \tag{1.26}$$

This property follows by noting that to order only a fraction of the demand of the last period $t + \ell$ is not optimal because it forces an order to be placed for the beginning of period $t + \ell$. Shifting the fraction of the demand $D_{t+\ell}$ from the order quantity $Q_t$ to the order quantity $Q_{t+\ell}$ does not change the ordering costs but diminishes the holding costs. A further restriction on the possible optimal values of $Q_t$ is obtained by the observation that it is not optimal to include the demand $D_{t+\ell}$ in the order of period $t$ if $h\ell D_{t+\ell} > a$, because the costs can then be reduced by placing an additional order in period $t + \ell$.

Wagner & Whitin [77] have developed a dynamic programming solution for this model. Define, for $t = 1, \ldots, H$,

- $f(t)$: the minimum cost incurred during the periods $t, \ldots, H$, given that the inventory level is zero at the end of period $t - 1$;
- $C(t, \ell)$: the cost of ordering at the beginning of period $t$ a quantity that is sufficient to fulfill the demand of the periods $t, \ldots, t + \ell$:

$$C(t, \ell) = a + h \sum_{u=1}^{\ell} u D_{t+u}, \quad \ell = 0, 1, \ldots, H - t, \quad t = 1, \ldots, H. \tag{1.27}$$

**Algorithm 1.2** [Wagner-Whitin]

Step 1: Set $f(H + 1) = 0$ and start with $t = H$.

Step 2: Apply the backward recursion

$$f(t) = \min_{0 \leq \ell \leq H-t} \{C(t, \ell) + f(t + \ell + 1)\}. \tag{1.28}$$

Step 3: If $t > 1$ decrease $t$ by 1 and repeat from Step 2; otherwise, stop.

The minimum total ordering and holding costs is $f(1)$; the optimal order quantities have to be recovered from the intermediate results of the recursion: the minimizing $\ell$ in (1.28) indicates the optimal number of additional periods for which to order if $t$ becomes a reorder instant. □

The computation time of the Wagner-Whitin algorithm strongly increases with the number of periods $H$. In some cases, the optimal order quantity for the first period is strongly influenced by variations in the perhaps still uncertain demand in the last period. To overcome these difficulties, Silver & Meal [62] have developed a heuristic for this model. They propose a forward recursion. Let $\tau$ be a reorder instant at which an order is placed for the $\ell$ periods $\tau, \ldots, \tau+\ell-1$. Then, the average cost over these periods is

$$\bar{C}(\tau,\ell) = \frac{1}{\ell}\left[a + h\sum_{u=1}^{\ell-1} uD_{\tau+u}\right], \quad \ell = 1,\ldots,H+1-\tau, \quad \tau = 1,\ldots,H. \tag{1.29}$$

The heuristic proposes to place the next order just after a local minimum of the average cost since the previous reorder instant has been reached. The order quantity for the first period only depends on the demand in the periods up to the period in which such a local minimum occurs.

**Algorithm 1.3** [Silver-Meal heuristic]

Step 1: Start with $\tau = 1$.

Step 2: Compute the average costs $\bar{C}(\tau,\ell)$ for $\ell = 1,2,\ldots$, until $\bar{C}(\tau,\ell^*+1) > \bar{C}(\tau,\ell^*)$ for some $\ell^* \geq 1$, or until $\tau+\ell^* = H+1$. Take $Q_\tau = D_\tau + \cdots + D_{\tau+\ell^*-1}$ as the order quantity for period $\tau$.

Step 3: Take $\tau+\ell^*$ as the new value of $\tau$ (the new reorder instant). Repeat from Step 2 if $\tau \leq H$; otherwise, stop.

The average cost $\bar{C}(\tau,\ell)$ is not computed if $D_{\tau+\ell-1} = 0$. Then, $\ell$ is increased until a period with nonzero demand is found. The new average cost is then compared to the previously computed average cost. □

**Remark 1.1** The Silver-Meal heuristic has been tested extensively, see Baker [9]. In many cases, the total cost of the generated order policy is less than 1% above the minimum. The heuristic works less well when demand strongly diminishes in later periods and when there are many periods with zero demand, cf. Silver et al. [64, Sct. 6.6.7]. Silver & Miltenburg [63] have developed a modification of the Silver-Meal heuristic for such cases. □

Another heuristic for this model is the part-period-balancing algorithm, cf. DeMatteis [22], Mendoza [52]. Let again $\tau$ be a reorder instant at which an order is placed for the periods $\tau,\ldots,\tau+\ell$. The cumulative holding cost over these periods is

$$C_h(\tau,\ell) = h\sum_{u=1}^{\ell} uD_{\tau+u}, \quad \ell = 0,1,\ldots,H-\tau, \quad \tau = 1,\ldots,H. \tag{1.30}$$

This heuristic tries to compensate ordering cost against future holding costs. A new order is placed as soon as the cumulative holding cost since the previous reorder instant exceed the ordering cost.

**Algorithm 1.4** [Part-period-balancing]

Step 1: Start with $\tau = 1$.

Step 2: Compute the cumulative holding costs $C_h(\tau,\ell)$ for $\ell = 1,2,\ldots$, until $C_h(\tau,\ell) > a$ for some $\ell$, or until $\tau+\ell = H+1$. Take $Q_\tau = D_\tau + \cdots + D_{\tau+\ell-1}$ as the order quantity for period $\tau$.

Step 3: Take $\tau+\ell$ as the new value of $\tau$ (the new reorder instant). Repeat from Step 2 if $\tau \leq H$; otherwise, stop.

If the foregoing procedure generates an order in period $H$, the solution can be improved by shifting the quantity $D_H$ to the previous reorder instant $\tau_p$ if $h(H-\tau_p)D_H < a$. □

**Remark 1.2** The above algorithms do not take a possible lead time into account. If there is a deterministic lead time $L > 0$ the algorithms should be applied a time interval $L$ before the beginning of the first period in order that the first quantity $Q_1$ is delivered in time. □

**Remark 1.3** The above algorithms assume that the initial stock is zero. If the initial stock is nonzero, say $I_0 > 0$, then the demand of the first few periods have to be diminished by the initial stock before the algorithms can be applied with the modified demand. For instance, if $D_1 \leq I_0 < D_1 + D_2$, the demand of the first period can be completely satisfied by the initial stock (and, hence, $Q_1 = 0$), while the modified demand for the second period becomes $\tilde{D}_2 = D_2 - (I_0 - D_1) > 0$. The algorithms are then applied from period 2 onwards. □

**Remark 1.4** The foregoing algorithms can all be applied with a rolling horizon. This may be useful when the demand in periods $2, \ldots, H$ is still subject to change, or if demand information over new periods $H+1, H+2, \ldots$ becomes available. In the first case, problems occur when the demand changes for periods for which the demand has already been ordered. In general, the solutions of the models with deterministic time-varying demand are very sensitive to small irregularities (breakage, loss) and upward changes in demand, and a safety stock may be required to prevent stockouts under such circumstances. In the latter case, a new solution with a new horizon is preferably determined at the first reorder instant after period 1. Blackburn & Millen [12] have found that the Silver-Meal heuristic may outperform the Wagner-Whitin algorithm in a rolling horizon environment because the solution found by the latter method is too sensitive to changes. ☐

### 1.2.6 Stochastic demand, continuous review

This section is concerned with inventory systems with stationary stochastic demand and continuous review policies. As in Section 1.2.1, $a$ denotes the ordering cost per order and $h$ denotes the holding cost per unit per unit of time. Lead times are assumed to be constant, and equal to $L$ units of time. Further, it is assumed that each customer demands one unit and that all demand occurring when the item is out of stock is backlogged. The expected demand per unit of time will be denoted by $E\{D\}$ and the standard deviation of the demand per unit of time by $\sigma\{D\}$. The most commonly used continuous review policies are:

- the $(s, Q)$ policy: whenever the inventory position drops at or below the reorder point $s$ a fixed quantity $Q$ is ordered;
- the $(s, S)$ policy: whenever the inventory position drops at or below the reorder point $s$ an order is placed of a size that brings the inventory position to the order-up-to level $S$. Policies with an order-up-to level are also called base-stock policies.

In case of unit demand per customer, the $(s, Q)$ and $(s, S)$ policies are equivalent when $Q = S - s$. In the analysis of continuous review policies an important role is played by the demand during the lead time. This random variable will be denoted by $D_L$. With constant lead times, the mean and the variance of $D_L$ are given by

$$E\{D_L\} = LE\{D\}, \qquad \sigma^2\{D_L\} = L\sigma^2\{D\}. \tag{1.31}$$

The safety stock, defined as $s - E\{D_L\}$, is needed to protect against the uncertainty in the demand during the lead time. In case of a fill rate constraint with target $\beta$, cf. Section 1.1, the reorder point $s$ has to be chosen such that the actual fill rate $\Psi$ exceeds the target fill rate:

$$\Psi = 1 - \frac{E\{B\}}{Q} = 1 - \frac{E\{[D_L - s]^+\} - E\{[D_L - S]^+\}}{S - s} \geq \beta; \tag{1.32}$$

here, the random variable $B$ denotes the amount of backlog accumulated during a reorder cycle, $[x]^+ \doteq \max\{0, x\}$ for all real $x$, $[D_L - s]^+$ is the amount of backlog at the end of a reorder cycle and $[D_L - S]^+$ is the amount of backlog at the beginning of a reorder cycle. By a balance argument, the expected demand during a reorder cycle is equal to $Q = S - s$. Since the expected number of orders per unit of time is $E\{D\}/Q$ and since the expected inventory on hand just before a delivery instant is $E\{[s - D_L]^+$ and just after a delivery instant is $E\{[S - D_L]^+$, the expected average cost per unit of time is

$$E\{C(s, S)\} = \frac{aE\{D\}}{S - s} + \tfrac{1}{2}h[E\{[s - D_L]^+\} + E\{[S - D_L]^+\}]. \tag{1.33}$$

Although the decision variables $s$ and $S$ should be simultaneously determined to minimize $E\{C(s, S)\}$ subject to the fill rate constraint (1.32), it is common practice to follow a sequential approach in which first the order quantity is determined according to the EOQ formula, that is,

$$Q = S - s = \sqrt{\frac{2aE\{D\}}{h}}, \tag{1.34}$$

usually rounded to the nearest integer, and then the reorder point $s$ is determined as the smallest integer such that, cf. (1.32),

$$E\{[D_L - s]^+\} - E\{[D_L - S]^+\} \leq (1 - \beta)Q = (1 - \beta)\sqrt{\frac{2aE\{D\}}{h}}. \tag{1.35}$$

This sequential approach performs well if $\sigma\{D_L\} < Q$, cf. Tijms [65]. The second term in the fill rate constraint, $E\{[D_L - S]^+\}$, is negligible if the target fill rate is large ($\beta > 0.9$) and the coefficient of variation of the demand during the lead time $D_L$ is small ($\sigma\{D_L\}/E\{D_L\} < 0.5$). Zheng & Federgruen [82] describe an efficient algorithm to search the optimal values of $s$ and $S$ simultaneously. See also Federgruen & Zipkin [29] and Axsäter [7, Ch. 5,6].

**Example 1.3** In the case that unit demand occurs according to a Poisson process at a rate of $\lambda = E\{D\}$ customers per unit of time, the random variable $D_L$ has a Poisson distribution with parameter $\lambda L$, cf. Appendix A.2.3. The constraint (1.35) then reads

$$\left[\sum_{j=s}^{\infty}(j-s)\frac{(\lambda L)^j}{j!} - \sum_{j=S}^{\infty}(j-S)\frac{(\lambda L)^j}{j!}\right] e^{-\lambda L} \le (1-\beta)\sqrt{\frac{2a\lambda}{h}}.$$

The infinite sums can be computed by the relation

$$\sum_{j=m}^{\infty}(j-m)\frac{(\lambda L)^j}{j!}\,e^{-\lambda L} = \lambda L - m + \sum_{j=0}^{m-1}(m-j)\frac{(\lambda L)^j}{j!}\,e^{-\lambda L}, \quad m = 0,1,2,\dots.$$

The expected average cost per unit of time (1.33) becomes

$$E\{C(s,S)\} = \frac{a\lambda}{S-s} + \tfrac{1}{2}h\left[\sum_{j=0}^{s}(s-j)\frac{(\lambda L)^j}{j!} + \sum_{j=0}^{S}(S-j)\frac{(\lambda L)^j}{j!}\right] e^{-\lambda L}.$$

Consider the following numerical example: $a = \$5$, $h = \$0.05$ per unit per day, and a Poisson demand rate of $\lambda = 10$ units per day. This implies that $Q = 44.72$ by (1.34), with corresponding reorder cycle $T = 4.47$.

Table 1.1: Feasible reorder points and order-up-to levels for $L = 5$, $\beta = 0.98$.

| $s$ | $S$ | $Q$ | $E\{[S-D_L]^+\}$ | $E\{[s-D_L]^+\}$ | $E\{[D_L-S]^+\}$ | $E\{[D_L-s]^+\}$ | $E\{C(s,S)\}$ | $\Psi$ |
|---|---|---|---|---|---|---|---|---|
| 56 | 100 | 44 | 50.00 | 6.82 | 0.00 | 0.82 | 2.5567 | 0.9815 |
| 56 | 101 | 45 | 51.00 | 6.82 | 0.00 | 0.82 | 2.5565 | 0.9819 |
| 56 | 102 | 46 | 52.00 | 6.82 | 0.00 | 0.82 | 2.5573 | 0.9823 |
| 56 | 103 | 47 | 53.00 | 6.82 | 0.00 | 0.82 | 2.5592 | 0.9827 |
| 56 | 104 | 48 | 54.00 | 6.82 | 0.00 | 0.82 | 2.5620 | 0.9830 |
| 56 | 105 | 49 | 55.00 | 6.82 | 0.00 | 0.82 | 2.5658 | 0.9834 |
| 56 | 106 | 50 | 56.00 | 6.82 | 0.00 | 0.82 | 2.5704 | 0.9837 |
| 56 | 107 | 51 | 57.00 | 6.82 | 0.00 | 0.82 | 2.5758 | 0.9840 |
| 55 | 107 | 52 | 57.00 | 6.03 | 0.00 | 1.03 | 2.5373 | 0.9802 |
| 55 | 108 | 53 | 58.00 | 6.03 | 0.00 | 1.03 | 2.5442 | 0.9806 |
| 55 | 119 | 64 | 69.00 | 6.03 | 0.00 | 1.03 | 2.6570 | 0.9839 |
| 54 | 119 | 65 | 69.00 | 5.29 | 0.00 | 1.29 | 2.6264 | 0.9802 |

Table 1.1 concerns the case of a lead time of $L = 5$ days and a target fill rate of $\beta = 0.98$. It contains for several values of the order quantity $Q$ around $Q = 44.72$ the policy with the minimum value of the reorder point $s$ that meets the fill rate constraint. Note that the expected average cost and the actual fill rate are rather flat functions of the order quantity $Q$ in this region. Further, note that the expected backlog at the beginning of a reorder cycle, $E\{[D_L - S]^+\}$, is negligible so that the expected stock level at the beginning of a reorder cycle is $E\{[S-D_L]^+\} \approx S - E\{D_L\} = S - 50$. The safety stock is $s - E\{D_L\} = 6$ in this region. The expected average cost has a local minimum at $Q = 45$. However, when the order quantity becomes $Q = 52$ the required reorder point drops from $s = 56$ to $s = 55$ and another local minimum of $C(55, 107) = \$2.5373$ is met which is less than $C(56, 101) = \$2.5565$. The next local minimum is found for $Q = 65$ but here the cost is higher: $C(54, 119) = \$2.6264$.

Table 1.2 concerns the case of a lead time of $L = 50$ days and a target fill rate of $\beta = 0.98$. Similar remarks can be made as those made with regard to Table 1.1. The safety stock is now $s - E\{D_L\} = 31$ in the region of $Q = 44.72$. Local minima of the expected average cost are found at $Q = 45$, $Q = 49$, $Q = 54$, $Q = 59$,

Table 1.2: Feasible reorder points and order-up-to levels for $L = 50$, $\beta = 0.98$.

| $s$ | $S$ | $Q$ | $E\{[S-D_L]^+\}$ | $E\{[s-D_L]^+\}$ | $E\{[D_L-S]^+\}$ | $E\{[D_L-s]^+\}$ | $E\{C(s,S)\}$ | $\Psi$ |
|---|---|---|---|---|---|---|---|---|
| 531 | 575 | 44 | 75.00 | 31.88 | 0.00 | 0.88 | 3.8084 | 0.9801 |
| 531 | 576 | 45 | 76.00 | 31.88 | 0.00 | 0.88 | 3.8082 | 0.9805 |
| 531 | 577 | 46 | 77.00 | 31.88 | 0.00 | 0.88 | 3.8090 | 0.9809 |
| 530 | 579 | 49 | 79.00 | 30.97 | 0.00 | 0.97 | 3.7696 | 0.9803 |
| 529 | 583 | 54 | 83.00 | 30.06 | 0.00 | 1.06 | 3.7525 | 0.9804 |
| 528 | 587 | 59 | 87.00 | 29.16 | 0.00 | 1.16 | 3.7515 | 0.9803 |
| 527 | 591 | 64 | 91.00 | 28.27 | 0.00 | 1.27 | 3.7631 | 0.9801 |

Table 1.3: Feasible reorder points and order-up-to levels for $L = 5$, $\beta = 0.50$.

| $s$ | $S$ | $Q$ | $E\{[S-D_L]^+\}$ | $E\{[s-D_L]^+\}$ | $E\{[D_L-S]^+\}$ | $E\{[D_L-s]^+\}$ | $E\{C(s,S)\}$ | $\Psi$ |
|---|---|---|---|---|---|---|---|---|
| 28 | 72 | 44 | 22.00 | 0.00 | 0.00 | 22.00 | 1.6865 | 0.5001 |
| 28 | 73 | 45 | 23.00 | 0.00 | 0.00 | 22.00 | 1.6862 | 0.5112 |
| 27 | 73 | 46 | 23.00 | 0.00 | 0.00 | 23.00 | 1.6620 | 0.5000 |
| 27 | 74 | 47 | 24.00 | 0.00 | 0.00 | 23.00 | 1.6639 | 0.5107 |
| 26 | 74 | 48 | 24.00 | 0.00 | 0.00 | 24.00 | 1.6417 | 0.5000 |
| 19 | 81 | 62 | 31.00 | 0.00 | 0.00 | 31.00 | 1.5815 | 0.5000 |
| 18 | 82 | 64 | 32.00 | 0.00 | 0.00 | 32.00 | 1.5813 | 0.5000 |
| 17 | 83 | 66 | 33.00 | 0.00 | 0.00 | 33.00 | 1.5826 | 0.5000 |

$Q = 64$, etcetera. The global minimum is $C(528, 587) = \$3.7515$. The safety stock is here $s - E\{D_L\} = 28$. Note that a ten-fold increase in the lead times leads to a 48% increase in cost.

Table 1.3 concerns the case of a lead time of $L = 5$ days and a low target fill rate of $\beta = 0.50$. Note that the expected backlog at the beginning of a reorder cycle, $E\{[D_L - S]^+\}$, is still negligible. In this case, also the expected stock level at the end of a reorder cycle, $E\{[s - D_L]^+\}$ is zero. The expected net stock at the end of a reorder cycle is negative so that this case with a low target fill rate corresponds to the model with planned stockouts, cf. Section 1.2.4. Also, the safety stock, $s - E\{D_L\} = s - 50$, is negative in these cases. The expected average cost and the actual fill rate are more sensitive to the order quantity $Q$. The cost function has a local minimum for every even order quantity $Q$. The global minimum is found for $Q = 64$: $C(18, 82) = \$1.5813$.

Table 1.4: Feasible reorder points and order-up-to levels for $L = 1$, $\beta = 0.50$.

| $s$ | $S$ | $Q$ | $E\{[S-D_L]^+\}$ | $E\{[s-D_L]^+\}$ | $E\{[D_L-S]^+\}$ | $E\{[D_L-s]^+\}$ | $E\{C(s,S)\}$ | $\Psi$ |
|---|---|---|---|---|---|---|---|---|
| $-12$ | 32 | 44 | 22.00 | 0.00 | 0.00 | 22.00 | 1.6864 | 0.5000 |
| $-12$ | 33 | 45 | 23.00 | 0.00 | 0.00 | 22.00 | 1.6861 | 0.5111 |
| $-13$ | 33 | 46 | 23.00 | 0.00 | 0.00 | 23.00 | 1.6620 | 0.5000 |
| $-13$ | 34 | 47 | 24.00 | 0.00 | 0.00 | 23.00 | 1.6638 | 0.5106 |
| $-14$ | 34 | 48 | 24.00 | 0.00 | 0.00 | 24.00 | 1.6417 | 0.5000 |
| $-21$ | 41 | 62 | 31.00 | 0.00 | 0.00 | 31.00 | 1.5815 | 0.5000 |
| $-22$ | 42 | 64 | 32.00 | 0.00 | 0.00 | 32.00 | 1.5813 | 0.5000 |
| $-23$ | 43 | 66 | 33.00 | 0.00 | 0.00 | 33.00 | 1.5826 | 0.5000 |

Table 1.4 concerns the case of a lead time of $L = 1$ days and a low target fill rate of $\beta = 0.50$. The main difference with the case of $L = 5$ days in Table 1.3 is that the reorder point has become negative. The safety stocks, $s - E\{D_L\}$, are the same for $L = 1$ and $L = 5$ at given values of $Q$. In all four numerical examples, a better order quantity is found than the quantity $Q = 45$ suggested by (1.34). □

**Remark 1.5** In case of a cycle service constraint with cycle service level $\alpha$, the reorder point $s$ has to be chosen such that

$$\Pr\{D_L \leq s\} \geq \alpha.$$

For the case of a Poisson demand process this condition reads

$$\sum_{j=0}^{s} \frac{(\lambda L)^j}{j!} e^{-\lambda L} \geq \alpha.$$

Note that this condition does not involve $Q$ or $S$. Hence, with this service level constraint first $s$ can be determined as the smallest integer satisfying the above constraint, and then a value for $S$ can be searched that minimizes the expected average cost per unit of time (1.33). □

### 1.2.7 Stochastic demand, periodic review

This section is concerned with inventory systems with stationary stochastic demand and periodic reviewpolicies. The notations and assumptions are otherwise the same as in Section 1.2.6. The review period will be denoted by $R$. The most commonly used periodic review policies are:

- the $(R, S)$ policy: with intervals of $R$ units of time an order is placed of a size that brings the inventory position to the order-up-to level $S$;
- the $(R, s, S)$ policy: with intervals of $R$ units of time the inventory level is reviewed; whenever the inventory position is then at or below the reorder point $s$ an order is placed of a size that brings the inventory position to the order-up-to level $S$.

The length of the review period is either assumed to be given by other management considerations or can be chosen as the optimal reorder cycle length (1.5), in analogy to the sequential approach in the continuous review case,

$$R = R^* \doteq \sqrt{\frac{2a}{hE\{D\}}}. \tag{1.36}$$

If the review period $R$ is much smaller than the value $R^*$, it is preferable to use an $(R, s, S)$ policy to avoid many small orders. Here, we will only discuss the $(R, S)$ policy. If an order is placed at a review instant, it is delivered after a lead time $L$ while the next order can only arrive after a time $L + R$. Hence, the current order quantity should be sufficient to protect against the uncertainty in the demand during a lead time plus a review period. This random variable will be denoted by $D_{L+R}$. With constant lead times, the mean and the variance of $D_{L+R}$ are, in analogy to (1.31), given by

$$E\{D_{L+R}\} = (L + R)E\{D\}, \quad \sigma^2\{D_{L+R}\} = (L + R)\sigma^2\{D\}. \tag{1.37}$$

When all stockouts are backlogged, and for a fixed review period $R$, the order-up-to level $S$ has to be determined such that a fraction $\beta$ of all demand can be sold directly from stock (fill rate constraint):

$$\Psi = 1 - \frac{E\{B\}}{E\{Q\}} = 1 - \frac{E\{[D_{L+R} - S]^+\} - E\{[D_L - S]^+\}}{RE\{D\}} \geq \beta; \tag{1.38}$$

here, $B$ stands for the accumulated backlog during a review period $R$, as in (1.32), and the expected value of an order quantity is $E\{Q\} = RE\{D\}$ by a balance argument. The safety stock for this policy is $S - E\{D_{L+R}\}$. The expected average cost per unit of time is

$$E\{C(R, S)\} = \frac{a}{R} + \tfrac{1}{2}h[E\{[S - D_{L+R}]^+\} + E\{[S - D_L]^+\}]. \tag{1.39}$$

Note that there is no decision variable left to minimize this objective function if $R$ is fixed and $S$ is set to satisfy the service level constraint (1.38). The expected inventory on hand is simply related to the expected backlog through

$$E\{[S - D_{L+R}]^+\} = S - (L + R)E\{D\} + E\{[D_{L+R} - S]^+\}. \tag{1.40}$$

**Example 1.4** The required order-up-to level $S$ will be determined for various distributions for the demand during the lead time plus review period. In all cases, the review period is $R = 4$ weeks and the lead time is $L = 1$ week. The mean demand per week is $E\{D\} = 20$, so that $RE\{D\} = 80$ and $E\{D_{L+R}\} = 100$. The cost factors are $a = \$5$ and $h = \$0.05$ per unit per week, so that $R = 4$ is the largest integer smaller than $R^* = \sqrt{20}$ according to (1.36). The order-up-to level $S$ is obtained by solving

$$E\{[D_{L+R} - S]^+\} - E\{[D_L - S]^+\} \leq (1 - \beta)RE\{D\} = 80(1 - \beta).$$

Table 1.5: Order-up-to levels, fill rate, average cost and backlog at end of cycle for $\sigma^2\{D\} = 125$.

| Target | Normal | | | | Gamma | | | |
|---|---|---|---|---|---|---|---|---|
| $\beta$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ |
| 0.90 | 105 | 0.9041 | 3.6918 | 7.67 | 105 | 0.9027 | 3.6946 | 7.78 |
| 0.95 | 116 | 0.9506 | 4.1487 | 3.95 | 118 | 0.9517 | 4.2465 | 3.86 |
| 0.96 | 120 | 0.9624 | 4.3260 | 3.01 | 122 | 0.9619 | 4.4263 | 3.05 |
| 0.97 | 124 | 0.9719 | 4.5062 | 2.25 | 126 | 0.9701 | 4.6098 | 2.39 |
| 0.98 | 129 | 0.9810 | 4.7380 | 1.52 | 133 | 0.9809 | 4.9382 | 1.53 |
| 0.99 | 137 | 0.9904 | 5.4193 | 0.77 | 143 | 0.9903 | 5.1493 | 0.77 |

In Table 1.5 the normal distribution is compared with the gamma distribution in a case with a low variation in the demand: $\sigma^2\{D\} = 125$. This implies, cf. (1.37), that $\sigma\{D_{L+R}\} = 25$ with a coefficient of variation $C_{L+R} = 0.25$, and that $\sigma\{D_L\} = 5\sqrt{5} \approx 11.18$, with a coefficient of variation $C_L \approx 0.56$. The expected backlogs $E\{[D_{L+R} - S]^+\}$ and $E\{[D_L - S]^+\}$ are computed with the aid of the normal loss function, cf. (A.21), with parameters $\mu = 100$, $\sigma = 25$ and $\mu = 20$, $\sigma = 11.18$, respectively. The probabilities of negative realizations are for these normally distributed random variables $3 \times 10^{-5}$ and 0.037, respectively. These expected backlogs are also computed with the aid of the gamma loss function, cf. (A.38), with parameters $\psi = 16$, $\lambda = \frac{4}{25}$ (hence, an Erlang distribution) and $\psi = 3.2$, $\lambda = \frac{4}{25}$, respectively. Observe that the normal distribution yields lower order-up-to levels than the gamma distribution due to the lighter tail of the normal distribution. In these cases, the expected backlog at the beginning of a cycle is negligible. In Table 1.6 the gamma distribution is compared with a mixture of two Erlang distributions in a case with a moderate variation in the demand: $\sigma^2\{D\} = 1125$. This implies that $\sigma\{D_{L+R}\} = 75$ with a coefficient of variation $C_{L+R} = 0.75$, and that $\sigma\{D_L\} = 15\sqrt{5} \approx 33.54$, with a coefficient of variation $C_L \approx 1.68$. The expected backlogs are computed with the aid of the gamma loss function with parameters $\psi = \frac{16}{9}$, $\lambda = \frac{4}{225}$ and $\psi = \frac{16}{45}$, $\lambda = \frac{4}{225}$, respectively. The parameters of the mixed Erlang distribution are, cf. (A.42), $K_1 = 1$, $K_2 = 2$, $\lambda_1 = \lambda_2 = 0.0188$, $p = 0.1213$, and, cf. (A.43), $K_1 = K_2 = 1$, $\lambda_1 = 0.1779$, $\lambda_2 = 0.0221$, $p = 0.6368$, respectively. The corresponding loss functions are given in (A.44) and (A.46). The expected backlog at the beginning of a cycle varies from about 0.3 for $\beta = 0.9$ to about 0.02 for $\beta = 0.99$; ignoring this quantity in (1.38) leads to order-up-to levels that are about 2 higher than those displayed in the table. Observe that the mixed Erlang distribution yields lower order-up-to levels than the gamma distribution. The normal distribution does not seem appropriate in this case since the probability of negative realizations of $D_{L+R}$ is 0.09. Use of the normal distribution gives much lower order-up-to levels, e.g., $S = 244$ for $\beta = 0.99$.

Table 1.6: Order-up-to levels, fill rate, average cost and backlog at end of cycle for $\sigma^2\{D\} = 1125$.

| Target | Gamma | | | | ME (Erlang-2 and Exponential) | | | |
|---|---|---|---|---|---|---|---|---|
| $\beta$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ |
| 0.90 | 185 | 0.9008 | 7.7132 | 8.23 | 184 | 0.9004 | 7.6633 | 8.25 |
| 0.95 | 231 | 0.9506 | 9.9047 | 4.07 | 229 | 0.9503 | 9.8045 | 4.08 |
| 0.96 | 245 | 0.9602 | 10.5841 | 3.27 | 243 | 0.9602 | 10.4835 | 3.26 |
| 0.97 | 264 | 0.9704 | 11.5123 | 2.43 | 261 | 0.9701 | 11.3624 | 2.44 |
| 0.98 | 290 | 0.9803 | 12.7913 | 1.61 | 287 | 0.9803 | 12.6408 | 1.60 |
| 0.99 | 333 | 0.9900 | 14.9207 | 0.81 | 329 | 0.9901 | 14.7204 | 0.80 |

In Table 1.7 the gamma distribution is compared with a hyperexponential distribution ($H_2$) in a case with a high variation in the demand: $\sigma^2\{D\} = 8000$. This implies that $\sigma\{D_{L+R}\} = 200$ with a coefficient of variation $C_{L+R} = 2$, and that $\sigma\{D_L\} = 40\sqrt{5} \approx 89.44$, with a coefficient of variation $C_L \approx 4.47$. The parameters of the gamma distributions are $\psi = \frac{1}{4}$, $\lambda = \frac{1}{400}$ and $\psi = \frac{1}{20}$, $\lambda = \frac{1}{400}$, respectively. The parameters of the hyperexponential distributions are $K_1 = K_2 = 1$, $\lambda_1 = 0.03673$, $\lambda_2 = 0.00327$, $p = 0.7390$, and $K_1 = K_2 = 1$, $\lambda_1 = 0.01964$, $\lambda_2 = 0.0036$, $p = 0.9447$, respectively. The expected backlog at the beginning of a cycle varies from about 1.4 for $\beta = 0.9$ to about 0.1 for $\beta = 0.99$; ignoring this quantity in (1.38) leads to order-up-to levels that are about 50 higher than those displayed in the table (for the hyperexponential distribution the difference decreases with increasing $\beta$). Observe that the hyperexponential distribution yields lower order-up-to levels than the gamma distribution for $\beta > 0.95$ (tail behavior) and higher levels for $\beta \leq 0.95$. The normal distribution is not appropriate in this case since the probability of negative realizations of $D_{L+R}$ is

Table 1.7: Order-up-to levels, fill rate, average cost and backlog at end of cycle for $\sigma^2\{D\} = 8000$.

| Target | Gamma | | | | $H_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\beta$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ | $S$ | $\Psi$ | $E\{C(R,S)\}$ | $E\{[S - D_{L+R}]^+\}$ |
| 0.90 | 636 | 0.9000 | 30.32 | 9.40 | 655 | 0.9000 | 31.27 | 9.40 |
| 0.95 | 863 | 0.9500 | 41.53 | 4.67 | 872 | 0.9501 | 41.98 | 4.63 |
| 0.96 | 938 | 0.9601 | 45.26 | 3.72 | 941 | 0.9600 | 45.40 | 3.69 |
| 0.97 | 1035 | 0.9700 | 50.08 | 2.79 | 1031 | 0.9701 | 49.88 | 2.75 |
| 0.98 | 1174 | 0.9800 | 57.00 | 1.85 | 1157 | 0.9800 | 56.15 | 1.82 |
| 0.99 | 1415 | 0.9900 | 69.03 | 0.92 | 1373 | 0.9900 | 66.93 | 0.90 |

0.3.

Table 1.8: Order-up-to levels and average cost for various $R$ at $\beta = 0.95$.

| | $R = 1$ | | $R = 2$ | | $R = 3$ | | $R = 4$ | | $R = 5$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $C_{L+R}$ | $S$ | $E\{C(R,S)\}$ | $S$ | $E\{C(R,S)\}$ | $S$ | $E\{C(R,S)\}$ | $S$ | $E\{C(R,S)\}$ | $S$ | $E\{C(R,S)\}$ |
| 0.25 | 62 | 6.63 | 80 | 4.55 | 99 | 4.19 | 118 | 4.25 | 137 | 4.47 |
| 0.75 | 168 | 11.95 | 189 | 10.01 | 210 | 9.75 | 231 | 9.90 | 251 | 10.18 |
| 2.00 | 790 | 43.07 | 814 | 41.29 | 839 | 41.23 | 863 | 41.53 | 888 | 42.06 |

Finally, it turns out that in all above discussed cases a shorter review period gives a lower expected average cost. Table 1.8 shows the required order-up-to levels $S$ and the expected average cost $E\{C(R,S)\}$ according to gamma distributed demand at a target fill rate of $\beta = 0.95$ for the coefficients of variations in the demand considered above and for various lengths of the review period $R$. The minimum cost is attained for $R = 3$. Observe the tendency that a shorter review period becomes cheaper when the coefficient of variation in the demand increases. As could be expected, the required order-up-to level is an increasing function of the review period. □

## 1.3 Interactions in Inventory Management

The basic models of inventory theory assume that there is no interaction between various items and various stocking points, and consider individual items at a single stocking point. However, in many practical cases interactions do exist, and an inventory policy may be far from optimal when such interactions are neglected. Interactions may be due to (see, e.g., Love [49, Ch. 5]):

- joint capacity constraints:
  - a common storage space used for various items;
  - a common labor force and equipment for the receipt and handling of various items;
  - a single budget available for the inventory of various items.
- correlated demand:
  - assembly requires parts in fixed proportions;
  - special offers or advertisements at a regional or national scale influence sales at many stocking points;
  - substitution effect (e.g., the same item, but of a different color);
  - complementarity effect (several units of items that fit together, or make up a whole).
- interwoven cost structure:
  - joint ordering cost for a family of items ordered from the same supplier;
  - joint shipping cost;
  - joint discount structure (on the total amount or the total quantity of an order);

  - joint handling cost in a warehouse.

- interwoven product structure:
  - raw material is used in the production of several items;
  - produced parts are used in several final products.
- interwoven distribution structure:
  - an item is kept on stock at several locations (within the same company or organization);
  - sold units may return to a stocking point because of customer dissatisfaction (e.g., clothing or items bought from a mail order company) or after replacement by a spare part when repaired.

In Chapter 2, coordinated replenishment strategies will be discussed for families of items with a joint cost structure. The considered interactions through cost structure are:

- joint ordering, production and/or transportation costs:
  - items/parts stem from the same supplier or can be shipped with the same transport;
  - items belong to the same product group (with limited mutual switch-over times) or are variants of each other (only differing in packing, enclosing, etcetera);
  - there is a fixed ordering cost ($A$), independent of the composition of the order (administration, shipment);
  - there are additional ordering costs ($a_i$), depending on whether item $i$ forms part of an order;
  - the key question is: should an item be ordered at a reorder instant or not (the order quantities are of minor importance).
- joint discount structure:
  - a supplier offers discount on purchasing prices or shipping cost if the total value or volume of an order exceeds some threshold;
  - example: a contribution in shipping cost is only passed on below a certain order amount;
  - the order quantities are important in this case.

Advantages of joint replenishment (or joint production) strategies may be:

- lower purchasing cost;
- lower shipping cost;
- lower handling cost;
- lower ordering cost;
- lower production cost (set-up, switch-over).

Disadvantages of joint replenishment (or joint production) strategies may be:

- higher average inventory levels (since orders may be put forward and EOQ quantities are adjusted);
- inventory management more complex;
- higher management cost (more involved reviews, more complex computations);
- higher peaks at the receipt of goods;
- reduced flexibility (e.g., in coping with unusual situations).

Chapter 3 is concerned with replenishment strategies for multi-echelon inventory systems where items are kept in stock at various storage locations. Storage locations often form part of complex good flows (supply chains), from raw materials via parts and semi-manufactured items to end products which reach customers possibly via several distribution nodes: this results in multi-echelon inventory systems. Coordination of inventory management at various stocking points is desirable (since orders of lower levels form the demand process for higher levels, whereby demand usually occurs in larger quantities at higher echelons) to avoid possible oscillating effects. Coordination of inventory management on various stocking points is possible as far as they are controlled by the same company or if good agreements exist between companies (e.g., to prevent hoarding).

In order to implement a multi-echelon replenishment strategy, proper information exchange between the echelons is important (e.g., via EDI: electronic data interchange):

- from lower echelons upward: concerning reorder frequency, demand patterns, special actions (locally);
- from higher echelons downward: concerning production or replenishment plan: when becomes how much available of which item; product advertisements (globally).

In multi-echelon inventory management a distinction is made between pull and push systems. In a pull system each stocking point determines its own reorder policy. In a push system information is kept up to date at a central level and stocks are controlled from a central point. Push systems

- require much and reliable information;
- realize cost savings by better anticipation and coordination;
- may give problems with responsibilities of local managers.

As a compromise, a central planning may be implemented that allows limited local modifications provided that the latter are reported in time.

Chapter 4 is devoted to replenishment strategies for inventory systems with capacity constraints and with returning, repairable items.

# Chapter 2

# Coordinated Replenishment

This chapter is devoted to models, algorithms and policies for inventory systems in which coordinated replenishment of items may be profitable. Consecutively, models with constant, with time-varying and with stochastic demand will be discussed. The following concepts will be used:

**family:** a set of items that possesses a joint cost structure (e.g., because they are ordered from the same supplier, or because they are delivered by the same transporter);

**group:** a set of items from a family that have the same constant reorder cycle time.

With an individual item it is optimal to have a fixed reorder cycle; with coordinated ordering of several items this is not necessarily the case; in general it is assumed as a simplification that a group of items has a fixed reorder cycle (this may be suboptimal).
A distinction will be made between indirect and direct grouping of items, cf. Chakravarty [16]:

**indirect grouping:** the reorder cycle of a group is an integer multiple of the family reorder cycle (this grouping aims at double saving on ordering cost, both within groups and across groups);

**direct grouping:** the reorder cycle of a group is not related to those of other groups within the same family (this grouping aims at single saving on ordering cost, only within groups).

## 2.1 Constant demand, indirect grouping

This section is concerned with indirect grouping policies for a family of items with constant demand. The analysis will be based on fluid approximations for the demand and the inventory level. The assumptions for the model with constant (known) demand for a family of $N$ items are:

- a family or joint ordering cost $A$, which is independent of the composition of the order;
- a supplementary ordering cost $a_i$ for item $i$, which only is incurred when this item is included in an order;
- the demand rate $D_i$ for item $i$ is a constant number of units per unit of time;
- a holding cost $h_i$ for item $i$ is incurred per unit per unit of time; often, $h_i = rv_i$, with $v_i$ the purchasing cost (value) of item $i$ and $r$ an item independent carrying charge;
- no back orders or stockouts are allowed;
- no capacity restrictions on order quantities (or on production lot sizes) exist;
- the delivery of orders (or of production lots) takes place as a whole;
- the lead time is negligible or constant and is equal for all items;
- a continuous review of stocks is applied.

The aim is to minimize the average ordering and holding cost per unit of time (purchasing costs are fixed since no back orders or shortages are allowed). The decision variables could be the order quantities for the $N$ items, but in case of coordinated replenishment it is more suitable to consider the reorder frequencies for the $N$ items. For the case of indirect grouping of items the decision variables are more specifically defined as:

- with time intervals of length $T$ a family order is placed (the family reorder cycle);
- item $i$ is included in every $k_i$th replenishment of the family ($k_i = 1, 2, \ldots$).

Hence, the reorder cycle of item $i$ has a length of $T_i = k_i T$. Items with the same reorder frequency $k_i$ are said to form a group. The objective function to be minimized reads for indirect grouping:

$$C(T, \mathbf{k}) = \frac{1}{T}\left[A + \sum_{i=1}^{N} \frac{a_i}{k_i}\right] + \tfrac{1}{2} T \sum_{i=1}^{N} k_i h_i D_i. \tag{2.1}$$

Here, a restrictive assumption is that $\min_i\{k_i\} = 1$ (otherwise, a modification of the objective function is required with respect to the term $A/T$ since family orders are not placed every $T$ units of time).

**Remark 2.1** A trivial upper bound on the minimum cost is obtained by using the EOQ formula with ordering cost $A + a_i$ for each item separately and by summing the minimum cost per item over all items. A lower bound is obtained by ignoring the term $A/T$, and by minimizing the resulting separable function in the variables $T_i = k_i T$. Hence,

$$\sum_{i=1}^{N} \sqrt{2 a_i h_i D_i} \leq C^* \leq \sum_{i=1}^{N} \sqrt{2(A + a_i) h_i D_i}.$$

If the minimum cost $C^*$ with coordinated replenishment is close to the upper bound, coordination may not be profitable. In such a case, the increase in management cost may be larger than the reduction in inventory cost obtained by coordination. ☐

The optimal family cycle for a given vector of reorder frequencies $\mathbf{k} = (k_1, \ldots, k_N)$ can be derived in a similar way as the basic EOQ-formula, cf. (1.5) and Appendix B.1:

$$\hat{T}(\mathbf{k}) = \sqrt{\frac{2[A + \sum_{j=1}^{N}(a_j/k_j)]}{\sum_{i=1}^{N} k_i h_i D_i}}. \tag{2.2}$$

The corresponding average inventory cost per unit of time is, cf. (1.6):

$$\hat{C}(\mathbf{k}) = \sqrt{2\left[A + \sum_{j=1}^{N} \frac{a_j}{k_j}\right] \sum_{i=1}^{N} k_i h_i D_i} = \hat{T}(\mathbf{k}) \sum_{i=1}^{N} k_i h_i D_i. \tag{2.3}$$

In a similar way, it follows for a given family reorder cycle $T$ that the frequencies

$$\hat{k}_i(T) = \sqrt{\frac{2 a_i}{h_i D_i T^2}}, \quad i = 1, \ldots, N, \tag{2.4}$$

would be optimal if they were integer. A rounding-off rule like (1.8) states: round $\hat{k}_i(T)$ to $k_i = \tilde{k}_i(T)$ if

$$k_i(k_i - 1) < \frac{2 a_i}{h_i D_i T^2} \leq k_i(k_i + 1), \quad i = 1, \ldots, N. \tag{2.5}$$

Clearly, the frequencies $\tilde{k}_i(T)$ are ordered according to ascending value of $a_i/(h_i D_i)$. The relative values of the optimal frequencies are independent of $A$. The actual optimal values do depend on $A$ through $T$. Further, under this rounding the family cycle $T$ is bounded by

$$\sqrt{\frac{2 a_i}{h_i D_i}} \sqrt{\frac{1}{\tilde{k}_i(T)(\tilde{k}_i(T) + 1)}} \leq T < \sqrt{\frac{2 a_i}{h_i D_i}} \sqrt{\frac{1}{\tilde{k}_i(T)(\tilde{k}_i(T) - 1)}}, \quad i = 1, \ldots, N. \tag{2.6}$$

**Remark 2.2** It is readily verified that the function $\hat{T}(\mathbf{k})$, cf. (2.2), is decreasing in each individual $k_i,\ i = 1,\ldots,N$. Hence, the maximum value of this function is at $\mathbf{k} = \mathbf{1} \doteq (1,1,\ldots,1)$. Hence, for all feasible vectors $\mathbf{k}$ it holds that

$$\hat{T}(\mathbf{k}) \le T_{\max} \doteq \hat{T}(\mathbf{1}) = \sqrt{\frac{2[A + \sum_{j=1}^{N} a_j]}{\sum_{i=1}^{N} h_i D_i}}.$$

A minimum value for the function $\hat{T}(\mathbf{k})$ is obtained from the lefthand inequality in (2.6) recalling the assumption that $\min_i\{k_i\} = 1$. This implies that for all feasible vectors $\mathbf{k}$ we have

$$\hat{T}(\mathbf{k}) \ge T_{\min} \doteq \min_{i=1,\ldots,N} \sqrt{\frac{a_i}{h_i D_i}}.$$

Smaller family cycles may be associated with policies with $\min_i\{k_i\} > 1$. □

When the family reorder cycle $T$ and the reorder frequencies $k_i,\ i = 1,\ldots,N$, have been determined, the order quantities follow by (1.3) as $Q_i = D_i k_i T,\ i = 1,\ldots,N$.

