Lectures in Supply-Chain Optimization

Arthur F. Veinott, Jr.

Management Science and Engineering 361

Department of Management Science and Engineering
Stanford University
Stanford, California 94305

Contents

Homework

1

Introduction to Supply-Chain Optimization

1 OVERVIEW

Supply Chains. The supply chains of large corporations involve hundreds of facilities (retailers, distributors, plants and suppliers) that are globally distributed and involve thousands of parts and products. As one example, one auto manufacturer has 12 thousand suppliers, 70 plants, operates in 200 countries and has annual sales of 8.6 million vehicles. As a second example, the US Defense Logistics Agency, the world's largest warehousing operation, stocks over 100 thousand products. The goals of corporate supply chains are to provide customers with the products they want in a timely way and as efficiently and profitably as possible. Fueled in part by the information revolution and the rise of e-commerce, the development of models of supply chains and their optimization has emerged as an important way of coping with this complexity. Indeed, this is one of the most active application areas of operations research and management science today. This reflects the realization that the success of a company generally depends on the efficiency with which it can design, manufacture and distribute its products in an increasingly competitive global economy.

Decisions. There are many decisions to be made in supply chains. These include

- what products to make and what their designs should be;
- how much, when, where and from whom to buy product;
- how much, when and where to produce product;
- how much and when to ship from one facility to another;
- how much, when and where to store product;
- how much, when and where to charge for products; and
- how much, when and where to provide facility capacity.

These decisions are typically made in an environment that involves uncertainty about product demands, costs, prices, lead times, quality, etc. The decisions are generally made in a multi-party environment that includes competition and often includes collaborative alliances.

Alliances lead to questions of how to share the benefits of collaboration and what information the various parties ought to share. In many supply chains, the tradition has been not to share information. On the other hand, firms in more than a few supply chains realize that there are important benefits from sharing information too, e.g., the potential for making supply chains more responsive and efficient.

Inventories. Typically firms carry inventories at various locations in a supply chain to buffer the operations at different facilities and in different periods. Inventories are the links between facilities and time periods. Inventories of raw materials, work-in-process, and finished goods are ubiquitous in firms engaged in production or distribution (by sale or circulation) of one or more products. Indeed, in the United States alone, 2000 manufacturing and trade inventories totaled 1,205 billion dollars, or 12% of the gross domestic product of 9,963 billion dollars that year (2001 *Statistical Abstract of the United States*, U.S. Department of Commerce, Bureau of the Census, Tables 756 and 640). The annual cost of carrying these inventories, e.g., costs associated with capital, storage, taxes, insurance, etc., is significant—perhaps 25% of the total investment in inventories, or 301 billion dollars and 3% of the gross domestic product.

Scope. The conventional types of inventories include raw materials, work in process, and finished goods. But there are many other types of inventories which, although frequently not thought of as inventories in the usual sense, can and have been usefully studied by the methods developed for the study of ordinary inventories. Among others, these include:

- plant capacity,
- equipment,
- space (airline seat, container, hotel room)
- circulating goods (cars, computers, books),
- cash and securities,
- queues,
- populations (labor, livestock, pests, wildlife),
- goodwill,
- water and even
- pollutants.

Thus the scope of applications of the methods of supply-chain optimization is considerably wider than may seem the case at first.

2 MOTIVES FOR HOLDING INVENTORIES

Since it is usually expensive to carry inventories, efficient firms would not do so without good reasons. Thus, it seems useful to examine the motives for holding inventories. As we do so, we give examples briefly illustrating many of the more common motives as well as the methods used to analyze them. We shall discuss only the case of a single facility and single party in the remainder of this section, reserving the complications arising from multiple facilities and parties for subsequent sections.

It is helpful to begin by formulating and studying an n-period supply-chain problem under certainty. Let x_i, y_i and s_i be respectively the nonnegative production, end-of-period inventory and given demand for a single product in period $i = 1, \ldots, n$. Let $c_i(z)$ and $h_i(z)$ be respectively the costs of producing and storing $z \geq 0$ units of the product in period i. We can and do assume without loss of generality that $c_i(0) = h_i(0) = 0$ for all i. The problem is to choose production and inventory schedules $x = (x_i)$ and $y = (y_i)$ respectively that minimize the n-period cost

$$C(x,y) \equiv \sum_1^n [\, c_i(x_i) + h_i(y_i)] \tag{1}$$

subject to the stock-conservation constraints

$$x_i + y_{i-1} - y_i = s_i, i = 1, \ldots, n \tag{2}$$

and nonnegativity of production and inventories (the last to assure that demands are met as they arise without backorders)

$$x, y \geq 0 \tag{3}$$

where for simplicity we set $y_0 \equiv y_n \equiv 0$.

Network-Flow Formulation. The problem (1)-(3) can be viewed as one of finding a minimum-nonlinear-cost network flow as Figure 1 illustrates. The variables are the flows in the arcs that they label, the exogenous demands (negative demands are "supplies") at nodes $1, \ldots, n$ are

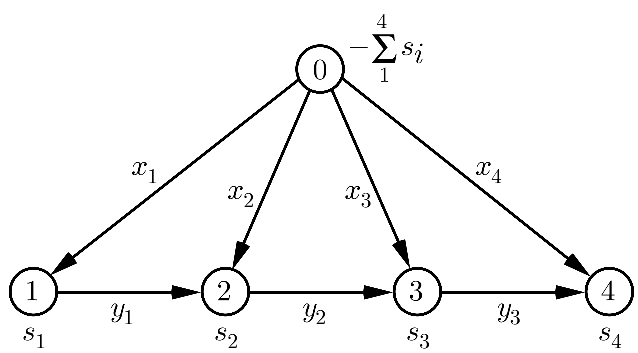


Figure 1. Production Planning Network

the given demands in those periods, and the supply at node zero is the total demand $\sum_i s_i$ in all periods. The stock-conservation constraints (2) are the flow-conservation constraints in the network. The flow-conservation constraint at node zero expresses the fact that total production in all periods equals total demand in all periods. The flow-conservation constraint at each other node $i > 0$ expresses the fact that the sum of the initial inventory and production in period i equals the sum of the demand and final inventory in that period.