### 2.1.1 Exact algorithmic solution

Goyal [33] has developed an algorithm for determination of the exact optimum of the objective function (2.1) under the assumption that $\min_i\{k_i\} = 1$. It is based on the observations that the optimal family cycle $T^*$ lies on the finite interval $[T_{\min}, T_{\max}]$, cf. Remark 2.2, and that the inequalities (2.6) divide this interval into a finite number of subintervals. As a consequence, the algorithm consists of finitely many steps.

**Algorithm 2.1** [Goyal]

Step 1: Compute the maximum value $T_{\max} = \hat{T}(\mathbf{1})$ and the minimum value $T_{\min}$ for the family cycle $T$, cf. Remark 2.2.

Step 2: Start with the optimal (rounded) frequencies $\tilde{k}_i(T_{\max})$; call this vector $\mathbf{k}_1$; compute the corresponding costs $\hat{C}(\mathbf{k}_1)$.

Step 3: Determine for each item $i$ the length of the family cycle $T_c(i)$ for which the value of $k_i = \tilde{k}_i(T_{\max})$ changes to $k_i + 1$ via the bounds (2.6) of $T = T_{\max}$:

$$T_c(i) = \sqrt{\frac{2a_i}{h_i D_i}} \sqrt{\frac{1}{k_i(k_i+1)}}. \tag{2.7}$$

Set $j = 1$ and start the recursion (steps 4 and 5).

Step 4: Increase $j$ by one. Let $T_c = \max_i\{T_c(i)\}$ be the first boundary that is met when $T$ decreases. If $T_c \le T_{\min}$ then goto step 6, otherwise continue with step 5.

Step 5: Let $i_c$ be the item for which $T_c(i_c) = \max_i\{T_c(i)\}$. Change $k_{i_c}(T)$ to $k_{i_c}(T) + 1$ to obtain the vector $\mathbf{k}_j$ and compute the corresponding minimum cost $\hat{C}(\mathbf{k}_j)$, cf. (2.3). Compute $T_c(i_c)$ with the new value of $k_{i_c}$. Return to step 4.

Step 6: The algorithm stops: the optimal policy is that policy $(\hat{T}(\mathbf{k}_j), \mathbf{k}_j)$ for which the lowest costs $\hat{C}(\mathbf{k}_j)$ have been found.

The cost function $\hat{C}(\mathbf{k}_j)$ is not in all cases convex in $j$. □

**Example 2.1** Consider a family consisting of $N = 3$ items, with family ordering cost $A = \$6$. Further data (per week) can be found in Table 2.1, the first four columns.
The optimal individual policies, cf. Section 1.2.1, for the three items are indicated in the columns with headers $T_i$, $Q_i$, and the corresponding costs under $C_i$. The minimum total average cost per week without coordination is $C = \$29$. Note that in this academic case the reorder instants can be synchronized in cycles of 42 weeks in which there are $21 + 7 + 2 = 30$ family reorder instants instead of $21 + 14 + 6 = 41$ without synchronization. In this way, $\frac{11}{42}A = \$1.57$ is saved on the average weekly cost.

Table 2.1: Data for Example 2.1; optimal individual and indirect grouping policies.

| Item $i$ | $a_i$ | $h_i$ | $D_i$ | $A+a_i$ | $h_iD_i$ | $T_i$ | $Q_i$ | $C_i$ | $k_i^*T^*$ | $Q_i^*$ | $\tilde{a}_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | \$ 3 | \$0.50 | 9 | \$ 9 | \$4.50 | 2 | 18 | \$ 9 | 2.05 | 18.47 | \$ 9.48 |
| 2 | \$ 3 | \$0.50 | 4 | \$ 9 | \$2.00 | 3 | 12 | \$ 6 | 2.05 | 8.21 | \$ 4.21 |
| 3 | \$43 | \$0.50 | 4 | \$49 | \$2.00 | 7 | 28 | \$14 | 6.16 | 24.63 | \$37.93 |

Next, consider coordinated replenishment by indirect grouping. Application of Goyal's Algorithm 2.1 starts with the computation of $T_{\max} = \sqrt{110/8.5} \approx 3.597$ and $T_{\min} = \sqrt{2/3} \approx 0.816$. Next, it determines for $T = T_{\max}$: $k_1 = 1$ ($\sqrt{51/495}$), $k_2 = 1$ ($\sqrt{51/220}$), $k_3 = 2$ ($\sqrt{731/220}$), cf. (2.4).

The successive iterations are:

$\hat{C}(1,1,2) = \$26.52$ [family cycle $\hat{T}(1,1,2) = 2.526$]

$T_c(1) = \sqrt{2/3} \approx 0.816$, $T_c(2) = \sqrt{3/2} \approx 1.225$, $T_c(3) = \sqrt{43/6} \approx 2.677$, $T_c = 2.677$, $i_c = 3$

$\hat{C}(1,1,3) = \$25.66$ [family cycle $\hat{T}(1,1,3) = 2.053$]

$T_c(3) = \sqrt{43/12} \approx 1.893$, $T_c = 1.893$, $i_c = 3$

$\hat{C}(1,1,4) = \$25.69$ [family cycle $\hat{T}(1,1,4) = 1.771$]

$T_c(3) = \sqrt{43/20} \approx 1.466$, $T_c = 1.466$, $i_c = 3$

$\hat{C}(1,1,5) = \$26.07$ [family cycle $\hat{T}(1,1,5) = 1.580$]

$T_c(3) = \sqrt{43/30} \approx 1.197$, $T_c = 1.225$, $i_c = 2$

$\hat{C}(1,2,5) = \$26.58$ [family cycle $\hat{T}(1,2,5) = 1.437$]

$T_c(2) = \sqrt{3/6} \approx 0.707$, $T_c = 1.197$, $i_c = 3$

$\hat{C}(1,2,6) = \$26.91$ [family cycle $\hat{T}(1,2,6) = 1.313$]

$T_c(3) = \sqrt{43/42} \approx 1.012$, $T_c = 1.012$, $i_c = 3$

$\hat{C}(1,2,7) = \$27.37$ [family cycle $\hat{T}(1,2,7) = 1.216$]

$T_c(3) = \sqrt{43/56} \approx 0.876$, $T_c = 0.876$, $i_c = 3$

$\hat{C}(1,2,8) = \$27.89$ [family cycle $\hat{T}(1,2,8) = 1.138$]

$T_c(3) = \sqrt{43/72} \approx 0.773$, $T_c = 0.816$, $i_c = 1$

Now, $T_c = T_c(1) = T_{\min}$ so that the iterations stop. The optimum reorder frequencies are $\mathbf{k}^* = (1,1,3)$. The optimum family cycle is $T^* = \hat{T}(1,1,3) = 2.053$. The minimum total average cost per week with coordination is $C^* = \hat{C}(1,1,3) = \$25.66$, that is, 11.5% less than without coordination (and without synchronization). The optimum item cycles and the optimum order quantities are listed in Table 2.1 under the headers $k_i^*T^*$ and $Q_i^*$, respectively. For later reference, the table also contains the values $\tilde{a}_i \doteq \frac{1}{2}h_iD_i(k_i^*T^*)^2$, $i = 1,2,3$, which represent the ordering costs that give $Q_i^*$ as optimum individual order quantity, respectively. □

**Remark 2.3** The optimal indirect grouping policy is not found by Goyal's algorithm if the minimum frequency is larger than 1. The latter may be the case when the family ordering cost $A$ is small with respect to the individual ordering costs $a_i$, $i = 1, \ldots, N$. See Van Eijs [70] for an extension of the algorithm to include strategies with minimum frequency larger than 1. Goyal's algorithm requires a computation time, which is strongly increasing with the number of items $N$ in the family; as a consequence, it is only applicable to moderately sized problems. The algorithm is, however, important for verifying the quality of heuristics. □

**Example 2.2** An example of a situation where the minimum frequency is larger than 1 in the optimal indirect grouping policy has been provided by Andres & Emmons [2]. Consider a family of $N = 2$ items. Let $A = \$1$, $a_1 = a_2 = \$50$, $h_1 = h_2 = \$1$, $D_1 = 400$ and $D_2 = 900$. Goyal's algorithm gives the following policy: $\mathbf{k}^* = (2,1)$, $T^* = \hat{T}(2,1) = 0.299$. The corresponding average cost per unit of time is $C^* = \hat{C}(2,1) = \$508.33$. However, the average cost per unit of time corresponding to the frequencies $\mathbf{k} = (3,2)$ is, cf. (2.1),

$$C(T,(3,2)) = \frac{1}{T}[\tfrac{2}{3}A + \tfrac{1}{3}a_1 + \tfrac{1}{2}a_2] + \tfrac{1}{2}T[3h_1D_1 + 2h_2D_2],$$

since only at four instants in each cycle of six potential reorder instants an order is placed. This cost function is minimal for, cf. (2.2),

$$\hat{T}(3,2) = \sqrt{\frac{2[\frac{2}{3}A + \frac{1}{3}a_1 + \frac{1}{2}a_2]}{3h_1D_1 + 2h_2D_2}} = \sqrt{\frac{\frac{2}{3}\cdot 127}{3000}} = \tfrac{1}{30}\sqrt{25.4} \approx 0.168.$$

The corresponding average cost per unit of time is $C = C(\hat{T}(3,2),(3,2)) = 3000 \cdot \hat{T}(3,2) = \$503.98$ which is less than $\hat{C}(2,1)$. Observe that the family ordering cost is much smaller than the item-dependent ordering costs in this example. ☐

**Remark 2.4** Goyal's algorithm is still applicable if $A = 0$ while $a_i > 0,\ i = 1,\ldots,N$. The minimum cost of the coordinated replenishment problem (with integer ordering frequencies) will then be larger than (or equal to) the sum of the minimum costs of the optimal individual policies (without integer constraints). ☐

**Exercise 2.1** *What is the increase in cost if the family cycle in Example 2.1 is rounded to $T = 2$ weeks? What do the order quantities become?*

**Exercise 2.2** *A retailer orders two items from the same importer. The family ordering cost is $A = \$1$, while the supplementary item ordering costs are $a_1 = \$3$ and $a_2 = \$8$, respectively. The purchasing costs are $v_1 = \$16$ per unit of item 1 and $v_2 = \$10$ per unit of item 2. The holding costs of both items per unit per week are a carrying charge $r = 0.005$ times the purchasing cost. The demand for both items is assumed to be deterministic and constant in time. The demand for item 1 is 25 units per week and the demand for item 2 is 40 units per week. First, consider the two items independently of each other and determine for both items the optimal order quantity, the optimal reorder cycle and the minimum average ordering and holding cost per week. How much can be saved on this cost by synchronizing the reorder instants of the two items in such a way that they occasionally coincide? Next, determine the optimal inventory policy in the class of indirect grouping, that is, determine the optimal family reorder cycle and the optimal reorder frequencies $k_1$ and $k_2$ with $\min\{k_1,k_2\} = 1$. Also, compute the corresponding minimum average ordering and holding cost per week.*

**Exercise 2.3** *Consider a family of four items that are supplied by the same distribution center. The family ordering cost is $A = \$10$, while the supplementary item ordering costs are $a_1 = \$5$, $a_2 = \$12$, $a_3 = \$12$, and $a_4 = \$8$. The holding costs per unit per month are $h_1 = \$1.50$, $h_2 = \$2.00$, $h_3 = \$2.00$, and $h_4 = \$1.50$. The monthly demands are $D_1 = 60$, $D_2 = 30$, $D_3 = 90$, and $D_4 = 80$. Assume deterministic demand. Implement Goyal's algorithm and determine the optimal inventory policy.*

### 2.1.2 Heuristic solutions

To avoid the long computation times required by Goyal's Algorithm 2.1 several heuristics have been developed to find good solutions for the indirect grouping problem. A distinction is made between one or two step heuristics and iterative heuristics.

First we will discuss one and two step heuristics. The one-step heuristic of Goyal & Belton [35] first determines as a reference item the item with the smallest individual cycle length. Then, the frequencies of the other items are chosen as the optimal frequencies given this cycle length. Finally, the family cycle length is chosen as the optimal cycle length given these frequencies.

**Algorithm 2.2** [Goyal & Belton]

Step 1: Compute for every item $i$ the value of the quotient $(A + a_i)/(h_iD_i)$. The reference item $i_r$ is the item with the smallest value. This item has by definition a frequency of 1: $k_{i_r} \doteq 1$.

Step 2: Determine an initial family cycle length $T^{(0)}$ as the optimal EOQ cycle length of the reference item $i_r$, cf. (1.5), with ordering cost $A + a_{i_r}$.

Step 3: Determine the reorder frequencies of the non-reference items $i \neq i_r$ by $k_i = \tilde{k}_i(T^{(0)})$, cf. (2.5).

Step 4: Determine the family cycle length $T^{(1)}$ as the optimal cycle given the vector of frequencies: $T^{(1)} = \hat{T}(\mathbf{k})$, cf. (2.2).

The corresponding cost is $\hat{C}(\mathbf{k})$. ☐

**Remark 2.5** This algorithm is an improvement of an algorithm proposed earlier by Silver [59] which determined the reference item on the basis of the values of the quotient $a_i/(h_iD_i)$. The latter is suggested by (2.4) but it ignores the fact that the reference item, being the item with the smallest individual cycle length, is the first responsible for the family ordering cost. □

Kaspi & Rosenblatt [45] extended the foregoing heuristic because the result of this heuristic strongly depends on the choice of the reference item. However, the item with the smallest individual cycle length may be an item with a minor contribution to the total family inventory costs, and still it gets an important influence on the reorder cycles of the other items. Therefore, a second step is added in which the reorder frequencies and the family cycle can be modified independently of the reference item.

**Algorithm 2.3** [Kaspi & Rosenblatt]

Step 1: Determine an initial policy $(T^{(1)}, \mathbf{k})$ with the aid of the heuristic of Goyal & Belton.

Step 2: Determine a better policy by first determining new reorder frequencies as $k_i = \tilde{k}_i(T^{(1)})$ for $i = 1, \ldots, N$, cf. (2.5), and by then modifying the family cycle to $T^{(2)} = \hat{T}(\mathbf{k})$.

Step 2 might be repeated to obtain an iterative heuristic. □

**Remark 2.6** A simulation study of Kaspi & Rosenblatt has revealed that the largest improvement occurs in the first application of Step 2. □

Next, we will discuss an iterative heuristic algorithm proposed by Goyal [34]. The general idea of iterative heuristics is to choose initial reorder frequencies, and alternatingly determine a family cycle length given the frequencies and new frequencies given the family cycle, until the policy does not change in two subsequent iterations. Such a procedure may end at a local minimum, or may start oscillating.
The heuristic of Goyal makes use of optimal family cycles $\check{T}_i(\mathbf{k})$ given the reorder frequencies $\mathbf{k}$, without regard to item $i$, cf. (2.2),

$$\check{T}_i(\mathbf{k}) = \sqrt{\frac{2[A + \sum_{j=1,\, j\neq i}^{N}(a_j/k_j)]}{\sum_{j=1,\, j\neq i}^{N} k_j h_j D_j}}, \quad i = 1, \ldots, N. \tag{2.8}$$

**Algorithm 2.4** [Iterative heuristic Goyal]

Step 1: Choose as initial vector of frequencies $\mathbf{k} = \mathbf{1}$.

Step 2: For $i = 1, \ldots, N$, successively perform the following two operations: compute $\check{T}_i = \check{T}_i(\mathbf{k})$ according to (2.8) and determine $\tilde{k}_i(\check{T}_i)$ with the aid of (2.5).

Step 3: Stop if the new vector $\mathbf{k}$ is the same as the previous one; then, compute the final family cycle $T = \hat{T}(\mathbf{k})$ according to (2.2); otherwise, repeat from Step 2.

Alternatively, the result of a one-step heuristic can be used to obtain initial frequencies in Step 1. □

Table 2.2: Data for Example 2.3; indirect grouping policies by heuristics.

| Item | $a_i$ | $h_i$ | $D_i$ | $(A+a_i)/(h_iD_i)$ | $\tilde{k}_i(T^{(0)})$ | $\tilde{k}_i(T^{(1)})$ |
|---|---|---|---|---|---|---|
| 1 | \$ 3 | \$0.50 | 9 | 2.0 | 1 | 1 |
| 2 | \$ 3 | \$0.50 | 4 | 4.5 | 1 | 1 |
| 3 | \$43 | \$0.50 | 4 | 24.5 | 3 | 3 |

**Example 2.3** In this example, the foregoing heuristic algorithms are applied to the family of $N = 3$ items described in Example 2.1 (Table 2.1). The values of $(A + a_i)/(h_iD_i)$ are displayed in Table 2.2. Clearly, the reference item is item $i_r = 1$. The initial cycle length is $T^{(0)} = 2$, cf. Table 2.1. Then, the Algorithm 2.2 of Goyal & Belton sets $k_1 = 1$ and takes $k_2$ and $k_3$ as the rounded values of $\tilde{k}_2(2) = \frac{1}{2}\sqrt{3}$ and $\tilde{k}_3(2) = \frac{1}{2}\sqrt{43}$, cf. (2.4), (2.5), respectively. The rounded frequencies are listed in Table 2.2. This heuristic finishes by taking $T = T^{(1)} = \hat{T}(1, 1, 3) = 2.053$, and, hence, obtains the optimal policy in this case, cf. Example 2.1. Observe that the cost contribution of the reference item is not small in comparison with that of the other items in

this case. The additional step of Algorithm 2.3 of Kaspi & Rosenblatt does not change the policy in this case ($T^{(2)} = T^{(1)}$). Algorithm 2.4 starts with $\mathbf{k} = \mathbf{1}$ and computes $\check{T}_1(1,1,1) = \sqrt{26}$, $\hat{k}_1(\sqrt{26}) = \sqrt{2/39}$, $\tilde{k}_1(\sqrt{26}) = 1$, $\check{T}_2(1,1,1) = 4$, $\hat{k}_2(4) = \frac{1}{4}\sqrt{3}$, $\tilde{k}_2(4) = 1$, $\check{T}_3(1,1,1) = 4\sqrt{3/13}$, $\hat{k}_3(4\sqrt{3/13}) = \frac{1}{4}\sqrt{559/3} \approx \sqrt{11.65}$, $\tilde{k}_3(4\sqrt{3/13}) = 3$, which is the optimal vector of frequencies. □

**Exercise 2.4** *A retailer purchases six items from the same regional warehouse. The joint ordering cost is* $A = \$10$. *The supplementary ordering cost* $a_i$*, the holding cost* $h_i$ *per unit per month and the constant demand rate* $D_i$ *per month are given in the table below, for items* $i = 1, \ldots, 6$.

| item $i$ | $a_i$ | $h_i$ | $D_i$ |
|---|---|---|---|
| 1 | \$2 | \$0.20 | 2000 |
| 2 | \$2 | \$0.80 | 500 |
| 3 | \$1 | \$0.40 | 1000 |
| 4 | \$2 | \$0.80 | 2000 |
| 5 | \$2 | \$0.20 | 500 |
| 6 | \$1 | \$0.40 | 2000 |

*Determine reorder policies according to the heuristic of Kaspi & Rosenblatt, and according to the iterative heuristic of Goyal.*

### 2.1.3 Powers-of-two policies

A subclass of the class of indirect grouping policies is formed by the so called powers-of-two policies. In such a strategy the reorder cycle time of each item is a nonnegative integer power of two times a given basic cycle or review period $R$. The objective function to be minimized becomes for this subclass of policies

$$C(R, \mathbf{k}) = \frac{1}{R}\left[\frac{A}{k_0} + \sum_{i=1}^{N} \frac{a_i}{k_i}\right] + \tfrac{1}{2}R \sum_{i=1}^{N} k_i h_i D_i; \tag{2.9}$$

here, $k_0 \doteq \min_i\{k_i\}$ is not necessarily equal to 1, and the frequencies are restricted to $k_i \in \{2^\ell;\ \ell = 0, 1, 2, \ldots\}$, $i = 1, \ldots, N$. Jackson et al. [42] have developed an efficient algorithm for determining optimal reorder frequencies given the review period $R$.

**Algorithm 2.5** [Powers-of-two policy]

Step 1: Order and renumber the items according to ascending values of $a_i/(h_i D_i)$.

Step 2: Determine $i_0$ as the largest index $i$ for which

$$\frac{A + \sum_{j=1}^{i} a_j}{\sum_{j=1}^{i} h_j D_j} \geq \frac{a_i}{h_i D_i}. \tag{2.10}$$

Step 3: Determine $k_0$ as the smallest power of two that is larger than or equal to

$$\frac{1}{R}\sqrt{\frac{A + \sum_{j=1}^{i_0} a_j}{\sum_{j=1}^{i_0} h_j D_j}}. \tag{2.11}$$

Step 4: Set $k_i = k_0$ for $i = 1, \ldots, i_0$.

Step 5: Determine, for $i = i_0 + 1, \ldots, N$, $k_i$ as the smallest power of two that is larger than or equal to

$$\frac{1}{R}\sqrt{\frac{a_i}{h_i D_i}}. \tag{2.12}$$

Optionally, the last three steps can be repeated for various values of $R$ to find a suitable value of $R$. □

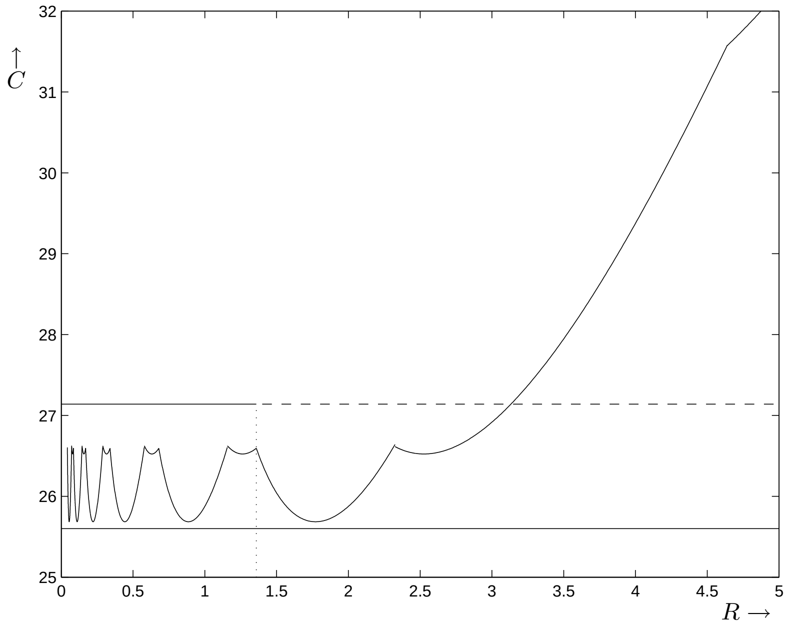

Figure 2.1: Minimum cost of power-of-two policies as function of the review period $R$.

Jackson et al. [42] have proved that the inventory cost (2.9) corresponding to the powers-of-two policy generated by their algorithm cannot exceed the minimum cost for the coordinated replenishment problem by more than 6%, provided that the quantity in (2.11) is larger than 1 (note that the review period $R$ can be decreased to satisfy this condition, if necessary). For this purpose they showed that a lower bound for the minimum cost over any joint replenishment policy is

$$C^* \geq C_{\text{lb}} \doteq \sqrt{2\left[A + \sum_{j=1}^{i_0} a_j\right] \sum_{j=1}^{i_0} h_j D_j} + \sum_{j=i_0+1}^{N} \sqrt{2 a_j h_j D_j}. \tag{2.13}$$

Based on the rounding to powers of two in (2.11) and (2.12) it can be shown that for the frequencies $\mathbf{k}^*$ determined by Algorithm 2.5 it holds that

$$C(R, \mathbf{k}^*) \leq (\sqrt{2} + \tfrac{1}{\sqrt{2}})\tfrac{1}{2} C_{\text{lb}} \leq \tfrac{3}{4}\sqrt{2}\, C^* \approx 1.06\, C^*, \quad \text{if} \ \ R \leq \sqrt{\left[A + \sum\nolimits_{j=1}^{i_0} a_j\right] \Big/ \sum\nolimits_{j=1}^{i_0} h_j D_j}\,. \tag{2.14}$$

Table 2.3: Powers-of-two policies for the family of Example 2.1.

| Item | $a_i$ | $h_i$ | $D_i$ | $a_i/(h_iD_i)$ | Index | LHS (2.10) | $k_i\,(R=1)$ | $k_i\,(R=2)$ | $k_i\,(R=4)$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | \$ 3 | \$0.50 | 9 | 0.667 | 1 | 2.000 | 2 | 1 | 1 |
| 2 | \$ 3 | \$0.50 | 4 | 1.500 | 2 | 1.846 | 2 | 1 | 1 |
| 3 | \$43 | \$0.50 | 4 | 21.500 | 3 | 6.471 | 8 | 4 | 2 |

**Example 2.4** In this example, the Powers-of-two Algorithm 2.5 is applied to the family of $N = 3$ items described in Example 2.1 (Table 2.1). The values of $a_i/(h_iD_i)$ are displayed in Table 2.3. Clearly, the items do not have to be renumbered in this case. The lefthand sides (LHS) of the inequalities (2.10) are also displayed in Table 2.3. It is seen that $i_0 = 2$ is the largest index for which (2.10) holds. The frequencies $k_i$, $i = 1, 2, 3$, are computed by (2.11): $k_1 = k_2 = k_0$, and by (2.12) for $k_3$, for a review period of $R = 1$, $R = 2$ and $R = 4$, respectively. The lower bound (2.13) has in this example the value $C_{\text{lb}} = 2\sqrt{39} + 2\sqrt{43} = \$25.60$. The 6% guaranty, $C(R, \mathbf{k}^*) < \$27.14$, holds for $R \leq \sqrt{1.846} \approx 1.359$. The cost $C(1, (2, 2, 8)) = \$25.88$ is guaranteed below this bound, the cost $C(2, (1, 1, 4)) = \$25.88$ is still below this bound, but the cost $C(4, (1, 1, 2)) = \$29.38$ is not. Figure 2.1 shows the minimum cost of the power-of-two policies as function of the review period $R$, its lower bound and the 6% upper bound. The cost function $C(R, \mathbf{k})$, cf. (2.9), has a minimum for each vector of reorder frequencies. For moderate values of $R$, the cost function follows

a repetitive pattern with alternatingly a local minimum at \$26.52, corresponding to vectors **k** which are powers-of-two multiples of the vector $(1, 1, 2)$, and a global minimum at \$25.69, corresponding to vectors **k** which are powers-of-two multiples of the vector $(1, 1, 4)$. The (scaled) repetitive pattern is due to the fact that the minimum costs corresponding to the vectors $2\mathbf{k}$, $4\mathbf{k}$, $8\mathbf{k}$, …, are attained at values of $R$ which are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, …, of the value of $R$ for which the minimum cost corresponding to the vector **k** is attained. For larger values of $R$, the minimum cost becomes an increasing function. For $2.3 < R < 4.6$, the best reorder frequencies are $(1, 1, 2)$, while for $R > 4.6$, the best reorder frequencies are $(1, 1, 1)$. □

Lee & Yao [48] present some properties of the cost function (2.9) and a search algorithm for the globally optimal powers-of-two policy in which also the length of the review period $R$ is optimized. An obvious improvement is to replace an arbitrarily chosen review period $R$ by the optimal cycle length given the reorder frequencies determined by Algorithm 2.5, cf. (2.2), (2.9):

$$\hat{T}(\mathbf{k}) = \sqrt{\frac{2[A/k_0 + \sum_{j=1}^{N}(a_j/k_j)]}{\sum_{i=1}^{N} k_i h_i D_i}}. \tag{2.15}$$

**Exercise 2.5** *Consider again the six-item coordinated replenishment problem described in Exercise 2.4. Determine optimal reorder frequencies in the class of powers-of-two policies for review periods* $R = 0.04$, $R = 0.12$, *and* $R = 0.24$ *month, respectively. Compute the average monthly costs corresponding to these policies, and compare them with the lower bound (2.13) and the upper bound (2.14) on the minimum cost.*

### 2.1.4 Production lot-size problems

A related problem to the coordinated replenishment problem with external supplier is the production lot-size problem for various items on a single machine. Suppose that $N$ items are produced on the same machine. Assume that a cyclic policy is applied, that is, there is a family production cycle of length $T$, and item $i$ is produced in every $k_i$th cycle. This lot-sizing problem is solvable with the foregoing algorithms with the following modifications, cf. Graves [37]:

- $A$ denotes the major family set-up cost, which accounts for machine opportunity cost during a possible family set-up time $U$ and for family set-up labor and material cost;
- $a_i$ denotes the minor set-up cost for item $i$, $i = 1, \ldots, N$, which accounts for machine opportunity cost during a possible item set-up time $u_i$ and for item set-up labor and material cost;
- if the produced items do not become available as a batch at the end of a production run, but at a rate $p_i$ for item $i$ during the production run, replace $h_i$ by $h_i(1 - D_i/p_i)$, $i = 1, \ldots, N$, in all formulas for the coordinated replenishment problem, as in Section 1.2.2;
- the production time for item $i$, $i = 1, \ldots, N$, is $\tau_i = Q_i/p_i = k_i T D_i/p_i$ plus possibly a set-up time $u_i$, and the maximum inventory level is $I_{i,\max} = (p_i - D_i)Q_i/p_i$, cf. (1.10);
- when the (optimal) cycle length $T$ and the frequencies $k_i$ have been determined, check whether the production schedule is feasible,

  $$U + \sum_{i=1}^{N}(u_i + \tau_i) < T, \tag{2.16}$$

  that is, whether the production capacity in the busiest cycle is sufficient;
- optionally, schedule items with $k_i > 1$ not simultaneously in the same period (scheduling items with $k_i > 1$ such that cycle lengths for all items are constant is not always trivial or even possible).

A necessary condition for feasibility of a single machine, multi-item lot sizing problem is

$$\sum_{i=1}^{N} D_i/p_i < 1. \tag{2.17}$$

If this condition is not fulfilled, the capacity of the machine is insufficient to satisfy all demand. This condition does not take into account any set-up times. However, if this condition is fulfilled, there exists a feasible production schedule with all production frequencies $k_i = 1$, $i = 1, \ldots, N$, and a large enough production cycle $T$ such that the set-up times $u_i$ are negligible with respect to the production times $\tau_i$, $i = 1, \ldots, N$.

Table 2.4: Production lot-size data and policies for the family of Example 2.5.

| Item | $a_i$ | $h_i$ | $D_i$ | $p_i$ | $u_i$ | $D_i/p_i$ | $k_i$ | $\tau_i$ | $Q_i$ | $k_i$ | $\tau_i$ | $Q_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | \$10 | \$0.40 | 100 | 400 | 0.01 | 0.250 | 1 | 0.310 | 124 | 1 | 0.310 | 124 |
| 2 | \$15 | \$0.50 | 150 | 1000 | 0.01 | 0.150 | 1 | 0.186 | 186 | 1 | 0.186 | 186 |
| 3 | \$50 | \$0.10 | 175 | 1000 | 0.01 | 0.175 | 2 | 0.434 | 434 | 2 | 0.434 | 434 |
| 4 | \$60 | \$0.10 | 250 | 2000 | 0.01 | 0.125 | 2 | 0.310 | 620 | 2 | 0.310 | 620 |
| 5 | \$60 | \$0.20 | 50 | 1000 | 0.01 | 0.050 | 3 | 0.186 | 186 | 4 | 0.248 | 248 |

**Example 2.5** Consider a family of five items that has to be manufactured on a single machine. The family set-up cost is $A = \$50$ and the family set-up time is $U = 0.05$ week. Item-related data can be found in Table 2.4. First note that the fractions $D_i/p_i$ of machine capacity required by the item $i$ add up to 0.75, so that condition (2.17) is satisfied. Application of Goyal's Algorithm 2.1 or any of the heuristics described in Section 2.1.2, with $h_i$ replaced by $h_i(1 - D_i/p_i)$, $i = 1, \ldots, N$, yields the production schedule with $k_1 = k_2 = 1$, $k_3 = k_4 = 2$, $k_5 = 3$ and $T = 1.241$ week, with minimum weekly cost $C = \$241.79$. However, this schedule does not satisfy condition (2.16), because the total production and set-up time in a period when all five items are produced is 1.527 week, which exceeds the length of a production period $T = 1.241$ week. But by producing item 3 in odd periods and item 4 in even periods, the total production and set-up time has a maximum of 1.207 week in odd periods and 1.083 week in even periods, which are both feasible in a production period of $T = 1.241$ week. This production schedule is displayed in Figure 2.2 for a cycle of six periods. A problem with this schedule is that the intervals between the start of production runs of item 5 are not equal (to $3T$), because the production times of items 3 and 4 are not equal. Delaying the start of the production run of item 5 in period 4 would require an additional family set-up time $U$ and add a family set-up cost $A = \$50$ per $6T$ weeks. A better alternative is to shorten the production run of item 5 in period 1 and to prolong the production run of item 5 in period 4. Still an other option is to search for powers-of-two policies which do not exhibit these kind of irregularities. Algorithm 2.5 yields, for a review period of $R = 1.241$ week, the production schedule with $k_1 = k_2 = 1$, $k_3 = k_4 = 2$, $k_5 = 4$, and weekly cost $C = \$243.65$ (see also Figure 2.2). This schedule is feasible if items 3 and 4 are produced in different periods and if item 5 is produced in periods when item 4 is produced; the total production and set-up time is then maximally equal to 1.145. □

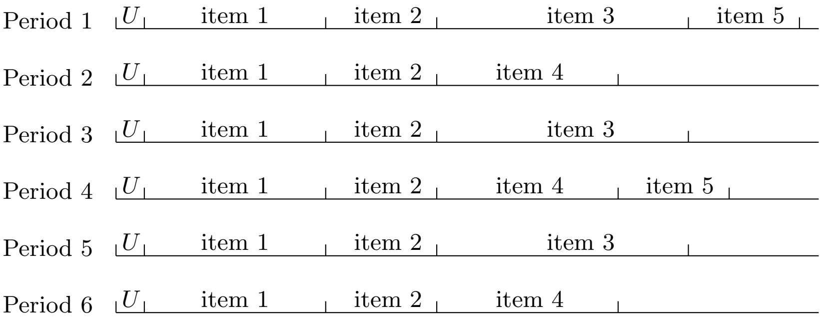

Figure 2.2: Production schedule for five items consisting of a cycle with six periods of length $T = 1.241$.

An other problem which can be solved with the aid of the algorithms for the coordinated replenishment problem will be discussed in Section 3.2.2.

**Exercise 2.6** *For the production schedule displayed in Figure 2.2, compute the modified production times and lot sizes for item 5 in periods 1 and 4 such that the inventory of this item reaches the level zero at the start of the production run. What is the increase in cost with respect to the (not realizable) minimum weekly cost* $C = \$241.79$. *What would be the actual cost if this schedule is applied with equal lot sizes in periods 1 and 4, implying that the inventory of this item does not reach the level zero at the start of the production run in period 4?*

**Exercise 2.7** *For the production scheduling problem described in Example 2.5, apply Algorithm 2.5 with a review period of* $R = 1.19$ *week and show that the corresponding cost is less than that with a review period of* $R = 1.241$ *week, while the production schedule can still be made feasible.*

**Exercise 2.8** *Consider a production scheduling problem for three items and a single machine with family set-up cost $20 and family set-up time 0.02 month. There are no item set-up times. The minor set-up cost $a_i$, the holding cost $h_i$ per unit per month, the demand rate $D_i$ per month and the production rate $p_i$ per month, $i = 1, 2, 3$, are given in the table below.*

| item $i$ | $a_i$ | $h_i$ | $D_i$ | $p_i$ |
|---|---|---|---|---|
| 1 | \$5 | \$0.40 | 400 | 1000 |
| 2 | \$25 | \$0.10 | 600 | 2000 |
| 3 | \$60 | \$0.25 | 50 | 500 |

*Determine a production schedule with the heuristic of Kaspi & Rosenblatt and verify whether this schedule can be made feasible. Search for a review period with which Algorithm 2.5 yields a production schedule that can be made feasible.*

### 2.1.5 All-units discounts

In Section 1.3 it has been noted that beside joint ordering costs also discounts may induce a company to coordinate replenishments of items that are ordered from the same supplier. In this section, indirect grouping for coordinated replenishments is considered for the case of joint ordering cost and a special type of discounts known as all-units discounts. More specifically, we will consider the case of an item-independent discount factor $d$, $0 < d < 1$. This means that the purchasing price per unit of item $i$ is as function of the family reorder cycle $T$ and the reorder frequencies $\mathbf{k}$, for $i = 1, \ldots, N$,

$$\begin{aligned} v_i(T,\mathbf{k}) &= v_{i0}, && \text{if } \textstyle\sum_{j=1}^N Q_j v_{j0} = T\sum_{j=1}^N k_j D_j v_{j0} < V, \\ v_i(T,\mathbf{k}) &= v_{i0}(1-d), && \text{if } \textstyle\sum_{j=1}^N Q_j v_{j0} = T\sum_{j=1}^N k_j D_j v_{j0} \geq V; \end{aligned} \tag{2.18}$$

here, $v_{i0}$ stands for the basic purchasing price of item $i$ (without discount), $Q_i = k_i T D_i$ is the order quantity of item $i$ at a certain reorder instant, $i = 1, \ldots, N$, and $V$ is a price break point for obtaining discount. Since the purchasing costs are no longer constant, they have to be included in the objective function. The holding costs are represented as $h_i = r v_i(T,\mathbf{k})$, $i = 1, \ldots, N$. Hence, the objective function becomes, with family cycle $T$, ordering frequencies $\mathbf{k}$ and $v_i(T,\mathbf{k})$ defined by (2.18),

$$C(T,\mathbf{k}) = \frac{1}{T}\left[A + \sum_{i=1}^N \frac{a_i}{k_i}\right] + \tfrac{1}{2}T\sum_{i=1}^N r v_i(T,\mathbf{k}) D_i k_i + \sum_{i=1}^N v_i(T,\mathbf{k}) D_i. \tag{2.19}$$

Silver & Peterson developed an heuristic for indirect grouping of items under the above described discount structure. They make the simplifying assumption that the discount is only awarded if the value of all orders exceeds the price break point $V$. Under this assumption, the crucial reorder instants are those at which only items $i$ with $k_i = 1$ are ordered. The heuristic is based on the observation that the reorder frequencies obtained by Algorithm 2.2 do not depend on the discount factor $d$.

**Algorithm 2.6** [Silver & Peterson]

Step 1: Apply Algorithm 2.2 (heuristic of Goyal & Belton) with $h_i = r v_{i0}(1-d)$, $i = 1, \ldots, N$. Call the resulting policy $(T_d^{(1)}, \mathbf{k})$. Let $\mathcal{J} \doteq \{j;\, k_j = 1\}$. Verify whether

$$T_d^{(1)} \sum_{j\in\mathcal{J}} D_j v_{i0} \geq V. \tag{2.20}$$

If this condition is satisfied, the value (without discount) of all orders exceeds the price break point $V$, and the current policy $(T_d^{(1)}, \mathbf{k})$ is feasible and such that the discount is obtained. If this condition is not satisfied, continue with the next steps.

Step 2: Increase the family cycle length $T_d^{(1)}$ to the value $T_d^{(V)}$ which is minimally required for obtaining discount, and, hence, is determined by:

$$T_d^{(V)} \sum_{j\in\mathcal{J}} D_j v_{i0} = V. \tag{2.21}$$

Compute the corresponding cost $C(T_d^{(V)}, \mathbf{k})$, cf. (2.19) with $v_i(T_d^{(V)}, \mathbf{k}) = v_{i0}(1-d)$, $i = 1, \ldots, N$.

Step 3: Redo Step 4 of Algorithm 2.2 with $h_i = rv_{i0}$, $i = 1, \ldots, N$, to obtain cycle length $T^{(1)}$. Compute the corresponding cost $C(T^{(1)}, \mathbf{k})$, cf. (2.19) with $v_i(T^{(1)}, \mathbf{k}) = v_{i0}$, $i = 1, \ldots, N$.

Step 4: Use the policy of Step 2 if $C(T_d^{(V)}, \mathbf{k}) < C(T^{(1)}, \mathbf{k})$, and use the policy of Step 3 otherwise.

The algorithm can be extended to situations with several discount factors and corresponding price break points, cf. Section 1.2.3. □

**Remark 2.7** If there are several items $i$ with $k_i > 1$, it may be possible to divide the reorder instants of these items such that the items in the above defined set $\mathcal{J}$ are never ordered alone. Then, the set $\mathcal{J}$ can be enlarged in the foregoing algorithm, and the discount price break point is more easily reached. However, when $N$ is large there are many ways to divide these items. Therefore, this possibility is not considered in the algorithm. □

**Example 2.6** Suppose that the holding costs in Example 2.1 stem from a carrying charge of $r = 0.002$ per dollar inventory per week times a purchasing price of \$250 for all three items. Further, suppose that the supplier of these items offers an all-units discount of $d = 0.04$ if the value of each order exceeds $V =$ \$10,000. This means that the discount purchasing price is \$240 per item. Application of Algorithm 2.2 with $h_i =$ \$0.48, $i = 1, 2, 3$, yields the same frequencies as with $h_i =$ \$0.50 (Example 2.3), namely $\mathbf{k} = (1, 1, 3)$, but a cycle length $T_d^{(1)} = 2.095$. The set of items with smallest order frequency is $\mathcal{J} = \{1, 2\}$. The value of the smallest orders is $250 \cdot (9 + 4) \cdot T_d^{(1)} =$ \$6,808.64, cf. (2.20). Hence, this policy is not feasible. The cycle length that is minimally required for obtaining discount is $T_d^{(V)} = 3.077$, cf. (2.21), and the corresponding cost is $C(T_d^{(V)}, \mathbf{k}) =$ \$4,107.02 per week, cf. (2.19). The cycle length according to Algorithm 2.2 with $h_i =$ \$0.50, $i = 1, 2, 3$, is $T^{(1)} = 2.053$. The corresponding cost is $C(T^{(1)}, \mathbf{k}) =$ \$4,275.66 per week. Hence, the policy proposed by Algorithm 2.6 is $\mathbf{k} = (1, 1, 3)$, and $T = T_d^{(V)} = 3.077$. □

**Exercise 2.9** *Consider the retailer and the six items from Exercise 2.4. Suppose that the holding costs stem from a carrying charge of* $r =$ \$0.02 *per dollar per month times the basic purchasing price* $v_{0i}$ *of item i,* $i = 1, \ldots, 6$. *Assume that the regional warehouse offers an all-units discount of* $d = 0.05$ *above the price break point* $V =$ \$20,000 *provided that the value of all orders exceeds this price break point. Determine an inventory policy with the aid of Algorithm 2.6.*

## 2.2 Constant demand, direct grouping

In this section we consider the same problem of a family of $N$ items with joint ordering cost structure and constant demand as in Section 2.1 but for the subclass of direct grouping policies. In this case, the $N$ items are divided into $M$ disjunct groups $\mathcal{G}_j$, $j = 1, \ldots, M$, and all members of group $\mathcal{G}_j$ have the same reorder cycle $T_j$, $j = 1, \ldots, M$. The reorder cycles of the various groups are determined independently of each other. Reorder instants of different groups may accidentally coincide, but such opportunities are not taken into account in the cost function.
The decision variables for the direct grouping problem are

- $M$: the number of groups;
- $\mathcal{G}_j$: the items in the $j$th groups, $j = 1, \ldots, M$;
- $T_j$: the group reorder cycle for items in $\mathcal{G}_j$, $j = 1, \ldots, M$.

The objective function to be minimized is, with $\mathbf{T} \doteq (T_1, \ldots, T_M)$,

$$C(\mathcal{G}_1, \ldots, \mathcal{G}_M, \mathbf{T}) = \sum_{j=1}^{M} \left[ \frac{1}{T_j} \left\{ A + \sum_{i \in \mathcal{G}_j} a_i \right\} + \tfrac{1}{2} T_j \sum_{i \in \mathcal{G}_j} h_i D_i \right]. \tag{2.22}$$

The optimal group cycles given the grouping $(\mathcal{G}_1, \ldots, \mathcal{G}_M)$ are readily determined in the standard way:

$$\hat{T}_j(\mathcal{G}_j) = \sqrt{\frac{2[A + \sum_{i \in \mathcal{G}_j} a_i]}{\sum_{i \in \mathcal{G}_j} h_i D_i}}, \quad j = 1, \ldots, M. \tag{2.23}$$

The corresponding minimum average inventory cost per unit of time for a given grouping (without accidentally coinciding orders) is:

$$\hat{C}(\mathcal{G}_1, \ldots, \mathcal{G}_M) = \sum_{j=1}^{M} \sqrt{2 \left[ A + \sum_{i \in \mathcal{G}_j} a_i \right] \sum_{i \in \mathcal{G}_j} h_i D_i}. \tag{2.24}$$

The number of possible groupings strongly increases with the size $N$ of the family. A reduction of the number of groupings that have to be considered for optimization is possible due to the so called consecutiveness property (cf. Chakravarty [14]):

- Renumber the items according to ascending value of $a_i/(h_i D_i)$.
- Then, the optimal groups $\mathcal{G}_1, \ldots, \mathcal{G}_M$ form ordered sets, that is, if item $a$ and item $b$ belong to the same group, say $\mathcal{G}_j$, then item $i$ belongs to $\mathcal{G}_j$ for all $i,\ a < i < b$.