One important motive for carrying inventories arises when there is a temporal increase in the marginal cost of supplying demand, i.e., $\dot{c}_i(s_i)$ increases in i over some interval. (The derivative here is with respect to the quantity, not time.) There are at least two ways in which this can happen.

Linear Costs and Temporal Increase in Unit Supply Cost. One is where the costs are linear, so there are unit costs c_i and h_i of production and storage in period i. Then it is optimal to hold inventory in a period if the unit production cost in that period is less than that in the following period and the unit storage cost is small enough. In any case, the problem (1)-(3) is then a minimum-linear-cost uncapacitated network-flow problem in which node zero is the source from which the demands at the other nodes are satisfied. Clearly a minimum-cost flow can be constructed by finding for each node $i > 0$ a minimum-cost chain (i.e., directed path) from node zero to i, and satisfying the demand at i by shipping along that chain. Let C_i be the resulting minimum cost. The C_i can be found recursively from the dynamic-programming forward equations ($h_0 \equiv C_0 \equiv \infty$)

$$C_i = \min(c_i, h_{i-1} + C_{i-1}), i = 1, \ldots, n. \tag{4}$$

This recursion expresses the fact that a minimum-cost chain from node zero to node i either consists solely of the production arc from node zero to i and incurs the cost c_i, or contains node $i-1$ and the storage arc joining nodes $i-1$ and i, and incurs the cost $h_{i-1} + C_{i-1}$. In short, the minimum-cost way to satisfy each unit of demand in period i is to choose the cheaper of two alternatives, viz., satisfy the demand by production in period i or by production at an optimally chosen prior period and storage to period i. The C_i are calculated by forward induction in the order $C_1, C_2, \ldots, C_n$.

In this process one records the periods $i_1 = 1 < i_2 < \cdots < i_p \le n$, say, in which it is optimal to produce, i.e., periods i for which $C_i = c_i$. Then if it is optimal to produce in a period i_k, say, it is optimal to produce an amount exactly satisfying all demands prior to the next period i_{k+1} ($i_{p+1} \equiv n+1$) in which it is optimal to produce, i.e.,

$$x_{i_k} = \sum_{j=i_k}^{i_{k+1}-1} s_j,\ k = 1, \ldots, p. \tag{5}$$

This means that it is optimal to produce only in periods in which there is no entering inventory, i.e.,

$$y_{i-1}x_i = 0, i = 1, \ldots, n. \tag{6}$$

The C_i can be computed in *linear time* with at most n operations where an *operation* is here an addition and a comparison. To see this observe from (4) that the computing C_i requires one addition (computing $h_{i-1} + C_{i-1}$) and one comparison (choosing the smaller of the two costs c_i and $h_{i-1} + C_{i-1}$). Because there are n such C_i to compute, the claim is verified.

Since the periods in which it is optimal to produce are independent of the demands, it is clear from (5) that optimal production in a period is increasing and linear in present and future demands with the rate of increase not exceeding that of the demands. Also, the magnitude of the change in optimal present production resulting from a change in forecasted demand in a future period diminishes the more distant the period of the change in forecasted demand.

This example illustrates several themes that will occur repeatedly throughout this course.

Network-Flow Models of Supply Chains. One is the fundamental idea in §4-§6 of formulating and solving supply-chain problems as minimum-cost network-flow problems with scale diseconomies or economies in the arc flow costs. This approach has several advantages. First, it unifies the treatment of many supply-chain models. Second, it extends the applicability of the methods to broad classes of problems outside of supply-chain management. Third, it facilitates use of the special structure of the associated graphs to characterize optimal flows and develop efficient methods of computing those flows.

Lattice Programming and Comparison of Optima. A second fundamental recurring theme is that of predicting the direction and relative magnitude of changes in optimal decision variables resulting from changes in parameters of an optimization problem without computation. The theory of lattice programming is developed for this purpose in §2. That theory is extended to substitutes, complements and ripples in minimum-convex-cost network flows in §4. This permits prediction of the direction and relative magnitude of the change of the optimal flow in an arc resulting from changing certain arc parameters, e.g., bounds on arc flows or parameters of arc flow costs, without computation. So pervasive are these qualitative results that they will be applied repeatedly in all subsequent sections of this course. As a concrete example, if there are scale diseconomies in production and storage costs, the effect of an increase in the storage cost in a given period is to reduce optimal storage in each period, reduce optimal production in or before the given period, and increase optimal production after the given period.

Dynamic Programming. A third major theme is that of solving supply-chain problems, and more generally, minimum-cost network-flow problems, by dynamic programming. In particular, the idea of solving such problems by means of a sequence of minimum-cost-chain problems arises repeatedly in §4-§6 where there are scale diseconomies or economies in arc flow costs. Also, dynamic-programming recursions are used often in §8 to characterize by induction the properties of optimal policies and the minimum-cost functions, as well as to compute them.

Complexity. A final theme that occurs throughout is that of developing efficient algorithms and estimating their running times in terms of problem size. The last is typically measured by parameters like the numbers of periods and products.

Temporal Increase in Demand and Scale Diseconomies in Supply. Another way in which there can be a temporal increase in the marginal cost of supplying demand arises where there is a temporal increase in demand and scale diseconomies in production. A temporal increase in demand may occur because of long-term growth or fluctuations, e.g., seasonality, thereof. Production scale diseconomies occur when there are alternate sources of supply, each with limited capacity, or when production at a plant in excess of normal capacity must be deferred to a second shift or to over-time with an attendant increase in unit labor costs. Figure 2 illustrates this possibility.