**Remark 2.8** The ordering described by the consecutiveness property also appears in the optimal indirect grouping, cf. (2.4), and in the Powers-of-two Algorithm 2.5. □

**Example 2.7** A collection of consecutive sets is, for instance:

$$\mathcal{G}_1 = \{1, 2, 3\}, \quad \mathcal{G}_2 = \{4\}, \quad \mathcal{G}_3 = \{5, 6, 7, 8\}, \quad \mathcal{G}_4 = \{9, 10\}.$$

Put on a row, the numbers 1,. . . ,10 are counted off. □

### 2.2.1 Exact solution by dynamic programming

After renumbering the items according to the consecutiveness property, the optimal grouping can be determined by dynamic programming, cf. Chakravarty & Goyal [17]. In each step of the recursion, an additional item is considered. All possible groupings of this item with previously considered items satisfying the consecutiveness property are evaluated, while the optimal grouping of the remaining items has already been determined. To formulate the dynamic programming recursion define:

- $f(n)$: the minimum cost after the grouping of $n$ items;
- $\ell$: the number of items in the same group with item $n$.

The initial value is $f(0) = 0$, and the recursion reads, using one term of the minimum cost determined in (2.24), for $n = 1, \ldots, N$,

$$f(n) = \min_{1 \le \ell \le n} \left\{ f(n-\ell) + \sqrt{2 \left[ A + \sum_{i=n-\ell+1}^{n} a_i \right] \sum_{i=n-\ell+1}^{n} h_i D_i} \right\}. \tag{2.25}$$

In spite of the reduction realized by the consecutiveness property, the computation time is still strongly increasing with $N$.

**Example 2.8** In this example, the dynamic programming approach is applied to the family of $N = 3$ items described in Example 2.1. As showed in Example 2.4, the items do not have to be renumbered to satisfy the consecutiveness property. The algorithm starts with $f(0) = 0$, then considers item 1:

$$f(1) = \sqrt{2[A + a_1]\, h_1 D_1} = \$9.00.$$

Next, it determines whether it is better to take items 1 and 2 apart or together:

$$f(2) = \min \left\{ \begin{array}{l} f(1) + \sqrt{2[A + a_2]\, h_2 D_2} = \$15.00. \\ \sqrt{2[A + a_1 + a_2]\, [h_1 D_1 + h_2 D_2]} = \$12.49. \end{array} \right.$$

It turns out that it is advantageous to join items 1 and 2. Finally, item 3 is considered:

$$f(3) = \min \left\{ \begin{array}{l} f(2) + \sqrt{2[A + a_3]\, h_3 D_3} = \$26.49. \\ f(1) + \sqrt{2[A + a_2 + a_3]\, [h_2 D_2 + h_3 D_3]} = \$29.40. \\ \sqrt{2[A + a_1 + a_2 + a_3]\, [h_1 D_1 + h_2 D_2 + h_3 D_3]} = \$30.58. \end{array} \right.$$

Hence, it follows that the optimal direct grouping of this family is a division into two groups: $\mathcal{G}_1 = \{1, 2\}$, $\mathcal{G}_2 = \{3\}$, with minimum costs per week $C^* = \$26.49$ and with group cycles found with (2.23) as $T_1 = 1.92$ and $T_2 = 7$. Observe that the cycle length of group $\mathcal{G}_1$ is smaller than the optimal individual cycle lengths of items 1 and 2 (Table 2.1). Further, note that for this example the minimum cost with direct grouping is higher than the minimum cost with indirect grouping (Example 2.1). □

In Example 2.11 it is shown that the optimal grouping of items may follow an irregular pattern as function of the family ordering cost $A$.

**Exercise 2.10** *Determine the optimal direct grouping of the family of two items described in Example 2.2.*

### 2.2.2 Heuristic solution

Bastian [11] has developed an heuristic algorithm for the direct grouping of items in a family.

**Algorithm 2.7** [Bastian]

Step 1: Renumber the items according to ascending value of $a_i/(h_iD_i)$.

Step 2: Start with $M = N$ groups consisting of 1 item each.

Step 3: Determine for each pair of neighboring groups the decrease in cost (saving) when they would be joined, cf. (2.24), that is, compute for $j = 1, \ldots, M - 1$,

$$\sqrt{2\left[A+\sum_{i\in\mathcal{G}_j} a_i\right]\sum_{i\in\mathcal{G}_j} h_iD_i}+\sqrt{2\left[A+\sum_{i\in\mathcal{G}_{j+1}} a_i\right]\sum_{i\in\mathcal{G}_{j+1}} h_iD_i}-\sqrt{2\left[A+\sum_{i\in\mathcal{G}_j\cup\mathcal{G}_{j+1}} a_i\right]\sum_{i\in\mathcal{G}_j\cup\mathcal{G}_{j+1}} h_iD_i}.$$

Step 4: Actually join the two groups with the largest saving, decrease $M$ by 1 and repeat from Step 3 if the largest saving is positive; otherwise, stop.

Step 5: Finally, determine the optimal reorder cycles for the established groups with the aid of (2.23).

The algorithm is readily adapted to situations in which the number of groups is prescribed; then, the algorithm stops if the desired number of groups has been reached, possibly while there still is positive saving or after a few steps with increasing cost. □

The heuristic of Bastian is a greedy algorithm: always accept the largest saving. Note that the cost difference in Step 3 only has to be computed in the second and later iterations as far as the group is involved that has been formed in the previous iteration.

**Example 2.9** In this example, the heuristic of Bastian is applied to the family of $N = 3$ items described in Example 2.1. As noted in Example 2.8, no renumbering is required. The algorithm starts with three groups: $\mathcal{G}_1 = \{1\}$, $\mathcal{G}_2 = \{2\}$ and $\mathcal{G}_3 = \{3\}$. The saving of joining the first two groups is $9 + 6 - 12.49 = \$2.51$, that of joining the latter two groups is $6 + 14 - 20.40 = -\$0.40$. Since only the saving of joining the first two groups is positive, those two groups are actually joined, and the new groups are $\mathcal{G}_1 = \{1, 2\}$, $\mathcal{G}_2 = \{3\}$. In the final step the saving of joining these two groups is computed: $12.49 + 14 - 30.58 = -\$4.09$. Since this saving is negative, the groups are not joined. In this example, the heuristic of Bastian finds the optimal direct grouping, and the reorder cycles are the same as in Example 2.8. □

**Exercise 2.11** *A company purchases three items a, b and c from the same supplier. With every order, a fixed amount of \$5 has to be paid. Further, item dependent ordering costs are incurred: \$30 for item a, \$24 for item b, and \$50 for item c. The demand for these items is deterministic and constant in time. The demand per year is 4800 units for item a, 3600 units for item b, and 2500 for item c. The holding costs are equal for the three items, and amount to \$20 per unit per year. First, determine for each item separately the optimal order quantity, the optimal reorder cycle and the minimum cost per year. Next, determine the optimal direct grouping of these items with the aid of dynamic programming; also, compute the optimal reorder cycles and the corresponding minimum cost per year. Finally, determine a direct grouping of these items with the aid of the heuristic of Bastian.*

### 2.2.3 Direct grouping with discounts

Consider the all-units discount structure described in Section 2.1.5. A discount factor $d$ is awarded to the purchasing price of all items in an order if the value of the order exceeds a certain price break point $V$. With direct grouping the assumption is that the discount can be awarded per group.

An important observation is that the consecutiveness property remains valid. The optimal direct grouping policy can still be determined with the aid of dynamic programming, but the recursion becomes more complex than in Section 2.2.1, cf. Chakravarty [15]. With $f(n)$ the minimum cost after grouping of $n$ items, including purchasing cost, the recursion becomes: for $n = 1, \dots, N$,

$$f(n) = \min_{1 \le \ell \le n} \{f(n-\ell) + G(n,\ell)\}; \tag{2.26}$$

here, $G(n,\ell)$ is the minimum cost of the group consisting of the $\ell$ items $n-\ell+1, \dots, n$. When determining $G(n,\ell)$, again three cases have to be distinguished (cf. Section 2.1.5):

- the globally optimal cycle length with discount; if this is not feasible then:
- the globally optimal cycle length without discount;
- and the price break point cycle length to obtain discount.

The application of the dynamic programming recursion (2.26) requires the repeated solution of these sub-problems. Combining the concepts behind (2.22) and (2.19) the average cost of the group consisting of the $\ell$ items $n-\ell+1, \dots, n$ are with group reorder cycle $T$:

$$G(n,\ell,T) = \frac{1}{T}\left\{A + \sum_{i=n-\ell+1}^{n} a_i\right\} + \tfrac{1}{2}T \sum_{i=n-\ell+1}^{n} r v_i(T) D_i + \sum_{i=n-\ell+1}^{n} v_i(T) D_i; \tag{2.27}$$

here, the purchasing price as function of the cycle length is: for $i = 1, \dots, N$,

$$\begin{aligned} v_i(T) &= v_{i0}, & &\text{if } T\textstyle\sum_{j=n-\ell+1}^{n} D_j v_{j0} < V, \\ v_i(T) &= v_{i0}(1-d), & &\text{if } T\textstyle\sum_{j=n-\ell+1}^{n} D_j v_{j0} \ge V. \end{aligned} \tag{2.28}$$

The globally optimal cycle lengths given that the group consists of items $n-\ell+1, \dots, n$ follow by (2.23).

**Example 2.10** The dynamic programming approach is applied to the family of $N = 3$ items described in Example 2.1 with the discount structure of Example 2.6. The algorithm starts with $f(0) = 0$, then considers item 1. With $h_1 = \$0.48$, the optimal individual cycle length is $T = 2.041$, and the value of the corresponding order quantity, \$4,592.79, is not feasible. The minimal cycle length for obtaining discount is $T = V/(v_{10}D_1) = 4.444$, and the corresponding average cost is $C = \$2{,}171.63$. With $h_1 = \$0.50$, the optimal individual cycle length is $T = 2$, and the corresponding average cost is $C = \$2{,}259.00$. Hence, the minimal cost for item 1 is

$$f(1) = G(1,1) = \$2{,}171.63.$$

Next, a similar analysis has to be performed for item 2 to compute $G(2,1)$, and for items 1 and 2 together to determine $G(2,2)$. First, consider item 2. With $h_2 = \$0.48$, the optimal individual cycle length is $T = 3.062$, and the value of the corresponding order quantity, \$3,061.86, is not feasible. The minimal cycle length for obtaining discount is $T = V/(v_{20}D_2) = 10$, and the corresponding average cost is $C = \$970.50$. With $h_2 = \$0.50$, the optimal individual cycle length is $T = 3$, and the corresponding average cost is $C = \$1{,}006.00$. Hence, $G(2,1) = \$970.50$. For items 1 and 2 together, the optimal cycle length with $h_1 = h_2 = \$0.48$ is $T = 1.961$, and the value of the corresponding order quantities, \$6,373.77, is not feasible. The minimal cycle length for obtaining discount is $T = V/(v_{10}D_1 + v_{20}D_2) = 3.077$, and the corresponding average cost is $C = \$3{,}133.50$. With $h_1 = h_2 = \$0.50$, the optimal cycle length is $T = 1.922$, and the corresponding average cost is $C = \$3{,}262.49$. Hence, $G(2,2) = \$3{,}133.50$. Then, the second dynamic programming step becomes

$$f(2) = \min \left\{ \begin{array}{l} f(1) + G(2,1) = \$3{,}142.12. \\ G(2,2) = \$3{,}133.50. \end{array} \right.$$

It is seen that it is best to join items 1 and 2 in one group. For the third step, the quantities $G(3,1)$, $G(3,2)$ and $G(3,3)$ have to be determined in a similar way. This is left to the reader. □

**Exercise 2.12** *Compute $f(3)$ for the family of three items considered in Example 2.10.*

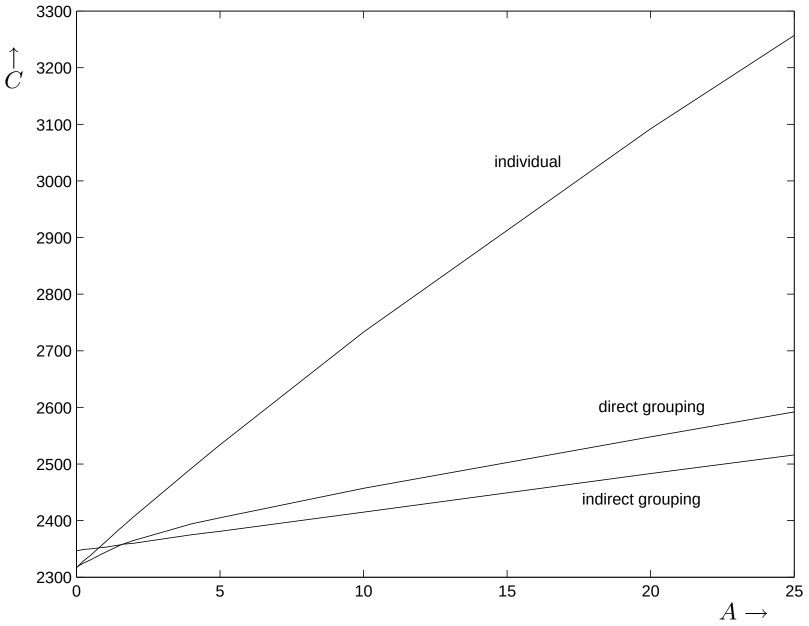

Figure 2.3: Minimum costs of individual, indirect grouping and direct grouping policies as functions of $A$.

### 2.2.4 Indirect versus direct grouping

Indirect grouping and direct grouping are two subclasses of the class of all possible policies for coordinated replenishments. The two extremal policies in this class are:

- order all items separately;
- order all items together.

All other policies are called mixed policies. The subclass of indirect grouping includes joint ordering of all items; the subclass of direct grouping includes both extremal policies. There exists no analytic way for determining which policy (from which subclass) is optimal in which situation. Therefore, simulation studies are required to compare the best policies in various subclasses. Here, simulation means that the parameters of a large number of inventory problems are sampled from some probability distributions (the problems are deterministic). It has turned out that the most important influence on the performance of the subclasses stems from the number of items in the family, $N$, and the ratio between the family ordering costs and the average individual ordering costs $A/\bar{a}$, with $\bar{a} \doteq \frac{1}{N}\sum_{i=1}^{N} a_i$. In general, indirect grouping gives higher savings with respect to separate ordering, between 30% and 70% with increasing ratio $A/\bar{a} \geq 1$. Direct grouping is only better if $A \ll \bar{a}$ but then the savings are generally less than 30%. See further Van Eijs et al. [73]. See also Silver et al. [64, Ch. 11].

**Example 2.11** Figure 2.3 shows the minimum costs of individually optimal policies without synchronization, of the optimal indirect grouping policies according to Algorithm 2.1 of Goyal, and of the optimal direct grouping policies according to dynamic programming as discussed in Section 2.2.1, for a family of $N = 8$ items, of which the data are summarized in Table 2.5, as functions of the family ordering cost $A$. Table 2.5 also contains the individually optimal reorder cycles $T_i(0),\ i = 1, \ldots, 8$ for $A = 0$. The optimal direct grouping policy consists of 8 single-item groups for $A = 0$; in the range from $A = 0.1$ to $A = 1.2$, items 2, 4 and 5 have joined a group, with a reorder cycle of about 0.15, and items 1 and 3 have joined a group, with a reorder cycle of about 0.18; from $A = 1.25$ to $A = 2.1$, items 1, 2, 3, 4 and 5 have joined one group, with a reorder cycle of about 0.16; from $A = 2.2$ to $A = 2.7$, items 2, 4, 5 and 6 have formed a group, with a reorder cycle of about 0.13, and items 1 and 3 have split off as a group, with a reorder cycle of about 0.19; from $A = 2.8$ to $A = 7.3$, these two groups have merged into one group, with a reorder cycle of about 0.14; from $A = 7.4$ to $A = 34$, items 7 and 8 have formed another group, with a reorder cycle of about 0.41; from $A = 35$ to $A = 70$, item 7 has made a transition to the large group, with a reorder cycle of about 0.17, while item 8 has formed a group on its own. From $A = 71$ onwards, also item 8 has joined the large group. Table 2.5 also contains the optimal reorder cycles $T_i(A),\ i = 1, \ldots, 8$ with direct grouping for $A = 0, 1, 2, 2.5, 5, 10, 50, 100$.

The optimal indirect grouping policy has reorder frequencies $k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = 1$, $k_7 = 2$ and $k_8 = 4$ in a broad range from $A = 0$ to about $A = 48.4$ where the reorder cycle increases from 0.143 to about 0.165; in the range from $A = 48.4$ to about $A = 189.4$, the only change is the reorder frequency of item 8: $k_8 = 3$; above $A = 189.4$, the reorder frequency of item 7 has decreased: $k_7 = 1$; in the range from $A = 216.7$ to just over $A = 1000$, the reorder frequency of item 8 has further decreased: $k_8 = 2$; for $A > 1000.2$, all reorder frequencies are equal to 1. Figure 2.3 shows that the optimal indirect grouping policy is worse than the individually optimal policies for $A$ smaller than about 0.8, and worse than the optimal direct grouping policy for $A$ smaller than about 1.6. For this family of items, direct grouping is optimal for $A$ smaller than about 1.6, and otherwise indirect grouping is optimal (for $A > 1000.2$, the optimal direct grouping policy is the same as the optimal indirect grouping policy). The difference between the individually optimal policies and the optimal policies with coordination, both indirect and direct grouping, strongly increases with increasing value of the family ordering cost $A$. □

Table 2.5: Data for Example 2.11; optimal reorder cycles for direct grouping policies.

| $i$ | $a_i$ | $h_i$ | $D_i$ | $a_i/(h_iD_i)$ | $T_i(0)$ | $T_i(1)$ | $T_i(2)$ | $T_i(2.5)$ | $T_i(5)$ | $T_i(10)$ | $T_i(50)$ | $T_i(100)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | \$15 | \$0.20 | 4000 | 0.019 | 0.194 | 0.184 | 0.157 | 0.188 | 0.144 | 0.147 | 0.175 | 0.205 |
| 2 | \$20 | \$1.00 | 2000 | 0.010 | 0.141 | 0.146 | 0.157 | 0.134 | 0.144 | 0.147 | 0.175 | 0.205 |
| 3 | \$22 | \$0.40 | 3600 | 0.015 | 0.175 | 0.184 | 0.157 | 0.188 | 0.144 | 0.147 | 0.175 | 0.205 |
| 4 | \$25 | \$0.80 | 3000 | 0.010 | 0.144 | 0.146 | 0.157 | 0.134 | 0.144 | 0.147 | 0.175 | 0.205 |
| 5 | \$28 | \$1.00 | 2500 | 0.011 | 0.150 | 0.146 | 0.157 | 0.134 | 0.144 | 0.147 | 0.175 | 0.205 |
| 6 | \$30 | \$1.50 | 3200 | 0.006 | 0.112 | 0.114 | 0.115 | 0.134 | 0.144 | 0.147 | 0.175 | 0.205 |
| 7 | \$35 | \$1.50 | 500 | 0.047 | 0.306 | 0.310 | 0.314 | 0.316 | 0.327 | 0.412 | 0.175 | 0.205 |
| 8 | \$40 | \$2.50 | 100 | 0.160 | 0.566 | 0.573 | 0.580 | 0.583 | 0.600 | 0.412 | 0.849 | 0.205 |

**Exercise 2.13** *Consider the family of eight items described in Example 2.11. Compare the policies and corresponding costs generated by the heuristic of Kaspi & Rosenblatt (algorithm 2.3) and the iterative heuristic of Goyal (algorithm 2.4) for $A = 5$.*

## 2.3 Time-varying demand

In this section a multi-item generalization of the model of Section 1.2.5 with time-varying deterministic demand will be discussed. Consider a family of $N$ items and a finite planning horizon of $H$ periods. The demand $D_{i,t}$ for item $i$ in period $t$ is assumed to be known ($i = 1, \ldots, N,\ t = 1, \ldots, H$). The cost factors are constant over the planning horizon. As in Sections 2.1, 2.2, $A$ denotes the family ordering cost, and $a_i$ denotes the supplementary ordering cost for item $i$, $i = 1, \ldots, N$. Orders are delivered at the beginning of a period. At the beginning of each period the demand for that period is set aside, and holding costs of $h_i$ dollar per unit of item $i$ per period are charged over the excess inventories. All demand must be fulfilled and no back orders are allowed. The decision variables are $Q_{i,t}$, the order quantity of item $i$ for the beginning of period $t$ ($i = 1, \ldots, N,\ t = 1, \ldots, H$). The inventory level of item $i$ at the end of period $t$ is denoted by $I_{i,t}$ ($i = 1, \ldots, N,\ t = 0, \ldots, H$). It is assumed that all initial inventory levels are zero: $I_{i,0} = 0$, $i = 1, \ldots, N$. The inventory levels can be described by the recursive equations; for $i = 1, \ldots, N$,

$$I_{i,t} = I_{i,t-1} + Q_{i,t} - D_{i,t}, \quad t = 1, \ldots, H. \tag{2.29}$$

These inventory levels must satisfy the requirements $I_{i,t} \geq 0$, $i = 1, \ldots, N,\ t = 1, \ldots, H$. The objective is to minimize the total cost of ordering and stock-keeping over $H$ periods:

$$C(\mathbf{Q}_1, \ldots, \mathbf{Q}_H) = \sum_{t=1}^{H} \left[ A\delta(\textstyle\sum_{i=1}^{N} Q_{i,t}) + \displaystyle\sum_{i=1}^{N} \{a_i\delta(Q_{i,t}) + h_i I_{i,t}\} \right], \tag{2.30}$$

here, $\mathbf{Q}_t \doteq (Q_{1,t}, \ldots, Q_{N,t})$ is the vector of order quantities for period $t$, and $\delta(.)$ is a dummy function, cf. (1.25). As in the single-item case the minimization of the cost function is facilitated by the observation that the optimal order quantities are zero or equal to the demand of a whole number of future periods. This observation implies the following properties of the optimal solution. For $i = 1, \ldots, N,\ t = 1, \ldots, H$,

1. item $i$ is only ordered for the beginning of period $t$ if the inventory has reached the level zero ($I_{i,t-1} = 0$);

2. the order quantity $Q_{i,t}$ is zero or equal to the demand in a whole number of future periods, that is, it is restricted to the set $\{0,\ D_{i,t},\ D_{i,t}+D_{i,t+1},\dots,\ D_{i,t}+\dots+D_{i,H}\}$;
3. the inventory level $I_{i,t-1}$ also is zero or equal to the demand in a whole number of future periods, that is, it is restricted to the same set $\{0,\ D_{i,t},\ D_{i,t}+D_{i,t+1},\dots,\ D_{i,t}+\dots+D_{i,H}\}$;
4. the demand $D_{i,t+\ell}$ is not included in the order in period $t$ if $h_i\ell D_{i,t+\ell} > A+a_i$ (then, to order in period $t+\ell$, if need be only item $i$, is cheaper than keeping $D_{i,t+\ell}$ $\ell$ periods in stock).

### 2.3.1 Exact solution

On the basis of the foregoing properties the minimization problem can be reformulated as an integer problem. Instead of the decision vectors $\mathbf{Q}_t$ and the similarly defined state vectors $\mathbf{I}_t$ we consider integer decision vectors $\mathbf{K}_t$ and integer state vectors $\mathbf{J}_t$ with components: for $i=1,\dots,N,\ t=1,\dots,H$,

- $K_{i,t}$: the order size for item $i$ for the beginning of period $t$ as number of periods demand for item $i$;
- $J_{i,t}$: the inventory level of item $i$ at the end of period $t$ as number of periods demand for item $i$.

The relevant values of these quantities vary with $t$, and are interrelated. Define: for $t=1,\dots,H$,

- $\mathcal{S}_t$: the relevant state space at the end of period $t-1$ consisting of vectors $\mathbf{J}_t$ with $0 \le J_{i,t} \le H-t+1$, $i=1,\dots,N$, and $J_{i,t}=0$ for at least one item $i$ (only reorder instants are relevant);
- $\mathcal{A}_t(\mathbf{J}_t)$: the relevant action space for the order at the beginning of period $t$ given the inventory levels $\mathbf{J}_t$; this space consists of vectors $\mathbf{K}_t$ with $K_{i,t}=0$ if $J_{i,t}>0$, and $0<K_{i,t}\le H-t+1$ if $J_{i,t}=0$, $i=1,\dots,N$.

Note that $\mathcal{S}_1=\{\mathbf{0}\}$. Further restrictions on the above spaces may be possible by property 4.

Table 2.6: State spaces (vectors) for the case $N=3$, $H=4$.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{S}_4$: | 000 | 100 | 010 | 001 | 110 | 101 | 011 | | | | | | | | | |
| $\mathcal{S}_3$: | 000 | 100 | 010 | 001 | 110 | 101 | 011 | 200 | 020 | 002 | 210 | 201 | 021 | 120 | 102 | 012 |
| | 220 | 202 | 022 | | | | | | | | | | | | | |
| $\mathcal{S}_2$: | 000 | 100 | 010 | 001 | 110 | 101 | 011 | 200 | 020 | 002 | 210 | 201 | 021 | 120 | 102 | 012 |
| | 220 | 202 | 022 | 300 | 030 | 003 | 310 | 301 | 031 | 130 | 103 | 013 | | | | |
| | 320 | 302 | 032 | 230 | 203 | 023 | 330 | 303 | 033 | | | | | | | |
| $\mathcal{S}_1$: | 000 | | | | | | | | | | | | | | | |

**Example 2.12** For the case $N=3$, $H=4$, the state spaces are displayed in Table 2.6. No states have been filtered out on the basis of property 4 because that step requires knowledge of the parameters of the model. As an example of an action space, consider $\mathcal{A}_2(2,0,0)$. Since an item is only ordered if its inventory level is zero, we have $K_1=0$. The inventory levels of items 2 and 3 are zero, so they must be ordered. For both items, the options are to order for 1, 2 or 3 periods. Hence, $\mathcal{A}_2(2,0,0)$ consists of 9 possible actions. □

The optimal solution can be determined with the aid of dynamic programming. To this end, define: for $t=1,\dots,H$,

- $f_t(\mathbf{J})$: the minimum cost incurred during periods $t,\dots,H$ starting from state $\mathbf{J}\in\mathcal{S}_t$ at the end of period $t-1$;
- $C_{t,\ell}(\mathbf{J},\mathbf{K})$: the cost incurred during periods $t,\dots,t+\ell-1$ given that the inventory levels at the end of period $t-1$ are $\mathbf{J}\in\mathcal{S}_t$ and given that the action $\mathbf{K}\in\mathcal{A}_t(\mathbf{J}_t)$ is taken:

$$C_{t,\ell}(\mathbf{J},\mathbf{K}) = A + \sum_{i=1}^{N}\left[a_i\delta(K_i) + h_i \sum_{u=1}^{J_i+K_i-1} \min\{u,\ell\}D_{i,t+u}\right]; \tag{2.31}$$

here, the next reorder instant $t+\ell$ is determined by

$$\ell \doteq \min_{i=1,\dots,N}\{J_i+K_i\}. \tag{2.32}$$

Note that an item $i$ with $J_i + K_i > \ell$ is not ordered in period $t+\ell$ and the demand for the periods $t+\ell, \ldots$ lies all $\ell$ periods on stock.
Silver [60] has developed a generalization of the Wagner-Whitin dynamic programming solution of 1-item problems to the case of a family of $N$ items.

**Algorithm 2.8** [Silver dynamic programming]

Step 1: Determine for each $t$ all relevant states $\mathbf{J}_t \in \mathcal{S}_t$ ($\mathbf{J}_1 = \mathbf{0}$) and all relevant actions $\mathbf{K}_t \in \mathcal{A}_t(\mathbf{J}_t)$.

Step 2: Set $f_{T+1}(\mathbf{0}) = 0$ and start with $t = H$.

Step 3: Apply the backward recursion: for $\mathbf{J} \in \mathcal{S}_t$,

$$f_t(\mathbf{J}) = \min_{\mathbf{K} \in \mathcal{A}_t(\mathbf{J}_t)} \left\{ f_{t+\ell}(\mathbf{J} + \mathbf{K} - \ell\mathbf{1}) + C_{t,\ell}(\mathbf{J},\mathbf{K}) \right\}; \tag{2.33}$$

Step 4: If $t > 1$ decrease $t$ by 1 and repeat from Step 3; otherwise, stop.

Finally, the optimal ordering scheme has to be interpreted (the integer vectors $\mathbf{K}_t$ have to be translated back to order quantities $\mathbf{Q}_t$). ☐

**Example 2.13** Consider the case $N = 2$, $H = 3$. Then for $t = H = 3$ there is no other choice than to order the lacking items, and only ordering costs are involved:

$$\begin{aligned} f_3(0,0) &= A + a_1 + a_2, \\ f_3(1,0) &= A + a_2, \\ f_3(0,1) &= A + a_1. \end{aligned}$$

For $t = H - 1 = 2$ there is some more choice, depending on the inventory levels:

$$f_2(0,0) = \min \begin{cases} C_{2,2}((0,0),(2,2)) = A + a_1 + a_2 + h_1 D_{1,3} + h_2 D_{2,3}, \\ f_3(0,1) + C_{2,1}((0,0),(1,2)) = f_3(0,1) + A + a_1 + a_2 + h_2 D_{2,3} \\ f_3(1,0) + C_{2,1}((0,0),(2,1)) = f_3(1,0) + A + a_1 + a_2 + h_1 D_{1,3}, \\ f_3(0,0) + C_{2,1}((0,0),(1,1)) = f_3(0,0) + A + a_1 + a_2; \end{cases}$$

$$f_2(1,0) = \min \begin{cases} f_3(0,1) + C_{2,1}((1,0),(0,2)) = f_3(0,1) + A + a_2 + h_2 D_{2,3}, \\ f_3(0,0) + C_{2,1}((1,0),(0,1)) = f_3(0,0) + A + a_2; \end{cases}$$

$$f_2(2,0) = \min \begin{cases} C_{2,2}((2,0),(0,2)) = A + a_2 + h_1 D_{1,3} + h_2 D_{2,3}, \\ f_3(1,0) + C_{2,1}((2,0),(0,1)) = f_3(1,0) + A + a_2 + h_1 D_{1,3}; \end{cases}$$

$$f_2(0,1) = \min \begin{cases} f_3(1,0) + C_{2,1}((0,1),(2,0)) = f_3(1,0) + A + a_1 + h_1 D_{1,3}, \\ f_3(0,0) + C_{2,1}((0,1),(1,0)) = f_3(0,0) + A + a_1; \end{cases}$$

$$f_2(0,2) = \min \begin{cases} C_{2,2}((0,2),(2,0)) = A + a_1 + h_1 D_{1,3} + h_2 D_{2,3}, \\ f_3(0,1) + C_{2,1}((0,2),(1,0)) = f_3(0,1) + A + a_1 + h_2 D_{2,3}. \end{cases}$$

Finally, for $t = 1$ there is only one relevant state but many options:

$$f_1(0,0) = \min \begin{cases} C_{1,3}((0,0),(3,3)) = A + a_1 + a_2 + h_1 D_{1,2} + 2h_1 D_{1,3} + h_2 D_{2,2} + 2h_2 D_{2,3}, \\ f_3(0,1) + C_{1,2}((0,0),(2,3)) = f_3(0,1) + A + a_1 + a_2 + h_1 D_{1,2} + h_2 D_{2,2} + 2h_2 D_{2,3}, \\ f_3(0,1) + C_{1,2}((0,0),(3,2)) = f_3(0,1) + A + a_1 + a_2 + h_1 D_{1,2} + 2h_1 D_{1,3} + h_2 D_{2,2}, \\ f_2(0,2) + C_{1,1}((0,0),(1,3)) = f_2(0,2) + A + a_1 + a_2 + h_2 D_{2,2} + h_2 D_{2,3}, \\ f_2(2,0) + C_{1,1}((0,0),(3,1)) = f_2(2,0) + A + a_1 + a_2 + h_1 D_{1,2} + h_1 D_{1,3}, \\ f_3(0,0) + C_{1,2}((0,0),(2,2)) = f_3(0,0) + A + a_1 + a_2 + h_1 D_{1,2} + h_2 D_{2,2}, \\ f_2(0,1) + C_{1,1}((0,0),(1,2)) = f_2(0,1) + A + a_1 + a_2 + h_2 D_{2,2}, \\ f_2(1,0) + C_{1,1}((0,0),(2,1)) = f_2(1,0) + A + a_1 + a_2 + h_1 D_{1,2}, \\ f_2(0,0) + C_{1,1}((0,0),(1,1)) = f_2(0,0) + A + a_1 + a_2. \end{cases}$$

Note that for the action $\mathbf{K} = (3,3)$ the holding costs are charged over all three periods so that the demands $D_{i,3}$ incur a holding cost of $2h_i$, $i = 1,2$. For the action $\mathbf{K} = (3,2)$ the holding costs are charged over two periods; this means that the demand $D_{1,3}$ incurs a holding cost of $2h_1$. For the action $\mathbf{K} = (3,1)$ the next reorder instant is $t = 2$ ($\ell = 1$) so that the holding costs are only charged over a single period. Therefore, the demand $D_{1,3}$ incurs a holding cost of $h_1$ over period 1; the holding cost over period 2 is included in $f_2(2,0)$.
☐

**Remark 2.9** The application of coordinated replenishment with time-varying demand with a rolling horizon is not as straightforward as in the 1-item case, cf. Remark 1.4. The foregoing dynamic programming solution and also the heuristics that will be discussed in the next section assume that the inventory levels of *all* items before the beginning of the first period are zero. The solution of the problem over $H$ periods may contain intermediate reorder instants at which only a few items are ordered, and perhaps none at which all items are ordered. This requires modifications of the algorithms. Also, cases with zero demand of some items in the first (few) periods requires modification of the ordering cost: $\delta(K_i)$ in (2.31) should be read as zero if the action $K_i$ only concerns periods of zero demand of item $i$, $i = 1, \ldots, N$. □

### 2.3.2 Heuristic algorithms

The exact solution discussed in the previous section is very time consuming in applications with many items and/or many periods. Therefore, several heuristics have been developed for this problem. To judge the quality of heuristics a good and simply computable lower bound for the minimum cost is useful. The idea to derive such a lower bound is to decompose the $N$-item problem into $N$ 1-item problems. To this end, the family ordering cost $A$ is divided over the $N$ items according to some weights $w_{i,t}$ ($\sum_{i=1}^N w_{i,t} = 1$). Since for all combinations of values of $Q_{i,t}$ and $w_{i,t}$ it holds that

$$\delta(\textstyle\sum_{i=1}^N Q_{i,t}) \geq \displaystyle\sum_{i=1}^N w_{i,t}\delta(Q_{i,t}), \tag{2.34}$$

it follows that the minimum of the cost function (2.30) has a lower bound which is a separable function:

$$\min_{Q_{i,t}} \sum_{t=1}^H \left[ A\delta(\textstyle\sum_{i=1}^N Q_{i,t}) + \displaystyle\sum_{i=1}^N \{a_i\delta(Q_{i,t}) + h_i I_{i,t}\} \right] \geq \sum_{i=1}^N \min_{Q_{i,t}} \sum_{t=1}^H \left[\{Aw_{i,t} + a_i\}\delta(Q_{i,t}) + h_i I_{i,t}\right]. \tag{2.35}$$

The lower bound can be computed by solving $N$ 1-item minimization problems. The lower bound depends on the choice of the weights so that various lower bounds could be generated by considering different combinations of weights.

The first heuristic to be discussed is a heuristic by Silver. It is a combination of the one-step heuristic for coordinated replenishment with constant demand (e.g., the Goyal & Belton heuristic) and the Silver-Meal heuristic (Algorithm 1.3) for 1-item problems with time varying demand.

**Algorithm 2.9** [Silver's heuristic]

Step 1: Determine the average demand over all periods of each item: $\bar{D}_i \doteq \frac{1}{H}\sum_{t=1}^H D_{i,t}$.

Step 2: Determine an indirect grouping of the items on the basis of these average demands, that is, determine a vector $\mathbf{k}$ of order frequencies according to Algorithm 2.2; divide the $N$ items into two groups: $i \in \mathcal{G}_1$ if $k_i = 1$ and $i \in \mathcal{G}_2$ if $k_i > 1$.

Step 3: Consider the items in $\mathcal{G}_1$ as one aggregated item with

$$\hat{a} \doteq A + \sum_{i\in\mathcal{G}_1} a_i, \qquad \hat{h} \doteq \sum_{i\in\mathcal{G}_1} h_i, \qquad \hat{D}_t \doteq \frac{1}{\hat{h}} \sum_{i\in\mathcal{G}_1} h_i D_{i,t}, \tag{2.36}$$

and apply the Silver-Meal heuristic to the aggregated item; this yields $H_1 \leq H$ reorder instants and determines the order quantities for items $i \in \mathcal{G}_1$.

Step 4: Apply the Silver-Meal heuristic to the items in $\mathcal{G}_2$ separately (with ordering costs $a_i$) with restriction to the $H_1$ reorder instants of $\mathcal{G}_1$.

Remember that periods with zero demand are skipped by the Silver-Meal heuristic. □

**Example 2.14** Consider a family consisting of $N = 3$ items that are ordered from the same supplier. The family ordering costs are $A = \$10$. The supplementary ordering costs $a_i$ and the holding cost $h_i$ can be found in Table 2.7, just like the demand $D_{i,t}$ over $H = 6$ weekly periods, $t = 1, \ldots, 6$, $i = 1, 2, 3$. Step 1 of Silver's heuristic is the computation of the average demand over all periods of each item: $\bar{D}_i$, $i = 1, 2, 3$. Step 2 is

Table 2.7: Demand in units per item per week; ordering and holding costs; reorder frequencies.

| Item | $a_i$ | $h_i$ | period $t$: | 1 | 2 | 3 | 4 | 5 | 6 | $\bar{D}_i$ | $\frac{A+a_i}{h_i\bar{D}_i}$ | $\frac{2a_i}{h_i\bar{D}_i(T^{(0)})^2}$ | $k_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 0.10 | $D_{1,t}$ | 10 | 0 | 25 | 40 | 30 | 15 | 20 | 7.50 | 3.13 | 2 |
| 2 | 2 | 0.30 | $D_{2,t}$ | 60 | 40 | 60 | 50 | 30 | 60 | 50 | 0.80 | — | 1 |
| 3 | 12 | 0.60 | $D_{3,t}$ | 20 | 25 | 20 | 40 | 25 | 20 | 25 | 1.47 | 1.00 | 1 |

the application of Algorithm 2.2: on the basis of the ratios $(A + a_i)/(h_i\bar{D}_i)$ the reference item is determined as $i_r = 2$. The initial family cycle becomes $T^{(0)} = \sqrt{\frac{8}{5}}$. Then, the reorder frequencies follow as $k_2 = 1$ and by (2.5), as $k_1 = 2$ and $k_3 = 1$. Hence, $\mathcal{G}_1 = \{2, 3\}$ and $\mathcal{G}_2 = \{1\}$. The data for the aggregated item of group $\mathcal{G}_1$ are: $\hat{a} = 24$, $\hat{h} = 0.90$, $\hat{h}\hat{D}_1 = 30$, $\hat{h}\hat{D}_2 = 27$, $\hat{h}\hat{D}_3 = 30$, $\hat{h}\hat{D}_4 = 39$, $\hat{h}\hat{D}_5 = 24$, $\hat{h}\hat{D}_6 = 30$. The Silver-Meal heuristic for the aggregated item yields: $\tau = 1$, $\bar{C}(1, 1) = 24$, $\bar{C}(1, 2) = \frac{1}{2}(24 + 27) > \bar{C}(1, 1)$, hence, $\tau = 2$, $\bar{C}(2, 1) = 24$, $\bar{C}(2, 2) = \frac{1}{2}(24 + 30) > \bar{C}(2, 1)$, hence, $\tau = 3$, $\bar{C}(3, 1) = 24$, $\bar{C}(3, 2) = \frac{1}{2}(24 + 39) > \bar{C}(3, 1)$, hence, $\tau = 4$, $\bar{C}(4, 1) = 24$, $\bar{C}(4, 2) = \frac{1}{2}(24 + 24) = 24 \leq \bar{C}(4, 1)$, $\bar{C}(4, 3) = \frac{1}{3}(24 + 24 + 2 \cdot 30) > \bar{C}(4, 2)$, so that $\tau = 6$. Hence, the Silver-Meal heuristic indicates that the items in group $\mathcal{G}_1$ should be ordered in every period except period 5. Next, the Silver-Meal heuristic is applied to the item in $\mathcal{G}_2$ in such a way that period 5 cannot become a reorder instant: $\tau = 1$, $\bar{C}(1, 1) = 5$, $\bar{C}(1, 2)$ is skipped since $D_{1,2} = 0$, $\bar{C}(1, 3) = \frac{1}{3}(5 + 2 \cdot 2.5) = \frac{10}{3} \leq \bar{C}(1, 1)$, $\bar{C}(1, 4)$ is skipped since demand for period 5 is coupled to period 4, $\bar{C}(1, 5) = \frac{1}{5}(5 + 2 \cdot 2.5 + 3 \cdot 4 + 4 \cdot 3) = \frac{34}{5} > \bar{C}(1, 3)$, hence, $\tau = 4$, $\bar{C}(4, 1)$ is skipped since demand for period 5 has to be ordered in period 4, $\bar{C}(4, 2) = \frac{1}{2}(5 + 3) = 4$, $\bar{C}(4, 3) = \frac{1}{3}(5 + 3 + 2 \cdot 1.5) = \frac{11}{3} < \bar{C}(4, 2)$. The ultimate order quantities are given in Table 2.8. The total cost over the 6 weeks is

$$5A + 2a_1 + 5a_2 + 5a_3 + h_1(2D_{1,3} + D_{1,5} + 2D_{1,6}) + h_2D_{2,5} + h_3D_{3,5} = \$165.$$

Note that if the Silver-Meal heuristic would not skip period 2 with zero demand for item 1, period 3 would become a reorder instant instead of period 4 and the total cost would be $168.50. □

Table 2.8: Order quantities according to Algorithm 2.9.

| period $t$: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $Q_{1,t}$ | 35 | – | – | 85 | – | – |
| $Q_{2,t}$ | 60 | 40 | 60 | 80 | – | 60 |
| $Q_{3,t}$ | 20 | 25 | 20 | 65 | – | 20 |

The next heuristic is called the coefficient method and has been proposed by Lambrechts et al. [47]. It is a generalization of the part-period-balancing heuristic (Algorithm 1.4) for individual items. Let $\tau_i$ denote the last reorder instant of item $i$ up to now, $i = 1, \ldots, N$. At period $u$ the heuristic computes coefficients $\alpha_{i,u}$ as the difference between the cumulative holding cost from the last reorder instant $\tau_i$ up to period $u$ and the ordering cost $a_i$:

$$\alpha_{i,u} = h_i \sum_{t=\tau_i}^{u} (t - \tau_i)D_{i,t} - a_i, \quad i = 1, \ldots, N. \tag{2.37}$$

Note that $\alpha_{i,\tau_i} = -a_i$ and that $\alpha_{i,u}$ is increasing in $u$. The principle underlying the heuristic is to compensate for ordering costs in a certain period by not ordering in some later periods. Orders are postponed as long as the coefficients $\alpha_{i,u}$ are negative or small. Since the compensation against later periods is not possible for the last period $H$ and, hence, the method would perform badly if the last period is to become a reorder instant, an additional step is added for period $H$ to see whether ordering in this period can be avoided.

**Algorithm 2.10** [Coefficient method]

Step 1: Start with $\tau_i = 1$ for each item $(i = 1, \ldots, N)$ and take $u = 2$.

Step 2: Compute the coefficients $\alpha_{i,u}$ for all items. Let $\mathcal{C} \doteq \{i;\ \alpha_{i,u} > 0\}$ be the set of items that are candidates for ordering. Actually order all items $i \in \mathcal{C}$ if

$$\sum_{i \in \mathcal{C}} \alpha_{i,u} > A. \tag{2.38}$$

If this condition is satisfied, then the order quantities become

$$Q_{i,\tau_i} = D_{i,\tau_i} + \cdots + D_{i,u-1}, \quad i \in \mathcal{C}, \tag{2.39}$$

and $\tau_i = u$ for $i \in \mathcal{C}$.

Step 3: If $u < H$ then increase $u$ by 1 and repeat from Step 2; otherwise, continue with the next step.

Step 4: If period $H$ has been indicated as reorder instant by Step 2, then for all items $i$ with $Q_{i,H} > 0$,

- Set $\tau_i$ equal to the last reorder instant of item $i$ before $H$. Add the demand $D_{i,H}$ to the order in period $\tau_i$ if
$$h_i(H - \tau_i)D_{i,H} < a_i. \tag{2.40}$$
- For all remaining items with $Q_{i,H} > 0$, let $\mathcal{C} \doteq \{i;\, Q_{i,H} > 0\}$ and shift the demand $D_{i,H}$ to the order in period $\tau_i$ if
$$\sum_{i\in\mathcal{C}} h_i(H - \tau_i)D_{i,H} < A + \sum_{i\in\mathcal{C}} a_i. \tag{2.41}$$

□

Table 2.9: Computations of coefficients for Algorithm 2.10.

| $u$ | $\alpha_{1,u}$ | $\alpha_{2,u}$ | $\alpha_{3,u}$ | decision |
|---|---|---|---|---|
| 1 | – | – | – | order items 1, 2 and 3 |
| 2 | −5 | 10 | 3 | order items 2 and 3 (13 > 10) |
| 3 | 0 | 16 | 0 | order item 2 (16 > 10) |
| 4 | 12 | 13 | 48 | order items 1, 2 and 3 (73 > 10) |
| 5 | −2 | 7 | 3 | order nothing (10 ≤ 10) |
| 6 | 1 | 43 | 27 | order items 1, 2 and 3 (71 > 10) |

**Example 2.15** The coefficient method will be illustrated for the data of Example 2.14, cf. Table 2.7. The results of the computations of the coefficients $\alpha_{i,u}$ can be found in Table 2.9, together with the corresponding ordering decisions. Since period $H = 6$ has been indicated as reorder instant by the first phase of the method (see the left part of Table 2.10) step 4 of the algorithm has to be performed. All three items are scheduled to be ordered in period $H = 6$ and period 4 is the previous reorder instant for all of them. The inequality $2h_iD_{i,6} < a_i$, cf. (2.40), is only satisfied for item 1. The demand $D_{1,6}$ is shifted to the order in period 4, saving \$2. The inequality $2h_2D_{2,6} + 2h_3D_{3,6} < A + a_2 + a_3$, cf. (2.41), is not satisfied so that period 6 remains a reorder instant for items 2 and 3. The modified ordering schedule can be found in the right part of Table 2.10. The total cost over the 6 weeks is

$$5A + 2a_1 + 5a_2 + 4a_3 + h_1(2D_{1,3} + D_{1,5} + 2D_{1,6}) + h_2D_{2,5} + h_3(D_{3,3} + D_{3,5}) = \$165.$$

Although the ordering schedule is different from that determined by Silver's heuristic, cf. Example 2.14, the costs are the same. Only an ordering cost of $a_3 = 12$ has been exchanged against a holding cost of $h_3D_{3,3} = 12$. □

Table 2.10: Order quantities according to Algorithm 2.10, after first phase and after final phase.

| period $t$: | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Q_{1,t}$ | 35 | – | – | 70 | – | 15 | 35 | – | – | 85 | – | – |
| $Q_{2,t}$ | 60 | 40 | 60 | 80 | – | 60 | 60 | 40 | 60 | 80 | – | 60 |
| $Q_{3,t}$ | 20 | 45 | – | 65 | – | 20 | 20 | 45 | – | 65 | – | 20 |

The last heuristic that will be discussed here has been proposed by Joneja [43]. It is called the cost-covering heuristic and is closely related to the coefficient method. It has been built on the alleged weakness of the coefficient method of ignoring items with a negative value of the coefficient $\alpha_{i,u}$, cf. (2.37), but close to zero, as candidates for joining a family order. Such items may trigger a next order in the very near future. Therefore, the cost-covering heuristic reconsiders items that were not included in a family order, and computes a second coefficient to determine whether an item should be included as yet to a previous order:

$$\beta_{i,u} = h_i(\tau_0 - \tau_i) \sum_{t=\tau_0}^{u} D_{i,t} - a_i, \quad u > \tau_0 \doteq \max_{1 \le i \le N} \tau_i, \quad i = 1, \ldots, N. \tag{2.42}$$

Observe that $\beta_{i,u} = -a_i < 0$ if $\tau_i = \tau_0$, that is, if item $i$ is already included in the last family order. If $\tau_i < \tau_0$, the coefficients $\beta_{i,u}$ represents the savings of shifting the quantity $D_{i,\tau_0} + \cdots + D_{i,u}$ over $\tau_0 - \tau_i$ periods by adding item $i$ to the family order at $\tau_0$.

**Algorithm 2.11** [Cost-covering heuristic]

Step 1: Start with $\tau_i = 1$ for each item ($i = 1, \ldots, N$) and take $u = 2$.

Step 2: Compute the coefficient $\beta_{i,u}$ for all items. For all $i$ with $\beta_{i,u} > 0$, item $i$ is included in the family order in period $\tau_0$, and $\tau_i$ becomes $\tau_0$.

Step 3: Compute the coefficient $\alpha_{i,u}$ for all items. Let $\mathcal{C} \doteq \{i;\, \alpha_{i,u} > 0\}$ be the set of items that are candidates for ordering. Actually order all items $i \in \mathcal{C}$ if

$$\sum_{i \in \mathcal{C}} \alpha_{i,u} > A. \tag{2.43}$$

If this condition is satisfied, then the order quantities become

$$Q_{i,\tau_i} = D_{i,\tau_i} + \cdots + D_{i,u-1}, \quad i \in \mathcal{C}, \tag{2.44}$$

and $\tau_i = u$ for $i \in \mathcal{C}$ and also $\tau_0 = u$.

Step 4: If $u < H$ then increase $u$ by 1 and repeat from Step 2; otherwise, stop: $\tau_i$ remains the last reorder instant of item $i$, $i = 1, \ldots, N$.

This heuristic does not treat period $H$ differently from other periods. □

Table 2.11: Computations of coefficients for Algorithm 2.11.

| $u$ | $\beta_{1,u}$ | $\beta_{2,u}$ | $\beta_{3,u}$ | decision | $\alpha_{1,u}$ | $\alpha_{2,u}$ | $\alpha_{3,u}$ | decision |
|---|---|---|---|---|---|---|---|---|
| 1 | – | – | – | — | – | – | – | order items 1, 2 and 3 |
| 2 | −5 | −2 | −12 | — | −5 | 10 | 3 | order items 2 and 3 (13 > 10) |
| 3 | −2.5 | −2 | −12 | no change | 0 | 16 | 0 | order item 2 (16 > 10) |
| 4 | 8 | −2 | 24 | add items 1 and 3 | −1 | 13 | 12 | order items 2 and 3 (25 > 10) |
| 5 | 2 | −2 | −12 | add item 1 | −2 | 7 | 3 | order nothing (10 ≤ 10) |
| 6 | −5 | −2 | −12 | — | 1 | 43 | 27 | order items 1, 2 and 3 (71 > 10) |

**Example 2.16** In this example, the cost-covering heuristic will be applied to the data of Example 2.14, cf. Table 2.7. The results of the computations of the coefficients $\beta_{i,u}$ and $\alpha_{i,u}$ can be found in Table 2.11, together with the corresponding ordering decisions. Observe that at $u = 4$ the items 1 and 3 are added to the order of period 3. This does not prevent, however, that period 4 becomes a reorder instant. At $u = 5$ item 1 is added to the order of period 4, and this does prevent period 5 of becoming a reorder instant. The ordering schedule can be found in Table 2.12. The total cost over the 6 weeks is

$$5A + 4a_1 + 5a_2 + 5a_3 + h_1 D_{1,5} + h_2 D_{2,5} + h_3 D_{3,5} = \$167.$$

This cost is \$2 higher than the costs of the previous two heuristics. □

Simulation studies have been performed in order to be able to determine the quality of the foregoing heuristics. A large number of problems have been generated, with varying proportion $A/\bar{a}$, horizon $H$, average demand and variance of demand over the $H$ periods. Each problem has been simulated with various realizations of the demand patterns. The problems have been solved exactly and with the heuristics.

Table 2.12: Order quantities according to Algorithm 2.11.

| period $t$: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $Q_{1,t}$ | 10 | – | 25 | 70 | – | 15 |
| $Q_{2,t}$ | 60 | 40 | 60 | 80 | – | 60 |
| $Q_{3,t}$ | 20 | 25 | 20 | 65 | – | 20 |

The coefficient method scores best on the average relative error over all problem instances, and on computation time. Only the maximum relative error is larger than for cost-covering heuristic. Joneja [43] has proved that the cost of the solution generated by the cost-covering heuristic is at most three times as high as the minimum cost. For the coefficient method, problem instances can be constructed for which the costs are an arbitrary factor higher than the minimum cost. The heuristic of Silver turns out to perform worst, especially in case of strongly fluctuating demand and with relatively small $A$. This may be related to the inflexibility of this heuristic with respect to items that are assigned to the group $\mathcal{G}_1$ and that are necessarily always jointly ordered. The average (maximum) relative errors found are 4% (54%) for the coefficient method, 8% (41%) for the cost-covering heuristic and 12% (300%) for the heuristic of Silver.
There exist alternative approaches; see Atkins & Iyogun [5], Iyogun [41], who also generalize the Silver-Meal heuristic. Further, multi-pass heuristics which improve initial solutions in later steps have been developed, cf. Kao [44]. These are generally better than single-pass heuristics but require more computation time.

**Exercise 2.14** *A manufacturer orders three part types from the same supplier. The joint ordering cost is \$30, the supplementary ordering costs are \$10 for all three part types. The holding costs are \$0.10, \$0.40 and \$0.40 per unit per week for part type 1, 2 and 3, respectively. The planned demand $D_{i,t}$ for part type $i$ in week $t$ has been listed in the table below for the three part types and for the next six weeks. All demand must be satisfied from stock.*

| week $t$: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $D_{1,t}$ | 36 | 36 | 36 | 36 | 36 | 36 |
| $D_{2,t}$ | 20 | 40 | 20 | 40 | 20 | 40 |
| $D_{3,t}$ | 8 | 4 | 10 | 10 | 4 | 8 |

*Determine an inventory policy for the three part types for the next six weeks with the aid of Silver's heuristic, with the aid of the coefficient method, and with the aid of the cost-covering heuristic, respectively. Compare the costs associated to the three policies.*

## 2.4 Stochastic demand

This section is devoted to coordinated replenishment problems with stochastic demand processes. The model assumptions for a family of $N$ items are:

- there is an infinite planning horizon;
- the ordering costs ($A,\ a_i,\ i = 1, \ldots, N$) are constant in time;
- the holding costs ($h_i,\ i = 1, \ldots, N$) are constant in time;
- the demand is stochastic with time independent parameters (e.g., a Poisson process for slow movers); the expected demand per unit of time is $E\{D_i\}$ and the standard deviation of the demand in a unit of time is $\sigma\{D_i\},\ i = 1, \ldots, N$;
- there is a constant lead time $L$ for orders as a whole, or there are item-dependent constant lead times $L_i,\ i = 1, \ldots, N$;
- any stockouts are backlogged (there is no stockout cost but a service level constraint for each item);
- there are no restrictions on order quantities.

The aim is the minimization of the expected average cost per unit of time subject to a service level constraint, cf. Section 1.1. Some policies for coordinated replenishment for a family of items that have been considered in the literature are (see Aksoy & Erenguc [1] and Goyal & Satir [36] for surveys):

- periodic review policies:
  - policies with a joint review period $R$ (with item-dependent order-up-to levels $S_i$: $(R, \mathbf{S})$, with item-dependent reorder points $s_i$ and order-up-to levels $S_i$: $(R, \mathbf{s}, \mathbf{S})$, with item-dependent reorder points $s_i$ and order quantities $Q_i$: $(R, \mathbf{s}, \mathbf{Q})$);
  - policies with coinciding review instants (with item-dependent reorder frequencies $k_i$ and order-up-to levels $S_i$: $(R, \mathbf{k}, \mathbf{S})$, related to indirect grouping);
  - policies with grouping and joint review periods $R_j$ per group (with item-dependent order-up-to levels $S_{ij}$: $(R_j, \mathbf{S}_j)$, related to direct grouping).
- continuous review policies:
  - policies that supplement orders with other items ("can-order" policies as introduced by Ballintfy [10], with item-dependent reorder points $s_i$, can-order levels $c_i$ and order-up-to levels $S_i$: $(\mathbf{s}, \mathbf{c}, \mathbf{S})$, or order quantities $Q_i$: $(\mathbf{s}, \mathbf{c}, \mathbf{Q})$);
  - policies that order on the basis of the total demand (place a family order when the total demand for all items since the last reorder instant exceeds $Q$, with item-dependent order-up-to levels $S_i$: $(Q, \mathbf{S})$, cf. Pantumsinchai [54]).

Section 2.4.1 is devoted to a continuous review "can-order" policy under Poisson demand processes. A heuristic for a periodic review policy related to indirect grouping will be discussed in Section 2.4.2. Literature on this subject includes Viswanathan [76] who considers an $(R, \mathbf{s}, \mathbf{S})$ policy, Wildeman et al. [78] who apply a global optimization procedure, Eynan & Kropp [26], Van Eijs [72] who considers joint ordering and transportation decisions, and Liu & Yuan [50] who study a system with correlated demand.

### 2.4.1 Can-order policies

This section is concerned with can-order policies. These are continuous review policies where the reorder decision can be taken after each transaction.
Every item has three decision variables: for $i = 1, \ldots, N$,

- $s_i$: the reorder point for item $i$ (the "must" order point);
- $c_i$: the join-order level for item $i$ (the "can" order level);
- $S_i$: order-up-to level for item $i$ (this level determines the order quantity).

These values are ordered as $s_i \leq c_i < S_i,\ i = 1, \ldots, N$. The corresponding policy is:

- order *when* the inventory position of an item $i$ falls at or below its level $s_i$;
- order all items $j$ with their inventory position at or below their level $c_j$ up to their level $S_j$.

The extremal values of the can-order level are:

- $c_i = s_i$: item $i$ is individually ordered
  (with ordering costs $A + a_i$; optimal in case $a_i \gg A$);
- $c_i = S_i - 1$: item $i$ is ordered at each reorder instant unless no demand has occurred
  (optimal in case $a_i \ll A$ and $h_i D_i$ relatively small).

The aim is the minimization of the expected average ordering and holding cost per unit of time subject to a service level constraint. The objective function can be formulated as:

$$\min \sum_{i=1}^{N} [M_i(A + a_i) + J_i a_i + h_i E\{\bar{I}_i\}]; \tag{2.45}$$

here, the auxiliary variables are: for $i = 1, \ldots, N$,

- $M_i$: expected number of times per year that item $i$ causes an order (by reaching its level $s_i$);

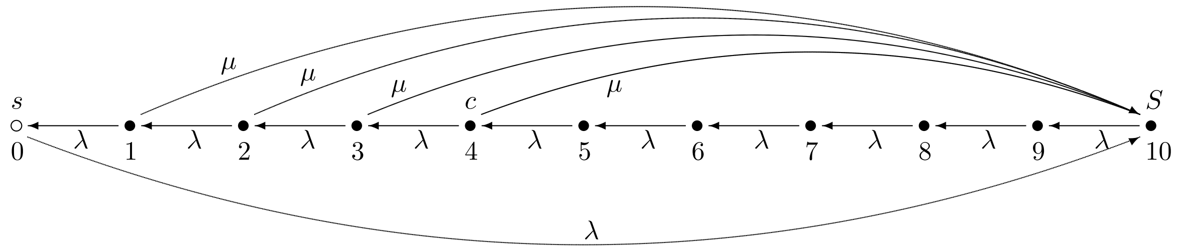

Figure 2.4: Transition diagram for the Markov inventory process of the can-order model.

- $J_i$: expected number of times per year that item $i$ supplements an order (when its inventory is at or below level $c_i$);
- $E\{\bar{I}_i\}$: expected average inventory level of item $i$.

These $3N$ auxiliary variables $M_i$, $J_i$ and $\bar{I}_i$, $i=1,\ldots,N$, are functions of the $3N$ decision variables $s_i$, $c_i$ and $S_i$, $i=1,\ldots,N$. The problem with $3N$ decision variables is too complex to solve as a whole. Therefore, the analysis will be restricted to the case of Poisson demand processes, with demand rate $\lambda_i$ units of item $i$ per unit of time, $i=1,\ldots,N$, and a decomposition assumption is introduced: the opportunities for item $i$ to join an order occur according to Poisson process with rate

$$\mu_i = \sum_{j=1,\, j\neq i}^{N} M_j, \quad i=1,\ldots,N. \tag{2.46}$$

This is a rough assumption, because the actual rates vary over time. By this assumption, the original problem with $3N$ decision variables is decomposed into $N$ problems with 3 decision variables each. The interaction between these $N$ problems is established via the opportunity rates $\mu_i$, $i=1,\ldots,N$.
The approximative approach of iterative solution globally consists of the following steps. First the case $L=0$ is considered. When the lead time $L=0$, orders are delivered immediately, and no safety stock is needed. Hence, we can take $s_i=0$ for $i=1,\ldots,N$ and still satisfy all demand directly from stock. Initial values for $M_i$ can be obtained on the basis of individual $(s_i,S_i)$ policies (in case $\mu_i=0$, $c_i=s_i=0$). Assuming that the demand per customer is 1 unit we have $S_i=Q_i$, the economic order quantity, and

$$M_i = \frac{E\{D_i\}}{Q_i} = \sqrt{\frac{h_i E\{D_i\}}{2(A+a_i)}} = \sqrt{\frac{h_i \lambda_i}{2(A+a_i)}}, \quad i=1,\ldots,N. \tag{2.47}$$

In the iteration steps, for each item successively, the opportunity rate $\mu_i$ is estimated on the basis of the last values of $M_j$, $j\neq i$, cf. (2.46); then, a 1-item problem is solved and a new value of $M_i$ is determined, $i=1,\ldots,N$. The iterations stop if the decision variables have not changed or if the cost saving is relatively small. Next, the case $L>0$ is considered. For each item separately, a safety stock is determined in such a way that the service constraint is satisfied. To this end, the values of the levels $s_i=0$, $c_i$ and $S_i$ found with $L=0$ are equally increased until the service constraint is met.
There exist various methods for solving the 1-item problems. We will discuss an approach developed by Silver [58]. For the ease of notation, the item index will be omitted in the following derivation. The assumptions of Silver for solving the 1-item problems are a zero lead time ($L=0$), a zero reorder point ($s=0$), a Poisson demand process with a single unit demand per customer (with demand rate $\lambda$ per unit of time), and a Poisson process of opportunities for joining an order (with rate $\mu$ per unit of time). For a single item, the expected total cost per unit of time is of the form:

$$C(c,S) = M(A+a) + Ja + h\bar{I} = MA + (M+J)a + hE\{\bar{I}\}, \tag{2.48}$$

with $M$, $J$ and $E\{\bar{I}\}$ functions of $c$ and $S$. Because both the demand and the opportunities for joining an order occur according to Poisson processes, the inventory level process is a Markov process. The state space is finite. Because an order is placed and immediately delivered as soon as the inventory level reaches $s=0$, the state $s=0$ is an instantaneous state, and can be omitted from the state space. Figure 2.4 shows the transition diagram of the inventory process for the case $c=4$, $S=10$. Let $p_j$ denote the stationary probability that the inventory level is $j$ units, $j=1,\ldots,S$. These state probabilities satisfy the following balance equations. The states between the levels $c$ and $S-1$ can only be reached from states with 1 unit more on stock and can only be left by the event of a demand (at rate $\lambda$):

$$\lambda p_j = \lambda p_{j+1}, \quad j = c+1, \ldots, S-1. \tag{2.49}$$

The states between the levels 1 and $c$ can only be reached from states with 1 unit more on stock but can be left by the event of a demand (at rate $\lambda$) and by the event of an opportunity to join an order (at rate $\mu$):

$$(\lambda + \mu) p_j = \lambda p_{j+1}, \quad j = 1, \ldots, c. \tag{2.50}$$

The state $S$ can be reached from the states between the levels 1 and $c$ by the event of an opportunity to join an order (at rate $\mu$) and from the state 1 by the event of a forced order (at rate $\lambda$, via an instantaneous visit to state $s=0$), and can only be left by the event of a demand (at rate $\lambda$):

$$\lambda p_S = \lambda p_1 + \mu \sum_{j=1}^{c} p_j. \tag{2.51}$$

Finally, the normalization of the probability distribution of the inventory level process implies that

$$\sum_{j=1}^{S} p_j = 1. \tag{2.52}$$

With the aid of the equations (2.49) and (2.50) it is possible to express all state probabilities in terms of the single probability $p_S$:

$$p_j = p_S, \quad j = c+1, \ldots, S; \qquad p_j = \rho^{c-j+1} p_S, \quad j = 1, \ldots, c; \tag{2.53}$$

here,

$$\rho \doteq \frac{\lambda}{\lambda + \mu}. \tag{2.54}$$

The normalization (2.52) determines the last unknown probability $p_S$ by substitution of (2.53):

$$p_S = \left[ S - c + \rho \, \frac{1-\rho^c}{1-\rho} \right]^{-1}. \tag{2.55}$$

Together, (2.53) and (2.55) specify the stationary distribution of the inventory level at an arbitrary instant. The quantities $M$, $J$ and $\bar{I}$ can now be specified as functions of $c$ and $S$. Since an order is caused by a particular item when it reaches the level $s=0$, and because the latter occurs at rate $\lambda$ when the inventory is in state 1, it follows that the expected number of orders per unit of time caused by the item is

$$M = \lambda p_1 = \lambda \rho^c p_S = \frac{\lambda \rho^c}{S - c + \rho \, \frac{1-\rho^c}{1-\rho}}. \tag{2.56}$$

Similarly, since opportunities for supplementing an order are assumed to occur at rate $\mu$ in the states $1, \ldots, c$, the expected number of orders per unit of time supplemented by the item is

$$J = \mu \sum_{j=1}^{c} p_j = \mu \rho \, \frac{1-\rho^c}{1-\rho} \, p_S = \frac{\lambda(1-\rho^c)}{S - c + \rho \, \frac{1-\rho^c}{1-\rho}}; \tag{2.57}$$

here, it has been used that $\mu\rho = \lambda(1-\rho)$ by (2.54). Together, (2.56) and (2.57) yield for the expected total number of orders per unit of time in which the item takes a part:

$$M + J = \frac{\lambda}{S - c + \rho \, \frac{1-\rho^c}{1-\rho}} = \lambda p_S. \tag{2.58}$$

Finally, the expected average inventory level is

$$E\{\bar{I}\} = \sum_{j=1}^{S} j p_j = \frac{\frac{1}{2}(S-c)(S+c+1) + \frac{\rho}{1-\rho}[c - \rho \, \frac{1-\rho^c}{1-\rho}]}{S - c + \rho \, \frac{1-\rho^c}{1-\rho}}. \tag{2.59}$$

Substitution of (2.56), (2.58) and (2.59) into (2.48) leads to an explicit expression of the expected total cost as function of the can-order level $c$ and the order-up-to level $S$:

$$C(c,S) = \frac{\lambda\rho^c A + \lambda a + \frac{1}{2}h(S-c)(S+c+1) + h\frac{\rho}{1-\rho}[c - \rho\,\frac{1-\rho^c}{1-\rho}]}{S - c + \rho\,\frac{1-\rho^c}{1-\rho}}. \tag{2.60}$$

The minimization of this cost function proceeds in two steps. For fixed $c$ it is a rather simple function of $S$. Taking the derivative with respect to $S$ and setting the latter equal to zero gives

$$\frac{d}{dS}\,C(c,S) = -C(c,S)\,p_S + h(S+\tfrac{1}{2})p_S = 0. \tag{2.61}$$

Solving this quadratic equation in $S$ formally gives

$$\hat{S}(c) = c - \rho\,\frac{1-\rho^c}{1-\rho} + \sqrt{\frac{2\lambda(a+\rho^c A)}{h} + \frac{2c\rho^{c+1}}{1-\rho} - \frac{\rho(1-\rho^c)(1+\rho^{c+1})}{(1-\rho)^2}}. \tag{2.62}$$

The expression of which the square root is taken is in some cases positive and then describes the optimal order-up-to level $\hat{S}(c)$ as function of $c$; otherwise it is negative, and then the cost function (2.60) has for fixed $c$ a boundary extreme at $S = c+1$. Next, substitution of the optimal value of $\hat{S}(c)$ into the cost function $C(c,S)$ gives a function of the single variable $c$. The determination of the optimal value of $c$ can be carried out by the "Golden-Section" method, among others; see Appendix B.2. Finally, substitution of the optimal can-order level $c$ into the function $\hat{S}(c)$ gives the optimal order-up-to level $S$.

**Remark 2.10** Application of the Golden-Section method requires the selection of an initial search interval. The lower bound is 0. An upper bound is rather arbitrary. One could try a multiple (1.5–2) of the items' EOQ value without coordination. Further, the Golden-Section method applies to real-valued functions. Note that the function $\hat{S}(c)$ can also be evaluated for noninteger values of $c$. Moreover, an integer value of $c$ does not give an integer value of $S$, in general. Hence, both $c$ and $S$ should be rounded afterwards. Alternatively, the Golden-Section method can be applied with rounded values, cf. Example B.1. ☐

Once the iterations for the case $L = 0$ have been carried out, a safety stock is determined for each item separately for the actual lead time $L > 0$. Let $c_0$ and $S_0$ denote the optimal values of the can-order level and the order-up-to level of an item found for $L = 0$. Starting from $s_0 = 0$, $c_0$ and $S_0$, the values of $s$, $c$ and $S$ are equally increased until the service constraint is met. Since the differences $c - s$ and $S - c$ do not change during this procedure, also the values of $M$ and $J$ do not change. Once the reorder point $s$ has been determined on the basis of the service measure, the final policy for the item is $(s, c_0+s, S_0+s)$. The resulting safety stock is equal to $s - \lambda L$ and gives rise to an extra holding cost of $h(s-\lambda L)$.

Next, the determination of the actual service level will be discussed for the two service level constraints described in Section 1.1. In both cases, the observation will be used that the stationary distribution of the inventory *position* for a $(s, c_0+s, S_0+s)$ policy is related to the stationary distribution of the inventory on hand for a $(0, c_0, S_0)$ policy by

$$\Pr\{I(s, c_0+s, S_0+s) = s+k\} = \Pr\{I(0, c_0, S_0) = k\}, \quad k = 1, 2, \ldots, S_0, \tag{2.63}$$

here, the latter distribution is given by (2.53) and (2.55). Further, the distribution of the inventory position $I_O$ just before an order for the item is placed is related to the distribution of the inventory position $I$ at arbitrary instants by Bayes' theorem. For the event that the item supplements an order it holds that

$$\Pr\{I_O = s+k\} = \frac{\mu p_k}{\mu(p_1 + \cdots + p_{c_0}) + \lambda p_1} = \frac{\mu p_k}{\lambda p_S} = (1-\rho)\rho^{c_0-k}, \quad k = 1, \ldots, c_0; \tag{2.64}$$

for the event that the item causes an order to be placed it holds that

$$\Pr\{I_O = s\} = \frac{\lambda p_1}{\mu(p_1 + \cdots + p_{c_0}) + \lambda p_1} = \frac{\lambda p_1}{\lambda p_S} = \rho^{c_0}. \tag{2.65}$$

Let $D_L$ denote the demand during the lead time $L$. For a cycle service level $\alpha$, consider the probability $\Pi$ of no stockout in a reorder cycle. By the law of total probability it holds that:

$$\Pi \doteq \Pr\{D_L \le I_O\} = \sum_{m=s}^{c} \Pr\{D_L \le m \mid I_O = m\}\Pr\{I_O = m\}. \tag{2.66}$$

Due to the lack of memory of the Poisson demand process it follows that $D_L$ is independent of $I_O$, and that $D_L$ has a Poisson distribution with mean $\lambda L$. Hence, the probability $\Pi$ is completely determined by (2.64), (2.65) and

$$\Pi = \sum_{k=0}^{c_0} \Pr\{I_O = s + k\} \sum_{j=0}^{s+k} \frac{(\lambda L)^j}{j!}\, e^{-\lambda L}. \tag{2.67}$$

Hence, the reorder point $s$ can be determined for which the service level constraint $\Pi \geq \alpha$ is met. One can start with $s = 0$ and increase $s$ by 1 until the constraint is met, or search in the neighborhood of $s = \lambda L$; see also Remark 1.5.

For a fill rate constraint with target $\beta$, consider the fraction of demand satisfied from stock (the actual fill rate) in a reorder cycle:

$$\Psi = 1 - E\{B\}/E\{Q\}; \tag{2.68}$$

here, $E\{B\}$ denotes the expected backlog that accumulates per cycle, and $E\{Q\}$ denotes the mean order quantity which is equal to the expected demand per cycle since there are no lost sales. The mean order quantity does not depend on the reorder point $s$, and, therefore, has to be computed only once:

$$E\{Q\} = \sum_{m=s}^{c} (S - m) \Pr\{I_O = m\} = \sum_{k=0}^{c_0} (S_0 - k) \Pr\{I_O = s + k\} = S_0 - (1 - \rho) \sum_{k=1}^{c_0} k \rho^{c_0 - k}. \tag{2.69}$$

The mean accumulated backlog during a cycle is given by

$$E\{B\} = E\{[D_L - I_O]^+\} - E\{[D_L - S]^+\}. \tag{2.70}$$

The last term, the mean backlog at the beginning of a cycle, can be computed in the standard manner, cf. Example 1.3. For the first term, the mean backlog at the end of a cycle, it follows again due to the independence of $D_L$ and $I_O$ that

$$E\{[D_L - I_O]^+\} = \sum_{m=s}^{c} \Pr\{I_O = m\} \sum_{j=m}^{\infty} (j - m) \frac{(\lambda L)^j}{j!}\, e^{-\lambda L} = \sum_{k=0}^{c_0} \Pr\{I_O = s + k\} \sum_{j=s+k}^{\infty} (j - s - k) \frac{(\lambda L)^j}{j!}\, e^{-\lambda L}, \tag{2.71}$$

with again $\Pr\{I_O = s + k\}$ given by (2.64) and (2.65). The infinite sum in (2.71) can be computed in a similar way as in Example 1.3. Hence, the reorder point $s$ can be determined for which the fill rate constraint $\Psi \geq \beta$ is met, in a similar way as for the cycle service constraint. The complete heuristic for determining a can-order policy under the assumptions imposed by Silver [58] is summarized below.

**Algorithm 2.12** [Can-order policy for Poisson demand]

Step 1: Take $L = 0$ and $s_i = 0$ for all $i$. Compute initial values of the mean number of orders caused by item $i$, $M_i$, based on deterministic demand, cf. (2.47), for $i = 1, \ldots, N$. Start with $i = 1$.

Step 2: Estimate the rate $\mu_i$ on the basis of the current values of $M_j$, $j \neq i$, $j = 1, \ldots, N$, cf. (2.46). Determine the optimal $c_i$ and $S_i$ that minimize the cost function $C(c, S)$ for item $i$, cf. (2.60). The new value of $M_i$ follows from (2.56).

Step 3: If $i < N$, increase $i$ by 1 and repeat from Step 2; otherwise, verify whether the solution has converged; if not, set $i = 1$ and repeat from Step 2; otherwise, fix for each item the last found values of $c_i$ and $S_i$ as $c_i^{(0)}$ and $S_i^{(0)}$, and continue with Step 4.

Step 4: Consider the actual lead time $L > 0$. Determine for each item separately the reorder point required for satisfying the service level constraint, that is, successively evaluate the policies $(s_i, c_i^{(0)} + s_i, S_i^{(0)} + s_i)$ for $s_i = 0, 1, 2, \ldots$, until the constraint is met.

In the last step, either (2.67) can be used for cycle service constraints, or (2.68)–(2.71) for fill rate constraints, together with (2.64) and (2.65). ☐

The estimated cost corresponding to the can-order policy — under the approximative assumption that opportunities for joining an order occur according to a Poisson process — is, cf. (2.48),

$$C(\mathbf{s}, \mathbf{c}, \mathbf{S}) = \sum_{i=1}^{N} [M_i A + (M_i + J_i) a_i + h_i E\{\bar{I}_i\}], \tag{2.72}$$

here, $M_i$, $J_i$, are the appropriate item-related numbers of orders according to (2.56) and (2.57), and the expected average inventory on hand $E\{\bar{I}_i\}$ is given by (omitting the item index $i$):

$$E\{\bar{I}\} = \tfrac{1}{2}(E\{[S - D_L]^+\} + E\{[I_O - D_L]^+\}). \tag{2.73}$$

This quantity can be computed in a similar way as the mean backlog, cf. (2.71).

Table 2.13: Data per month and individual deterministic policies for Example 2.17.

| item | $a_i$ | $h_i$ | $\lambda_i$ | $Q_i$ | $M_i$ | $T_i$ |
|---|---|---|---|---|---|---|
| 1 | \$10 | \$1 | 10 | 34.64 | 0.289 | 3.46 |
| 2 | \$ 5 | \$1 | 20 | 46.90 | 0.426 | 2.35 |
| 3 | \$15 | \$1 | 30 | 62.45 | 0.480 | 2.08 |

**Example 2.17** Consider a family consisting of $N = 3$ items. The family ordering cost is $A = \$50$. The supplementary ordering costs $a_i$, the holding costs $h_i$ per unit per month and the average demand per month, $E\{D_i\} = \lambda_i$, are listed in Table 2.13, $i = 1, 2, 3$. The lead time is $L = 0.1$ month. Management imposes a target fill rate of $\beta = 0.99$ for all three items. The economic order quantities $Q_i$, assuming deterministic demand and individual ordering, and the corresponding number of orders per month, $M_i$, cf. (2.47), and the reorder cycle $T_i$, $i = 1, 2, 3$, are also displayed in Table 2.13. The iteration starts with item 1, with opportunity rate $\mu_1 = M_2 + M_3 = 0.426 + 0.480 = 0.907$. The value of the parameter $\rho_1$ is $10/(10 + 0.907) = 0.917$, cf. (2.54). The cost function (2.60) for $L = 0$ is minimized by $c_1 = 13$ and $S_1 = 28$. The new expected number of orders caused by item 1 is $M_1 = 0.144$, cf. (2.56). Observe that this value is much smaller than the initial value of 0.289. This is explained by the observation that offering opportunities for supplementing orders diminishes the number of times an item reaches its reorder point. Then, item 2 is considered, with opportunity rate $\mu_2 = M_1 + M_3 = 0.144 + 0.480 = 0.624$ and, hence, $\rho_2 = 20/(20 + 0.624) = 0.970$. The cost function (2.60) for $L = 0$ is minimized by $c_2 = 26$ and $S_2 = 40$. The new expected number of orders caused by item 2 is $M_2 = 0.284$. Next, item 3 is considered, with opportunity rate $\mu_3 = M_1 + M_2 = 0.144 + 0.284 = 0.428$. In this way, the iterative procedure continues. The results are summarized in Table 2.14. After the third round of iterations the procedure stops because the values of $c_i$ and $S_i$, $i = 1, 2, 3$, after round 3 are the same as after round 2. The final values of $M_i$, $i = 1, 2, 3$, indicate that about 44% of all orders are caused by item 3, 36% by item 2 and 20% by item 1. The expected total number of orders per month is $M_1 + M_2 + M_3 = 0.837$ so that the expected time between orders is $\frac{1}{0.837} = 1.195$ month.

Table 2.14: Iteration steps of Algorithm 2.12 for Example 2.17.

| iteration | item | $\mu_i$ | $\rho_i$ | $c_i$ | $S_i$ | $C_i(c_i, S_i)$ | $M_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.907 | 0.917 | 13 | 28 | 28.40 | 0.144 |
|  | 2 | 0.624 | 0.970 | 26 | 40 | 40.68 | 0.284 |
|  | 3 | 0.428 | 0.986 | 28 | 58 | 58.88 | 0.381 |
| 2 | 1 | 0.665 | 0.938 | 14 | 29 | 29.63 | 0.170 |
|  | 2 | 0.551 | 0.973 | 26 | 41 | 41.32 | 0.296 |
|  | 3 | 0.465 | 0.985 | 28 | 58 | 58.58 | 0.371 |
| 3 | 1 | 0.666 | 0.938 | 14 | 29 | 29.63 | 0.169 |
|  | 2 | 0.540 | 0.974 | 26 | 41 | 41.41 | 0.298 |
|  | 3 | 0.468 | 0.985 | 28 | 58 | 58.56 | 0.370 |

In the last step of Algorithm 2.12, the policies of round 3 are adjusted to a target fill rate of $\beta = 0.99$ for a lead time of $L = 0.1$ month. For $i = 1, 2, 3$, the actual fill rate $\Psi_i$ is computed according to (2.68)–(2.71) together with (2.64), (2.65), and $s_i$ is increased until $\Psi_i \geq 0.99$. The intermediate results are summarized in Table 2.15.

The final values of the parameters of the can-order policy can be found in Table 2.16, together with the corresponding mean order quantities $E\{Q_i\}$, the expected mean inventory on hand $E\{\bar{I}_i\}$, the mean accumulate backlog $E\{B_i\}$, the mean reorder cycle $E\{T_i\}$, the mean number of orders per month $M_i$ and $J_i$ caused respectively supplemented by item $i$, and the estimated monthly cost $C_i$ attributable to item $i$, $i = 1, 2, 3$. Moreover, the safety stocks $V_i$ and the expected times $\tau_{c_i} = (S_i - c_i)/\lambda_i$ to reach the can-order level $c_i$ from the order-up-to level $S_i$ and $\tau_{s_i} = (S_i - s_i)/\lambda_i$ to reach the reorder point $s_i$ from the order-up-to level $S_i$

Table 2.15: Final step of Algorithm 2.12 for Example 2.17 with target fill rates.

| $s_i$ | $\Psi_1$ | $\Psi_2$ | $\Psi_3$ |
|---|---|---|---|
| 0 | 0.9825 | 0.9693 | 0.9621 |
| 1 | 0.9936 | 0.9827 | 0.9742 |
| 2 | | 0.9918 | 0.9844 |
| 3 | | | 0.9916 |

(disregarding opportunities for joining an order) are included in this table for item $i$, $i = 1, 2, 3$. Note that these quantities are independent of the target fill rates. Observe that item 2, with the smallest supplementary ordering cost, is fastest in reaching its can-order level, while item 3, with the smallest individual cycle time $T_i$, is fastest in reaching its reorder point (on the average, when all items simultaneously start from their order-up-to levels).

Table 2.16: Final can-order policy for Example 2.17 with target fill rates.

| item | $\lambda_i L$ | $s_i$ | $c_i$ | $S_i$ | $E\{Q_i\}$ | $E\{\bar{I}_i\}$ | $E\{B_i\}$ | $E\{T_i\}$ | $M_i$ | $J_i$ | $C_i$ | $V_i$ | $\tau_{c_i}$ | $\tau_{s_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 15 | 30 | 23.92 | 17.11 | 0.15 | 2.39 | 0.169 | 0.249 | \$29.76 | 0 | 1.50 | 2.90 |
| 2 | 2 | 2 | 28 | 43 | 33.51 | 24.38 | 0.28 | 1.68 | 0.298 | 0.299 | \$42.29 | 0 | 0.75 | 2.05 |
| 3 | 3 | 3 | 31 | 61 | 52.55 | 31.95 | 0.44 | 1.75 | 0.370 | 0.201 | \$59.02 | 0 | 1.00 | 1.93 |

Alternatively, the policies of round 3 of Algorithm 2.12 can be adjusted to a cycle service constraint of, say, $\alpha = 0.98$ for a lead time of $L = 0.1$ month. For $i = 1, 2, 3$, the probability $\Pi_i$ is computed according to (2.67) with (2.64) and (2.65), and $s_i$ is increased until $\Pi_i \geq 0.98$. The intermediate results are summarized in Table 2.17.

Table 2.17: Final step of Algorithm 2.12 for Example 2.17 with cycle service constraint.

| $s_i$ | $\Pi_1$ | $\Pi_2$ | $\Pi_3$ |
|---|---|---|---|
| 0 | 0.7337 | 0.5520 | 0.3628 |
| 1 | 0.8901 | 0.6955 | 0.4679 |
| 2 | 0.9668 | 0.8353 | 0.6191 |
| 3 | 0.9922 | 0.9275 | 0.7680 |
| 4 | | 0.9733 | 0.8789 |
| 5 | | 0.9916 | 0.9451 |
| 6 | | | 0.9781 |
| 7 | | | 0.9922 |

The required reorder points $s_1 = 3$, $s_2 = 5$ and $s_3 = 7$ are larger than those for the target fill rates above. Note that the final probabilities $\Pi_i$ all exceed 0.99 in this case. □

From the foregoing discussion it can be concluded that the determination of the values of the decision variables of a can-order policy is complex and approximative, even under strong assumptions of Poisson demand processes and a demand of 1 unit per customer. For other demand processes (e.g., of fast movers), modifications are needed which do not always seem tractable. Moreover, simulation is needed for a real performance analysis of a can-order policy (without the assumption of Poisson opportunity processes for joining orders). It has been found that the model costs of can-order policies overestimate the real costs, and that the model service levels of can-order policies underestimate the real service levels. In some circumstances ($A/\bar{a}$ large), the assumption of Poisson opportunity processes for joining orders is demonstrably bad: the opportunities occur more frequently towards the end of a reorder cycle; it is then better to choose $c_i = S_i - 1$ and to determine $(s_i, S_i)$ without the Poisson assumption. Van Eijs [71] has developed a heuristic for this situation. It should be noted that the fact that can-order policies sometimes perform badly is mostly due to the applied method of analysis, and not always to policy itself. However, Ignall [39] showed for a two-item case that can-order policies are not necessarily the optimal policy for a continuous-review coordinated replenishment problem. In Silver [61] Algorithm 2.12 has been extended to can-order policies under compound Poisson demand processes. Federgruen et al. [28] determine the can-order parameters for the 1-item problem in the case of a compound Poisson demand process by a policy iteration algorithm for semi-Markov decision problems.

**Exercise 2.15** *Consider a single item with a Poisson demand process at a rate of 2 units per day. Assume that this item is a member of a group in which opportunities for joining an order occur according to a Poisson process at a rate of once every 4 days. Suppose that this item is reordered according to a can-order policy with* $s = 2$, $c = 5$ *and* $S = 10$, *with negligible lead time* $L = 0$. *Determine the stationary distribution of the inventory level process and the average inventory level. Compute the expected number of orders per day caused by this item and the expected number of orders per day supplemented by this item.*

**Exercise 2.16** *Show that the mean order quantity* $E\{Q\}$ *of an item, cf. (2.69), is equal to* $\lambda/(M + J)$, *cf. (2.58), for this item. Explain this relation.*

**Exercise 2.17** *Give a computable expression for the mean backlog at the end of a cycle,* $E\{[D_L - I_O]^+\}$, *of an item, cf. (2.71), for the case of a reorder point of* $s = 0$. *Derive a recursion for this quantity for a reorder point of* $s > 0$ *in terms of that for the reorder point* $s - 1$.

**Exercise 2.18** *Consider a family of* $N = 2$ *items, with joint ordering cost* $A = \$40$ *and supplementary ordering costs* $a_1 = \$10$, $a_2 = \$15$, *with holding costs per unit per week* $h_1 = \$5$, $h_2 = \$8$, *with Poisson demand rates* $\lambda_1 = 1$ *and* $\lambda_2 = 2$ *items per week, and with a lead time of* $L = 1$ *week. Determine the can-order policy according to Algorithm 2.12 at a target fill rate of* $\beta = 0.98$ *for both items. Compute for both items the probability that no backlog occurs in a cycle with this policy and the mean order quantity.*

**Exercise 2.19** *A shop sells two items which are ordered from the same wholesale dealer. The joint ordering cost is \$15, the additional ordering cost of item 1 is \$12 and that of item 2 is \$18. The lead time is 2 days. The holding cost is \$0.10 per unit per day for item 1 and \$0.08 per unit per day for item 2. The shop uses a can-order policy with parameters* $(s_1, c_1, S_1)$ *for item 1 and with parameters* $(s_2, c_2, S_2)$ *for item 2. First, suppose that the demand is deterministic, 9 units per day for item 1 and 20 units per day for item 2. Compute the optimal reorder cycle for the case that both items are always jointly ordered. Determine the values of the can-order parameters that realize this reorder cycle in such a way that all demand is satisfied from stock. Next, suppose that the demand is stochastic. The manager of the shop has chosen the following parameter values:* $s_1 = 20$, $c_1 = 40$, $S_1 = 80$, $s_2 = 40$, $c_2 = 80$, $S_2 = 160$. *The can-order policy is applied as follows: every evening after closing time the inventories are reviewed; if one of the items has an inventory level at or below its reorder point, an order is placed according to the can-order parameters; ordered goods are delivered 2 evenings later. Demand which cannot be satisfied from stock is lost. At the beginning of day 1, 30 units are on stock of item 1 and 100 units of item 2. The realizations of the demand over the next 10 days are given in the table below.*

| day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $i = 1$ | 8 | 12 | 15 | 10 | 10 | 5 | 10 | 15 | 10 | 10 |
| $i = 2$ | 20 | 25 | 20 | 15 | 25 | 40 | 25 | 20 | 20 | 15 |

*Simulate the inventory levels of the two items over these 10 days. Compute the total ordering cost over these 10 days. Also, compute the total holding cost, charged over the remaining inventories at the end of each day before delivery of a possible order, over these 10 days. How many units demand of the two items are lost?*

### 2.4.2 Periodic review policies

Atkins & Iyogun [4] have proposed a heuristic for the determination of the parameters of an $(R, \mathbf{k}, \mathbf{S})$ periodic review policy, with

- $R = T$: the family reorder cycle (the review period),
- $k_i$ (integer): the reorder frequency for item $i$, $i = 1, \ldots, N$,
- $S_i$: the order-up-to level for item $i$, $i = 1, \ldots, N$,

for a coordinated replenishment model with family ordering cost and stationary stochastic demand. The aim is the minimization of the expected average ordering and holding cost per unit of time subject to a prescribed service level (a target fill rate $\beta_i$ for item $i$, $i = 1, \ldots, N$).

Globally, the heuristic consists of the following two phases. In the first phase, the stochastic nature of the demand is ignored, and review periods are determined based on the average demand. In the second phase, safety stocks are determined for each item separately, such that the service level constraints for the stochastic demands are met. More specifically, in the first phase weights $w_i,\ i = 1, \ldots, N$, are determined for the allocation of the family ordering cost $A$ to the $N$ items, and then a family cycle $T$ and reorder frequencies $k_i$ are determined on the basis of the expected demand $E\{D_i\}$ of all items and with ordering costs $w_iA + a_i$ for item $i,\ i = 1, \ldots, N$. In the second phase, for every item separately the order-up-to level $S_i$ is determined on the basis of an $(R_i, S_i)$ policy with review period $R_i = k_iT$ fixed (from the first phase), taking into account the stochastic nature of the demand, and with ordering cost $w_iA + a_i$ for item $i,\ i = 1, \ldots, N$.