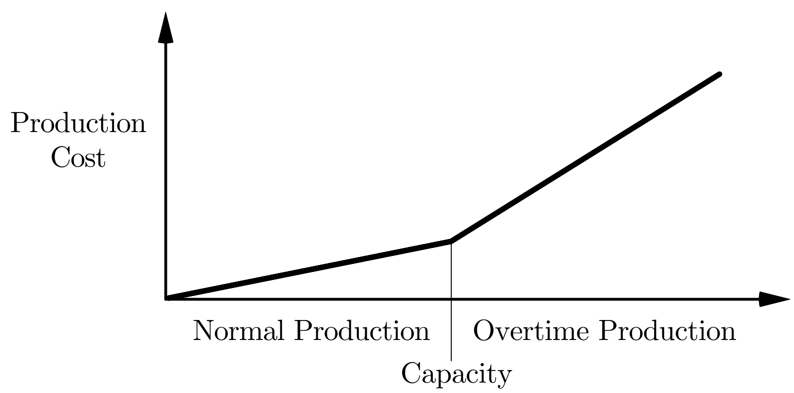


FIGURE 2. Production Cost with Scale Diseconomies

As a simple example, consider a toy maker who faces respectively no demand and a demand for $s > 0$ toys in the first and second halves of a year. Suppose also that the toy maker can produce $z \geq 0$ toys in either half of the year at a cost $c(z)$ with $c(\cdot)$ being convex. Then the marginal cost $\dot{c}(z)$ of producing z units is increasing in z, i.e., there are scale diseconomies in production. If storage costs are neglected (because, for example, they might be fixed), then the toy maker's problem is to choose nonnegative amounts x_1 and x_2 of toys to produce in the first and second halves of the year that minimize

$$C(x_1, x_2) \equiv c(x_1) + c(x_2)$$

subject to

$$x_1 + x_2 = s.$$

The minimum-cost schedule is $x^0 = (\frac{s}{2}\ \ \frac{s}{2})$, i.e., to produce equal numbers of toys in each half of the year. To see this, observe that if x is any feasible schedule, then so is its permutation $x' = (x_2, x_1)$. Moreover, since $x^0 = \frac{1}{2}x + \frac{1}{2}x', C$ is convex and $C(x) = C(x')$,

$$C(x^0) \le \frac{1}{2}C(x) + \frac{1}{2}C(x') = C(x).$$

Observe that it is optimal to produce in the first half even though there is no demand in that half in order to avoid producing at a high marginal cost in the last half if one instead did not produce in the first half. Thus, it is optimal to carry inventories in the first half. There is another important property of the optimal production schedule that deserves mention because it will play an important role in subsequent generalizations of this problem to invariant network flows in §5. It is that the optimal schedule is independent of the production cost function c, provided only that c is convex.

Scale Economies in Supply

Scale economies in supply provided another important motive for holding inventories. Scale economies occur because of the availability of quantity discounts or setup costs associated with production/procurement as Figure 3 illustrates. Scale economies can make it attractive to combine

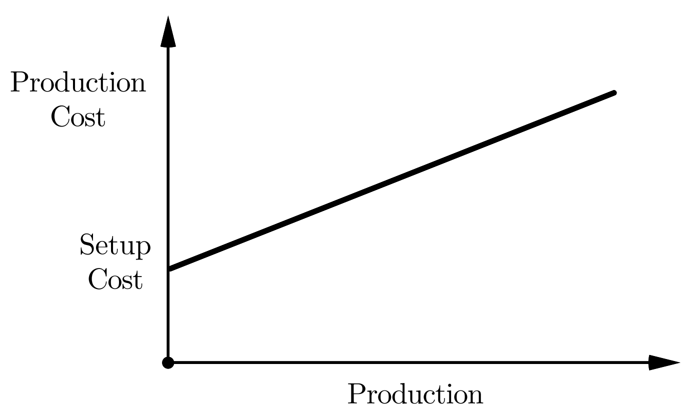


FIGURE 3. Production Cost with Scale Economies

orders for one or more products placed at different points in time because of the reduction in average unit purchase cost that ensues. On the other hand, the process of combining orders for one or more products in this way does have a cost, viz., one is led to place some orders before they are needed, thereby creating inventories. This leads one to seek a balance between the extremes of frequently ordering small quantities (which entails high ordering costs) and occasionally ordering large quantities (which entails large holding costs). Scale economies are naturally reflected by concavity of the cost function C say. When that is so, the minimum of C over a convex polytope occurs at an extreme point thereof. For if x is an element of the polytope and $e^1, \ldots, e^k$ are its extreme points, then there exist numbers $\alpha^1, \ldots, \alpha^k \ge 0$ with $\sum_i \alpha^i = 1$ such that $x = \sum_i \alpha^i e^i$. Then by the concavity of C,

$$C(x) \ge \sum_1^k \alpha^i C(e^i) \ge \min_{1 \le i \le k} C(e^i).$$

In §6 we characterize the extreme points of the set of nonnegative network flows in terms of the structure of the graph and the number of demand nodes, i.e., nodes with nonzero demands. We also develop a send-and-split algorithm for searching the extreme points to find one that is optimal. In general graphs, the running time of send-and-split is polynomial in the number of nodes and arcs, and exponential in the number of demand nodes. In planar graphs in which the demand nodes all lie on the outer face, e.g., Figure 1, the running time is also polynomial in the number of demand nodes.

Dynamic Economic-Order-Interval Problem. As an example of these ideas, it is not difficult to see that for the production-planning graph of Figure 1, the subgraph induced by the arcs with positive flow in an extreme flow is a collection of chains that share only node zero and are directed away from it. This implies that there is no node whose induced subgraph contains two arcs with a common head, or equivalently, (6) holds. Thus the extreme flows in this problem have the property that one orders in a period only if there is no entering inventory just as for the case of linear costs. Of course, this is to be expected because linear functions are concave. However, the periods in which it is optimal to order with concave costs are not generally independent of the size of the demands as they are with linear costs. For this reason, the algorithm given in (4) is no longer valid for general concave functions. Nevertheless, there is an alternate dynamic-programming algorithm for searching the schedules with the property (6). To describe the method, let c_{ij} be the sum of the costs of ordering and storing in periods $i{+}1, \ldots, j$ when one orders in period $i{+}1$ an amount equaling the total demand $\sum_{i+1}^{j} s_k$ in periods $i{+}1, \ldots, j$. Let C_j be the minimum cost in periods $1, \ldots, j$ when there is no inventory on hand at the end of period j. Then the C_j can be calculated from the dynamic-programming forward recursion ($C_0 \equiv 0$)

$$C_j = \min_{0 \le i < j} (C_i + c_{ij}), j = 1, \ldots, n. \tag{7}$$

The running time of the algorithm is easily seen to be quadratic in n because the c_{ij} can be computed for each j in the order $i = j{-}1, \ldots, 0$. Incidentally, like the recursion (4), the recursion (7) also finds the minimum costs of chains from node 0 to nodes $1, \ldots, n$. However, in the present case, the graph is not the one in Figure 1, but rather one in which c_{ij} is the cost of traversing arc (i, j). Also, the running time in the present case is $O(n^2)$ as compared with $O(n)$ for the case of linear costs. Actually, recent research has revealed the surprising fact that the running time of this algorithm can be reduced to $O(n)$ in the important special case in which the concave cost is of the setup type illustrated in Figure 3.

Stationary Economic-Order-Interval Problem. When the demands and costs are stationary, it is natural to hope that it will be optimal to place orders at equally spaced points in time. Unfortunately, this is not possible because of the discrete number of opportunities one has to order. For example, if it is optimal to order twice in a three-period problem, then it is not possible to order at equally-spaced points in time because that would entail ordering midway through the second period which is not permitted. This difficulty does not arise if instead we consider the continuous-time approximation of the model in which there is a constant demand rate $s > 0$ per unit time and a storage cost $h > 0$ per unit stored per unit time. Also suppose that the scale economies in procurement is of the simplest type, viz., a setup cost $K > 0$ incurred each time an order is placed. Then it is natural to expect that an optimal schedule will still entail ordering only when inventory runs out as is the case in discrete time. The Invariance Theorem for network flows alluded to above implies that it is indeed optimal to order at equally-spaced points in time, though the schedule does depend on the length of the (finite) time horizon. If one is in fact interested in a relatively long time interval, then it is natural to consider the related problem of minimizing the long-run average cost per unit time. Since the demand rate and costs are stationary, it is not difficult to show, with the aid of a dynamic-programming argument, that there is a stationary optimal policy, i.e., the (order) intervals between successive orders are all equal, say, to $T > 0$. A variant of this stationary economic-order-interval problem was apparently first formulated and solved by Ford Harris in 1913 and is widely used in practice.

As we have seen above, it suffices to optimize over the class of policies in which one orders the quantity sT each time the inventory runs out. The long-run average cost incurred by such a policy is the sum of the average setup cost and the average holding cost per unit time. To compute this sum, observe that the average number of setups per unit time is $\frac{1}{T}$ and the average inventory on hand is $\frac{sT}{2}$. Thus the long-run average cost per unit time, denoted $A(T)$, is

$$A(T) = \frac{K}{T} + \frac{hsT}{2}.$$

Minimizing this expression with respect to T gives the celebrated *square root formula* for the *economic order interval* T^*, viz.,

$$T^* = \sqrt{\frac{2K}{hs}}.$$

Observe that T^* increases as the setup cost K increases, thereby reducing the average number of setups per unit time. Similarly, T^* decreases as the unit holding-cost rate h increases, thereby reducing the average inventory on hand. Also T^* decreases as the demand rate s increases. It is notable that the *economic order quantity* sT^* increases in proportion to the square root of the demand rate. Thus quadrupling the demand rate merely doubles the economic order quantity.

Uncertainty in Demand or Supply

Uncertainty in demand or supply provides an important motive for holding inventories when it is costly to adjust inventories quickly. Uncertainty is frequently in the level and/or timing of demand by customers whose needs are often not known in advance by the supply-chain manager. Uncertainty may also be in the supply, e.g., because of strikes, equipment breakdowns, or uncertain vendor procurement lead times. In each of these situations, it is often desirable to maintain inventories to buffer uncertain demands or supplies. The level of inventories that one maintains should reflect the extent of the uncertainty and the costs of over and under supply. The former may include storage and disposal costs, while the latter may include the costs of foregone opportunities for sales and/or perhaps penalties for failure to meet delivery schedules. The model (1)-(3) that we used to study dynamic supply-chain problems under certainty must be modified to handle the case in which demands are uncertain. In particular, we must modify either the stock-conservation equation (2) or the nonnegativity of inventories in (3) because it is possible that the stock on hand after ordering in a period will be insufficient to satisfy the uncertain demand arising in the period. We begin by discussing the single-period problem.

Spares Provisioning. Spares provisioning provides a simple example of the uncertain-demand motive for holding inventories. Spares are provided in many industries, e.g., when an automobile, aircraft, or electronics manufacturer produces spare parts for use in subsequent repairs of the equipment. The uncertain-demand motive for holding inventories is present in such problems because the future demand for spares is uncertain at the time the item is originally produced and it is often much cheaper to produce spares during the original production of the item than to do so subsequently.

As an example, consider a journal that must decide how many copies of each issue to print. At the time of its initial printing, the journal knows its total number of subscriptions, but must decide how many extra copies $y \geq 0$ to print in excess of its known subscriptions to provide for the uncertain demand $D \geq 0$ for back issues by future subscribers. The marginal cost $c > 0$ of printing each extra copy of the journal during the initial print run is low. The revenue from each copy of the journal that is subsequently purchased as a back issue is $r > c$, with r being much larger than c, say ten times larger. Since the demand for back issues is moderate in comparison with that for subscriptions and the fixed cost of a print run is high, it does not pay to reprint the journal subsequently. Thus, demand for back issues that can not be satisfied from the initial print run is lost.