Next, we elaborate the heuristic in more detail. For a given allocation of the family ordering costs $A$ with weights $w_i$ it holds that

$$T_i = k_iT = \sqrt{\frac{2(w_iA + a_i)}{h_iE\{D_i\}}}, \quad i = 1, \ldots, N. \tag{2.74}$$

A reformulation of this relation leads to

$$w_i = \frac{1}{2A}\,[h_iE\{D_i\}T_i^2 - 2a_i], \quad i = 1, \ldots, N. \tag{2.75}$$

The heuristic chooses $w_i = 0$ if $k_i > 1$. Let $\mathcal{I} \doteq \{i \,|\, k_i = 1\}$ be the set of items that are included in every family order. Then it follows from the normalization of the weights and the fact that $T_i = T$ for $i \in \mathcal{I}$:

$$1 = \sum_{i=1}^{N} w_i = \frac{1}{2A}\sum_{i\in\mathcal{I}}[h_iE\{D_i\}T^2 - 2a_i]. \tag{2.76}$$

This normalization leads to the optimal reorder cycle for the group of items $\mathcal{I}$, cf. (2.23), which will be used as review period:

$$R^2 = T^2 = 2\left[A + \sum_{i\in\mathcal{I}} a_i\right] \Bigg/ \sum_{i\in\mathcal{I}} h_iE\{D_i\}\ . \tag{2.77}$$

The complete heuristic for determining an $(R, \mathbf{k}, \mathbf{S})$ periodic review policy is summarized below.

**Algorithm 2.13** [Atkins & Iyogun]

Step 1: Start with $w_i = 0$ for all $i$. Order the items such that $T_1 \le T_2 \le \ldots \le T_N$, that is, in ascending value of $a_i/(h_iE\{D_i\})$, cf. (2.74). Set $m = 1$.

Step 2: Increase $w_1, \ldots, w_m$ (and correspondingly $T_1, \ldots, T_m$) until $T_1 = \cdots = T_m = T_{m+1}$ (use (2.75) for this purpose).

Step 3: If $\sum_{i=1}^{m} w_i < 1$ and $m < N$, increase $m$ by 1 and repeat from Step 2; otherwise, define the set $\mathcal{I} \doteq \{1, \ldots, m\}$.

Step 4: Compute the family cycle $R = T$ for this set $\mathcal{I}$ according to (2.77). Determine the weights $w_i$ for $i = 1, \ldots, m$ with this family cycle $T$ and with $k_i = 1$ for $i = 1, \ldots, m$ (by definition of the set $\mathcal{I}$), according to (2.75).

Step 5: Determine $k_i$ for $i = m + 1, \ldots, N$ by rounding $T_i/T$, computed from (2.74) with $w_i = 0$, $i = m + 1, \ldots, N$, to the nearest integer.

Step 6: Determine for every item separately the order-up-to level $S_i$ on the basis of a target fill rate $\beta_i$ at a fixed review period $R_i = k_iT$, $i = 1, \ldots, N$, e.g., with the assumption of gamma distributed demand during the lead time plus review period $L + R_i$.

The items $i = m + 1, \ldots, N$ do not have a share in the family ordering cost. Step 2 is to be skipped when $m = N$; if $m$ has reached the value $N$, then $\mathcal{I} \doteq \{1, \ldots, N\}$. □

The resulting safety stocks are

$$V_i = S_i - (L + R_i)E\{D_i\}, \quad i = 1, \ldots, N. \tag{2.78}$$

The expected average inventory levels are with this policy bounded by: for $i = 1, \ldots, N$,

$$E\{\bar{I}_i\} \geq \tfrac{1}{2}\left[S_i - LE\{D_i\} + S_i - (L + R_i)E\{D_i\}\right] = S_i - LE\{D_i\} - \tfrac{1}{2}R_iE\{D_i\} = V_i + \tfrac{1}{2}R_iE\{D_i\}.$$

The expected order quantities are $E\{Q_i\} = R_iE\{D_i\},\ i = 1, \ldots, N$. The total expected cost is

$$E\{C\} = \sum_{i=1}^{N}\left[\frac{w_iA + a_i}{R_i} + h_iE\{\bar{I}_i\}\right] = \frac{A}{T} + \frac{1}{T}\sum_{i=1}^{N}\frac{a_i}{k_i} + \sum_{i=1}^{N}h_iE\{\bar{I}_i\}. \tag{2.79}$$

**Remark 2.11** The first phase of Algorithm 2.13 (the first five steps) can be replaced by any other heuristic for the deterministic indirect grouping problem or by the exact Algorithm 2.1, cf. Section 2.1, using the average demand as the deterministic demand. On the other hand, the first five steps of Algorithm 2.13 form an alternative heuristic for the deterministic indirect grouping problem. This heuristic can often be improved by computing the optimal cycle length given the reorder frequencies determined in Steps 4 and 5 after Step 5 (like Step 4 of the Goyal & Belton heuristic 2.2). □

Table 2.18: Data per week and optimal individual policies without family ordering cost for Example 2.18.

| item | $a_i$ | $v_i$ | $E\{D_i\}$ | $\sigma\{D_i\}$ | $T_i$ | $Q_i$ | index |
|---|---|---|---|---|---|---|---|
| $a$ | \$3.00 | \$8.00 | 9 | 9 | 4.17 | 37.50 | 3 |
| $b$ | \$3.00 | \$12.50 | 40 | 25 | 1.58 | 63.25 | 2 |
| $c$ | \$3.00 | \$3.52 | 4 | 2 | 9.42 | 37.69 | 4 |
| $d$ | \$3.00 | \$33.30 | 60 | 24 | 0.79 | 47.46 | 1 |

**Example 2.18** Consider a family consisting of $N = 4$ items, indicated by $a, b, c, d$. The family ordering cost is $A = \$20$. The supplementary ordering costs $a_i$, the purchasing prices $v_i$ and the averages $E\{D_i\}$ and the standard deviations $\sigma\{D_i\}$ of the demands per week are listed in Table 2.18, $i = 1, 2, 3, 4$. The carrying charge is $r = \$0.0048$ per dollar per week, and the holding costs are $h_i = rv_i$ per unit per week. The lead time is $L = 1$ week. According to Algorithm 2.13, first the reorder cycles $T_i$ are computed with ordering cost $a_i$ (that is, with $w_i = 0$), and deterministic demand $E\{D_i\}$, cf. (2.74). These values of $T_i$ and the corresponding order quantities $Q_i$ are shown in Table 2.18. Moreover, the items have been indexed in ascending value of $T_i$: $\{1, 2, 3, 4\} = \{d, b, a, c\}$; see the column with the header "index". The application of the heuristic continues as follows. First, increase $w_1$ until $T_1 = T_2$ ($T_d = T_b = 1.58$). Then, by (2.75),

$$w_1 = w_d = \tfrac{1}{2A}\left[rv_dE\{D_d\}T_b^2 - 2a_d\right] = 0.45 < 1;$$

now 45% of the family ordering cost $A = \$20$ is allotted to item $d$. Next, increase $w_1$ and $w_2$ until $T_1 = T_2 = T_3$ ($T_d = T_b = T_a = 4.17$). Then,

$$w_1 = w_d = \tfrac{1}{2A}\left[rv_dE\{D_d\}T_a^2 - 2a_d\right] = 4.01;$$

$$w_2 = w_b = \tfrac{1}{2A}\left[rv_bE\{D_b\}T_a^2 - 2a_b\right] = 0.89.$$

Here, the procedure stops because $w_1 + w_2 > 1$. The result is $m = 2$, and $\mathcal{I} = \{1, 2\} = \{d, b\}$. The family cycle for this set of items $\mathcal{I}$ is, cf. (2.77):

$$T = \sqrt{2[A + a_d + a_b]/[rv_dE\{D_d\} + rv_bE\{D_b\}]} \approx 2.0825.$$

The final weights $w_1$ and $w_2$ are computed at $T = 2.0825$ from (2.75):

$$w_1 = w_d = \tfrac{1}{2A}\left[rv_dE\{D_d\}T^2 - 2a_d\right] = 0.89;$$

$$w_2 = w_b = \tfrac{1}{2A}\left[rv_bE\{D_b\}T^2 - 2a_b\right] = 0.11;$$

this means that 89% of $A$ is allotted to item $d$ and 11% of $A$ is allotted to item $b$. No part of $A$ is allotted to the other items, because they only act as supplements to orders. Finally, $k_3$ and $k_4$ are computed with the final value of $T$ from (2.74) with $w_3 = w_4 = 0$:

$$k_3 = k_a = T_a/T \approx 2; \qquad k_4 = k_c = T_c/T = 4.53 \approx 5.$$

The above reorder frequencies are the same as those generated by Goyal's Algorithm 2.1, only the reorder cycle obtained by the heuristic of Atkins & Iyogun (2.0825) is slightly longer than optimal (2.0776). Finally, the order-up-to levels $S_i$ are determined in Step 6 at a target fill rate $\beta_i = 0.98$ and with a fixed review period $R_i = T_i$, separately for $i = 1, \ldots, 4$. The results are summarized in Table 2.19 for a family reorder cycle of $T = 2.0825$ weeks. The column with the header "$C_i$" contains the cost that is fully attributable to item $i$, $i = 1, \ldots, 4$, that is, $a_i/R_i + h_i E\{\bar{I}_i\}$. It does not include a part of the family ordering cost which amounts to $A/T = \$9.60$. The expected total weekly cost is $E\{C\} = \$42.31$ for this policy, cf. (2.79), which is slightly better than the expected total weekly cost with a family reorder cycle of $T = 2.0776$ weeks ($E\{C\} = \$42.37$). This is due to the sequential approach to fix the reorder cycles first and to determine the order-up-to levels later, cf. Example 1.3. ☐

Table 2.19: Summary of results by Algorithm 2.10.

| item | index | $w_i$ | $k_i$ | $R_i$ | $E\{Q_i\}$ | $S_i$ | $V_i$ | $E\{\bar{I}_i\}$ | $a_i/R_i$ | $h_iV_i$ | $h_i[E\{\bar{I}_i\} - V_i]$ | $C_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | 3 | 0.00 | 2 | 4.17 | 37.48 | 84 | 37.52 | 56.61 | 0.72 | 1.44 | 0.73 | 2.89 |
| $b$ | 2 | 0.11 | 1 | 2.08 | 83.30 | 198 | 74.70 | 117.18 | 1.44 | 4.48 | 2.55 | 8.47 |
| $c$ | 4 | 0.00 | 5 | 10.41 | 41.65 | 52 | 6.35 | 27.53 | 0.29 | 0.11 | 0.36 | 0.75 |
| $d$ | 1 | 0.89 | 1 | 2.08 | 124.95 | 241 | 56.05 | 119.76 | 1.44 | 8.96 | 10.18 | 20.58 |

Atkins & Iyogun [4] compare the heuristic $(R, \mathbf{k}, \mathbf{S})$ policy and the can-order $(\mathbf{s}, \mathbf{c}, \mathbf{S})$ policy with a simply computable lower bound for the minimal cost as a gauge. This lower bound, derived in Atkins & Iyogun [3], is based on the same idea of allotting the family ordering cost $A$ to the $N$ items according to weights $w_i$, $i = 1, \ldots, N$. Then, by inequality (2.34) we have for the unknown minimum cost $C_{\min}$:

$$C_{\min} \geq \sum_{i=1}^{N} \min C_i; \tag{2.80}$$

here, $C_i$ is the cost associated to a corresponding 1-item problem with ordering costs $w_i A + a_i$, $i = 1, \ldots, N$. Hence, to compute the above lower bound, $N$ 1-item problems have to be solved. Atkins & Iyogun [4] conducted experiments with Poisson demand processes and stockout costs. They found that:

- the parameters of a $(R, \mathbf{k}, \mathbf{S})$ policy are easier to determine than those of a can-order policy;
- the can-order policy outperforms the periodic review $(R, \mathbf{k}, \mathbf{S})$ policy when $A/\bar{a}$ is small, but otherwise the $(R, \mathbf{k}, \mathbf{S})$ policy outperforms the can-order policy;
- in a typical case, the $(R, \mathbf{k}, \mathbf{S})$ policy is 12% and the can-order policy 28% above the lower bound (while independent ordering is 64% above that bound).

Finally, note that the heuristic $(R, \mathbf{k}, \mathbf{S})$ policy is not dependent on the assumption of Poisson demand processes and constant lead times, and is therefore more suitable in case of "fast movers". Viswanathan [76] shows that an $(R, \mathbf{s}, \mathbf{S})$ policy, in which a good common review period $R$ is found by a stepwise search and the reorder point $s_i$ and the order-up-to level $S_i$ are determined separately for each item $i$ given the review period $R$ and using only the supplementary ordering cost $a_i$, outperforms the $(R, \mathbf{k}, \mathbf{S})$ policy by a few percent in many cases.

**Exercise 2.20** *Consider again the six-item coordinated replenishment problem described in Exercise 2.4. Determine a reorder policy according to the deterministic part of the heuristic of Atkins & Iyogun.*

**Exercise 2.21** *Apply the deterministic part of the heuristic of Atkins & Iyogun to the three-item family considered in Example 2.1. Compare the ordering cost allotted to item i* ($w_iA + a_i$) *with the substitute ordering cost* $\tilde{a}_i$ *in Table 2.1 for* $i = 1, 2, 3$ *and comment on the differences. Compute the order-up-to levels for the case that the lead time is* $L = 1$ *week, that the mean demand per week is equal to the demand given in Table 2.1, and the standard deviations of the weekly demands are* $\sigma\{D_1\} = 3$, $\sigma\{D_2\} = 2$, $\sigma\{D_3\} = 1$, *and that the target fill rates are 90% for all three items. What are the corresponding safety stocks?*

**Exercise 2.22** *Determine a periodic review policy according to the heuristic of Atkins & Iyogun for the family of three items considered in Example 2.17. Compute the corresponding mean order quantities.*

**Exercise 2.23** *Determine a periodic review policy according to the heuristic of Atkins & Iyogun for the two-item family considered in Exercise 2.18. Compute the corresponding mean order quantities.*

# Chapter 3

# Multi-echelon Inventory Systems

This chapter is concerned with the management of flows of goods on parts of their way from producer to buyer. Important issues in such supply chains are the requirements to have the goods of the right quality at the right time at the right location. The main subsystems of physical distribution are:

**inventory management:** this includes demand forecasting, the choice of an ordering system, which consists of rules for determining the ordering quantities and the reorder instants;

**warehouse management:** this includes determining its function in the distribution channel, the choice concerning the number of warehouses and their locations, the choice of buying or renting warehouse space, the choice of the lay-out of the warehouse and the equipment in the warehouse, the choice of the internal transport and material handling system, and the policies concerning the order picking system;

**transport management:** this includes the decision of using own transport or putting transport out to contract, the choice of the means of transport, a route planning system, and the choice of the sizes of the pallets to be used, and whether to buy or to rent them.

The decisions in the area of physical distribution can also be classified according to their scope:

**strategic:** this includes the choice of the distribution channels (directly or via distribution centers to retailers or indirectly via wholesale dealers), the decision concerning the distribution spread (the percentage of shops in the trade that sell the goods), the choice between a “push” or a “pull” policy, and the determination of the desired service levels;

**tactical:** this includes the decisions concerning the number and the locations of the warehouses, the choice of the handling- and storage system, the choice of a storage policy, and choices concerning the means of transportation;

**operational:** this includes the decisions concerning when and how much to order or to produce, the planning of the order picking, and the routing of the transport.

Section 3.1 contains a classification of multi-echelon structures and a discussion of the functions of a distribution center. Section 3.2 is devoted to inventory policies for linear and divergent two-echelon systems with constant demand. Section 3.3 discusses algorithms for determining pull and push policies for two-echelon systems with time-varying demand. Finally, Section 3.4 is concerned with a periodic review policy for divergent two-echelon systems with stochastic demand.

## 3.1 Multi-echelon structures

In practice, there exist many echelon structures, both in the physical distribution and between departments in a factory. Even within a company or organization several echelon structures may exist next to each other. For instance, a chain of supermarkets may have different echelon structures for durable goods and for perishable goods.

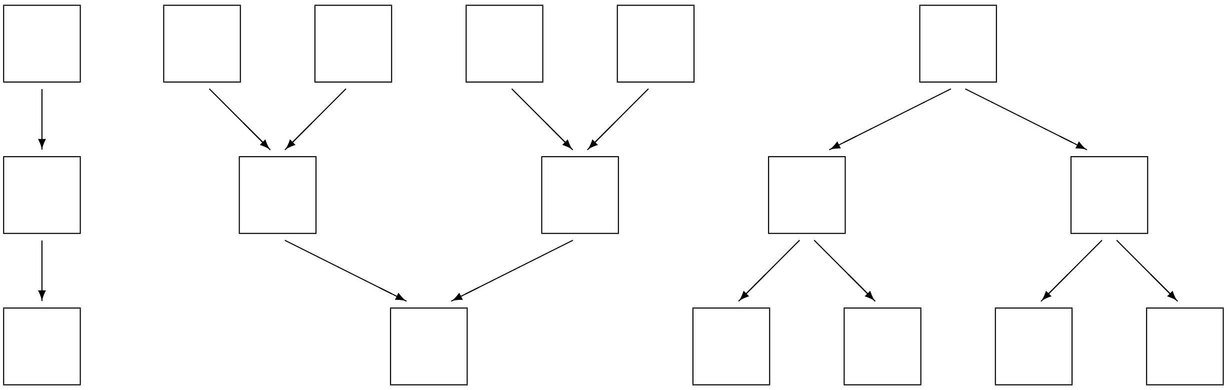

Figure 3.1: A linear (l.), a convergent (m.) and a divergent (r.) supply chain consisting of three echelons.

### 3.1.1 Classification

A classification of echelon structures is the following:

**linear chains:** all units follow the same path or route (this may occur in an environment with only sequential working on the units); usually, they are part of more complex chains;

**converging chains:** they occur in factories where parts are assembled, and in distribution from various factories to a single distribution center;

**diverging chains:** they occur in factories where raw materials are cut into various part types and where semi-manufactured items are made into various end products, and in distribution from a single distribution center to various retailers;

**general networks:** they consist of both converging and diverging parts.

It is customary to speak of high and low echelons. The lowest echelons are the ones closest to the customers (buyers). Usually, external demand only occurs at the lowest echelons, while external orders are only placed at the highest echelons. The goods move from the highest to the lowest echelons, while the demand works on from the lowest to the highest echelons.

In multi-echelon inventory systems a distinction is made between pull and push systems. In pull system each stocking point has its own ordering policy, and stocking points at higher echelons have to adjust their inventory management to the uncontrollable demand from the lower echelons. In push system information on the inventory positions of the stocking points at all echelons is kept centrally, and the inventory management is done centrally. The latter requires much, reliable and up-to-date information. A cost saving is possible by better anticipation and coordination. However, problems may arise with the responsibilities of local managers.

### 3.1.2 Distribution centers

This section is devoted to the function and some aspects of distribution centers (DCs). An important function of a DC is that it restricts the number of transport lines between a number of factories and a number of retailers. For instance, if a group of 25 factories supply a group of 100 retailers, then direct transport would mean 2,500 transport lines, while indirect transport via a single DC reduces the number of transport lines to 125. Moreover, direct transport will consist of relatively low volumes, and will therefore be relatively expensive. Further, the space for storage is generally cheaper in a DC than at the retailers because a DC can be located in an area outside city centers. Also, transferring goods to other means of transport (e.g., from truck, train or ship to a smaller truck or van that can reach a shopping area) may be a function of a DC. Further, a DC may serve for regrouping and repacking of goods which may arrive at the DC in large bulks and depart from the DC in much smaller quantities. Standard activities in a distribution center include the receipt and the control of goods, cross-docking (transferring goods from one means of transport to another without storage), material handling, storage, order picking, grouping of items (per order), repacking in smaller quantities, combining of orders for transport, administration of incoming and outgoing goods, and regular control of the inventories. Finally, Value Added Logistics (VAL) activities may be performed in a DC. These activities aim at postponing product differentiation and lead to a larger flexibility and lower stocks. VAL activities are most profitable if there is a large distance between a factory and its market, as is the case with a factory in East Asia and retailers in Europe. Since VAL activities may require costly equipment, the number of distribution centers in which they are performed will be small, e.g., a single VAL distribution center for the whole of Europe. Value Added Logistics activities can be performed in a distribution center owned by the producer but can also be put out to contract to logistic service industries (this is called public warehousing). Examples of VAL activities are specific labeling (e.g., according to the rules of countries), supplementing manuals (in specific languages), making to measure, mixing of ingredients, installing components or software, assembling of parts, fitting of accessories, testing, quality control, repairing, examining of returns, reconditioning of clothes, specific repacking (e.g., according to rules of countries), dispatching and administrating orders.

To validate the right of existence of a DC the cost savings by the reduction of the number of transport lines and the lower holding cost have to be weighed against the cost of maintaining the DC and of the extra handling cost in the DC.

Next, we discuss in more detail the material handling systems and the order picking systems in a warehouse. These are often expensive systems which constitute a significant part of the costs of companies. Therefore, efficiency is an important aspect. Ways to arrive at efficiency are by eliminating, combining and simplifying movements in a warehouse. As far as possible it is best to work with a unit load (e.g., defined as a quantity which can be stacked on a pallet). Further, the reliability and the safety of these systems are important. Moreover, they should be connected to an information system so that the locations of all units can be tracked at any time.

A rough classification of material handling systems is:

**conveyors:** they are relatively static in routing, allow a limited variation in products; examples of use are: for bottles (washing, filling, closing, labelling), and for luggage at airports;

**cranes:** they are mobile in more directions, they are more flexible in case of disturbances, and they take up less floor space than conveyors; examples of use are: for bottles (putting in crates), for storing in storage carousels, and for computer controlled storage;

**vehicles:** they are still more mobile and have a larger range of action, they are still more flexible than cranes, they may be manually or automatically guided; examples of use are: order picking in warehouses.

Order picking systems deal with collecting a limited number of units of some set of items for a number of customers. It generally concerns relatively small quantities and requires much labor and/or much capital. Examples where order picking systems are used are in mail-order companies, in factories (parts for production) and in distribution centers. Possible objectives in optimizing the use of order picking systems are:

- minimization of the total order-picking cost per unit of time;
- minimization of the total order-picking time per unit of time;
- minimization of the average lead time to the customers;
- maximization of the average number of picked orders or items per unit of time.

Important aspects of order picking systems are:

- the lay-out of the warehouse and the division of goods over the warehouse;
- a good information system;
- a good routing system;
- the zoning of the warehouse (and the orders);
- the combining of orders (batching);
- a redivision of goods over the warehouse if some demand patterns change.

A systematic planning and design of order picking systems includes:

- a strategic planning: what type of customers, what service levels, what demand patterns?

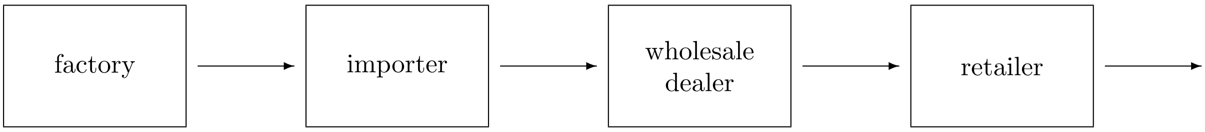

Figure 3.2: A linear supply chain consisting of four echelons.

- a classification items: poisonous, inflammable, pilferage, special sizes;
- an ABC analysis: important for the locations in a warehouse;
- the determination of the required storage capacity;
- the determination of a storage policy (what items are stored where in the warehouse? how to divide the warehouse in to zones?);
- the determination of the degree of automatization (this concerns the actual order picking as well as the information system);
- the choice of a material handling system;
- the choice of an information system;
- the determination of a batching and routing policy.

Hausman et al. [38] derive an optimal storage assignment for automatic warehousing systems. Wilson [79] considers the location of stock in a warehouse as function of the order quantity and the demand of the various products. Tomkins & White [66] is devoted to facilities planning. Ratliff & Rosenthal [55] solve a special case of an order-picking problem as a travelling salesman problem. Other studies on order-picking systems include Goetschalckx & Ratliff [31, 32]. Eastman [23] is concerned with material handling systems. Robeson & House [56] discuss distribution issues.

## 3.2 Constant demand

In this section, models and policies will be discussed for multi-echelon systems with constant demand at the stocking points of the lowest echelon. In general, the demand in higher echelons is formed by the orders of lower echelons. Hence, the demand pattern in higher echelons is not continuous but stepwise! This subject is also discussed in Silver et al. [64, Ch. 12].

### 3.2.1 Linear chains

We start this section with an example of the oscillation effect that may occur in supply chains when not enough care is taken in choosing reorder policies and in exchanging information between echelons.

**Example 3.1** Consider a linear supply chain consisting of a factory, an importer, a wholesale dealer and a retailer, cf. Figure 3.2. Suppose that at all echelons the following simple ordering rule is applied. Every week an order is placed that is immediately delivered provided that there are no stockouts at the higher echelon. The order-up-to level is set at 110% of the demand of the previous week ($S = 1.1 \cdot D$) so that a safety stock is used of 10% of the demand of the previous week ($V = 0.1 \cdot D$). The order quantity is the difference between the order-up-to level and the inventory position: $Q = S - I$. Assume that the demand at the retailer has been $D = 100$ during the last several weeks. In this stationary situation, the order-up-to level has converged to $S = 110$ at all echelons, with $V = 10$, $Q = 100$. Now suppose that the demand at the retailer has increased to $D = 110$ in a certain week due to a special offer. This has the following effect on the order quantities:

| | | | | | |
|---|---|---|---|---|---|
| Retailer: | $D = 110$ | $S = 121$ | $V = 11$ | $I = 110 - 110 = 0$ | $Q = 121 - 0 = 121$ |
| Wholesale: | $D = 121$ | $S = 133$ | $V = 12$ | $I = 110 - 121 = -11$ | $Q = 133 + 11 = 144$ |
| Importer: | $D = 144$ | $S = 158$ | $V = 14$ | $I = 110 - 144 = -34$ | $Q = 158 + 34 = 192$ |
| Factory: | $D = 192$ | $S = 211$ | $V = 19$ | $I = 110 - 192 = -82$ | $Q = 211 + 82 = 293$ |

Note that the order quantities have increased by 21%, 44%, 92% and 193%, respectively, due to a 10% increase in demand! Next, suppose that in the weeks after the special offer the demand is back at $D = 100$. This has the following effect on the order quantities:

Retailer: $D = 100$ $S = 110$ $V = 10$ $I = 0 + 121 - 100 = 21$ $Q = 110 - 21 = 89$
Wholesale: $D = 89$ $S = 98$ $V = 9$ $I = -11 + 144 - 89 = 44$ $Q = 98 - 44 = 54$
Importer: $D = 54$ $S = 59$ $V = 5$ $I = -34 + 192 - 54 = 104$ $Q = 0$ $(59 - 104 < 0)$
Factory: $D = 0$ $S = 0$ $V = 0$ $I = -82 + 293 = 211$ $Q = 0$

Observe that the retailer is back in the stationary situation with the order-up-to level, but that the higher echelons show a counterreaction with small or zero order quantities. In the factory, 193 units have been produced above the usual quantity of 100, while only 82 are forwarded to the importer as backlog of the previous week. So, the factory has had high production cost, perhaps with working overtime, and now ends up with a high inventory level. The effect of a demand of $D = 100$ in the next week is:

Retailer: $D = 100$ $S = 110$ $V = 10$ $I = 21 + 89 - 100 = 10$ $Q = 110 - 10 = 100$
Wholesale: $D = 100$ $S = 110$ $V = 10$ $I = 44 + 54 - 100 = -2$ $Q = 110 + 2 = 112$
Importer: $D = 112$ $S = 123$ $V = 11$ $I = 104 + 0 - 112 = -8$ $Q = 123 + 8 = 131$
Factory: $D = 131$ $S = 144$ $V = 13$ $I = 211 + 0 - 131 = 80$ $Q = 144 - 80 = 64$

The wholesale dealer is back at $S = 110$, but the higher echelons exhibit an upward movement. At the factory, this upward movement is quenched by the fact that negative order quantities are not feasible. The next weeks, the oscillations fade away:

Retailer: $D = 100$ $S = 110$ $V = 10$ $I = 10 + 100 - 100 = 10$ $Q = 110 - 10 = 100$
Wholesale: $D = 100$ $S = 110$ $V = 10$ $I = -2 + 112 - 100 = 10$ $Q = 110 - 10 = 100$
Importer: $D = 100$ $S = 110$ $V = 10$ $I = -8 + 131 - 100 = 23$ $Q = 110 - 23 = 87$
Factory: $D = 87$ $S = 96$ $V = 9$ $I = 80 + 64 - 87 = 57$ $Q = 96 - 57 = 39$

After two more weeks, the steady state will be reached for the whole supply chain. □

Possible causes of the oscillation effect may be special offers, advertisements, an announcement of a price change, but also stochastic fluctuations and holidays. The oscillation effect can be limited by better exchange of information between echelons, by exponential smoothing in forecasting future demand to avoid overreaction on small changes in the demand pattern, by removing echelons from the supply chain if feasible, and by limiting the (necessity of) safety stocks, e.g., by better quality control. For (linear) chains it is important that ordering decisions in a stocking point consider beside the inventory position at the stocking point itself at least the inventory positions in all lower stocking points and the demand in the lowest stocking point. This consideration led Clark & Scarf [18] to the introduction of the concept of echelon inventory. The echelon stock in a stocking point is defined as the number of units that are present at the stocking point itself plus those that have already passed the stocking point but have not yet been committed to outside customers. Hence, the echelon stock includes the inventory on hand at lower echelons and the inventory in transit to lower echelons. As with the usual inventory concept we can distinguish for the echelon stock at a stocking point:

**echelon inventory on hand:** the number of units present at the stocking point or at lower echelons, or in transit to lower echelons; this quantity plays a role in determining holding costs;

**net echelon inventory (stock):** the echelon inventory on hand minus the amount of backlog at the lowest echelon;

**echelon inventory position:** the net echelon inventory plus the number of units on order from a higher echelon but not yet delivered; this quantity may be required for determining a reorder instant;

**echelon safety stock:** the average echelon inventory position just before a reorder instant minus the average demand during the lead time to the stocking point; this quantity is related to the service level constraint or the cost of stockouts or losses.

As mentioned in Section 1.3, a distinction is made in multi-echelon inventory systems between pull and push policies. A more refined classification is:

- pure pull policies:

  - use only local information;
  - lead to increasing variability at higher echelons;
  - require high safety stocks;

- pull policies based on echelon stock:
  - use information from lower echelons;
  - lead to limited variability;
  - require limited safety stocks;
- push policies:
  - requires central information;
  - apply central inventory management;
  - require central safety stocks;
  - may apply reshuffling of stocks on echelon level.

A push policy takes inventories and costs of all echelons into account.

Next, we will consider a linear chain consisting of two echelons, and, hence, consisting of two stocking points. The parameters of the model are:

- $D$: the demand per unit of time at the lower echelon;
- $a_1$, $a_2$: the ordering cost for the higher/lower echelon, respectively;
- $h_1$, $h_2$: the holding cost per unit per unit of time at the higher/lower echelon, respectively; in some cases, $h_1 = rv_1$ and $h_2 = rv_2$ with $v_1$, $v_2$ the purchasing cost of a unit upon arrival at the higher/lower echelon, respectively, and $r$ the carrying charge;
- $L_1$, $L_2$: the lead time to the higher/lower echelon, respectively.

The demand at the lower echelon is assumed to be deterministic and continuous in time. Also, the lead times are deterministic. In most situations, it will hold that $h_2 > h_1$ because storage is usual more expensive at the lower echelons and because the value of the goods will increase ($v_2 > v_1$) when they move down the echelons if only since more shipping cost has been invested. Further, we may have $a_1 > a_2$ when external ordering is more expensive than internal ordering. However, if transport charges are included in the internal ordering cost distances and volumes may also play a role.

The decision variables are:

- $Q_1$, $Q_2$: the order quantity at the higher/lower echelon, respectively;
- $s_1$, $s_2$: the reorder point at the higher/lower echelon, respectively.

First, the case that the lead time from the higher to the lower echelon is negligible ($L_2 = 0$) will be considered. As a preliminary property, we will argue that it is optimal to have $Q_1 = nQ_2$ for some integer $n$, $n \geq 1$. We will show with the aid of an example that a policy with $Q_1/Q_2$ not equal to a positive integer can be improved upon.

**Example 3.2** Suppose that a policy with $Q_1 = \frac{3}{2}Q_2$ is considered for a two-echelon linear chain. Then, the quantity $\frac{1}{2}Q_2$ lies on stock without use. Keep $Q_2$ fixed, so that the cost incurred at the lower echelon is fixed. The cost at the higher echelon in 3 periods of the lower echelon is for $Q_1 = \frac{3}{2}Q_2$: $2a_1 + \frac{3}{2}h_1Q_2$. Consider first alternative 0: order alternatingly $Q_1 = Q_2$ and $Q_1 = 2Q_2$. The cost at the higher echelon in 3 periods of the lower echelon is then: $2a_1 + h_1Q_2$, that is a saving of $\frac{1}{2}h_2Q_2$ with respect to the original policy. Hence, the policy with $Q_1 = \frac{3}{2}Q_2$ cannot be optimal. But the policy of alternative 0 can be further improved upon. For comparison, the cost at the higher echelon in 6 periods of the lower echelon is $4a_1 + 2h_1Q_2$. Next, consider alternative 1: order always $Q_1 = Q_2$. The cost at the higher echelon in 6 periods of the lower echelon is $6a_1 + 0h_1Q_2$. This is better than alternative 0 if $a_1 < h_1Q_2$. Finally, consider alternative 2: order always $Q_1 = 2Q_2$. The cost at the higher echelon in 6 periods of the lower echelon is $3a_1 + 3h_1Q_2$. This is better than alternative 0 if $a_1 > h_1Q_2$. Hence, $Q_1 = nQ_2$ with either $n = 1$ or $n = 2$ is cheaper than alternatingly ordering $Q_1 = Q_2$ and $Q_1 = 2Q_2$. In this way, it can be shown that the optimal policy is of the form $Q_1 = nQ_2$, with $n$ a positive integer. □

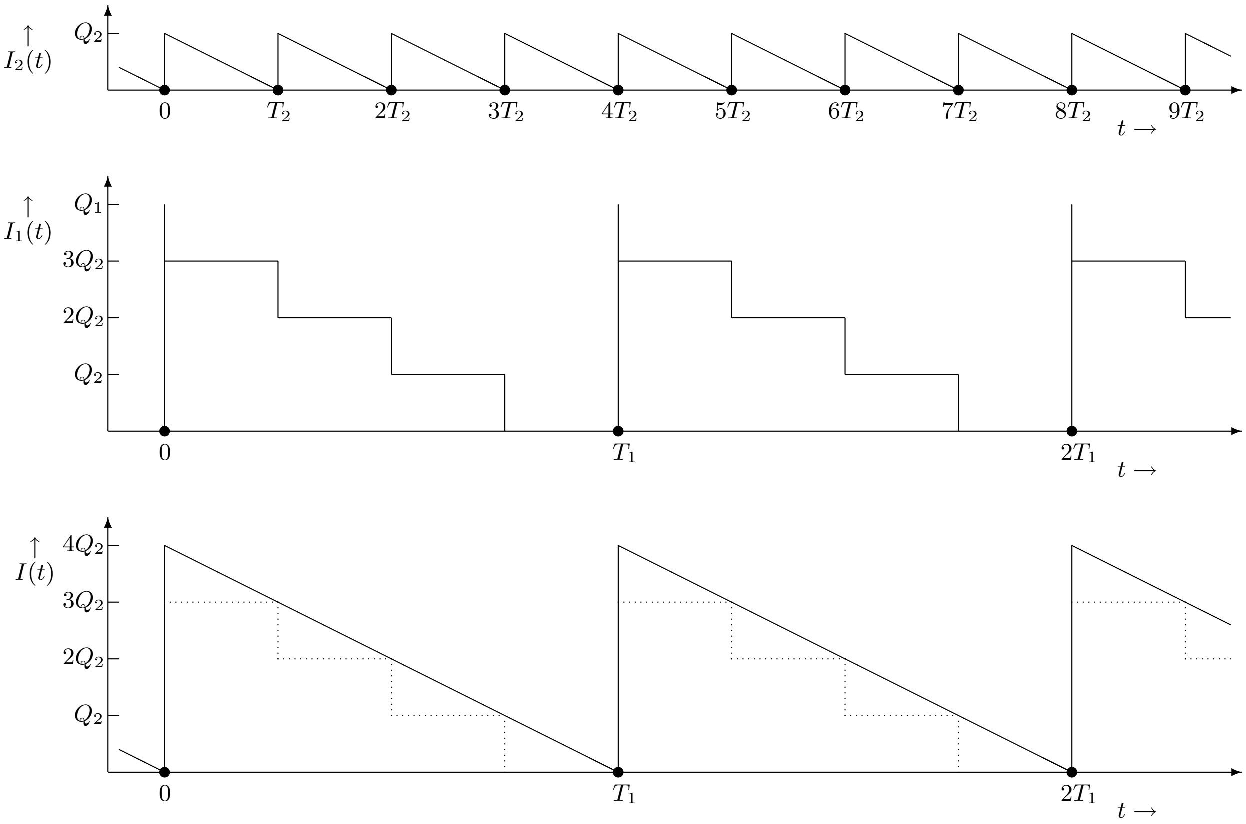

Figure 3.3: Inventory on hand at lower echelon (top), at higher echelon (middle) and in system (bottom).

Next, the average cost per unit of time with a push policy with $Q_1 = nQ_2$ will be derived. At the lower echelon the inventory level $I_2(t)$ follows a common saw-tooth pattern, cf. the upper graph in Figure 3.3, so that the average stock at the lower echelon is found to be $\frac{1}{2}Q_2$. The inventory level at the higher echelon, however, behaves like a step function: when an order of $Q_1$ units is delivered, a fraction $\frac{1}{n}$ is immediately forwarded to the lower echelon. During the first cycle at the lower echelon the inventory level at the higher echelon is constant and equal to $\frac{n-1}{n}Q_1 = (n-1)Q_2$. Afterwards, a second fraction $\frac{1}{n}$ of $Q_1$ is forwarded to the lower echelon, so that during the second cycle at the lower echelon the inventory level at the higher echelon is $\frac{n-2}{n}Q_1 = (n-2)Q_2$, etcetera. This continues until the end of the cycle at the higher echelon. The middle graph of Figure 3.3 shows the inventory on hand $I_1(t)$ at the higher level for the case $n = 4$. Hence, the average inventory level at the higher echelon is:

$$\frac{1}{n}\sum_{j=0}^{n-1}\frac{j}{n}Q_1 = \frac{\frac{1}{2}n(n-1)}{n}\frac{Q_1}{n} = \tfrac{1}{2}(n-1)Q_2 = \tfrac{1}{2}(Q_1 - Q_2). \tag{3.1}$$

This implies that the average cost per unit of time with a push policy is:

$$C(Q_1, Q_2) = D\left[\frac{a_1}{Q_1} + \frac{a_2}{Q_2}\right] + \tfrac{1}{2}\left[h_1(Q_1 - Q_2) + h_2Q_2\right], \tag{3.2}$$

or, with $Q_1 = nQ_2$,

$$C(n, Q_2) = \frac{D}{Q_2}\left[\frac{a_1}{n} + a_2\right] + \tfrac{1}{2}\,Q_2[nh_1 + h_2 - h_1]. \tag{3.3}$$

**Remark 3.1** An alternative derivation and interpretation of the cost function follows by noting that the echelon inventory level $I(t) \doteq I_1(t) + I_2(t)$ at the higher echelon follows a saw-tooth pattern, cf. the lower graph in Figure 3.3, so that the average echelon inventory level is $\frac{1}{2}Q_1$. Associating a holding cost $h_1$ to this inventory and noting that the echelon inventory at the higher echelon includes the inventory at the lower echelon, an additional holding cost $h_2 - h_1$ has to be associated to the average inventory level of $\frac{1}{2}Q_2$ at the lower echelon. So, the total holding cost becomes $\frac{1}{2}\,h_1Q_1 + \frac{1}{2}\,(h_2 - h_1)Q_2$ which is the same as the last term in (3.2) and (3.3). □

The optimal $Q_2$ at a fixed value of $n$ is:

$$\hat{Q}_2(n) = \sqrt{\frac{2(\frac{1}{n}a_1 + a_2)D}{nh_1 + h_2 - h_1}}, \tag{3.4}$$

with corresponding minimum average cost per unit of time

$$\hat{C}(n) = \sqrt{2(\tfrac{1}{n}a_1 + a_2)(nh_1 + h_2 - h_1)D}. \tag{3.5}$$

Since the square root function is monotonously increasing, the optimal real-valued minimum of this cost function is obtained by solving the equation

$$\tfrac{d}{dn}\,\hat{C}^2(n) = 2D\tfrac{d}{dn}\{\tfrac{1}{n}a_1(h_2 - h_1) + nh_1a_2 + a_2(h_2 - h_1) + a_1h_1\} = 2D\{-\tfrac{1}{n^2}a_1(h_2 - h_1) + h_1a_2\} = 0.$$

Hence, the optimal real-valued $n$ minimizing the foregoing cost function is:

$$n^*_{\text{push}} = \sqrt{\frac{a_1(v_2 - v_1)}{a_2v_1}} = \sqrt{\frac{a_1(h_2 - h_1)}{a_2h_1}}. \tag{3.6}$$

The optimal push policy is obtained by rounding $n^*_{\text{push}}$ to a positive integer in such a way that the cost (3.5) is minimal.

Alternatively, consider a pull policy for the same situation. The lower echelon determines $Q^*_2$ on the basis of the local cost as the economic order quantity, cf. (1.2), (1.6),

$$Q^*_2 = \sqrt{\frac{2a_2D}{h_2}}, \qquad C^*_2 = \sqrt{2a_2h_2D}. \tag{3.7}$$

The higher echelon must *follow* with $Q_1 = nQ^*_2$ for some $n \geq 1$, with additional cost

$$C_1(n) = a_1\,\frac{D}{Q_1} + \tfrac{1}{2}\,h_1(Q_1 - Q^*_2) = a_1\,\frac{D}{nQ^*_2} + \tfrac{1}{2}\,h_1(n-1)Q^*_2. \tag{3.8}$$

Note that applying the EOQ formula to higher echelons does not work due to the fact that demand does not occur continuously in time, cf. (3.1). The optimal real-valued $n$ for given $Q^*_2$ is:

$$n^*_{\text{pull}} = \frac{1}{Q^*_2}\sqrt{\frac{2a_1D}{h_1}} = \sqrt{\frac{a_1h_2}{a_2h_1}}, \tag{3.9}$$

with corresponding cost

$$C^*_1 = \sqrt{2a_1h_1D} - \tfrac{1}{2}\,h_1Q^*_2 = \sqrt{2a_1h_1D} - \tfrac{1}{2}\,\frac{h_1}{h_2}\sqrt{2a_2h_2D}. \tag{3.10}$$

Again, the optimal pull policy is obtained by rounding $n^*_{\text{pull}}$ to a positive integer in such a way that the cost (3.8) is minimal. For both policies it is noted that if $n^* = 1$, the higher echelon has no stocking function and only serves for cross-docking. Observe that both ratios $n^*_{\text{push}}$, cf. (3.6), and $n^*_{\text{pull}}$, cf. (3.9), are independent of the demand rate $D$. The difference between these ratios is largest when $h_2 - h_1$ is small in comparison to $h_2$ while $a_1/a_2$ is large.

**Example 3.3** Consider a linear chain consisting of two echelons, with parameters $a_1 = \$20$, $a_2 = \$1$, $v_1 = \$1$, $v_2 = \$2$, $D = 1000$, $r = \$0.25$, so that $h_1 = \$0.25$, $h_2 = \$0.50$. First, we determine the optimal push policy. From (3.6) it follows that $n^*_{\text{push}} = \sqrt{20}$ so that $n^*_{\text{push}} = 4$ or $n^*_{\text{push}} = 5$. The corresponding costs according to (3.5) are $\hat{C}(4) = \sqrt{500(5+1)(4+1)} = 50\sqrt{6} \approx \$122.47$ and $\hat{C}(5) = \sqrt{500(4+1)(5+1)} = 50\sqrt{6} \approx \$122.47$. Since there is no difference, both values are optimal. The corresponding optimal order quantities are $\hat{Q}^*_2(4) = 40\sqrt{6} \approx 98$, $Q^*_1 = 4\hat{Q}^*_2(4) = 392$, and $\hat{Q}^*_2(5) = \frac{100}{3}\sqrt{6} \approx 82$, $Q^*_1 = 5\hat{Q}^*_2(5) = 410$, respectively.