The goal is to choose the number y of extra copies of the journal to print in the initial run that minimizes the expected net cost $G(y)$ of provisioning for back issues. Of course, $G(y) = cy - r\mathrm{E}(D \wedge y)$. Observe that $G(\,\cdot\,)$ is convex since linear functions are convex and concave, the minimum of two linear functions is concave, the negative of a concave function is convex, and sums of convex functions are convex. To simplify the exposition, assume that D has a continuous distribution $\Phi(\,\cdot\,)$. Then $G(\,\cdot\,)$ attains its minimum on the positive real line at y^0, say, and $\dot{G}(y^0) = 0$. This follows from the facts that $\dot{G}(y) = c - r\mathrm{E}(\frac{\partial}{\partial y}(D \wedge y)) = c - r[1 - \Phi(y)]$ is continuous with $\dot{G}(0) = c - r < 0$ and that $\lim_{y\to\infty} \dot{G}(y) = c > 0$. Thus y^0 satisfies

$$(8) \qquad 1 - \Phi(y^0) = \frac{c}{r}.$$

i.e., y^0 should be chosen so that the stock-out probability $1 - \Phi(y^0)$ is $\frac{c}{r}$. If $r = 10c$, then the desired stock-out probability is .1. Observe that the optimal stock-out probability falls, or equivalently the optimal starting stock y^0 rises, as c falls and as r rises. Also y^0 rises if the demand distribution Φ is replaced by another, say Φ', that is *stochastically larger*, i.e., $\Phi(z) \ge \Phi'(z)$ for all z. As we shall see in §7, this notion of stochastic order turns out to be very useful in many other problems as well because it is often the appropriate way of comparing the locations of two distributions.

Dynamic Supply Problem. Spares provisioning is an example of a single-period problem in which inventories left over at the end of a period have no value. It is more frequently the case that such inventories can be used to satisfy demands in subsequent periods. This leads to the dynamic supply problem discussed in §8. There we use dynamic programming to study the problem. For example, on letting $C_i(x)$ be the minimum expected cost in periods $i, \ldots, n$ given the initial stock of a single product on hand before ordering in period i is x, one gets the dynamic-programming recursion ($C_{n+1} \equiv 0$)

$$(9) \qquad C_i(x) = \min_{y \ge x} \Big\{ c(y - x) + G_i(y) + \mathrm{E}C_{i+1}(y - D_i) \Big\}, i = 1, \ldots, n$$

where $y \ge x$ is the starting stock on hand after ordering a nonnegative amount with immediate delivery in period i, $c(z)$ is the cost of ordering $z \ge 0$ units of the product in a period, $G_i(y)$ is the convex expected holding and shortage cost in period i when y is the starting stock in the period, D_i is the nonnegative demand in period i, and unsatisfied demands in period i are backordered. The form of the optimal starting stock $y(x)$ after ordering in period i as a function of the initial stock x in the period depends crucially on the form of the ordering cost function as we now discuss.

Scale Diseconomies in Supply. If $c(\,\cdot\,)$ is convex, then $y(x)$ is increasing in x and the order quantity $y(x) - x$ is decreasing in x. The intuitive rationale is that increasing the initial inventory to $x' \le y(x)$ reduces the marginal cost of ordering to $y(x)$ so there is an incentive to raise $y(x')$ above $y(x)$. But the size of the order at x' should not exceed that at x since that would raise the marginal cost of the order at x' above its level at x and so create an incentive to reduce $y(x')$.

Scale Economies in Supply. If $c(\,\cdot\,)$ is concave and of setup type, then $y(\,\cdot\,)$ is an (s,S) policy, i.e., there are numbers $s \le S$ such that it is optimal to order up to S if $x < s$, i.e., $y(x) = S$, while it is optimal not to order if $x \ge s$, i.e., $y(x) = x$. The intuitive rationale is that if it is optimal to order at all low enough initial inventory levels x, it is optimal to order to a level S independent of x. This is because once a decision to order is made, the setup cost is incurred and the remaining costs are independent of x. Choosing s less than S prevents too-frequent ordering and attendant costly setups. One can think of $S - s$ as a "minimum" economic-order-quantity.

Scale-Independent Supply Costs and Myopic Policies. If $c(\,\cdot\,)$ is linear, then $y(x) = x \vee y^0$ for some number y^0. To see why, observe that $c(\,\cdot\,)$ is both convex and concave of setup type with zero setup cost, so $y(\,\cdot\,)$ is both increasing and (s,S). But clearly $y(x)$ is increasing in a neighborhood of $x = s$ only if $s = S$. The economic rational for choosing $s = S$ is that the absence of setup costs implies the absence of incentives to avoid frequent orders. Thus $y(x) = x \vee y^0$ is (s,S) with $y^0 = s = S$. In this event, as we show in §9, the optimal policy in the first period of an n-period problem is often *myopic*, i.e., is independent of n, and so is optimal for the one-period problem. For example, this is so if also $G_i(y) = G(y)$ for all i and $c(\,\cdot\,) = 0$.[1] Then it is optimal to choose y^0 myopically, i.e., to minimize $G(\,\cdot\,)$. The rationale for this is that instead choosing y^0 so $G(y^0)$ exceeds $\min_y G(y)$ increases costs in the first period while at the same time producing no compensating reduction in future expected costs.

Temporal Increase in Demand with Costly Temporal Fluctuation in Supply

If it is costly to change the rate of supply, e.g., because of costs of hiring/firing, recruitment, training, etc., and if there is a temporal fluctuation in demand, then there may be a motive to carry inventories. For example, one may prefer to build up inventories in anticipation of a temporary increase in demand rather than incur the costs of increasing the size of the labor force and maintaining or reducing it later.

[1]By modifying the salvage value of stock and backorders left at the end of period n, it is possible to reduce a problem with a linear component cz of ordering costs to the same problem in which the linear component is zero. All that is required is to assume instead that $C_{n+1}(x) \equiv -cx$. This means that each unit of surplus stock at the end of period n is disposed of with a refund of the cost c and each unit of unfilled backorders at the end of period n is filled at the cost c. Then the n-period ordering cost totals $c(D_1 + \cdots + D_n - x)$ and is independent of the policy used. Thus there is no loss in generality in assuming that the linear component of the ordering cost is zero, i.e., $c = 0$.

Display

Retailers often display items in order to induce customers to purchase them. This provides a motive for such retailers to carry inventories.

Unavoidable

Sometimes inventories are unavoidable. For example, this may happen for a variety of reasons, e.g., return of sales, waiting lines, pollutants, pests, etc.

Summary

To sum up, we have discussed the following motives for carrying inventories:

- temporal increase in marginal cost of supplying demand, e.g., with linear costs and a temporal increase in unit supply cost or a temporal increase in demand and scale diseconomies in supply;
- scale economies in supply;
- uncertainty in demand or supply;
- temporal increase in demand with costly temporal fluctuation in supply;
- display; and
- unavoidable.

Among the methods that have proved useful to study these problems are

- network-flow and Leontief-substitution models of supply chains;
- linear-cost network flows and Leontief-substitution systems;
- graph theory;
- lattice programming, and substitutes and complements in convex-cost network flows;
- invariant convex-cost network flows;
- concave-cost network flows;
- dynamic programming, both deterministic and stochastic;
- complexity, i.e., analysis of running times;
- continuous-time approximation of discrete-time models;
- fast approximate solution with guaranteed effectiveness;
- stochastic order and
- total positivity.

This page is intentionally left blank.

2

Lattice Programming and Supply Chains: Comparison of Optima

[Ve65b], [TV68], [TV73a], [TV73b], [To76], [To78], [Ve89b], [VeFo]

1 INTRODUCTION

Supply-chain managers and other decision makers are often interested in understanding qualitatively how they should respond to changes in conditions. Indeed, some authorities have argued that obtaining understanding is *the* goal of models for decision making. They reason that such models are always approximate and so the numbers one obtains from numerical calculations with them are mainly useful to obtain understanding—not answers. Even those who do not accept this viewpoint—and many do not—agree that the development of understanding of situations is an important goal of modeling.

Since data gathering and computation are expensive—particularly for large scale optimization problems—the question arises whether it is possible to develop a theory of optimization that would provide a qualitative understanding of the solution of an optimization problem without data gathering or computation. The answer is it is, and we shall do so in the remainder of this section.

To that end, it may be useful first to give a few examples of qualitative questions arising in supply-chain management. Similar examples abound in other fields as well.

• Is it optimal to increase or decrease production of a single product in a given period if in some (possibly different) period demand increases? Production or storage costs increase? Production or storage capacity increases?

• Does increasing the initial stock of one product imply that it is optimal to increase or decrease production of other products in a given period?

• Do rising selling prices imply higher or lower inventories?

• Is it optimal to raise or lower selling prices (resp., current production) as labor costs rise?

• Do optimal production levels change more or less than changes in demand? Production/storage capacity? Initial inventories?

• Do changes in demand in a (future) period have a larger or smaller impact on optimal current production than do like changes in demand in subsequent periods?

At first glance, it might seem that the above questions have obvious answers. For example, it is plausible to expect that an increase in forecasted future demand would lead one to increase planned production in each period before the increase in forecasted demand. But this is false in general. To see why, suppose there are concave production costs and linear storage costs. Then, at low enough demand levels, it is optimal to produce enough in the first period to satisfy all subsequent demands in order to take advantage of the scale-economies in production. On the other hand, at high enough demand levels, it is optimal to produce only enough in each period to satisfy the demand in that period in order to avoid large storage costs. Thus, the optimal production level in the first period is not increasing in future demands. Having said this, one may ask whether the optimal production level in the first period is ever increasing in subsequent demands. As we shall show in the sequel, the answer is that it is, provided that all production and storage costs are convex.

This example illustrates that answers to the above questions cannot be given unequivocally without additional information about the structure of the problem—in this case, the form of the cost functions. Moreover, this is the usual state of affairs. Indeed, the answer to each of the above questions is that it depends.

The goal of this section is to develop a qualitative theory of optimization, called *lattice programming* (for reasons that will soon become clear), to answer such questions. As we shall see, these questions can usually be reduced to that of determining when some optimal solution $s = s_t^0 \in \Re^n$ of the mathematical program

$$\min_{s \in L_t} f(s,t) \tag{1}$$

is increasing in the parameter $t \in \Re^m$. Moreover, we show that this will be the case provided that

• $L \equiv \{(s,t) : s \in L_t\}$ is a "sublattice" of $\Re^{n+m}$,

• f is "subadditive" on L, and

• L_t is compact and $f(\cdot\,,t)$ is lower semicontinuous thereon for each t.

In order to appreciate the usefulness of this result, we shall need to define and characterize sublattices of Euclidean space and subadditive functions thereon. Before doing this, a few broader comments are in order.

How can the result be used?

There are several ways in which the result can be used. It can give a qualitative understanding of the solution to a problem without computation. It can be used to simplify needed computations. And it can be used to design simple approximations of optimal decision rules that have many properties of those rules in situations where they are prohibitively expensive to calculate.

Convex vs. Lattice Programming

It is of interest to compare the goals and results of convex programming with those of lattice programming. Most of conventional mathematical programming is concerned with conditions assuring that local optimality implies global optimality, and for this reason is rooted in the theory of convex sets. By contrast, we are here concerned with the order of optimal solutions and so are led to a development based on lattices.

As one reflection of the difference between these theories, consider what happens when some variables are required to be integer, e.g., where one must order in multiples of a given batch size like a case, a box, etc. This destroys convexity, but preserves sublattices. Thus the presence of integrality constraints enormously complicates the results of convex programming. By contrast, the monotonicity results of lattice programming carry over without change to their integer counterparts.

Applicability

In the sequel we shall develop a portion of the theory of lattice programming and give its applications to supply chains. The theory also has broad applicability to other fields of operations research, e.g., reliability, queueing, marketing, distribution, mining, networks, etc., and to other disciplines like statistics and economics.

2 SUBLATTICES IN $\Re^n$

Upper and Lower Bounds

The set $\Re^n$ of n-tuples of real numbers is partially ordered by the usual less-than-or-equal-to relation $\le$, i.e., $r \le s$ in $\Re^n$ if $s - r \ge 0$. Call $s \in \Re^n$ a *lower bound* (resp., *upper bound*) of a subset S of $\Re^n$ if $s \le r$ (resp., $s \ge r$) for all $r \in S$. If $L \subseteq S$, call $s \in \Re^n$ the *greatest lower bound* (resp., *least upper bound*) of L in S if $s \in S$, s is a lower (resp., upper) bound of L and if $r \le s$ (resp., $r \ge s$) for every lower (resp., upper) bound r of L in S. Figure 4 illustrates these concepts where S is the plane $\Re^2$.