Next, we determine the optimal pull policy. From (3.7) it follows that $Q^*_2 = 20\sqrt{10} \approx 63$ with $C^*_2 = 10\sqrt{10} \approx 31.62$. Further, (3.9) leads to $n^*_{\text{pull}} = \sqrt{40}$ so that $n^*_{\text{pull}} = 6$ or $n^*_{\text{pull}} = 7$. The corresponding costs according to (3.8) are $C_1(6) = 52.91 + 39.38 = 92.29$ and $C_1(7) = 45.35 + 47.25 = 92.60$. Hence, $n^*_{\text{pull}} = 6$ is optimal and $Q^*_1 = 6Q^*_2 = 378$. The total cost with the optimal pull policy is $C = 31.62 + 92.29 = \$123.91$ which is not much worse than the total cost with the optimal push policy. $\square$

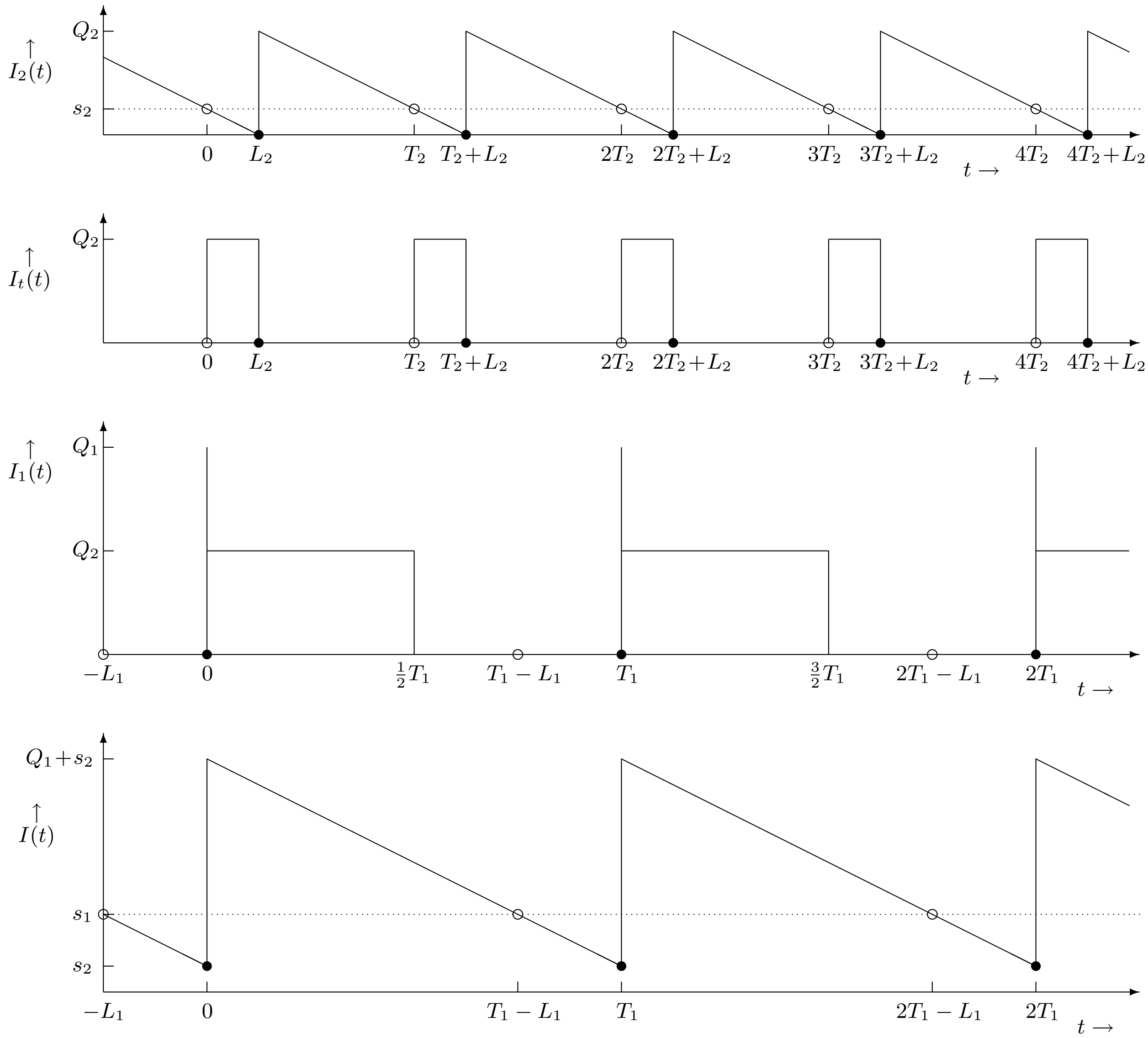

Figure 3.4: Inventory on hand at lower echelon, in transit, at higher echelon, in system (from top to bottom).

Finally, the consequences of a nonnegligible lead time $L_2 > 0$ between the higher and the lower echelon will be studied. By the usual arguments, cf. Section 1.2.1, it is still optimal to have orders arrive at the lower echelon when the inventory level reaches zero. This implies that orders must be placed at the higher echelon when the inventory position at the lower echelon is at the level, cf. (1.9),

$$s_2 = D \cdot L_2. \tag{3.11}$$

Moreover, this means that orders must arrive at the higher echelon at instants when the inventory position at the lower echelon is at the level $s_2$ so that part of the arriving order can immediately be transferred to the lower echelon. This situation is illustrated for the case $Q_1 = 2Q_2$ in Figure 3.4 which shows the saw-tooth pattern of the inventory level $I_2(t)$ at the lower echelon, the step pattern of the inventory $I_t(t)$ in transit between the echelons and of the inventory level $I_1(t)$ at the higher echelon, and the saw-tooth pattern of the echelon inventory level $I(t) \doteq I_1(t) + I_t(t) + I_2(t)$. Observe that the echelon stock $I(t)$ is replenished when it reaches the level $s_2$ due to the time lag $L_2$ before the goods can be used to satisfy external demand at the lower echelon. The open (closed) dots in the upper two graphs of Figure 3.4 indicate the ordering (delivery) instants at the lower echelon, those in the lower two graphs the ordering (delivery) instants at the higher echelon. Note that the reorder instants at the higher echelon do not correspond to a unique inventory level of $I_1(t)$. However, they do correspond to a unique level of the echelon inventory position $I(t)$. The echelon reorder point at the higher echelon is

$$s_1 = s_2 + D \cdot L_1 = D \cdot (L_1 + L_2). \tag{3.12}$$

The average number of units in transit is

$$\bar{I}_t = \frac{Q_2 L_2 + 0(T_2 - L_2)}{T_2} = \frac{Q_2}{T_2} L_2 = DL_2, \tag{3.13}$$

which is independent of the order quantities $Q_1$ and $Q_2$. Hence, if a holding cost is incurred for units in transit, a fixed amount is added to the cost function (3.3) and the same order quantities are optimal as those in the corresponding case with $L_2 = 0$.

**Example 3.4** Consider a linear chain consisting of two echelons, with parameters $a_1 = \$32$, $a_2 = \$1$, $h_1 = \$8$, $h_2 = \$9$, $D = 1800$ per year, $L_1 = 0.04$ year and $L_2 = 0.01$ year. From (3.6) it follows that $n^*_{\text{push}} = 2$. The corresponding optimal order quantities are $\hat{Q}^*_2(2) = 60$ and $Q^*_1 = 2\hat{Q}^*_2(2) = 120$. The average cost per year for the optimal push policy is $\hat{C}(2) = \$1020$, cf. (3.5). From (3.9) it follows that $n^*_{\text{pull}} = 6$. The optimal order quantity at the lower echelon is $Q^*_2 = 20$. Hence, the optimal order quantity at the higher echelon is $Q^*_1 = 6Q^*_2 = 120$. The cost per year for the optimal pull policy is $C = 180 + 880 = \$1060$, which is \$40 (3.9%) more than the minimum cost. For both policies the echelon reorder point for the higher echelon is $s_1 = 1800 \cdot 0.05 = 90$ and the reorder point for the lower echelon is $s_2 = 1800 \cdot 0.01 = 18$. ☐

**Exercise 3.1** *Consider the following two-echelon inventory problem. A retailer has a shop in the city center and a depot at the industrial site of the same city. Space in the city center is scarce. Therefore, the holding cost is much higher there. For a certain item, the holding cost is \$3 per unit per year at the shop and \$1 per unit per year at the depot. The internal ordering cost of the shop is \$10 per order. The external ordering cost of the depot is \$80 per order. The demand for this item at the shop is 6000 units per year. This demand is approximately deterministic and constant over time. All demand must be satisfied from stock. Lead times are deterministic. First determine the optimal order quantity for the shop leaving the cost incurred at the depot out of account. Compute the minimum average cost at the shop according this policy. Given this policy at the shop, determine the optimal order quantity and the minimum average cost for the depot. Next, consider the two echelons combined. Determine the optimal order quantities for the shop and the depot. Compute the minimum average cost per year corresponding to this policy. Determine for both policies for both echelons the reorder points for the case that the lead time to the depot is 10 days and the lead time from the depot to the shop is* $\frac{1}{2}$ *day (assume 300 working-days in a year).*

### 3.2.2 Diverging chains

Consider a diverging distribution chain consisting of two echelons. $N$ retailers order a certain item from the same distribution center, while the distribution center orders this item from an external supplier. The parameters of the model are:

- $a_0$: the ordering cost for the distribution center;
- $a_i$: the ordering cost for retailer $i$, $i = 1, \ldots, N$;
- $h_0$: the holding cost per unit per unit of time for the distribution center;
- $h_i$: the holding cost per unit per unit of time for retailer $i$, $i = 1, \ldots, N$;
- $D_i$: the demand per unit of time at retailer $i$, $i = 1, \ldots, N$.

The aggregated demand per unit of time at the distribution center is $D_0 = \sum_{i=1}^{N} D_i$.
Consider the following subclass of push policies. The distribution center uses a reorder cycle of $R$ units of time and retailer $i$ orders $k_i$ times per cycle of length $R$ from the distribution center, $i = 1, \ldots, N$. Then, the order quantity $Q_0$ of the distribution center and the order quantity $Q_i$ of retailer $i$ are determined by

$$Q_0 = D_0 R; \qquad Q_i = D_i R / k_i, \quad i = 1, \ldots, N. \tag{3.14}$$

The objective function is the total ordering and holding cost per unit of time. It can be derived in a similar way as that for a linear chain, cf. (3.2), noting that the distribution center has similar holding costs related to each retailer, cf. (3.1):

$$C(R,\mathbf{k}) = \frac{a_0}{R} + \sum_{i=1}^{N} \frac{k_i a_i}{R} + \tfrac{1}{2} R h_0 \sum_{i=1}^{N} \frac{k_i - 1}{k_i} D_i + \tfrac{1}{2} R \sum_{i=1}^{N} \frac{h_i D_i}{k_i}. \tag{3.15}$$

This function can be rewritten as

$$C(R,\mathbf{k}) = \frac{1}{R}\left[a_0 + \sum_{i=1}^{N} k_i a_i\right] + \tfrac{1}{2} R \left[h_0 D_0 + \sum_{i=1}^{N} \frac{h_i - h_0}{k_i} D_i\right]. \tag{3.16}$$

This cost function has the same structure as that of the indirect grouping coordinated replenishment problem, cf. (2.1), only the role of the parameters with respect to the integers $k_i$, $i = 1, \ldots, N$, is reversed. The cost function (3.16) can be transformed into (2.1) by the following correspondences, cf. Das & Goyal [20]. The two-echelon diverging supply chain with $N$ retailers is equivalent to a coordinated replenishment problem for $N+1$ items (the distribution center is related to an item labeled 0) with parameters: for $i = 1, \ldots, N$,

$$T = \frac{1}{R}; \quad \hat{A} = 0; \quad \hat{a}_i = \tfrac{1}{2}(h_i - h_0) D_i; \quad \hat{a}_0 = \tfrac{1}{2} h_0 D_0; \quad \tfrac{1}{2}\hat{h}_i \hat{D}_i = a_i; \quad \tfrac{1}{2}\hat{h}_0 \hat{D}_0 = a_0. \tag{3.17}$$

Note that the reorder frequency of the distribution center is forced to be $k_0 = 1$. After the above transformation, the optimal policy can be determined with the aid of Goyal's Algorithm 2.1 or a policy can be generated by any heuristic for the indirect grouping coordinated replenishment problem, with the modification that $k_0$ is kept equal to 1.

**Remark 3.2** It may seem odd that a coordinated replenishment problem is formulated with family ordering cost $\hat{A} = 0$. The cost function (3.16), however, does not become separable in this case, because the integer constraints on the frequencies $k_i$, $i = 1, \ldots, N$, are essential for its validity. □

It is quite difficult to determine a pull policy for a diverging chain. Each stocking point at the lowest echelon can readily determine its optimal reorder quantity according to the EOQ-formula, but without any coordination this will lead to irregular, acyclic demand patterns at higher echelons. This makes it cumbersome to formulate the average inventory levels at higher echelons and to determine optimal reorder policies for higher echelons.

Table 3.1: Data per year and a push policy for Example 3.5.

| Retailer $i$ | $a_i$ | $h_i$ | $D_i$ | $\hat{a}_i$ | $\hat{h}_i\hat{D}_i$ | $\hat{a}_i/(\hat{h}_i\hat{D}_i)$ | $k_i$ | $Q_i$ | $T_i$ |
|---|---|---|---|---|---|---|---|---|---|
| a | \$1.80 | \$0.50 | 2900 | \$580 | \$3.60 | 161.1 | 3 | 133 | 0.0458 |
| b | \$2.00 | \$1.10 | 1850 | \$925 | \$4.00 | 231.3 | 3 | 85 | 0.0458 |
| c | \$1.20 | \$0.90 | 2750 | \$1100 | \$2.40 | 458.3 | 4 | 95 | 0.0344 |
| d | \$3.20 | \$0.30 | 1600 | \$160 | \$6.40 | 25.0 | 1 | 220 | 0.1375 |
| e | \$3.10 | \$0.90 | 3200 | \$1280 | \$6.20 | 206.5 | 3 | 147 | 0.0458 |
| f | \$2.70 | \$0.30 | 1400 | \$140 | \$5.40 | 25.9 | 1 | 193 | 0.1375 |

**Example 3.5** Consider a two-echelon diverging distribution chain consisting of a distribution center and $N = 6$ retailers. At the distribution center, the ordering cost is $a_0 = \$10$ and the holding cost is $h_0 = \$0.10$ per unit per year. The ordering cost $a_i$, the holding cost $h_i$ per unit per year and the demand $D_i$ per year at retailer $i$, $i = 1, \ldots, 6$, can be found in Table 3.1. The aggregated demand per year at the distribution center is $D_0 = 13{,}700$. The corresponding coordinated replenishment problem concerns $N + 1 = 7$ items, has family ordering cost $\hat{A} = \$0$, supplementary ordering costs $\hat{a}_0 = \$685$ and $\hat{a}_i$, $i = 1, \ldots, N$, given in Table 3.1, and holding cost times demand $\hat{h}_0\hat{D}_0 = \$20$ and $\hat{h}_i\hat{D}_i$, $i = 1, \ldots, N$, given in Table 3.1, cf. (3.17). We determine an ordering policy for the transformed problem according to the heuristic of Kaspi & Rosenblatt (Algorithm 2.3). First, the items are indexed in ascending value of $(\hat{A} + \hat{a}_i)/(\hat{h}_i\hat{D}_i) = \hat{a}_i/(\hat{h}_i\hat{D}_i)$. This leads to the reference "item" being retailer d (for the distribution center, $\hat{a}_0/(\hat{h}_0\hat{D}_0) = 34.3$). The initial family cycle based on this reference item is $T^{(0)} = 5\sqrt{2} \approx 7.071$. The reorder frequencies for the other retailers, based on this cycle length, are $k_{\text{a}} = k_{\text{b}} = k_{\text{e}} = 3$, $k_{\text{c}} = 4$, $k_{\text{f}} = 1$ ($k_0 = 1$ and $k_{\text{d}} = 1$ are kept fixed). Given

these frequencies, the optimal cycle length is $T^{(1)} = 7.270$. The policy improvement step does not alter the frequencies. Hence, the family cycle becomes $T^{(2)} = 7.270$, so that the review period is $R = 0.1375$ year ($\approx 7.15$ week) for the two-echelon system. The order quantity of the distribution center becomes $Q_0 = 1884$. The reorder frequency $k_i$ and the order quantity $Q_i$ of retailer $i$, $i = 1, \ldots, N$, are included in Table 3.1. The corresponding average yearly cost is $\hat{C} = \$601.99$ which is slightly higher than the minimum cost according to Goyal's Algorithm 2.1: $C^* = \$601.62$.

Table 3.2: Optimal pull policies at the lower echelon for Example 3.5.

| Retailer | $Q_i^*$ | $T_i^*$ | $M_i^*$ |
|---|---|---|---|
| a | 144.5 | 0.0498 | 20.07 |
| b | 82.0 | 0.0443 | 22.56 |
| c | 85.6 | 0.0311 | 32.11 |
| d | 184.8 | 0.1155 | 8.66 |
| e | 148.5 | 0.0464 | 21.55 |
| f | 158.7 | 0.1134 | 8.82 |

Table 3.2 contains the individual optimal reorder quantity $Q_i^*$, and the corresponding optimal cycle $T_i^*$ and number of orders per year, $M_i^*$, for retailer $i$, $i = 1, \ldots, 6$, according to the basic EOQ models. Because the reorder cycles are asynchronous, these optimal pull policies lead to an acyclic demand pattern at the distribution center. □

The application of push policies in divergent multi-echelon chains requires synchronization between the retailers such that an order is shipped to each retailer at the instant when an order from the external supplier arrives at the distribution center (partial cross-docking). In case of nonnegligible lead times $L_0$ to the distribution center and $L_i$ from the distribution center to retailer $i$, $i = 1, \ldots, N$, the distribution center should place an order when the echelon inventory position reaches the echelon reorder point, cf. (3.12),

$$s_0 = D_0 L_0 + \sum_{i=1}^{N} D_i L_i = \sum_{i=1}^{N} D_i (L_0 + L_i). \tag{3.18}$$

When the order arrives at the distribution center, the inventory position at retailer $i$ will be $s_i = D_i L_i$, $i = 1, \ldots, N$.

**Exercise 3.2** *Consider a two-echelon distribution chain for a non-food product consisting of a local warehouse and five supermarkets. At the warehouse, the ordering cost is \$20 and the holding cost is \$0.20 per unit per year. The ordering cost $a_{2i}$, the holding cost $h_{2i}$ per unit per year and the demand $D_{2i}$ per year at supermarket $i$, $i = 1, \ldots, 5$, can be found in the table below. Determine an ordering policy for the transformed problem according to the heuristic of Kaspi & Rosenblatt and interpret the results in terms of the distribution chain. Compare the found push policy with the individual optimal policies.*

| Supermarket $i$ | $a_{2i}$ | $h_{2i}$ | $D_{2i}$ |
|---|---|---|---|
| 1 | \$8 | \$0.60 | 2000 |
| 2 | \$10 | \$0.80 | 5000 |
| 3 | \$10 | \$0.50 | 4000 |
| 4 | \$12 | \$0.60 | 5000 |
| 5 | \$15 | \$0.50 | 4000 |

## 3.3 Time-varying demand

In this section, multi-echelon inventory systems are considered with known but time varying demand in the stocking points at the lowest echelon of diverging chains. The optimal pull policy is easy to find for general diverging chains by the following procedure:

**Algorithm 3.1** [Pull policy]

Step 1: Apply the Wagner-Whitin Algorithm 1.2 to the stocking points at the lowest echelon.

Step 2: Aggregate the order quantities of the lowest echelon to obtain the demand process for the one but lowest echelon.

Step 3: Apply the Wagner-Whitin algorithm to the stocking points at the one but lowest echelon.

Step 4: Continue in this way until the highest echelon is reached.

Alternatively, the Silver-Meal heuristic can be used. □

The pull policy ignores cost dependencies between the echelons. Therefore, the aim is to determine a (sub)optimal push policy that takes cost dependencies into account. An integral approach by dynamic programming is rather difficult. Here, a sequential approach (from the lowest to the highest echelon) will be discussed for a linear chain. The parameters of the model with a horizon of $H$ periods are:

- $D_{2t},\ t = 1, \ldots, H$: the demand per period at the lower echelon;
- $a_1,\ a_2$: the ordering cost at the higher/lower echelon, respectively;
- $h_1,\ h_2$: the holding cost per unit per period at the higher/lower echelon, respectively.

The decision variables are:

- $Q_{1t},\ t = 1, \ldots, H$: the order quantity at the higher echelon in period $t$;
- $Q_{2t},\ t = 1, \ldots, H$: the order quantity at the lower echelon in period $t$.

The demand at the higher echelon is $D_{1t} = Q_{2t},\ t = 1, \ldots, H$. Blackburn & Millen [13] have developed a heuristic for an assembly system, that is, for a converging chain. For the case of a linear chain with two echelons it is based on an ideal ratio $n$ between the order quantities at the higher and at the lower echelon derived from the constant demand model, cf. (3.6):

$$n = \max\left\{1, \sqrt{\frac{a_1(h_2 - h_1)}{a_2 h_1}}\right\}. \tag{3.19}$$

Note that $n$ is taken minimally equal to 1, but is otherwise not rounded to an integer. Motivated by cost function (3.3) coordination between the echelons is extorted by adding a fraction $\frac{1}{n}$ of the ordering cost $a_1$ to each order placed at the lower echelon and by adding $(n-1)h_1$ to the holding cost at the lower echelon. Hence, the following modified cost parameters are defined for the lower echelon:

$$\hat{a}_2 \doteq a_2 + a_1/n, \qquad \hat{h}_2 = nh_1 + h_2 - h_1. \tag{3.20}$$

The heuristic of Blackburn & Millen [13], confined to a linear chain consisting of two echelons, proceeds as follows.

**Algorithm 3.2** [Blackburn & Millen]

Step 1: Determine the ideal ratio $n$ according to (3.19).

Step 2: Apply the Wagner-Whitin algorithm to the lower echelon with modified costs $\hat{a}_2$, $\hat{h}_2$, cf. (3.20), to determine $Q_{2t},\ t = 1, \ldots, H$.

Step 3: Use $D_{1t} = Q_{2t},\ t = 1, \ldots, H$, and apply the Wagner-Whitin algorithm to the higher echelon with costs $a_1$, $h_1$, to determine $Q_{1t},\ t = 1, \ldots, H$. Here, the ratio $n$ is abandoned.

Alternatively, the Silver-Meal heuristic can be used. □

**Example 3.6** Let $a_1 = \$20$, $a_2 = \$4$, $v_1 = \$20$, $v_2 = \$30$, $r = \$0.02/\$/\text{month}$, so that $h_1 = \$0.40$, $h_2 = \$0.60$. The demand in the next 8 months $D_{2t},\ t = 1, \ldots, 8$, is given in Table 3.3. For a push policy according to Blackburn & Millen compute $n = \sqrt{\frac{20\times 10}{4\times 20}} = \frac{1}{2}\sqrt{10} \approx 1.58 > 1$ by (3.19) so that the modified costs become $\hat{a}_2 = 4 + \frac{20}{1.58} = 16.65$ and $\hat{h}_2 = 0.2 + 1.58 \times 0.4 = 0.832$, cf. (3.20). Application of the Silver-Meal heuristic at the lower echelon with the modified costs yields $Q_{2t},\ t = 1, \ldots, 8$, cf. Table 3.3. The estimated

Table 3.3: Push policy according to Blackburn & Millen and Silver-Meal for Example 3.6.

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $D_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{2t}$ | 50 | 0 | 60 | 55 | 0 | 45 | 25 | 25 |
| $Q_{1t}$ | 50 | 0 | 60 | 100 | 0 | 0 | 50 | 0 |

Table 3.4: Push policy according to Blackburn & Millen and Wagner-Whitin for Example 3.6.

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $D_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{2t}$ | 50 | 0 | 60 | 55 | 0 | 45 | 25 | 25 |
| $Q_{1t}$ | 50 | 0 | 60 | 55 | 0 | 95 | 0 | 0 |

cost at the lower echelon based on the modified costs is $6\hat{a}_2 + (10 + 15)\hat{h}_2 = \$120.71$. Application of the Silver-Meal heuristic at the higher echelon with demand $D_{1t} = Q_{2t}$ and costs $a_1$, $h_1$ gives $Q_{1t},\ t = 1, \ldots, 8$, cf. Table 3.3. The real cost at the higher echelon is $4a_1 + (90 + 25)h_1 = \$126$ and the real cost at the lower echelon is $6a_2 + (10 + 15)h_2 = \$39$. The total real cost is \$165. The realized ratio between the order frequencies is $n = 6/4 = 1.50$. Application of the Wagner-Whitin algorithm gives the results as displayed in Table 3.4, with cost $4a_1 + (25 + 50)h_1 = \$110$ at the higher echelon, cost $6a_2 + (10 + 15)h_2 = \$39$ at the lower echelon and total cost \$149. Note that there are alternative optimal solutions: $Q_{16} = 70$, $Q_{17} = 0$, $Q_{18} = 25$, and $Q_{16} = 70$, $Q_{17} = 25$, $Q_{18} = 0$.

Now consider a pull policy. Application of the Silver-Meal heuristic at the lower echelon with the unmodified costs $a_2$ and $h_2$ yields $Q_{2t},\ t = 1, \ldots, 8$, cf. Table 3.5. The cost at the lower echelon is $8a_2 + 0h_2 = \$32$. Application of the Silver-Meal heuristic at the higher echelon with demand $D_{1t} = Q_{2t}$ and costs $a_1$ and $h_1$ gives $Q_{1t},\ t = 1, \ldots, 8$, cf. Table 3.5. The cost at the higher echelon is $4a_1 + (10 + 40 + 2 \times 15 + 25)h_1 = \$122$. The total cost is \$154. The realized ratio between the order frequencies is $n = 8/4 = 2$. Observe that the Silver-Meal heuristic leads in this case to a pull policy that is cheaper than the push policy obtained with this heuristic! Application of the Wagner-Whitin algorithm gives the same results for the lower echelon but different results for the higher echelon, cf. Table 3.6. The optimal pull policy has cost $4a_1 + (10 + 15 + 25 + 50)h_1 = \$120$ at the higher echelon, and total cost \$152 which is 2% higher than that of the optimal push policy. Again, there are alternative optimal solutions. □

Table 3.5: Pull policy for Example 3.6 by Silver-Meal.

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $D_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{1t}$ | 50 | 0 | 115 | 0 | 0 | 70 | 0 | 25 |

Table 3.6: Pull policy for Example 3.6 by Wagner-Whitin.

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $D_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{2t}$ | 40 | 10 | 60 | 40 | 15 | 45 | 25 | 25 |
| $Q_{1t}$ | 50 | 0 | 60 | 55 | 0 | 95 | 0 | 0 |

**Exercise 3.3** *A firm owns a shop in the center of a city and a depot near the harbor. Goods are first delivered at the depot where they are checked, repacked and kept in storage if necessary. For a certain item the demand in the shop is deterministic but the amount varies from week to week. The demand for the next six weeks is given in the following table.*

| week | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| demand | 24 | 16 | 32 | 24 | 12 | 50 |

*The ordering cost at the depot is \$60, the ordering cost at the shop is \$15 and mainly consists of shipping cost. Orders are delivered at the first day of a week. Holding costs are only charged for units on stock at the* *end of a week. The holding costs are \$1 per unit per week at the shop and \$0.20 per unit per week at the depot. Determine a pull policy with the aid of the Silver-Meal heuristic for this two-echelon inventory system. Determine a push policy with the aid of the Blackburn & Millen heuristic combined with the Silver-Meal heuristic. Compute the total cost over six weeks corresponding to both policies. Show that these policies can simply be improved upon.*

## 3.4 Stochastic demand

This section is concerned with two-echelon diverging chains with independent, stationary stochastic demand at the stocking points of the lower echelon. This represents a situation of a central depot or distribution center and $N$ local stocking points or retailers. The aim is to develop an good integral control, that is, a push policy, cf. Sections 1.3, 3.2.1. The means to this goal are multi-echelon ordering rules, also called base-stock control, cf. Clark & Scarf [18]. Such strategies are optimal for diverging chains provided that the demand at the retailers is not too strongly varying, that is, there is no strong "imbalance" between the retailers. The approach is this section is based on Van der Heijden [68]. This problem is equivalent to a hierarchical production planning problem for a family of items with common components (ignoring production capacity constraints).
The inventory policy is based on periodic review of the stock, applied as follows:

- the central depot uses an $(R, S_0)$ policy: with fixed intervals $R$ the *echelon*-stock is ordered up to $S_0$;
- the order of the central depot is delivered after a lead time $L_0$;
- the delivery instant at the central depot is the reorder instant for the retailers;
- retailer $i$ uses an $(R, S_i)$ policy: the stock is ordered up to $S_i,\ i = 1, \ldots, N$;
- the order of retailer $i$ is delivered after a lead time $L_i,\ i = 1, \ldots, N$.

This basic procedure may require some modifications:

- if the total demand during the lead time $L_0$ is more than expected, then the stock in the central depot will be too small to fully satisfy the orders of the retailers; in such a case,
  - retailer $i$ is first replenished up to the expected demand in the period $L_i + R$; the remaining stock is divided over the retailers by ratios $p_i,\ i = 1, \ldots, N$, *or*
  - the total understock is divided over the retailers by ratios $f_i,\ i = 1, \ldots, N$;
- if the total demand during the lead time $L_0$ is less than expected, then the stock in the central depot will be larger than the sum of the order quantities of the retailers; in such a case,
  - the remaining stock could stay in the central depot, *or*
  - the total overstock is divided over the retailers by ratios $f_i,\ i = 1, \ldots, N$.

In the sequel, the following model will be considered:

- there is a stockless central depot (it only serves for cross-docking and reallocation of goods);
- there is full information at the central depot to apply a push policy for fair allocation of stock over the retailers and for lower stocks;
- no stock redistribution is possible between retailers;
- all shortages are backlogged; at each retailer, a fill rate constraint is imposed, cf. (1.38);
- there are no capacity constraints.

The parameters of the model are:

- $L_0$: the lead time from the external supplier to the central depot;
- $L_i$: the lead time from the central depot to retailer $i,\ i = 1, \ldots, N$;

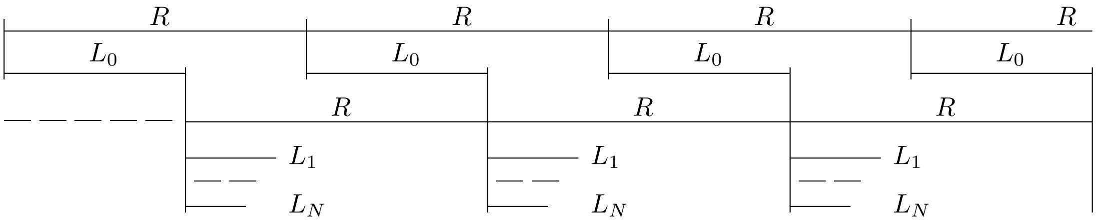

Figure 3.5: The review periods at the two echelons shifted in time (depot above, retailers below).

- $R$: the review period (this period is of equal length for both echelons, but it is shifted in time by an amount $L_0$ for the retailers in comparison to the central depot; see Figure 3.5);
- $D_{i,t}$: the demand at retailer $i$ during a period $t$; this is a random variable with time-independent mean $\mu_i \doteq E\{D_{i,t}\}$ and standard deviation $\sigma_i \doteq \sigma\{D_{i,1}\}$, $i = 1, \ldots, N$;
- $h_i$: the holding cost per unit per unit of time at retailer $i$, $i = 1, \ldots, N$;
- $\beta_i$: the target fill rate at retailer $i$, $i = 1, \ldots, N$.

It will be assumed that demand is independent across time periods and that demand is independent across retailers. Since the review period is given, the ordering and shipping costs are fixed. The lead times are assumed to be known and constant, and the approach below especially concerns situations where the lead time to the central depot is large in comparison to the lead times to the retailers ($L_0 \gg L_1, \ldots, L_N$). The decision variables are:

- $S_0$: the order-up-to level of the central depot;
- $S_i^*$: the maximum order-up-to level of retailer $i$, $i = 1, \ldots, N$;
- $f_i$: the rationing fraction for allocating units to retailer $i$, $i = 1, \ldots, N$.

Here, the variables are to be chosen such that $\sum_{i=1}^N S_i^* = S_0$ and $\sum_{i=1}^N f_i = 1$. The control mechanism of the inventory policy to be discussed uses the following auxiliary random variables:

- $z_i$: the inventory position of retailer $i$ just before a reorder instant of the retailers, $i = 1, \ldots, N$;
- $S_i$: the inventory position of retailer $i$ just after a reorder instant of the retailers, $i = 1, \ldots, N$;
- $Q$: the order quantity of the central depot.

When an order arrives at the central depot, the echelon inventory position of the central depot is $S_0 - \tilde{D}_{L_0}$, with $\tilde{D}_{L_0} \doteq \sum_{i=1}^N D_{i,L_0}$ the total demand at all retailers during the lead time $L_0$. The amount that is allocated to retailer $i$ by the central depot at such an instant is $S_i - z_i$ with

$$S_i = S_i^* - f_i \tilde{D}_{L_0}, \quad i = 1, \ldots, N. \tag{3.21}$$

Hence, the maximum order-up-to level of retailer $i$ is reduced by a fraction $f_i$ of the total demand $\tilde{D}_{L_0}$ during the lead time to the central depot $L_0$, for $i = 1, \ldots, N$. The echelon inventory position just before the central depot places an order is $S_0 - Q$. The echelon inventory on hand just before the order of the central depot arrives is:

$$\sum_{j=1}^N z_j = S_0 - Q - \sum_{i=1}^N D_{i,L_0} = S_0 - Q - \tilde{D}_{L_0}.$$

Hence, the allocation to retailer $i$ can alternatively be described by $S_i - z_i$, with

$$S_i = S_i^* - f_i \left[ S_0 - Q - \sum_{j=1}^N z_j \right], \quad i = 1, \ldots, N.$$

The echelon inventory position just after the delivery of an order at the central depot is:

$$\sum_{i=1}^N S_i = S_0 - \tilde{D}_{L_0} = \sum_{j=1}^N z_j + Q.$$

A problem in the foregoing allocation arises if $S_i - z_i < 0$ for some retailer $i = i_0$. This phenomenon is indicated by *imbalance*. The cause for imbalance is

- little demand at retailer $i_0$ (the remaining amount of stock $z_{i_0}$ is relatively high);
- more demand at other retailers (the total demand $\tilde{D}_{L_0}$ is such that $S_{i_0}$ is smaller than $z_{i_0}$, cf. (3.21)).

In case of imbalance, the allocation rule would require redistribution of stock from retailer $i_0$ to some other retailers, which would lead to extra shipping cost. Let us study the occurrence of imbalance. The inventory position at retailer $i$ just after the order quantity $Q$ of the central depot has been reallocated is, cf. (3.21),

$$S_i^{(1)} = S_i^* - f_i \tilde{D}_{L_0^{(1)}}, \quad i = 1, \ldots, N. \tag{3.22}$$

Here, it is assumed that there is no imbalance at the outset. The inventory position at retailer $i$ just before the next allocation instant, that is, a review period $R$ later, is

$$z_i^{(2)} = S_i^* - f_i \tilde{D}_{L_0^{(1)}} - D_{i,R}, \quad i = 1, \ldots, N. \tag{3.23}$$

This inventory level must be raised to the target level

$$S_i^{(2)} = S_i^* - f_i \tilde{D}_{L_0^{(2)}}, \quad i = 1, \ldots, N. \tag{3.24}$$

Imbalance occurs if the target level $S_i^{(2)}$ is smaller than the actual inventory level $z_i^{(2)}$ before allocation. Hence, the imbalance due to retailer $i$ is:

$$\Omega_i \doteq [z_i^{(2)} - S_i^{(2)}]^+ = [f_i \tilde{D}_{L_0^{(2)}} - f_i \tilde{D}_{L_0^{(1)}} - D_{i,R}]^+, \quad i = 1, \ldots, N. \tag{3.25}$$

This definition of imbalance was introduced by De Kok [21]; Zipkin [83] used another type of imbalance measure. It is seen that the imbalance only depends on the rationing factors $f_i$, and not on the order-up-to levels $S_i^*$, $i = 1, \ldots, N$. To avoid the occurrence of imbalance as much as possible, the ratios $f_i$ will be chosen in such a way that the total expected imbalance is minimal. To this end, Van der Heijden [68] assumes that the random variables, cf. (3.23), (3.24),

$$Y_i \doteq z_i^{(2)} - S_i^{(2)} = f_i \tilde{D}_{L_0^{(2)}} - f_i \tilde{D}_{L_0^{(1)}} - D_{i,R}, \quad i = 1, \ldots, N, \tag{3.26}$$

approximately possess a normal distribution. Note that the random variables $Y_i$, $i = 1, \ldots, N$, can take positive and negative values, but are not symmetric around their mean. The mean of $Y_i$ is independent of $f_i$:

$$E\{Y_i\} = E\{z_i^{(2)} - S_i^{(2)}\} = -E\{D_{i,R}\} = -R\mu_i, \quad i = 1, \ldots, N. \tag{3.27}$$

Note that the terms on the righthand side of (3.26) contain overlapping demand since the second lead time $L_0^{(2)}$ partly coincides with the review period $R$ at the retailers and possibly with the previous lead time $L_0^{(1)}$. To determine the variance of $Y_i$, two cases are distinguished, cf. Figure 3.6:

- $R > L_0$: then the contributions to $Y_i$ are:
  period $(0, L_0)$: $-f_i \tilde{D}_{L_0^{(1)}}$,
  period $(L_0, R)$: $-D_{i,R-L_0}$,
  period $(R, L_0 + R)$: $f_i \sum_{j \neq i} D_{j,L_0} - (1 - f_i) D_{i,L_0}$;
- $R < L_0$: then the contributions to $Y_i$ are:
  period $(0, R)$: $-f_i \tilde{D}_R$,
  period $(R, L_0)$: $0$ (since $L_0^{(1)}$ and $L_0^{(2)}$ overlap in this period),
  period $(L_0, L_0 + R)$: $f_i \sum_{j \neq i} D_{j,R} - (1 - f_i) D_{i,R}$.

With $M \doteq \min\{R, L_0\}$, the variance of $Y_i$ can be written as

$$\sigma^2\{Y_i\} = [R - 2f_i M]\, \sigma_i^2 + 2f_i^2 M \sum_{j=1}^{N} \sigma_j^2, \quad i = 1, \ldots, N. \tag{3.28}$$

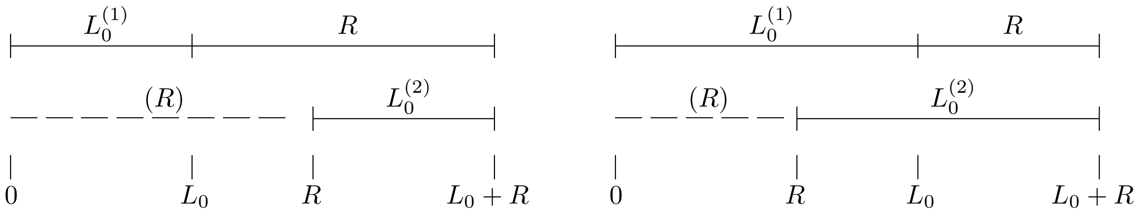

Figure 3.6: Review periods at the retailers and lead times to the depot: two cases.

The mean imbalance $E\{\Omega_i\} = E\{[Y_i]^+\}$ is increasing with the variance $\sigma^2\{Y_i\}$ when $Y_i$ has a normal distribution, $i = 1, \ldots, N$. It follows by differentiation of (3.28) that the variance $\sigma^2\{Y_i\}$ is minimal with respect to $f_i$ at $\hat{f}_i = \frac{1}{2}\sigma_i^2 / \sum_{j=1}^N \sigma_j^2$, $i = 1, \ldots, N$. However, these values of $f_i$, $i = 1, \ldots, N$, are not feasible with respect to the constraint $\sum_{i=1}^N f_i = 1$. Therefore, the constrained optimization problem has to be solved with the Lagrange multiplier method. In order to minimize the function

$$\sum_{i=1}^N E\{\Omega_i\} + \eta \left[1 - \sum_{i=1}^N f_i\right], \tag{3.29}$$

with $\eta$ the Lagrange multiplier, the following equations must be satisfied:

$$\frac{d}{df_j} \sum_{i=1}^N E\{\Omega_i\} = \eta, \quad j = 1, \ldots, N; \qquad \sum_{i=1}^N f_i = 1. \tag{3.30}$$

Under the assumption of normally distributed random variables $Y_i$, the total expected imbalance reads, cf. (3.25), (3.26),

$$\sum_{i=1}^N E\{\Omega_i\} = \sum_{i=1}^N E\{[Y_i]^+\} = \sum_{i=1}^N \sigma\{Y_i\} NL(R\mu_i / \sigma\{Y_i\}); \tag{3.31}$$

here, $NL(.)$ denotes the normal loss function defined by (A.18). The derivatives of this function with respect to the rationing factors $f_i$ can be determined by means of the chain rule. From (3.28) and the properties of the normal loss function, cf. Appendix A.2.1, it follows that, for $j = 1, \ldots, N$,

$$\frac{d}{df_j} \sum_{i=1}^N E\{\Omega_i\} = \frac{d}{d\sigma\{Y_j\}} E\{[Y_j]^+\} \frac{d}{df_j} \sigma\{Y_j\} = \frac{M f_{0,1}(R\mu_j / \sigma\{Y_j\})}{\sigma\{Y_j\}} \left[-\sigma_j^2 + 2 f_j \sum_{i=1}^N \sigma_i^2\right]; \tag{3.32}$$

here, $f_{0,1}(.)$ denotes the density of the standard normal distribution. Since $\sigma\{Y_j\}$ depends on $f_j$, cf. (3.28), it is not possible to determine an explicit expression for the rationing fraction $f_j$ for which $dE\{\Omega_j\}/df_j = \eta$ for a given value of $\eta$. Therefore, a numerical search procedure must be applied to find this root, e.g., the bisection method for finding a root of an equation, cf. Appendix B.3, with initial search interval $[\hat{f}_j, 1]$, for $j = 1, \ldots, N$. Further, in order to find the value of $\eta$ for which the fractions $f_j$ sum to one another application of the bisection method is required. In this way, the minimum total expected imbalance with respect to the fractions $f_j$, $j = 1, \ldots, N$, can be obtained numerically. Because of this total expected imbalance minimizing property, the above rationing rule is referred to as "Balanced Stock" rationing.

In the next step of the approach by Van der Heijden [68] it is assumed that the fractions $f_i$ are fixed according to the above procedure, and the maximum order-up-to levels $S_i^*$ are determined such that the target fill rates $\beta_i$ are reached, $i = 1, \ldots, N$. In this step, it is assumed that by the choice of the rationing fractions the occurrence of imbalance can be ignored. As a consequence, retailer $i$ is allocated a positive order quantity $Q_i = S_i - z_i$ at each order arrival instant at the central depot, with mean, cf. (3.27),

$$E\{Q_i\} = E\{S_i - z_i\} = E\{D_{i,R}\} = R\mu_i, \quad i = 1, \ldots, N. \tag{3.33}$$

The mean stockout $B_i$ accumulated during a review period $R$ at retailer $i$ is, cf. (1.38),

$$E\{B_i\} = E\{[D_{i,L_i+R} - S_i]^+\} - E\{[D_{i,L_i} - S_i]^+\}, \quad i = 1, \ldots, N. \tag{3.34}$$

In the present case, both $D_{i,L_i+R}$, $D_{i,L_i}$ and $S_i$ are random variables, since $S_i$ depends on $\tilde{D}_{L_0}$, for $i = 1,\ldots,N$, cf. (3.21). In order to reach the target fill rates, the maximum order-up-to levels $S_i^*$ have to be determined such that, cf. (1.38),

$$E\{[D_{i,L_i+R} + f_i\tilde{D}_{L_0} - S_i^*]^+\} - E\{[D_{i,L_i} + f_i\tilde{D}_{L_0} - S_i^*]^+\} \leq (1-\beta_i)R\mu_i, \quad i = 1,\ldots,N. \tag{3.35}$$

Note that the smallest $S_i^*$, $i = 1,\ldots,N$, that satisfy the above inequalities, can be determined for each retailer separately (if imbalance is ignored). Such a problem involves random variables

$$X_i^{(1)} = D_{i,L_i+R} + f_i\tilde{D}_{L_0}, \quad X_i^{(2)} = D_{i,L_i} + f_i\tilde{D}_{L_0}, \quad i = 1,\ldots,N. \tag{3.36}$$

Since the lead time $L_0$ to the central depot occurs before $L_i + R$, it follows that $\tilde{D}_{L_0}$ and $D_{i,L_i+R}$ are independent random variables. The means of these variables are:

$$E\{X_i^{(1)}\} = (L_i + R)\mu_i + f_i L_0 \sum_{j=1}^{N} \mu_j, \quad E\{X_i^{(2)}\} = L_i\mu_i + f_i L_0 \sum_{j=1}^{N} \mu_j, \quad i = 1,\ldots,N. \tag{3.37}$$

The variances of these variables are:

$$\sigma^2\{X_i^{(1)}\} = (L_i + R)\sigma_i^2 + f_i^2 L_0 \sum_{j=1}^{N} \sigma_j^2, \quad \sigma^2\{X_i^{(2)}\} = L_i\sigma_i^2 + f_i^2 L_0 \sum_{j=1}^{N} \sigma_j^2, \quad i = 1,\ldots,N. \tag{3.38}$$

The distributions of $X_i^{(1)}$ and $X_i^{(2)}$ cannot readily be obtained for general probability distributions of the demand per period at the various retailers. Therefore, Van der Heijden [68] approximates the distributions of $X_i^{(1)}$ and $X_i^{(2)}$ by mixtures of Erlang distributions with mean and variance given by (3.37) and (3.38), respectively, cf. Appendix A.2.5. With these approximations, the smallest maximum order-up-to levels $S_i^*$ satisfying (3.35) can be determined numerically. The computation of the decision variables for the replenishment strategy is summarized below.

**Algorithm 3.3** [Balanced Stock rationing]

Step 1: Compute the rationing fractions $f_i$, $i = 1,\ldots,N$, such that the expected total imbalance is minimal, with the aid of the Lagrange multiplier method. This step requires a numerical procedure to find the value of the Lagrange multiplier $\eta$ for which $\sum_{i=1}^{N} f_i = 1$. For each considered value of $\eta$, it requires a numerical procedure to find the value of $f_i$ for which $\frac{d}{df_i}\, E\{\Omega_i\} = \eta$, for each $i$, $i = 1,\ldots,N$.

Step 2: Given the values of the fractions $f_i$, $i = 1,\ldots,N$, compute the smallest maximum order-up-to levels $S_i^*$ such that the target fill rates $\beta_i$ are reached, for $i = 1,\ldots,N$. The actual fill rates are based on an approximation by a mixture of Erlang distributions.

The order-up-to level at the central depot follows as $S_0 = \sum_{i=1}^{N} S_i^*$. □

Note that there are no decision variables left to minimize the total holding cost. This issue could be addressed in an extension of the present model with a stock-keeping central depot, cf. Van der Heijden [69]. Then, items could be stored at the central depot in case the demand $\tilde{D}_{L_0}$ during the lead time to the central depot is much less than average and if the holding cost $h_0$ per item per unit of time at the central depot is lower than those at the retailers. In the present model, the total cost is fixed after application of Algorithm 3.3. The expected average inventory on hand at retailer $i$ is, cf. (1.39):

$$E\{\bar{I}_i\} = \tfrac{1}{2}\,[E\{[S_i - D_{i,L_i}]^+\} + E\{[S_i - D_{i,L_i+R}]^+\}], \quad i = 1,\ldots,N. \tag{3.39}$$

Since for any real $x$, $[x]^+ = x + [-x]^+$, it follows with (3.21) that

$$E\{[S_i - D_{i,L_i+R}]^+\} = E\{S_i\} - (L_i + R)\mu_i + E\{[D_{i,L_i+R} + f_i\tilde{D}_{L_0} - S_i^*]^+\}, \quad i = 1,\ldots,N, \tag{3.40}$$

and

$$E\{[S_i - D_{i,L_i}]^+\} = E\{S_i\} - L_i\mu_i + E\{[D_{i,L_i} + f_i\tilde{D}_{L_0} - S_i^*]^+\}, \quad i = 1,\ldots,N, \tag{3.41}$$

with

$$E\{S_i\} = S_i^* - f_i L_0 \sum_{j=1}^{N} \mu_j, \quad i = 1,\ldots,N. \tag{3.42}$$

Hence, the expected average inventory on hand $E\{\bar{I}_i\}$ at retailer $i$ can readily be computed since the expectations on the right-hand sides of (3.40) and (3.41) have already been calculated when $S_i^*$ was determined satisfying the inequalities (3.35), $i = 1, \ldots, N$. Further, the average safety stock at retailer $i$ is, cf. Section 1.2.7:

$$E\{V_i\} = E\{S_i\} - (R + L_i)\mu_i = S_i^* - f_i L_0 \sum_{j=1}^{N} \mu_j - (R + L_i)\mu_i, \quad i = 1, \ldots, N. \tag{3.43}$$

Finally, the average inventory in transit to retailer $i$ is $L_i\mu_i$, independent of the policy (due to the assumption of complete backlogging).

Table 3.7: Data and results for Example 3.7 (in weeks).

| | Central depot | Retailer 1 | Retailer 2 | Retailer 3 |
|---|---|---|---|---|
| Mean demand $\mu_i$ | (1200) | 100 | 300 | 800 |
| St. deviation demand $\sigma_i$ | (412) | 85 | 50 | 400 |
| Target fill rate $\beta_i$ | | 0.95 | 0.90 | 0.70 |
| Lead time $L_i$ | 2 | 1 | 1 | 1 |
| Rationing fraction $f_i$ | | 0.031 | 0.189 | 0.780 |
| Maximum order-up-to level $S_i^*$ | 5094 | 480 | 1109 | 3505 |
| Mean on-hand inventory $E\{\bar{I}_i\}$ | | 258 | 220 | 587 |
| Mean stockout $E\{B_i\}$ | | 5 | 30 | 240 |
| Mean safety stock $E\{V_i\}$ | | 205 | 55 | 34 |
| Mean imbalance $E\{\Omega_i\}$ | | 4.86 | 0.19 | 1.11 |
| Simulated mean on-hand inventory | | 257 | 219 | 585 |
| Simulated actual fill rate | | 0.948 | 0.898 | 0.698 |
| Simulated mean imbalance | | 0.5 | 0.4 | 0.0 |

**Example 3.7** Van der Heijden [68] contains a numerical example of which the data can be found in Table 3.7. The review period is $R = 1$ week. The most critical retailer is retailer 1, since the coefficient of variation $\sigma_1/\mu_1 = 0.85$ and the target fill rate $\beta_1$ are maximal over all retailers. This retailer gets the smallest rationing fraction. Simulation indicates that the actual fill rates are slightly below the targets (in the order of 0.002). The simulated mean imbalance turns out to be much smaller than the mean imbalance predicted by the normal loss function in the model. □

**Remark 3.3** De Kok [21] introduced another rationing policy for the present multi-echelon inventory problem called Consistent Appropriate Share (CAS) rationing. In this approach, items are allocated to the retailers based on safety stock ratios; see also Verrijdt & De Kok [74, 75]. At the order arrival instant of the central depot, the items are allocated to the retailers such that their inventories are raised to the levels

$$S_i = E\{D_{i,L_i+R}\} + p_i \left[ S_0 - \tilde{D}_{L_0} - \sum_{j=1}^{N} E\{D_{j,L_j+R}\} \right], \quad i = 1, \ldots, N;$$

that is, retailer $i$ gets allocated its expected demand during the period $L_i + R$ plus a fraction of the remaining amount of items, $i = 1, \ldots, N$. These fractions are chosen proportional to the safety stocks:

$$p_i = cK_i\sigma\{D_{i,L_i+R}\}, \quad i = 1, \ldots, N;$$

with $K_i$ a safety factor such that the target fill rate $\beta_i$ is reached, $i = 1, \ldots, N$, and $C$ a normalization constant such that $\sum_{i=1}^{N} p_i = 1$. This allocation falls in the framework of the more general allocation rule (3.21) by identifying

$$f_i = p_i, \qquad S_i^* = E\{D_{i,L_i+R}\} + p_i \left[ S_0 - \sum_{j=1}^{N} E\{D_{j,L_j+R}\} \right], \quad i = 1, \ldots, N.$$

The fraction $p_i$ is relatively large if $\sigma\{D_{i,L_i+R}\}$ is large, meaning high uncertainty in demand, and/or if $K_i$ is large, meaning a high target fill rate $\beta_i$, $i = 1, \ldots, N$. But the allocation to a retailer with relatively high fraction $f_i = p_i$ is most sensitive to variations in the total demand $\tilde{D}_{L_0}$, cf. (3.21). Hence, most sensitivity is directed to retailers of which the inventory is already difficult to control. This explains why CAS rationing yields strong imbalance and erroneous results in some cases where simulation results do not agree with model predictions, in particular with respect to the actual fill rates, even after improvements of the method for cases when some fractions $p_i$ are negative, cf. Van der Heijden [68]. For instance, CAS rationing assigns a large fraction to retailer 1 ($p_1 > 0.95$) in Example 3.7. The Balanced Stock rationing is better in avoiding negative allocations and is better in predicting the actual performance of a system, while it leads to lower mean stock levels than CAS rationing. □

Table 3.8: Data and results for Example 3.8 (in weeks) with review period $R = 1$ week.

| | Depot | Retailer 1 | Retailer 2 | Retailer 3 | Retailer 4 | Retailer 5 |
|---|---|---|---|---|---|---|
| Mean demand $\mu_i$ | (800) | 100 | 100 | 100 | 400 | 100 |
| St. deviation demand $\sigma_i$ | (240) | 50 | 100 | 50 | 200 | 50 |
| Target fill rate $\beta_i$ | | 0.95 | 0.95 | 0.90 | 0.95 | 0.95 |
| Lead time $L_i$ | 2 | 0.5 | 0.5 | 0.5 | 0.5 | 1.0 |
| Rationing fraction $f_i$ | | 0.078 | 0.116 | 0.078 | 0.650 | 0.078 |
| Maximum order-up-to level $S_i^*$ | 3715 | 354 | 555 | 323 | 2063 | 419 |
| Mean on-hand inventory $E\{\bar{I}_i\}$ | | 132 | 272 | 104 | 632 | 148 |
| Mean stockout $E\{B_i\}$ | | 5 | 5 | 10 | 20 | 5 |
| Mean safety stock $E\{V_i\}$ | | 80 | 219 | 49 | 422 | 95 |
| Mean imbalance $E\{\Omega_i\}$ | | 0.60 | 7.40 | 0.60 | 1.27 | 0.60 |

**Example 3.8** To study the effect of various parameter values, consider the following numerical example with five retailers of which the data can be found in Table 3.8. The review period is $R = 1$ week. The parameters of retailer 1 form the base case. Retailer 2 has a higher coefficient of variation of the demand, and is the most critical retailer. Retailer 3 has a lower target fill rate. Retailer 4 has a higher mean demand with the same coefficient of variation. Retailer 5 has a higher lead time. In this case, retailer 4 gets the largest rationing fraction, then retailer 2 (in the order of decreasing standard deviation of the demand, as suggested by the optimal solution to the unconstrained problem, cf. above (3.29)), and retailers 1, 3 and 5 get the smallest rationing fraction which is the same for these three retailers because the rationing fractions do not depend on the lead times and the target fill rates. To study the effect of the review period, Table 3.9 contains results for the same two-echelon system but with a review period of $R = 2$ weeks. Increasing the review period shifts the rationing fractions somewhat in favor of retailer 4 with the highest standard deviation of the demand. Further, the predicted mean imbalance decreases considerably, but remains largest for retailer 2 with the highest coefficient of variation of the demand. □

Table 3.9: Results for Example 3.8 (in weeks) with review period $R = 2$ weeks.

| | Depot | Retailer 1 | Retailer 2 | Retailer 3 | Retailer 4 | Retailer 5 |
|---|---|---|---|---|---|---|
| Rationing fraction $f_i$ | | 0.075 | 0.095 | 0.075 | 0.680 | 0.075 |
| Maximum order-up-to level $S_i^*$ | 4448 | 444 | 629 | 403 | 2469 | 504 |
| Mean on-hand inventory $E\{\bar{I}_i\}$ | | 179 | 332 | 143 | 800 | 189 |
| Mean stockout $E\{B_i\}$ | | 10 | 10 | 20 | 40 | 10 |
| Mean safety stock $E\{V_i\}$ | | 74 | 227 | 33 | 380 | 84 |
| Mean imbalance $E\{\Omega_i\}$ | | 0.08 | 4.15 | 0.08 | 0.17 | 0.08 |

The inventory policy described in this section can be extended in several directions, cf. Van der Heijden [68, 69]. It can be generalized to divergent multi-echelon systems with more than two echelons. It can deal with nonstationary demand by adjusting the maximum order-up-to levels $S_i^*$, $i = 1, \ldots, N$ (the fractions $f_i$ should be recomputed only if the demand or lead time characteristics change considerably). Finally, the model can be extended with a stock keeping central depot, but then the expressions for the actual fill rates and the procedure for determining the rationing fractions become more complicated and require more approximative assumptions.

Literature on this subject includes Eppen & Schrage [24] who consider a model with penalty cost, Federgruen [27] who applies stochastic dynamic programming, Inderfurth [40] who provides a survey on safety stocks in divergent multi-echelon systems, Odanaka et al. [53] who consider a multi-echelon production system, Yoo et al. [81] who are concerned with distribution requirement planning, Korugan & Gupta [46] who study a multi-echelon system with returns, Axsäter & Zhang [8] and Axsäter [6] who provide an exact analysis of a continuous review system with (compound) Poisson demand, and Ettl et al. [25] who solve a supply network with base-stock control and service level constraints by the conjugate gradient method.

**Exercise 3.4** *Complete the details of the derivation of the variance* $\sigma^2\{Y_i\},\ i = 1, \ldots, N$, *in (3.28).*

**Exercise 3.5** *Complete the details of the derivation of the derivative of the total expected imbalance with respect to the rationing fractions in (3.32).*

**Exercise 3.6** *Consider the results presented in Table 3.7. Explain why the mean physical stock is only 585 for retailer 3 under Balanced Stock rationing while the maximum order-up-to level for this stocking point is 3505.*

**Exercise 3.7** *Consider the results presented in Tables 3.8, 3.9. Explain why the mean stockout seems to double for each retailer when the review period doubles, while the maximum order-up-to levels and the mean physical stocks only increase with a much smaller percentage.*

**Exercise 3.8** *Consider a two-echelon system for an item with a central depot and two retailers. The lead time to the central depot is* $L_0 = 2$ *weeks and the lead time to the retailers are* $L_1 = 0.1$ *and* $L_2 = 0.2$ *week, respectively. The weekly demand has a mean of 20 units and a standard deviation of 5 units at both retailers. Determine ordering policies according to Algorithm 3.3 at target fill rates of 0.90 for both retailers, for a review period of* $R = 1$ *week as well as for a review period of* $R = 2$ *weeks.*

# Chapter 4

# Other Inventory Systems with Interactions

This chapter is devoted to some other inventory problems with interaction. Section 4.1 is concerned with the inventory control of a set of items that have to share a common storage space. Section 4.2 deals with a single item of which units may fail and then are replaced by another unit while the failed unit may be repaired later on.

## 4.1 Inventory capacity constraint

In this section a set of $N$ items is considered which are ordered independently but which have to share a common storage space. The demand of all items is assumed to be deterministic and continuous in time. The parameters of the model are:

- $a_i$: the ordering cost for item $i,\ i = 1, \ldots, N$;
- $h_i$: the holding cost per unit per unit of time for item $i,\ i = 1, \ldots, N$;
- $D_i$: the demand per unit of time for item $i,\ i = 1, \ldots, N$;
- $\phi_i$: the space occupied by a unit of item $i,\ i = 1, \ldots, N$;
- $\Phi$: the total inventory capacity.

The decision variables are:

- $Q_i$: the order quantity for item $i,\ i = 1, \ldots, N$.

The average cost per unit of time is found as a simple generalization of the single item case, cf. (1.1),

$$C(\mathbf{Q}) \doteq C(Q_1, \ldots, Q_N) = \sum_{i=1}^{N} \left[ a_i \frac{D_i}{Q_i} + \tfrac{1}{2} h_i Q_i \right]. \tag{4.1}$$

This cost function has to be minimized subject to the capacity constraint

$$\sum_{i=1}^{N} \phi_i \, Q_i \leq \Phi. \tag{4.2}$$

This constraint is based on the supposition that a space $\phi_i Q_i$ has to be reserved for the maximum inventory level of item $i,\ i = 1, \ldots, N$, since the ordering of the various items is not coordinated.

**Algorithm 4.1** [Lagrange constrained optimization]

Step 1: Compute the unconstrained optimal order quantities, cf. (1.2),

$$Q_i^* = \sqrt{2 a_i D_i / h_i}, \quad i = 1, \ldots, N.$$

Step 2: Verify whether the capacity constraint (4.2) is satisfied. If so, stop. Otherwise, continue with Step 3.

Step 3: Solve the constrained optimization problem with the aid of a Lagrange multiplier, that is, minimize the following function with respect to the vector $\mathbf{Q}$ and the scalar $\eta$:

$$f(\mathbf{Q},\eta) = \sum_{i=1}^{N} \left[ a_i \frac{D_i}{Q_i} + \tfrac{1}{2} h_i Q_i \right] + \eta \left[ \sum_{i=1}^{N} \phi_i \, Q_i - \Phi \right]. \tag{4.3}$$

The last step requires a numerical search procedure (e.g., bisection). □

The partial derivative of the function $f(\mathbf{Q},\eta)$ with respect to $Q_i$ is

$$\frac{d}{dQ_i} f(\mathbf{Q},\eta) = -a_i \frac{D_i}{Q_i^2} + \tfrac{1}{2} h_i + \eta \phi_i, \quad i = 1, \ldots, N.$$

This implies that the optimal order quantity for fixed $\eta$ is

$$\hat{Q}_i(\eta) = \sqrt{\frac{2 a_i D_i}{h_i + 2\eta\phi_i}}, \quad i = 1, \ldots, N. \tag{4.4}$$

Further, the inventory capacity is fully used (consider the partial derivative of $f(\mathbf{Q},\eta)$ with respect to $\eta$) so that

$$\sum_{i=1}^{N} \phi_i \sqrt{\frac{2 a_i D_i}{h_i + 2\eta\phi_i}} = \Phi. \tag{4.5}$$

This relation implicitly determines the scalar $\eta$. The value of $\eta$ has to be determined with the aid of a numerical procedure, e.g., by the bisection method for finding a root of an equation, cf. Appendix B.3. To this end, it is useful to have an upper and lower bound on the optimal value of $\eta$. Note that by (4.4),

$$\eta = \frac{a_i D_i}{\phi_i Q_i^2} - \frac{h_i}{2\phi_i}, \quad i = 1, \ldots, N.$$

Since $Q_i = \frac{1}{N}\Phi/\phi_i$, $i = 1, \ldots, N$, is a feasible solution, it follows that the optimal value of $\eta$ lies on the interval

$$0 \le \eta \le \max_{i=1,\ldots,N} \left\{ \frac{a_i D_i \phi_i N^2}{\Phi^2} - \frac{h_i}{2\phi_i} \right\}. \tag{4.6}$$

Table 4.1: Data for Example 4.1 and results according to Algorithm 4.1.

| | product $i$: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| | $a_i$ | \$10 | \$10 | \$10 | \$40 | \$10 | \$10 | \$10 | \$40 |
| | $h_i$ | \$0.20 | \$0.80 | \$0.20 | \$0.80 | \$0.20 | \$0.80 | \$0.20 | \$0.80 |
| | $\phi_i$ | 1.0 | 1.0 | 1.0 | 1.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| | $D_i$ | 400 | 400 | 100 | 400 | 400 | 400 | 100 | 400 |
| $\Phi \ge 3000$ | $Q_i$ | 200.0 | 100.0 | 100.0 | 200.0 | 200.0 | 100.0 | 100.0 | 200.0 |
| | $C_i$ | \$40.00 | \$80.00 | \$20.00 | \$160.00 | \$40.00 | \$80.00 | \$20.00 | \$160.00 |
| | $T_i$ | 0.500 | 0.250 | 1.000 | 0.500 | 0.500 | 0.250 | 1.000 | 0.500 |
| $\Phi = 2400$ | $Q_i$ | 174.7 | 96.3 | 87.3 | 192.7 | 133.5 | 87.3 | 66.8 | 174.7 |
| | $C_i$ | \$40.37 | \$80.06 | \$20.18 | \$160.11 | \$43.31 | \$80.73 | \$21.66 | \$161.47 |
| | $T_i$ | 0.437 | 0.241 | 0.873 | 0.482 | 0.334 | 0.218 | 0.668 | 0.437 |
| $\Phi = 1800$ | $Q_i$ | 136.8 | 88.2 | 68.4 | 176.5 | 84.9 | 68.4 | 42.5 | 136.8 |
| | $C_i$ | \$42.92 | \$80.63 | \$21.46 | \$161.25 | \$55.61 | \$85.84 | \$27.80 | \$171.68 |
| | $T_i$ | 0.342 | 0.221 | 0.684 | 0.441 | 0.212 | 0.171 | 0.424 | 0.342 |
| $\Phi = 1200$ | $Q_i$ | 91.5 | 71.7 | 45.7 | 143.4 | 49.8 | 45.7 | 24.9 | 91.5 |
| | $C_i$ | \$52.88 | \$84.47 | \$26.44 | \$168.94 | \$85.29 | \$105.75 | \$42.64 | \$211.51 |
| | $T_i$ | 0.229 | 0.179 | 0.457 | 0.359 | 0.125 | 0.114 | 0.249 | 0.229 |

**Example 4.1** Table 4.1 contains data ($a_i,\ h_i,\ \phi_i,\ D_i,\ i = 1, \ldots, 8$) for a set of eight products that have to share storage space. The unconstrained optimal order quantities require a total storage space of 3000 volume units. Table 4.1 further displays the optimal order quantities computed according to Algorithm 4.1 at storage capacity constraints of 80%, 60% and 40% of the maximally required storage space of 3000 volume units. The table also shows the corresponding ordering and holding cost per item, $C_i \doteq a_i D_i/Q_i + \frac{1}{2} h_i Q_i$, and the reorder cycle $T_i = Q_i/D_i$ for each item, $i = 1, \ldots, 8$. There is no clear structure in the decrease of the order quantities: the order quantity of some items decreases more strongly when the storage capacity decreases from 80% to 60% than when it decreases from 100% to 80%, for other items it is the other way about. The total average cost increases from \$600.00 at $\Phi = 3000$ via \$607.89 at $\Phi = 2400$ and \$647.20 at $\Phi = 1800$ to \$777.91 at $\Phi = 1200$. □

The inventory optimization problem with a constraint on the total budget available for investment in inventories can be solved with a similar algorithm as Algorithm 4.1. Dagpunar [19] discusses such an optimization problem including a joint family ordering cost.

**Exercise 4.1** *Consider two items with the same ordering cost* $a$*, the same holding cost* $h$ *and the same volume* $\phi$*. Only the demand rates differ:* $D_1 = 4D_2$*. Determine the ratio between the optimal order quantities* $Q_1^*$ *and* $Q_2^*$ *both in the unconstrained and in the constrained case, with inventory capacity* $\Phi < 3\phi\sqrt{2aD_2/h}$*.*

**Exercise 4.2** *Consider four items with the same ordering cost* $a = \$25$ *and the same demand rate* $D = 200$*. The holding costs are* $h_1 = h_4 = \$1$, $h_2 = h_3 = \$4$ *per unit per unit of time, and the areas occupied per unit are* $\phi_1 = \phi_2 = 1$, $\phi_3 = \phi_4 = 4$ *square meters. Units cannot be piled. The available area for storage is* $\Phi = 700$ *square meters. Determine the optimal order quantities* $Q_i^*,\ i = 1, 2, 3, 4$*. Can you find an ad hoc solution such that the four items are ordered according to their individual economic order quantities? Finally, determine the optimal order quantities for the case that the available area for storage is* $\Phi = 500$ *square meters.*

**Exercise 4.3** *Prove that equation (4.5) indeed has a root in* $\eta$ *on the interval (4.6) if Step 3 of Algorithm 4.1 is executed.*

## 4.2 Repairable items

This section is devoted to the inventory management of spare parts that occur in machines like computers, copiers and printers. It is assumed that a unit that fails is replaced by a spare part and that the failed units are later repaired, if possible. Further, it is assumed that repaired units are as good as new ones. A certain fraction of the failed units cannot be repaired. Therefore, new spare parts have to be purchased from an external supplier from time to time. The parameters of the model are:

- $D$: the demand per unit of time of units for repair;
- $w$: the fraction of failed units that is no longer repairable (the waste rate);
- $a_O$: the ordering cost for placing an order at the external supplier;
- $a_R$: the set-up cost for starting a repair batch;
- $h_G$: the holding cost per unit per unit of time for good (new and repaired) units;
- $h_F$: the holding cost per unit per unit of time for failed, but repairable units;
- $L_O$: the lead time from the external supplier;
- $L_R$: the lead time of a repair batch.

The demand of units for replacement is assumed to be deterministic and continuous in time. Also, the lead times are deterministic. In most situations, it will hold that $h_G > h_F$. The following inventory policy will be studied, cf. Schrady [57]. The order quantity for new items is always the same, and also the repair lot size is always the same. The decision variables are:

- $Q_O$: the order quantity from the external supplier;

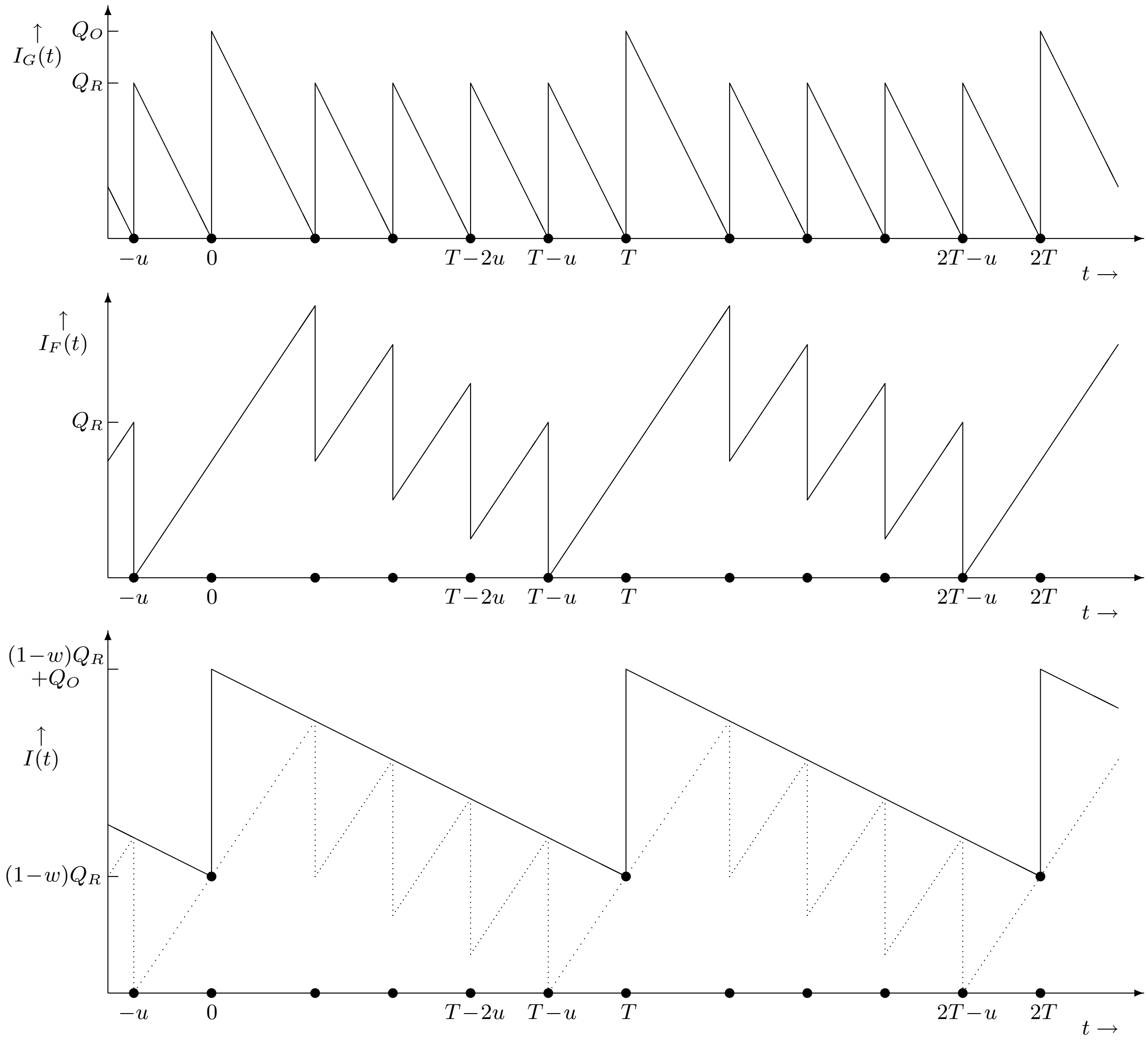

Figure 4.1: Inventory on hand of good units (top), of repairable units (middle) and of usable units (bottom).

- $Q_R$: the lot size for repairs;
- $s_O$: the reorder point for an external order;
- $s_R$: the point for starting a repair batch.

The policy is such that in between two external orders $n$ repair batches are performed, for some integer $n$, $n \geq 1$. This puts a constraint on the ratio of $Q_O$ and $Q_R$. First, the case will be considered that the lead time of a repair batch is negligible ($L_R = 0$).
Let $T$ denote the time between the arrival of two consecutive orders of new spare parts. In a cycle of length $T$, the number of incoming good units must be equal to the number of outgoing good units:

$$Q_O + nQ_R = DT. \tag{4.7}$$

Further, in a cycle of length $T$, the number of incoming units from the external supplier must be equal to the number of outgoing units as waste:

$$Q_O = wDT. \tag{4.8}$$

Together, these balance relations imply that

$$nQ_R = (1-w)DT = (1-w)Q_O/w. \tag{4.9}$$

For the time interval $u$ between two consecutive repair batches within a reorder cycle it must hold that

$$Q_R = Du. \tag{4.10}$$

As a consequence, it follows with (4.9) that the fraction of time that the system runs on repaired units is

$$\frac{nu}{T} = \frac{nQ_R}{DT} = 1 - w. \tag{4.11}$$

Figure 4.1 (upper graph) shows the inventory on hand $I_G(t)$ of good units for the case $n = 4$. This inventory level follows a somewhat irregular saw-tooth pattern, since an external order is succeeded by $n$ repair batches. In between replenishments, this inventory decreases at rate $D$. The average inventory on hand is $\frac{1}{2}Q_O$ during a fraction $(T - nu)/T$ of a cycle and is $\frac{1}{2}Q_R$ during a fraction $nu/T$ of a cycle. Hence, the average inventory on hand of good units follows with (4.11) as:

$$\bar{I}_G = \frac{\frac{1}{2}Q_O(T - nu) + \frac{1}{2}Q_R nu}{T} = \tfrac{1}{2}wQ_O + \tfrac{1}{2}(1 - w)Q_R. \tag{4.12}$$

The middle graph of Figure 4.1 shows the inventory on hand $I_F(t)$ of failed but repairable units. In between repair batches, this inventory increases at rate $(1 - w)D$. This inventory level decreases by $Q_R$ at instants when the inventory on hand $I_G(t)$ of good units reaches the level 0. The lower graph of Figure 4.1 shows the total inventory on hand $I(t) \doteq I_G(t)+I_F(t)$ of good and repairable units. In between external replenishments, this inventory decreases at rate $wD$. This inventory level follows a saw-tooth pattern which is lifted above the level 0. The latter is due to the fact that a number of repairable units has accumulated when the inventory on hand $I_G(t)$ of good units reaches the level 0 just before the arrival of an external order. This amount is $(1-w)Du = (1-w)Q_R$, cf. (4.10). The average total inventory of good and repairable units on hand is $\frac{1}{2}Q_O$ plus this basic amount:

$$\bar{I}_G + \bar{I}_F = (1 - w)Q_R + \tfrac{1}{2}Q_O. \tag{4.13}$$

Hence, the average inventory of failed but repairable units simply follows as

$$\bar{I}_F = (1 - w)Q_R + \tfrac{1}{2}Q_O - \tfrac{1}{2}wQ_O - \tfrac{1}{2}(1 - w)Q_R = \tfrac{1}{2}(1 - w)Q_R + \tfrac{1}{2}(1 - w)Q_O. \tag{4.14}$$

The foregoing implies that the average cost per unit of time with this policy is:

$$C(Q_O, Q_R) = D\left[\frac{a_O w}{Q_O} + \frac{a_R(1 - w)}{Q_R}\right] + \tfrac{1}{2}h_G[wQ_O + (1 - w)Q_R] + \tfrac{1}{2}h_F(1 - w)[Q_O + Q_R], \tag{4.15}$$

or, with $Q_R = (1 - w)Q_O/(nw)$, cf. (4.9),

$$C(Q_O, n) = \frac{wD}{Q_O}[a_O + na_R] + \tfrac{1}{2}Q_O\left[h_G\left\{w + \frac{(1 - w)^2}{nw}\right\} + h_F(1 - w)\left\{1 + \frac{1 - w}{nw}\right\}\right]. \tag{4.16}$$

For fixed $n$, the optimal external order quantity $Q_O$ is:

$$\hat{Q}_O(n) = w\sqrt{\frac{2nD(a_O + na_R)}{nw[h_G w + h_F(1 - w)] + (1 - w)^2[h_G + h_F]}}, \tag{4.17}$$

with corresponding minimum average cost per unit of time

$$\hat{C}(n) = \sqrt{2wD(a_O + na_R)\left[h_G w + h_F(1 - w) + \frac{(1 - w)^2}{nw}(h_G + h_F)\right]}. \tag{4.18}$$

Since the square root function is monotonously increasing, the optimal real-valued minimum of this cost function is obtained by solving the equation

$$\begin{aligned}\tfrac{d}{dn}\hat{C}^2(n) &= 2D\tfrac{d}{dn}\{[\tfrac{1}{n}a_O + a_R](1 - w)^2(h_G + h_F) + [a_O + na_R]w[h_G w + h_F(1 - w)]\}\\ &= 2D\{-\tfrac{1}{n^2}a_O(1 - w)^2(h_G + h_F) + a_R w[h_G w + h_F(1 - w)]\} = 0.\end{aligned}$$

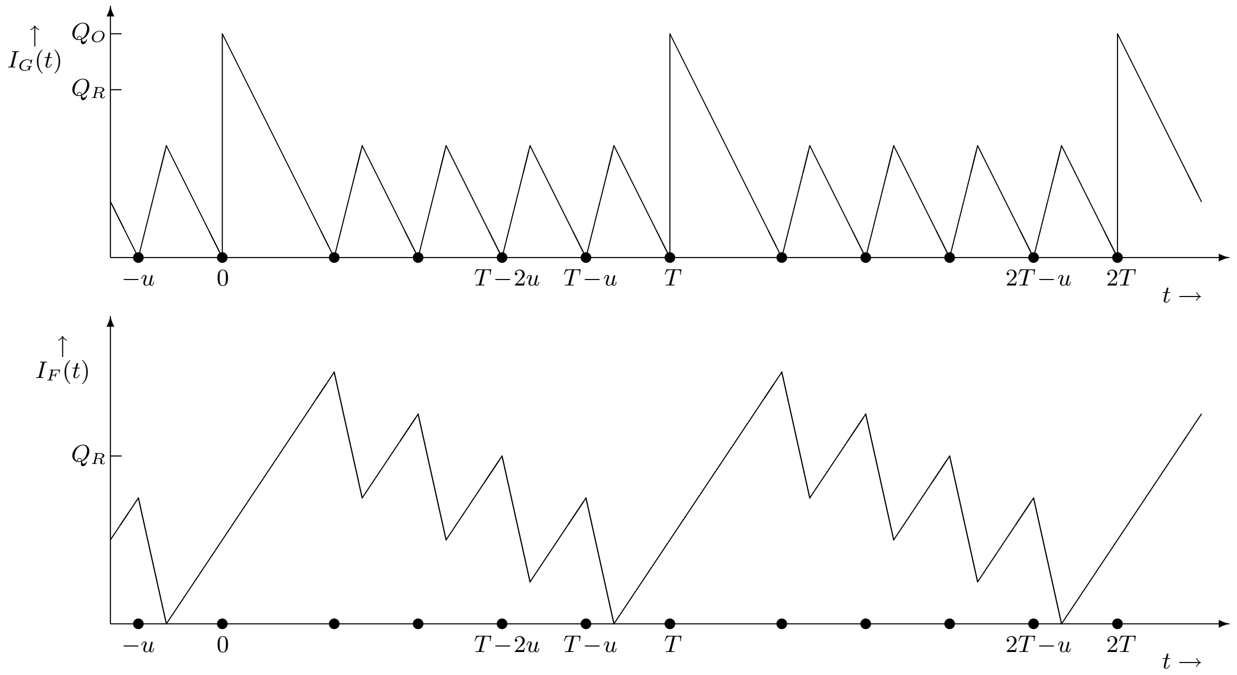

Figure 4.2: Inventory on hand of good units (top) and of repairable units (bottom) with release during repair.

Hence, the optimal real-valued $n$ minimizing the foregoing cost function is:

$$n^* = (1-w)\sqrt{\frac{a_O(h_G + h_F)}{a_R w[h_G w + h_F(1-w)]}}. \tag{4.19}$$

Note that this ratio does not depend on the demand rate $D$. The optimal policy is obtained by rounding $n^*$ to a positive integer in such a way that the cost (4.18) is minimal. Then, the optimal order quantities follow with (4.17) and (4.9).

**Remark 4.1** Without rounding it follows from (4.17) and (4.9) that

$$\hat{Q}_O(n^*) = \sqrt{\frac{2a_O w D}{h_G w + h_F(1-w)}}, \qquad Q_R = \frac{1-w}{n^* w}\hat{Q}_O(n^*) = \sqrt{\frac{2a_R D}{h_G + h_F}}.$$

The above approximately optimal values indicate that the optimal order quantity hardly depends on $a_R$ and that the optimal repair lot size hardly depends on $a_O$ and $w$. In the extreme case that the holding cost for failed units is negligible, $h_F = 0$, the above quantities reduce to

$$n^* = \frac{1-w}{w}\sqrt{\frac{a_O}{a_R}}, \qquad \hat{Q}_O(n^*) = \sqrt{\frac{2a_O D}{h_G}}, \qquad Q_R = \frac{1-w}{n^* w}\hat{Q}_O(n^*) = \sqrt{\frac{2a_R D}{h_G}},$$

that is, the quantities are approximately given by EOQ formulas independent of the waste rate $w$. □

**Example 4.2** The demand for a spare part of a certain type of copier is $D = 5$ per day. The fraction of units that is irreparable is $w = \frac{1}{5}$. The ordering cost is $a_O = \$20$ and the repair set-up cost is $a_R = \$20$. The holding cost per unit per day is $h_G = \$0.10$ for good units and $h_F = \$0.01$ for repairable units. From (4.19) it follows that $n^* = \frac{4}{5}\sqrt{\frac{0.55}{0.028}} = \sqrt{\frac{88}{7}} \approx 3.55$. From (4.18) it is obtained that $\hat{C}(3) = \sqrt{4 \cdot 40(0.028 + \frac{1}{3} \cdot 0.352)} \approx 4.822$ and $\hat{C}(4) = \sqrt{5 \cdot 40(0.028 + \frac{1}{4} \cdot 0.352)} \approx 4.817$ so that $n = 4$ is slightly better. The corresponding order quantity is $\hat{Q}_O(4) = 41.52$, cf. (4.17), and the corresponding repair lot size is $Q_R = \frac{1-w}{4w}\hat{Q}_O(4) = 41.52$, cf. (4.9). The corresponding reorder cycle is $T = Q_O/(wD) = 41.52$ days, cf. (4.8), and the corresponding repair cycle is $u = Q_R/D = 8.30$ days, cf. (4.10). □

Next, consider the consequences of lead times. A lead time $L_O$ for external orders has the standard consequence that an order must be placed $L_O$ units of time before the order must arrive. This reorder instant can only be translated into a unique reorder point for the inventory position of the total number of good and failed but repairable units, cf. Figure 4.1. In this sense, the reorder point becomes

$$s_O = (1-w)Q_R + wDL_O. \tag{4.20}$$

The consequence of a lead time for a repair batch is more important, even if we disregard the possibility that the item must compete with other items for repair capacity. If the repairs are performed close to the warehouse at a rate of $r$ units per unit of time and repaired units become available during the repair run, the inventory patterns change to a pattern as for the production lot-size model in Section 1.2.2. This is shown in Figure 4.2 where the upper graph contains the triangular pattern (apart from the external order) of the inventory on hand $I_G(t)$ of good units and the lower graph contains the triangular pattern of the inventory on hand $I_F(t)$ of failed but repairable units. However, if the repairs are performed away from the warehouse and leave the warehouse at the beginning of the lead time $L_R$, the whole graph of the inventory on hand $I_F(t)$ for failed but repairable parts in Figure 4.1 has to be shifted upwards by $(1-w)DL_R$, the number of repairable items that return to the warehouse during the lead time $L_R$. The average cost per unit of time, cf. (4.15), (4.16), increases by the amount $h_F(1-w)DL_R$, assuming that the holding cost during the repair lead time $L_R$ remains $h_F$. In this case, a repair batch has to start when the inventory position of good units reaches the reorder point $s_R = DL_R$, except if $I_G(t) + I_F(t) < Q_R$ since then an external order has to be placed. Moreover, the reorder point in (4.20) also has to be increased by $(1-w)DL_R$ to

$$s_O = (1-w)Q_R + wDL_O + (1-w)DL_R. \tag{4.21}$$

This adaptation is illustrated in Figure 4.3, where the dotted lines in the middle graph indicate the batch of failed items that is in repair.

**Example 4.3** If the lead time for orders is $L_O = 5$ days and the lead time for repair batches is $L_R = 10$ days in the situation of Example 4.2, the reorder point becomes $s_O = 78.22$, cf. (4.21), and the starting point for a repair batch becomes $s_R = DL_R = 50$. The reorder point lies between the minimum and maximum inventory levels of the total number of usable units on hand ($(1-w)Q_R = 73.22$ and $(1-w)Q_R + Q_O = 114.74$, respectively; see also Figure 4.3). The starting point for a repair batch lies above the maximum level of the inventory on hand of good units at the end of a repair cycle ($Q_R = 41.52$) so that this point has to be applied to the inventory position of the good units. It means that at some time intervals, two repair batches are in progress. This also follows from the fact that $L_R = 10 > u = 8.30$. The average total cost increases by $h_F(1-w)DL_R = 0.4$ to $\hat{C}(4) \approx 5.217$. ☐

**Exercise 4.4** *Express the maximum inventory level of repairable units, cf. the middle graph of Figure 4.1, in terms of $Q_O$, $Q_R$, and the parameters of the model.*

**Exercise 4.5** *The demand for a spare part of a type of laser printer is $D = 1$ per day. The fraction of units that is irreparable is $w = \frac{1}{5}$. The ordering cost is $a_O = \$20$ and the repair set-up cost is $a_R = \$45$. The holding cost per unit per day is $h_G = \$0.40$ for good parts and $h_F = \$0.20$ for repairable parts. Assume first that the lead times are negligible. Determine the optimal external order quantity, the optimal repair lot size and the corresponding reorder cycle. Compute the maximum inventory levels of good parts, of repairable parts and of all good and repairable parts together. Then, modify the results for a lead time $L_O = 5$ days from the external supplier and a lead time $L_R = 3$ days for a repair batch which is transported as a whole to and from a workshop away from the warehouse.*

**Exercise 4.6** *Adapt the analysis of this section to the situation that the repairs are performed in the warehouse at a rate of $r$ units per unit of time and repaired units become available during the repair run, cf. the inventory patterns shown in Figure 4.2.*

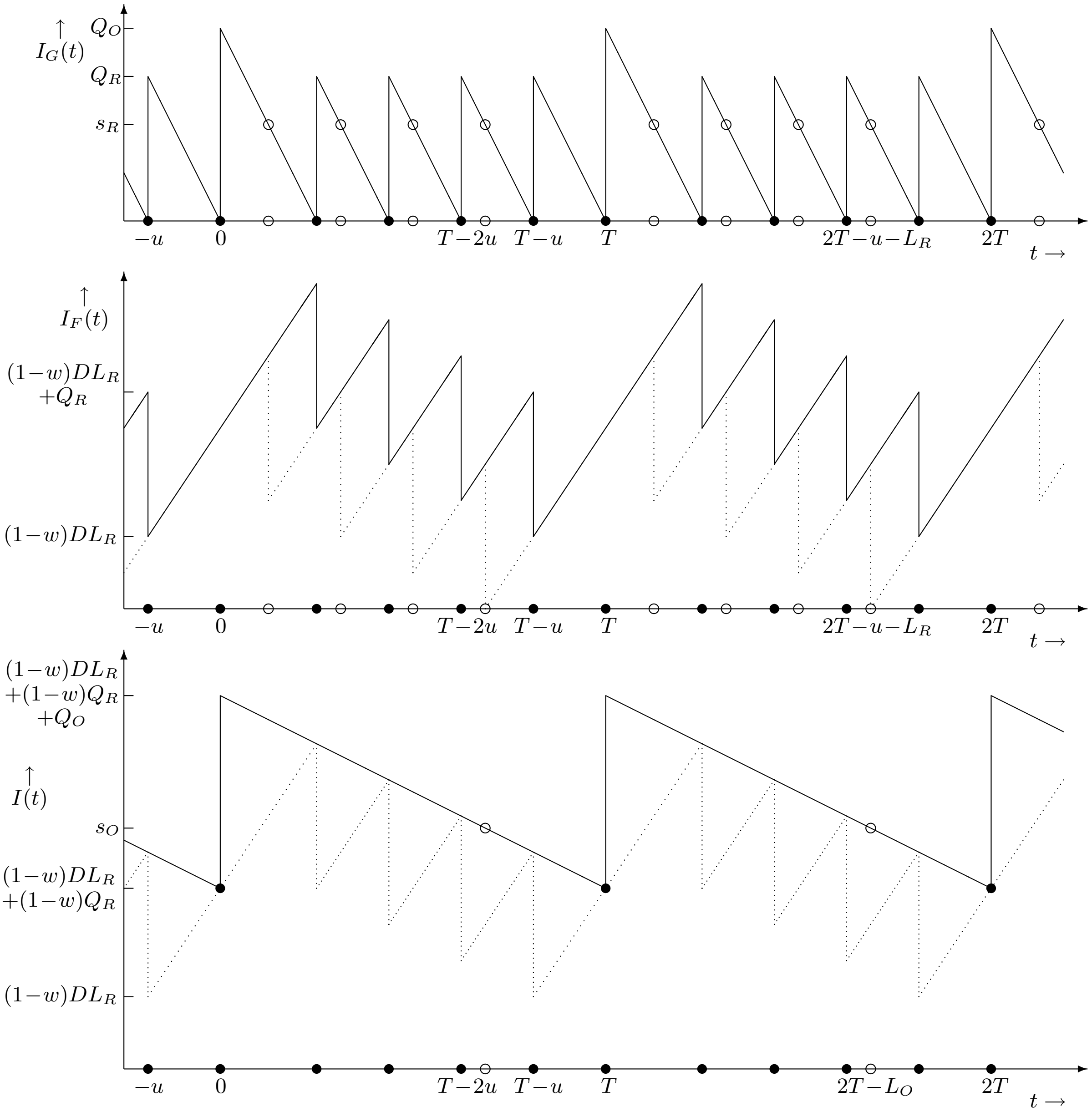

Figure 4.3: Inventory on hand of good units (top), of repairable units (middle) and of usable units (bottom).

# Appendix A

# Review of Probability Theory

## A.1 Random variables

A random variable is a function defined on a sample space, that is, an assignment of a real number to each element of the sample space $\Omega$. The (cumulative) *distribution function* $F_Y(y)$ of a random variable $Y$ is a nondecreasing function which tends to 0 as $y \downarrow -\infty$ and to 1 as $y \uparrow \infty$. It defines the probability of the event that $Y \leq y$, that is, $F_Y(y) = \Pr\{Y \leq y\}$, $y \in \mathbb{R}$. The range of a random variable (a subset of $\mathbb{R}$) is also called the support of its distribution function. If a distribution function $F_Y(y)$ is continuous for all $y \in \mathbb{R}$ and differentiable at almost all $y \in \mathbb{R}$ then it is called *absolutely continuous*, $Y$ is then called a *continuous* random variable and $f_Y(y) \doteq F'_Y(y)$ is called the *density* of $Y$. Absolutely continuous distributions are determined by their density. The integral of the density over the real axis is equal to one. If a distribution function $F_X(x)$ is a jump function then the corresponding random variable $X$ is called a *discrete* random variable. Let $x_1, x_2, \ldots$ be the jump points of $F_X(x)$, that is, the values which $X$ may assume. The function

$$\Pr\{X = x_j\} \doteq F_X(x_j+) - F_X(x_j-), \quad j = 1, 2, \ldots,$$

is called the probability distribution of the discrete random variable $X$.