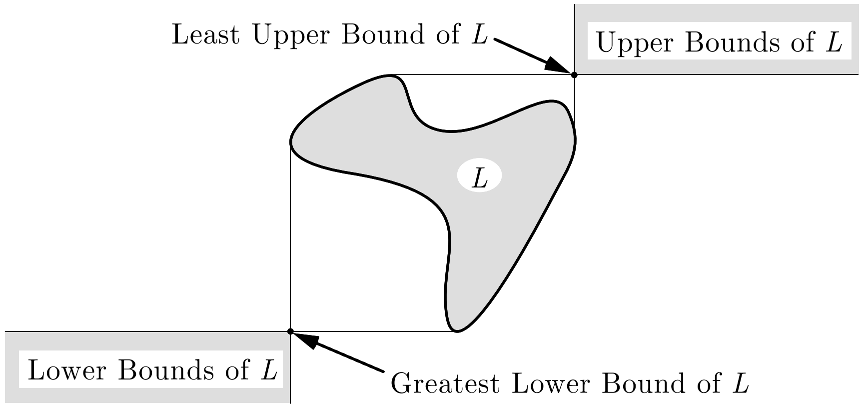


Figure 1. Upper and Lower Bounds of a Set

Sublattices

Call a subset L of a set $S \subseteq \Re^n$ a *sublattice* of S if every pair r, s of points in L has a greatest lower bound in S, denoted $r \wedge s$ and called their *meet*, a least upper bound in S, denoted $r \vee s$ and called their *join*, and both the meet and join are in L. If L is a sublattice of itself, call L a *lattice*.

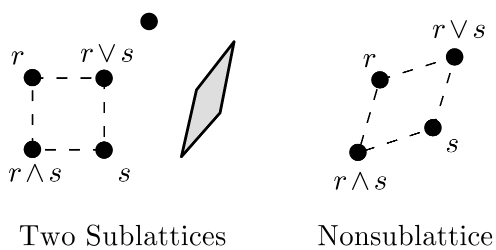


Figure 2a. Lattices in the Plane

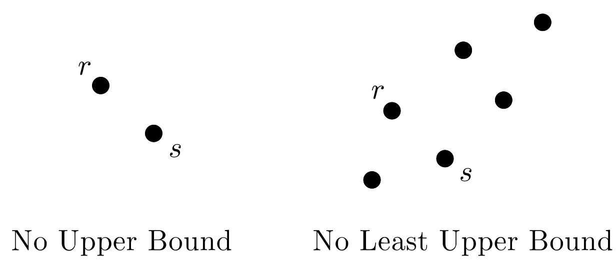


Figure 2b. Nonlattices in the Plane

Example 1. Chain. A *chain*, i.e., a set $L \subseteq \Re^n$ for which $r, s \in L$ imply either $r \leq s$ or $r \geq s$, is a sublattice of $\Re^n$. Evidently, $\Re$ is a chain, as is any subset thereof, e.g., the integers.

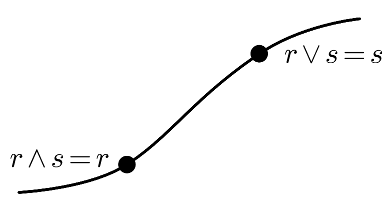


Figure 3. A Chain

Example 2. Products of Lattices and Sublattices. The (direct) product of a family of lattices (resp., sublattices) is a lattice (resp., sublattice). Meets and joins in the product are taken coordinate-wise. In particular, $\Re^n$ is a lattice with $r, s \in \Re^n$ implying $r \wedge s = \min(r, s)$ and $r \vee s = \max(r, s)$. The set of integer vectors in $\Re^n$ is a sublattice of $\Re^n$. Moreover, if $S \subseteq \Re^m$ is a lattice (resp., sublattice), then so is the *cylinder* $S \times \Re^n$ in $\Re^{m+n}$ with *base* S.

Example 3. Sublattice-Preserving Functions. Call a function f from a lattice $S \subseteq \Re^n$ to $\Re^m$ *sublattice preserving* if $f(r \wedge s) = f(r) \wedge f(s)$ and $f(r \vee s) = f(r) \vee f(s)$ for all $r, s \in S$ with the meet and join of $f(r)$ and $f(s)$ being taken in $\Re^m$. Then the image $f(L)$ of every sublattice L of S under f is a sublattice of $\Re^m$. If S is a product of n chains $S_1, \ldots, S_n$ of real numbers, then f is sublattice preserving if $f(s) = (f_1(s_{\tau_1}), \ldots, f_m(s_{\tau_m}))$ for $s \in S$ for some increasing real-valued functions $f_1, \ldots, f_m$ on $S_{\tau_1}, \ldots, S_{\tau_m}$ respectively and some function $\tau_.$ from the set $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$. Thus f is sublattice preserving if the action of f permutes, deletes, duplicates, stretches or shrinks (not necessarily uniformly) the coordinates. Indeed, it can be shown that these are the only sublattice-preserving functions.

Example 4. Sections and Projections of Sublattices. If $S \subseteq \Re^m$, $T \subseteq \Re^n$, and L is a sublattice of $S \times T$, the *section* $L_t \equiv \{s \in S : (s,t) \in L\}$ of L at $t \in T$ and the *projection* $\pi_S L \equiv \bigcup_{t \in T} L_t$ of L on S are sublattices of S as Figure 4 illustrates.