### A.1.1 Moments and expectation

The *expected value* of the random variable $Y$ with density $f_Y(y)$ is defined by

$$E\{Y\} = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy. \tag{A.1}$$

The expected value $E\{Y\}$ is also called the *mean* or the *average* of the random variable $Y$. If a random variable $Y$ has range $[0, \infty)$ then its expected value can alternatively be formulated as:

$$E\{Y\} = \int_0^{\infty} \int_0^{y} du\, f_Y(y)\, dy = \int_0^{\infty} \int_u^{\infty} f_Y(y)\, dy\, dt = \int_0^{\infty} \Pr\{Y > u\}\, du. \tag{A.2}$$

Analogously, the mean of a discrete random variable $X$ with range $\mathbb{N}$ can be written as

$$E\{X\} \doteq \sum_{x=0}^{\infty} x \Pr\{X = x\} = \sum_{x=1}^{\infty} \sum_{n=1}^{x} \Pr\{X = x\} = \sum_{n=1}^{\infty} \sum_{x=n}^{\infty} \Pr\{X = x\} = \sum_{n=1}^{\infty} \Pr\{X \geq n\}. \tag{A.3}$$

The $k$th moment of the distribution of a random variable $Y$ is defined as

$$E\{Y^k\} = \int_{-\infty}^{\infty} y^k\, f_Y(y)\, dy, \quad k = 1, 2, \ldots. \tag{A.4}$$

For the special case that $Y = X$ is a random variable with range $\mathbb{N}$ this implies

$$E\{X^k\} = \sum_{x=0}^{\infty} x^k \Pr\{X = x\}, \quad k = 1, 2, \ldots. \tag{A.5}$$

The *variance* $\sigma^2\{Y\}$ of a random variable $Y$ is defined as

$$\sigma^2\{Y\} \doteq E\{(Y - E\{Y\})^2\} = E\{Y^2\} - E^2\{Y\}; \tag{A.6}$$

here, $E^2\{Y\}$ denotes the square of the mean $E\{Y\}$. The square root $\sigma\{Y\}$ of the variance is called the *standard deviation* of the random variable $Y$. The *coefficient of variation* of the random variable $Y$ is defined as

$$C_Y \doteq \frac{\sigma\{Y\}}{E\{Y\}} = \frac{1}{E\{Y\}}\sqrt{E\{Y^2\} - E^2\{Y\}}. \tag{A.7}$$

The mean and the variance of a random variable possess the following properties: for any real $a$ and $b$,

$$E\{aY + b\} = a\,E\{Y\} + b, \qquad \sigma^2\{aY + b\} = a^2\,\sigma^2\{Y\}. \tag{A.8}$$

### A.1.2 Collections of random variables

Let $Y_1$ and $Y_2$ be two random variables defined on a common sample space. Their joint distribution function $\Pr\{Y_1 \le y_1, Y_2 \le y_2\}$, $y_1, y_2 \in \mathbb{R}$, represents the probability that both $Y_1 \le y_1$ and $Y_2 \le y_2$. The (marginal) distribution function of $Y_1$ ($Y_2$) is obtained by taking the limit $y_2 \to \infty$ ($y_1 \to \infty$) of the joint distribution function for fixed $y_1$ ($y_2$). The density of a pair of continuous random variables is denoted by $f_{Y_1,Y_2}(y_1, y_2)$. The *cross moment* of two random variables $Y_1$ and $Y_2$ is defined as the expectation of the product $Y_1Y_2$:

$$E\{Y_1Y_2\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1y_2\, f_{Y_1,Y_2}(y_1, y_2)\, dy_2\, dy_1. \tag{A.9}$$

The *covariance* $\mathrm{cov}\{Y_1, Y_2\}$ of two random variables $Y_1$ and $Y_2$ is defined as

$$\mathrm{cov}\{Y_1, Y_2\} \doteq E\{(Y_1 - E\{Y_1\})(Y_2 - E\{Y_2\})\} = E\{Y_1Y_2\} - E\{Y_1\}E\{Y_2\}. \tag{A.10}$$

The *correlation coefficient* $\rho\{Y_1, Y_2\}$ of two random variables $Y_1$ and $Y_2$ is defined as

$$\rho\{Y_1, Y_2\} \doteq \frac{\mathrm{cov}\{Y_1, Y_2\}}{\sigma\{Y_1\}\sigma\{Y_2\}}. \tag{A.11}$$

For any pair of random variables $Y_1$ and $Y_2$, $|\rho\{Y_1, Y_2\}| \le 1$. Moreover, $\rho\{Y_1, Y_2\} = \pm 1$ only if there exist constants $a$ and $b$ such that $Y_2 = aY_1 + b$.

The concept of joint distribution function carries over to collections of more than two random variables. Let $Y_i$, $i = 1, \ldots, n$, be a collection of random variables defined on the same sample space. Then it generally holds that

$$E\left\{\sum_{i=1}^{n} Y_i\right\} = \sum_{i=1}^{n} E\{Y_i\}, \qquad \sigma^2\left\{\sum_{i=1}^{n} Y_i\right\} = \sum_{i=1}^{n}\sigma^2\{Y_i\} + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathrm{cov}\{Y_i, Y_j\}. \tag{A.12}$$

Two random variables $Y_1$ and $Y_2$ are *independent* if for all $y_1, y_2 \in \mathbb{R}$,

$$\Pr\{Y_1 \le y_1, Y_2 \le y_2\} = \Pr\{Y_1 \le y_1\}\Pr\{Y_2 \le y_2\}. \tag{A.13}$$

If $Y_1$ and $Y_2$ are independent then

$$E\{Y_1Y_2\} = E\{Y_1\}\,E\{Y_2\}, \qquad \mathrm{cov}\{Y_1, Y_2\} = 0. \tag{A.14}$$

The reverse of this statement is not true, in general. The mean and the variance of two independent random variables $Y_1$ and $Y_2$ possess the following properties: for any real $a$, $b$ and $c$,

$$E\{aY_1 + bY_2 + c\} = a\,E\{Y_1\} + b\,E\{Y_2\} + c, \qquad \sigma^2\{aY_1 + bY_2 + c\} = a^2\,\sigma^2\{Y_1\} + b^2\,\sigma^2\{Y_2\}. \tag{A.15}$$

The concept of independence carries over to finite collections of random variables.

## A.2 Probability distributions

This section contains an overview of some useful probability distributions and their properties.

### A.2.1 Normal distributions

The normal distributions form a two-parameter family of continuous distributions with range $(-\infty, \infty)$. The density of a normal $\mathcal{N}(\mu, \sigma)$ distribution with mean $\mu$ and variance $\sigma^2$ is, for $\sigma > 0$,

$$f_{\mu,\sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}[(y-\mu)/\sigma]^2}, \quad -\infty < y < \infty. \tag{A.16}$$

The normal distribution with mean 0 and standard deviation 1, $\mathcal{N}(0,1)$, is called the standard normal distribution. If the random variable $Z$ has the standard normal distribution and $Y$ the $\mathcal{N}(\mu,\sigma)$ normal distribution then

$$Y \sim \mu + \sigma Z, \qquad Z \sim (Y-\mu)/\sigma. \tag{A.17}$$

An important property of normal distributions is that the sum of independent normally distributed random variables is normally distributed. For example, if $D_1 \sim \mathcal{N}(\mu_1, \sigma_1)$, $D_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ and $f$ is a constant, then $D_1 + fD_2 \sim \mathcal{N}(\mu_1 + f\mu_2, \sqrt{\sigma_1^2 + f^2\sigma_2^2})$.
The normal loss function is defined by

$$NL(y) \doteq E\{[Z-y]^+\} = \int_y^\infty (u-y) f_{0,1}(u)\, du = \frac{1}{\sqrt{2\pi}} \int_y^\infty (u-y)\, e^{-\frac{1}{2}u^2}\, du, \quad -\infty < y < \infty. \tag{A.18}$$

This function has the properties

$$NL(y) = f_{0,1}(y) - \frac{y}{\sqrt{2\pi}} \int_y^\infty e^{-\frac{1}{2}u^2}\, du, \quad -\infty < y < \infty, \tag{A.19}$$

and by the symmetry of the density of the standard normal distribution,

$$NL(y) = NL(-y) - y, \quad -\infty < y < \infty. \tag{A.20}$$

For a normal $\mathcal{N}(\mu,\sigma)$ distributed random variable $Y$ it holds that

$$E\{[Y-y]^+\} = \int_y^\infty (u-y) f_{\mu,\sigma}(u)\, du = \sigma NL([y-\mu]/\sigma), \quad -\infty < y < \infty. \tag{A.21}$$

The normal loss function has been tabulated; see e.g. Winston [80, Sect. 17.7]. Alternatively, it is related to the density and the distribution function (by means of integration by parts): for $-\infty < y < \infty$,

$$E\{[Y-y]^+\} = \sigma^2 f_{\mu,\sigma}(y) + (\mu - y) \int_y^\infty f_{\mu,\sigma}(u)\, du = \sigma f_{0,1}([y-\mu]/\sigma) + (\mu - y) \int_{(y-\mu)/\sigma}^\infty f_{0,1}(v)\, dv. \tag{A.22}$$

### A.2.2 Exponential distributions

The distribution function of a nonnegative exponentially distributed random variable $Y$ is by definition

$$\Pr\{Y \le y\} = 1 - e^{-\lambda y}, \quad y \ge 0, \tag{A.23}$$

for some positive rate (parameter) $\lambda$. The density is $\lambda e^{-\lambda y}$. The mean, the standard deviation and the coefficient of variation of an exponentially distributed random variable $Y$ are:

$$E\{Y\} = \frac{1}{\lambda}, \qquad \sigma\{Y\} = \frac{1}{\lambda}, \qquad C_Y = \sqrt{\frac{1}{\lambda}}. \tag{A.24}$$

An important property of the exponential distribution is its so called lack of memory. This property is expressed by the following formula:

$$\Pr\{Y > u + y \mid Y > u\} = \frac{e^{-\lambda(u+y)}}{e^{-\lambda u}} = e^{-\lambda y} = \Pr\{Y > y\}, \quad y, u \ge 0. \tag{A.25}$$

It states that the conditional probability that $Y$ exceeds the value $u + y$ given that $Y$ exceeds the value $u$ is equal to the unconditional probability that $Y$ exceeds the value $y$, for every positive $u$ and $y$. In other words, if the random variable $Y$ represents the duration of some process, and if this process is observed to be still in progress after some time $u$ then the time until completion has the same distribution as it had at the beginning of the process. As a consequence, the process can be considered as if it started anew at the time of observation $u$. The future of the process after time $u$ does not depend on its past until time $u$ given the observation that it is still not completed at time $u$.

Another important property of the exponential distribution is the fact that the minimum of a number of independent, exponentially distributed random variables $Y_j$ with parameters $\lambda_j,\ j = 1, \ldots, n$, is exponentially distributed with parameter the sum of the parameters of the variables $Y_j$, that is, for any number $n$ of variables,

$$\Pr\{\min\{Y_1, \ldots, Y_n\} \le y\} = 1 - e^{-(\lambda_1 + \cdots + \lambda_n)y}, \quad y \ge 0. \tag{A.26}$$

This property is readily verified by noting that, for $y \ge 0$,

$$\Pr\{\min\{Y_1, \ldots, Y_n\} > y\} = \Pr\{Y_1 > y, \ldots, Y_n > y\} = \prod_{j=1}^{n} \Pr\{Y_j > y\} = \prod_{j=1}^{n} e^{-\lambda_j y}. \tag{A.27}$$

### A.2.3 Poisson distributions and Poisson processes

A random variable with a Poisson distribution has the set $\mathbb{N}$ of nonnegative integers as its range. Such a random variable is defined by

$$\Pr\{X_a = n\} = \frac{a^n}{n!}\, e^{-a}, \quad n = 0, 1, 2, \ldots, \tag{A.28}$$

with $a$ a positive parameter. Its first two moments, variance and coefficient of variation are

$$E\{X_a\} = a, \quad E\{X_a^2\} = a^2 + a, \quad \sigma^2\{X_a\} = a, \quad C_{X_a} = \sqrt{1/a}. \tag{A.29}$$

A Poisson process $N(t),\ t \ge 0$, with rate $\lambda$, is a counting process such that the number of events in a time interval of length $t$ has the Poisson distribution with parameter $a = \lambda t$. The distribution of the time between successive events is exponential with mean $1/\lambda$.

### A.2.4 Gamma and Erlang distributions

The gamma distributions form a two-parameter family of continuous distributions with range $(0, \infty)$. The density of a gamma $\mathcal{G}(\psi, \lambda)$ distribution with shape parameter $\psi$ and scale parameter $\lambda$ is, for $\psi > 0$, $\lambda > 0$,

$$f_{Y_{\psi,\lambda}}(y) = \frac{\lambda^\psi}{\Gamma(\psi)}\, y^{\psi-1} e^{-\lambda y}, \quad y > 0. \tag{A.30}$$

The gamma function $\Gamma(\psi)$ is defined by

$$\Gamma(x) \doteq \int_0^\infty e^{-u}\, u^{x-1} du, \quad x > 0. \tag{A.31}$$

For integer-valued $n$ it holds that

$$\Gamma(n+1) = n!, \quad n \in \mathbb{N}. \tag{A.32}$$

The above density has a maximum at $y = (\psi - 1)/\lambda$ if $\psi > 1$ and at 0 otherwise. The moments of a gamma distribution are

$$E\{Y_{\psi,\lambda}^k\} = \frac{1}{\lambda^k} \prod_{j=0}^{k-1} (\psi + j) = \binom{\psi + k - 1}{k} \frac{k!}{\lambda^k}, \quad k = 1, 2, \ldots. \tag{A.33}$$

In particular, the first two moments, variance and coefficient of variation are

$$E\{Y_{\psi,\lambda}\} = \frac{\psi}{\lambda}, \quad E\{Y_{\psi,\lambda}^2\} = \frac{\psi(1+\psi)}{\lambda^2} = \frac{1+\psi}{\psi}\, E^2\{Y_{\psi,\lambda}\}, \quad \sigma^2\{Y_{\psi,\lambda}\} = \frac{\psi}{\lambda^2}, \quad C_{Y_{\psi,\lambda}} = \sqrt{\frac{1}{\psi}}. \tag{A.34}$$

The densities of gamma distributions are displayed in Figure A.1 for several values of the shape parameter $\psi$. The means of all these distributions are equal to 1, that is, $\lambda = \psi$ in all cases.

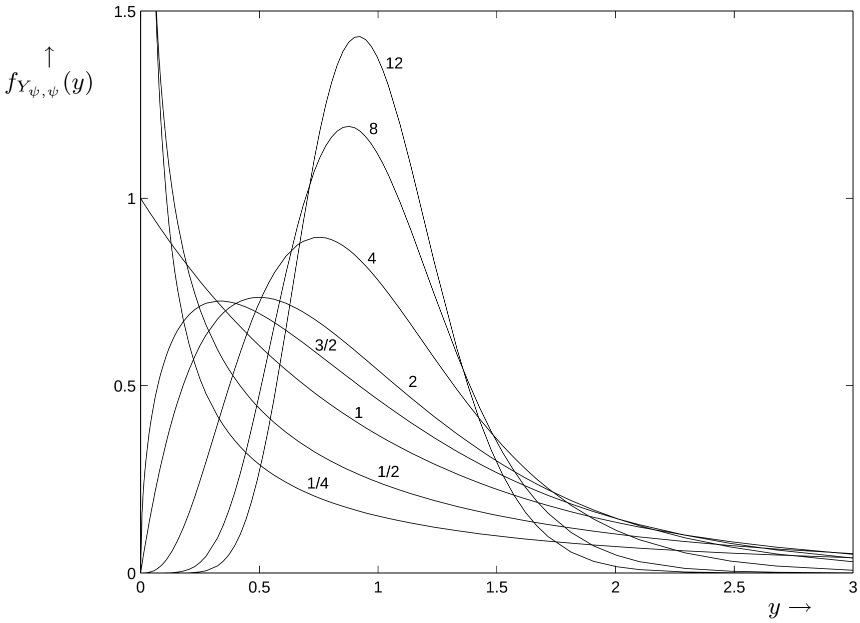

Figure A.1: Densities of gamma distributions with varying shape parameter $\psi$ and fixed mean of 1.

Special cases of this class of distributions are the Erlang $\mathcal{E}(K,\lambda)$ distributions for which $K$ is an integer. With the aid of repeated integration by parts, the cumulative distribution function of the Erlang distribution with shape parameter $K$ can be explicitly written as

$$\Pr\{Y_{K,\lambda} \leq y\} = 1 - \sum_{j=0}^{K-1} \frac{(\lambda y)^j}{j!} e^{-\lambda y}, \quad y \geq 0. \tag{A.35}$$

An important property of the Erlang $\mathcal{E}(K,\lambda)$ distribution is that it emerges as the $K$-fold convolution of the exponential distribution with parameter $\lambda$. This means that a random variable with an $\mathcal{E}(K,\lambda)$ distribution can be considered as the sum of $K$ independent, identically, exponentially distributed phases.
The gamma loss function is defined by

$$GL_{\psi,\lambda}(y) \doteq \int_y^\infty (u-y)\, f_{Y_{\psi,\lambda}}(u)\, du = \frac{\lambda^\psi}{\Gamma(\psi)} \int_y^\infty (u-y)\, u^{\psi-1} e^{-\lambda u}\, du, \quad y \geq 0. \tag{A.36}$$

For Erlang distributions (with an integer shape parameter $\Psi = K$) this function can be simplified with the aid of repeated integration by parts to

$$GL_{K,\lambda}(y) = \frac{\lambda^K}{(K-1)!} \int_y^\infty (u-y)\, u^{K-1} e^{-\lambda u}\, du = e^{-\lambda y} \sum_{j=0}^{K-1} (K-j)\, \frac{\lambda^{j-1} y^j}{j!}, \quad y \geq 0. \tag{A.37}$$

For general $\psi$, the gamma loss function can be expressed after a single integration by parts and with the property of gamma functions that $\Gamma(\psi+1) = \psi\Gamma(\psi)$, as, for $y \geq 0$,

$$GL_{\psi,\lambda}(y) = \frac{\psi}{\lambda} \Pr\{Y_{\psi+1,\lambda} > y\} - y \Pr\{Y_{\psi,\lambda} > y\} = \frac{\psi}{\lambda} \int_y^\infty f_{Y_{\psi+1,\lambda}}(u)\, du - y \int_y^\infty f_{Y_{\psi,\lambda}}(u)\, du. \tag{A.38}$$

The latter two integrals are known as incomplete gamma functions.

### A.2.5 Mixtures of Erlang distributions

A mixture of two Erlang distributions, to be denoted by $\mathcal{ME}(K_1, \lambda_1, K_2, \lambda_2, p)$, has five parameters and range $(0,\infty)$. The density of a mixture of two Erlang distributions is

$$f(y) = p\, \frac{\lambda_1^{K_1} y^{K_1-1}}{(K_1-1)!}\, e^{-\lambda_1 y} + (1-p) \frac{\lambda_2^{K_2} y^{K_2-1}}{(K_2-1)!}\, e^{-\lambda_2 y}, \quad y > 0; \tag{A.39}$$

here, $K_1$ and $K_2$ are positive integers, $\lambda_1$ and $\lambda_2$ are positive scale parameters, and $p$, $0 < p < 1$, is a weight. The mean of a random variable $Y$ with a mixture of two Erlang distributions is

$$E\{Y\} = p\,\frac{K_1}{\lambda_1} + (1-p)\,\frac{K_2}{\lambda_2}, \tag{A.40}$$

and its variance is

$$\sigma^2\{Y\} = p\,\frac{K_1(K_1+1)}{\lambda_1^2} + (1-p)\,\frac{K_2(K_2+1)}{\lambda_2^2} - E^2\{Y\}. \tag{A.41}$$

Mixtures of two Erlang distributions are often used to approximate the distributions of random variables with range $(0,\infty)$ of which only the mean $E\{Y\}$ and the variance $\sigma^2\{Y\}$ are known. For the choice of the parameters, a distinction is made whether the coefficient of variation $C = \sigma\{Y\}/E\{Y\}$ is smaller than 1 or not. If the coefficient of variation $C < 1$, then the parameters are chosen as follows: $K_1$ is the largest integer smaller than $1/C^2$, $K_2 = K_1 + 1$, and $\lambda_1 = \lambda_2$ and $p$ are determined such that the mean and the variance agree:

$$p = \frac{1}{1+C^2}\left[K_2C^2 - \sqrt{K_2(1+C^2) - K_2^2C^2}\right], \qquad \lambda_1 = \lambda_2 = (K_2 - p)/E\{Y\}. \tag{A.42}$$

If the coefficient of variation $C > 1$, then the parameters are chosen as follows: $K_1 = K_2 = 1$, $\lambda_1 \neq \lambda_2$ and $p$ are determined in such a way that the mean and the variance agree and the third moment is equal to that of a gamma distribution:

$$\lambda_1 = \frac{2}{E\{Y\}}\left[1 + \sqrt{\frac{C^2 - \frac{1}{2}}{1+C^2}}\right], \qquad \lambda_2 = \frac{4}{E\{Y\}} - \lambda_1, \qquad p = \frac{\lambda_1(1 - \lambda_2 E\{Y\})}{\lambda_1 - \lambda_2}. \tag{A.43}$$

Mixtures with $K_1 = K_2 = 1$ are mixtures of two exponential distributions. Such mixtures are called hyperexponential distributions and are indicated by the symbol $\mathrm{H}_2$.

The loss function corresponding to a mixture of two Erlang distributions simply follows from (A.37) as

$$E\{[Y-y]^+\} = p\,e^{-\lambda_1 y}\sum_{j=0}^{K_1-1}(K_1 - j)\,\frac{\lambda_1^{j-1}y^j}{j!} + (1-p)\,e^{-\lambda_2 y}\sum_{j=0}^{K_2-1}(K_2 - j)\,\frac{\lambda_2^{j-1}y^j}{j!}, \quad y \ge 0. \tag{A.44}$$

For the special case $K_1 = K$, $K_2 = K+1$ and $\lambda_1 = \lambda_2 = \lambda$ this loss function becomes

$$E\{[Y-y]^+\} = e^{-\lambda y}\left[\sum_{j=0}^{K-1}(K-j)\,\frac{\lambda^{j-1}y^j}{j!} + (1-p)\sum_{j=0}^{K}\frac{\lambda^{j-1}y^j}{j!}\right], \quad y \ge 0. \tag{A.45}$$

The loss function corresponding to a hyperexponential distribution ($K_1 = K_2 = 1$) is still simpler:

$$E\{[Y-y]^+\} = \frac{p}{\lambda_1}\,e^{-\lambda_1 y} + \frac{1-p}{\lambda_2}\,e^{-\lambda_2 y}, \quad y \ge 0. \tag{A.46}$$

# Appendix B

# Optimization methods

## B.1 Stationary points

In inventory theory objective functions of the following form are often encountered:

$$g(x) = \frac{a}{x} + bx, \quad x > 0; \tag{B.1}$$

here, $a$ and $b$ are positive constants. The derivative of this function is

$$g'(x) = -\frac{a}{x^2} + b, \quad x > 0. \tag{B.2}$$

This derivative has a unique positive zero at

$$x = x^* \doteq \sqrt{a/b}. \tag{B.3}$$

The second derivative of $g(x)$ at $x^*$ is positive:

$$g''(x) = \frac{2a}{x^3} \qquad \Rightarrow \quad g''(x^*) = \frac{2b\sqrt{b}}{\sqrt{a}}. \tag{B.4}$$

Hence, the function $g(x)$ has a minimum at $x^*$ with the value

$$g(x^*) = a\sqrt{b/a} + b\sqrt{a/b} = 2\sqrt{ab}. \tag{B.5}$$

## B.2 Golden section

The “Golden-Section” method is an interval elimination method for finding the minimum $x^*$ of a function $g(x)$ with no local minima beside the global minimum. This method assumes that $g(x)$ is strictly decreasing for $x < x^*$ and strictly increasing for $x > x^*$. An initial interval $(a, b)$ is divided into three subintervals $(a, c)$, $(c, d)$ and $(d, b)$ with relative lengths $z^2$, $1 - 2z^2$, $z^2$, respectively, with $z = \frac{1}{2}(\sqrt{5} - 1) \approx 0.618$. This implies that $z^2 = \frac{1}{2}(3 - \sqrt{5}) = 1 - z \approx 0.382$, $c = a + z^2(b - a)$ and $d = b - z^2(b - a) = a + z(b - a)$. If $g(c) < g(d)$, the interval $(d, b)$ is discarded, and the procedure is repeated for the interval $(a, d)$; otherwise, the interval $(a, c)$ is discarded, and the procedure is repeated for the interval $(c, b)$. The lengths of the new intervals are $d - a = b - c = z(b - a)$. Hence, the procedure continues with intervals which are a factor $z \approx 0.618$ smaller with each step, until the minimum has been found with sufficient accuracy.

**Example B.1** Consider an objective function of the form (B.1):

$$g(x) = \frac{500}{x} + \frac{x}{20}, \quad x > 0.$$

As initial points we take $a = 10$ and $b = 1000$ where one of the terms of $g(x)$ dominates the other: $g(10) = g(1000) = 50 + 0.5 = 50.5$. Table B.1 shows the results of the golden section procedure with and without rounding to integer order quantities. The procedure can be continued to end at the optimum $x = 100$. Note the flatness of the objective function near its minimum. $\square$

Table B.1: Results of golden section procedure for the function of Example B.1.

| Point $c$ | $g(c)$ | Point $d$ | $g(d)$ | Rem. interval | Point $c$ | $g(c)$ | Point $d$ | $g(d)$ | Remaining interval |
|---|---|---|---|---|---|---|---|---|---|
| 388 | 20.69 | 622 | 31.90 | (10, 622) | 388.15 | 20.70 | 621.85 | 31.90 | (10.00, 621.85) |
| 244 | 14.25 | 388 | 20.69 | (10, 388) | 243.71 | 14.24 | 388.15 | 20.70 | (10.00, 388.15) |
| 154 | 10.95 | 244 | 14.25 | (10, 244) | 154.44 | 10.96 | 243.71 | 14.24 | (10.00, 243.71) |
| 99 | 10.00 | 154 | 10.95 | (10, 154) | 99.27 | 10.00 | 154.44 | 10.96 | (10.00, 154.44) |
| 65 | 10.94 | 99 | 10.00 | (65, 154) | 65.17 | 10.93 | 99.27 | 10.00 | (65.17, 154.44) |
| 99 | 10.00 | 120 | 10.17 | (65, 120) | 99.27 | 10.00 | 120.34 | 10.17 | (65.17, 120.34) |
| 86 | 10.11 | 99 | 10.00 | (86, 120) | 86.24 | 10.11 | 99.27 | 10.00 | (86.24, 120.34) |
| 99 | 10.00 | 107 | 10.02 | (86, 107) | 99.27 | 10.00 | 107.32 | 10.02 | (86.24, 107.32) |
| 94 | 10.02 | 99 | 10.00 | (94, 107) | 94.29 | 10.02 | 99.27 | 10.00 | (94.29, 107.32) |
| 99 | 10.00 | 102 | 10.00 | (94, 102) | 99.27 | 10.00 | 102.34 | 10.00 | (94.29, 102.34) |
| 97 | 10.00 | 99 | 10.00 | (97, 102) | 97.37 | 10.00 | 99.27 | 10.00 | (97.37, 102.34) |

## B.3 Bisection

The method of interval bisection is an alternative adaptive search method for finding the minimum $x^*$ of a function $g(x)$ with no local minima beside the global minimum. This method employs information about the value of the derivative of the function $g(x)$. It assumes that $g'(x)$ is negative for $x < x^*$ and positive for $x > x^*$. The method starts with an initial interval $(a, b)$ which should contain $x^*$. The derivative $g'(x)$ is evaluated at the middle point $c = \frac{1}{2}(a + b)$. If $g'(c) > 0$, the interval $(c, b)$ is discarded, and the procedure continues with the interval $(a, c)$. If $g'(c) < 0$, the interval $(a, c)$ is discarded, and the procedure continues with the interval $(c, b)$. Each application of the method halves the length of the interval in which the minimum $x^*$ must lie. With the same initial interval, the method of interval bisection is faster than the Golden-Section method in terms of the number of iterations. However, the computation of the derivative $g'(x)$ may be more involved than that of $g(x)$.

Table B.2: Results of bisection procedure for the function of Example B.2.

| Middle point $c$ | $g'(c)$ | Remaining interval | Middle point $c$ | $g'(c)$ | Remaining interval |
|---|---|---|---|---|---|
| 505 | 0.0480 | (10, 505) | 505.00 | 0.0480 | (10.00, 505.00) |
| 258 | 0.0425 | (10, 258) | 257.50 | 0.0425 | (10.00, 257.50) |
| 134 | 0.0222 | (10, 134) | 133.75 | 0.0220 | (10.00, 133.75) |
| 72 | −0.0465 | (72, 134) | 71.88 | −0.0468 | (71.88, 133.75) |
| 103 | 0.0029 | (72, 103) | 102.81 | 0.0027 | (71.88, 102.81) |
| 88 | −0.0146 | (88, 103) | 87.34 | −0.0155 | (87.34, 102.81) |
| 96 | −0.0043 | (96, 103) | 95.08 | −0.0053 | (95.08, 102.81) |
| 100 | 0.0000 |  | 98.95 | −0.0011 | (98.95, 102.81) |
|  |  |  | 100.88 | 0.0009 | (98.95, 100.88) |
|  |  |  | 99.91 | −0.0001 | (99.91, 100.88) |

**Example B.2** Consider again the objective function of Example B.1. This function has derivative:

$$g'(x) = -\frac{500}{x^2} + \frac{1}{20}, \quad x > 0.$$

As initial points we again take $a = 10$ and $b = 1000$ where the derivative has opposite sign: $g'(10) = -4.9500$, $g'(1000) = 0.0495$. Table B.2 shows the results of the bisection procedure with and without rounding. Clearly, this procedure converges faster than the golden section procedure in Table B.1. □

The method of interval bisection can also be used to search a root of an equation or the zero of a function. In this case, it is assumed that the function is continuous and that it possesses a single zero on the initial search interval so that the sign of the function at the two end points is different. To find a root of a function, the sign of the function at the middle point is considered. If this sign is the same as that at the left end point, then the left half of the interval is discarded; otherwise, the right half of the interval is discarded.

# Bibliography

[1] Aksoy, Y., S. Erenguc. Multi-item inventory models with coordinated replenishments: a survey, *Int. J. Production Management* **8** (1988), 63–73.

[2] Andres, F., H. Emmons. A multi-product inventory system with interactive set-up costs, *Management Science* **21** (1975), 1055–1063.

[3] Atkins, D.R., P.O. Iyogun. A lower bound for a class of inventory/production problems, *Operations Research Letters* **6** (1987), 63–67.

[4] Atkins, D.R., P.O. Iyogun. Periodic versus 'can-order' policies for coordinated multi-item inventory systems, *Management Science* **34** (1988), 791–796.

[5] Atkins, D.R., P.O. Iyogun. A heuristic with lower bound performance guarantee for the multi-product dynamic lot-size problem, *IIE Transactions* **20** (1988), 369–373.

[6] Axsäter, S. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand, *Operations Research* **48** (2000), 686–696.

[7] Axsäter, S. *Inventory Control*, Springer, New York, 2006 (2nd ed.).

[8] Axsäter, S., W.F. Zhang. A joint replenishment policy for multi-echelon inventory control, *Int. J. Production Economics* **59** (1999), 243–250.

[9] Baker, K.R. Lot-sizing procedures and a standard date set: a reconciliation of the literature, *J. Manufacturing and Operations Management* **2** (1989), 199–221.

[10] Ballintfy, J.L. On a basic class of inventory systems, *Management Science* **10** (1964), 287–297.

[11] Bastian, M. Joint replenishment in multi-item inventory systems, *J. Oper. Res. Soc.* **37** (1986), 1113–1120.

[12] Blackburn, J., R. Millen. Heuristics lot-sizing performance in a rolling-schedule environment, *Decision Sciences* **11** (1980), 691–701.

[13] Blackburn, J., R. Millen. Improved heuristics for multi-stage requirements planning systems, *Management Science* **28** (1982), 44–56.

[14] Chakravarty, A.K. Multi-item inventory aggregation into groups, *J. Oper. Res. Soc.* **32** (1981), 19–26.

[15] Chakravarty, A.K. Joint inventory replenishments with group discounts based on invoice value, *Management Science* **30** (1984), 1105–1112.

[16] Chakravarty, A.K. An optimal heuristic for coordinated multi-item inventory replenishments, *J. Oper. Res. Soc.* **36** (1985), 1027–1039.

[17] Chakravarty, A.K., S.K. Goyal. Multi-item inventory grouping with dependent set-up cost and group overhead cost, *Engineering Costs and Production Economics* **10** (1986), 13–23.

[18] Clark, A.J., H. Scarf. Optimal policies for a multi-echelon inventory problem, *Management Science* **6** (1960), 475–490.

[19] Dagpunar, J.S. On the relationship between constrained and unconstrained multi-item replenishment problems, *Int. J. Oper. Prod. Management* **9** (1989), 77–81.

[20] Das, C., S.K. Goyal. Economic ordering policy for deterministic two-echelon distribution systems, *Engineering Costs and Production Economics* **21** (1991), 227–231.

[21] De Kok, A.G. Hierarchical production planning for consumer goods, *Eur. J. Oper. Res.* **45** (1990), 55–69.

[22] DeMatteis, J. An economic lot-sizing technique I: the part-period algorithm, *IBM Syst. J.* **7** (1968), 30–38.

[23] Eastman, R.M. *Materials Handling*, Marcel Dekker, New York, 1987.

[24] Eppen, G., L. Schrage. Centralized ordering policies in a multi-warehouse system with lead times and random demand, in: L.B. Schwarz (ed.) *Multi-level Production-Inventory Control Systems: Theory and Practice*, North-Holland, Amsterdam, 1981.

[25] Ettl, M., G.E. Feigin, G.Y. Lin, D.D. Yao. A supply network model with base-stock control and service requirements, *Operations Research* **48** (2000), 216–232.

[26] Eynan, A., D.H. Kropp. Periodic review and joint replenishment in stochastic demand environments, *IIE Transactions* **30** (1998), 1025–1033.

[27] Federgruen, A. Centralized planning models for multi-echelon inventory systems under uncertainty, in: S.C. Graves, A.H.G. Rinnooy Kan, P.H. Zipkin (eds.) *Logistics of Production and Inventory*, North-Holland, Amsterdam, 1993.

[28] Federgruen, A., H. Groenevelt, H.C. Tijms. Coordinated replenishment in a multi-item inventory system with compound Poisson demands, *Management Science* **30** (1984), 344–357.

[29] Federgruen, A., P. Zipkin. An efficient algorithm for computing optimal (s,S) policies, *Operations Research* **32** (1984), 1268–1285.

[30] Fogarty, D.W., J.H. Blackstone, T.R. Hoffmann. *Production & Inventory Management*, South-Western Publ. Co., Cincinnati, Ohio, 1991, 2nd ed..

[31] Goetschalckx, M., H.D. Ratliff. Order picking in an aisle, *IIE Transactions* **20** (1988), 53–62.

[32] Goetschalckx, M., H.D. Ratliff. Sequencing picking operations in a man-aboard order picking system, *Material Flow* **4** (1988), 255–263.

[33] Goyal, S.K. Determination of optimum packaging frequency of items jointly replenished, *Management Science* **21** (1974), 436–443.

[34] Goyal, S.K. Optimum ordering policy for a multi-item single supplier system, *Operational Research Quarterly* **25** (1974), 293–298.

[35] Goyal, S.K., A.S. Belton. On a simple method of determining order quantities in joint replenishment under deterministic demand, *Management Science* **25** (1979), 604.

[36] Goyal, S.K., A.T. Satir. Joint replenishment inventory control: deterministic and stochastic models, *Eur. J. Oper. Res.* **38** (1989), 2–13.

[37] Graves, S.C. On the deterministic demand multi-product single-machine lot scheduling problem, *Management Science* **25** (1979), 276–280.

[38] Hausman, W.H., L.B. Schwarz, S.C. Graves. Optimal storage assignment in automatic warehousing systems, *Management Science* **22** (1976), 629–638.

[39] Ignall, E. Optimal continuous review policies for two product inventory systems with joint set-up cost, *Management Science* **15** (1967), 277–279.

[40] Inderfurth, K. Safety stocks in multistage inventory systems: a survey, *Int. J. Production Economics* **35** (1994), 321–329.

[41] Iyogun, P.O. Heuristic methods for the multi-product dynamic lot size problem, *J. Oper. Res. Soc.* **42** (1991), 889–894.

[42] Jackson, P., W. Maxwell, J. Muckstadt. The joint replenishment problem with a powers-of-two restriction, *IIE Transactions* **17** (1985), 25–32.

[43] Joneja, D. The joint replenishment problem: new heuristic and worst case performance bounds, *Operations Research* **38** (1990), 711–723.

[44] Kao, P.C. A multi-product dynamic lot-size model with individual and joint set-up costs, *Operations Research* **27** (1979), 279–289.

[45] Kaspi, M., M.J. Rosenblatt. An improvement of Silver's algorithm for the joint replenishment problem, *IIE Transactions* **15** (1983), 264–267.

[46] Korugan, A., S.M. Gupta. A multi-echelon inventory system with returns, *Computers & Industrial Engineering* **35** (1998), 145–148.

[47] Lambrecht, M.R., H. Vanderveken, J. vander Eecken. On the joint replenishment problem under deterministic time varying demand patterns, *Onderzoeksrapport* nr. 7907, K.U. Leuven, 1979.

[48] Lee, F.-C., M.-J. Yao. A global optimum search algorithm for the joint replenishment problem under power-of-two policy, *Computers & Operations research* **30** (2003), 1319–1333.

[49] Love, S. *Inventory Control*, Mc Graw Hill, New York, 1979.

[50] Liu, L., X.M. Yuan. Coordinated replenishment in inventory systems with correlated demands, *Eur. J. Oper. Res.* **123** (2000), 490–503.

[51] McGee, V.E., D.F. Pyke. Periodic production scheduling at a fastener manufacturer, *Int. J. Production Economics* **46–47** (1996), 65–87.

[52] Mendoza, A. An economic lot-sizing technique II: mathematical analysis of the part-period algorithm, *IBM Syst. J.* **7** (1968), 39–46.

[53] Odanaka, T., K. Yamaguchi, T. Masui. Multi-echelon inventory production system solution, *Computers & Industrial Engineering* **27** (1994), 201–204.

[54] Pantumsinchai, P. A comparison of three joint ordering inventory policies, *Decision Science* **23** (1992), 111–127.

[55] Ratliff, H.D., A.S. Rosenthal. Order picking in a rectangular warehouse: a solvable case of the traveling salesman problem, *Operations Research* **31** (1983), 507–516.

[56] Robeson, J.F., R.G. House. *The Distribution Handbook*, The Free Press, New York, 1985.

[57] Schrady, D.A. A deterministic inventory model for repairable items, *Naval Research Logistics Quarterly* **14** (1967), 391–398.

[58] Silver, E.A. A control system for coordinated inventory replenishment, *Int. J. Production Res.* **12** (1974), 647–671.

[59] Silver, E.A. A simple method of determining order quantities in joint replenishment under deterministic demand, *Management Science* **22** (1976), 1351–1361.

[60] Silver, E.A. Coordinated replenishment of items under time-varying demand: dynamic programming formulation, *Naval Research Logistics* **26** (1979), 141–151.

[61] Silver, E.A. Establishing reorder points in the (S,c,s) coordinated control system under compound Poisson demand, *Int. J. Production Res.* **19** (1981), 743–750.

[62] Silver, E.A., H.C. Meal. A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment, *Production and Inventory Management*, **14** (1973), 64–74.

[63] Silver, E., J. Miltenburg. Two modifications of the Silver-Meal lot sizing heuristic, *INFOR* **22** (1984), 56–68.

[64] Silver, E.A., D.F. Pyke, R. Peterson. *Inventory Management and Production Planning and Scheduling*, Wiley, New York, 1998 (3rd ed.).

[65] Tijms, H.C. *Stochastic Models: an algorithmic approach.* Wiley, Chichester, 1994.

[66] Tomkins, J., J. White. *Facilities Planning*, Wiley, New York, 1984.

[67] Van der Duyn Schouten, F.A., M.J.G. Van Eijs, R.M.J. Heuts. Coordinated replenishment systems with discount opportunities, *Int. J. Production Research* **32** (1994), 2879–2896.

[68] Van der Heijden, M.C. Supply rationing in multi-echelon divergent systems, *Eur. J. Oper. Res.* **101** (1997), 532–549.

[69] Van der Heijden, M. Near cost-optimal inventory control policies for divergent networks under fill rate constraints, *Int. J. Production Economics* **63** (2000), 161–179.

[70] Van Eijs, M.J.G. A note on the joint replenishment problem under constant demand, *J. Oper. Res. Soc.* **44** (1993), 185–192.

[71] Van Eijs, M.J.G. On the determination of the control parameters of the optimal can-order policy, *Zeitschrift für Operations Research* **39** (1994), 289.

[72] Van Eijs, M.J.G. Multi-item inventory systems with joint ordering and transportation decisions, *Int. J. Production Economics* **35** (1994), 285–292.

[73] Van Eijs, M.J.G., R.M.J. Heuts, J.P.C. Kleijnen. Analysis and comparison of two strategies for multi-item inventory systems with joint replenishment costs, *Eur. J. Oper. Res.* **59** (1992), 405–412.

[74] Verrijdt, J.H.C.M., A.G. De Kok Distribution planning for a divergent $N$-echelon network without intermediate stocks under service restrictions, *Int. J. Production Economics* **38** (1995), 225–243.

[75] Verrijdt, J.H.C.M., A.G. De Kok Distribution planning for a divergent depotless two-echelon network under service restrictions, *Eur. J. Oper. Res.* **89** (1996), 341–354.

[76] Viswanathan, S. Periodic review (s,S) policies for joint replenishment inventory systems, *Management Science* **43** (1997), 1447–1454.

[77] Wagner, H.M., T.M. Within. Dynamic version of the economic lot size model, *Management Science* **5** (1958), 89–96.

[78] Wildeman, R.E., J.B.G. Frenk, R. Dekker. An efficient optimal solution method for the joint replenishment problem, *Eur. J. Oper. Res.* **99** (1997), 433–444.

[79] Wilson, H.G. Order quantity, product popularity and the location of stock in a warehouse, *IIE Transactions* **9** (1977), 230–237.

[80] Winston, W.L. *Operations Research: Applications and Algorithms.* PWS-Kent Publ. Co., Boston, 3rd ed., 1994.

[81] Yoo, Y., W. Kim, J. Rhee. Efficient inventory management in multi-echelon distribution systems, *Computers & Industrial Engineering* **33** (1997), 729–732.

[82] Zheng, Y.S., A. Federgruen. Finding optimal (s,S) policies is as simple as evaluating a single policy, *Operations Research* **39** (1991), 654–665 (corrections: *Operations Research* **40** (1992), 192).

[83] Zipkin, M.C. On the imbalance of inventories in multi-echelon systems, *Math. Oper. Res.* **9** (1984), 402–423.

# Index

# List of symbols

This list only contains the most important notations. Some symbols may locally have a different meaning. Random variables are indicated by “r.v.”.

$\doteq$ is by definition

$\approx$ is approximately equal to

$[\ldots]^+$ is $\max\{0,\ldots\}$

$\mathbf{1}$ vector of ones

$a$ individual item ordering cost

$A$ family ordering cost

$B$ backlog per cycle

$c$ can-order level

$\mathbf{c}$ vector of can-order levels

$C(\ldots)$ cost function

$C_Y$ coefficient of variation of r.v. $Y$

$d$ discount factor

$D$ demand per unit of time (fixed or r.v.)

$D_L$ demand during lead time (r.v.)

$D_{L+R}$ demand during lead time and review period (r.v.)

$E\{\ldots\}$ mean of a r.v.

$\mathcal{G}$ set of items forming a group

$I_{\{.\}}$ indicator function

$h$ holding cost

$H$ planning horizon

$I$ inventory level

$I_O$ inventory level at reorder instant

$k$ reorder frequency

$\mathbf{k}$ vector of reorder frequencies

$L$ lead time (fixed or r.v.)

$M$ number of orders per unit of time

$n$ ratio between order quantities

$N$ number of items in a family (Ch. 2, Sect. 4.1)

$N$ number of stocking points at an echelon (Ch. 3)

$Q$ order quantity (fixed or r.v.)

$p$ production rate

$r$ carrying charge

$R$ review period

$s$ reorder point

$\mathbf{s}$ vector of reorder points

$S$ order-up-to level

$\mathbf{S}$ vector of order-up-to levels

$T$ reorder cycle

$u$ set-up time

$U$ family set-up time

$v$ purchasing price

$V$ price break point

$w$ waste rate

$\alpha$ cycle service level

$\beta$ target fill rate

$\Pi$ actual probability of no backlog

$\sigma\{\ldots\}$ standard deviation of a r.v.

$\tau$ production time

$\phi$ volume of items

$\Phi$ inventory capacity

$\Psi$ actual fill rate