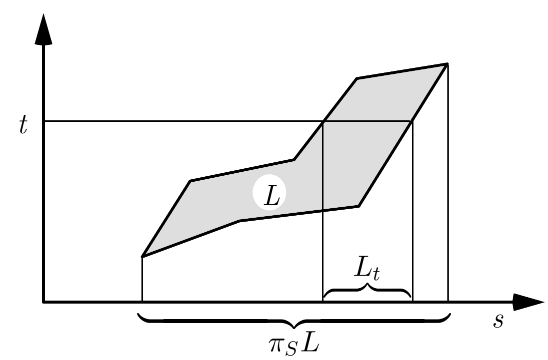


Figure 4. Sections and Projections of a Sublattice

Example 5. Intersections of Sublattices. The intersection of any family of sublattices of a lattice is a sublattice thereof. In particular the integer vectors in a sublattice of $\Re^n$ also form a sublattice thereof.

Example 6. Sublattices of Finite Products of Chains. If S is a finite product of chains $S_1, \ldots, S_n$, $f_{ij}(s_i, s_j)$ is a $+\infty$ or real-valued function and, for $i \neq j$, $f_{ij}(s_i, s_j)$ is decreasing in s_i and increasing in s_j, then the set of vectors $s \in S$ that satisfy the system of inequalities

(2) $$f_{ij}(s_i, s_j) \le 0 \text{ for all } 1 \le i, j \le n$$

is a sublattice of S. Indeed, Appendix 1 shows that every sublattice of S arises in this way.

To see why the set of solutions of (2) is a sublattice, observe that the set L_{ii} of vectors $s \in S$ that satisfy $f_{ii}(s_i, s_i) \le 0$ is a cylinder with base a chain in S_i, so L_{ii} a sublattice of S. Similarly, the set L_{ij} of vectors $s \in S$ that satisfy $f_{ij}(s_i, s_j) \le 0$, where $j \neq i$, is a cylinder with base $B_{ij} \equiv$

$\{(s_i, s_j) \in S_i \times S_j \mid f_{ij}(s_i, s_j) \le 0\}$. Also, as we show below, B_{ij} is a sublattice of $S_i \times S_j$, so L_{ij} is a sublattice of S. Now since the set of vectors $s \in S$ that satisfy the system (2) is the intersection $L \equiv \bigcap_{i,j} L_{ij}$ of sets L_{ij} each of which is a sublattice, the set L is a sublattice.

It remains to show that B_{ij} is a sublattice as Figure 5 illustrates. To that end, suppose (s_i, s_j), $(\sigma_i, \sigma_j) \in B_{ij}$. Without loss of generality, assume that $s_i \le \sigma_i$ and $s_j \ge \sigma_j$. Now the join and meet of (s_i, s_j) and (σ_i, σ_j) are respectively (σ_i, s_j) and (s_i, σ_j), so because $f_{ij}(\,\cdot\,,\,\cdot\,)$ is decreasing in the first variable and increasing in the second, it follows that $f_{ij}(\sigma_i, s_j) \vee f_{ij}(s_i, \sigma_j) \le f_{ij}(s_i, s_j) \le 0$. Thus, $(\sigma_i, s_j),\ (s_i, \sigma_j) \in B_{ij}$, so B_{ij} is a sublattice of $S_i \times S_j$ as claimed.

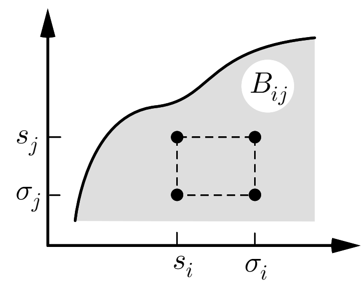


Figure 5. B_{ij} a Sublattice

Polyhedral Sublattices. One special case of (2) arises where the f_{ij} are *affine*, i.e., linear plus a constant. In that event it follows that the set of vectors $s \in \Re^n$ that satisfy the system of linear inequalities

(3) $$As \le b,$$

where A is an $m \times n$ matrix and b is an m-column vector, is a sublattice of $\Re^n$ provided that each row of A has at most one positive and at most one negative element. Constraints of this type are *dual-weighted-network-flow* constraints. Indeed Appendix 1 shows that every polyhedral sublattice of $\Re^n$ arises in this way. Ordinary dual-network-flow constraints are the special case in which each row of A has at most one $+1$, at most one -1 and zeros elsewhere.

Least and Greatest Elements

Every finite lattice has a least (resp., greatest) element, viz., the meet (resp., join) of its elements. This extreme element may be constructed by taking the meet (resp., join) of two elements, then the meet (resp., join) of that element with a third element, etc. Since the lattice is finite, the process terminates in finitely many steps with the desired least (resp., greatest) element. This process breaks down in infinite lattices, and indeed they need not have least or greatest elements—for example, that is so of $(0, 1)$ and $\Re$—unless appropriate closedness and boundedness hypotheses are imposed.

PROPOSITION 1. Lattices with Least and Greatest Elements. *If L is a nonempty closed lattice in $\Re^n$ that is bounded below* (*resp., above*)*, then L has a least* (*resp., greatest*) *element.*

Proof. It suffices to establish the result reading without parentheses. Choose $m \in L$. Now L has a lower bound l ($\le m$). Since $f(s) \equiv \sum_i s_i$ is a continuous function on the compact set $L \cap [l,m]$, f assumes its minimum thereon at, say, $r \in L$. If r is not the least element of L, there is an $s \in L$ such that $r \not\le s$. Then $r > r \wedge s \in L$, so $f(r \wedge s) < f(r)$, contradicting the fact r minimizes f over $L \cap [l,m]$. ∎

3 ASCENDING MULTI-FUNCTIONS

In order to discover when an optimization problem has an optimal solution that is increasing in a problem parameter, it turns out to be convenient to consider first when the set of all optimal solutions is "increasing" in the parameter. For this purpose we need a suitable notion of "increasing" for a *multi-function*, i.e., a set-valued function $S_.$ from a set $T \subseteq \Re^m$ into the set 2^S of all nonempty subsets of a set $S \subseteq \Re^n$.

To be useful, it is necessary that the definition of an "increasing" multi-function $S_.$ has the following two properties. First, it must have a *selection* (i.e., a function $s_.$ from T to S for which $s_t \in S_t$ for all $t \in T$) that is increasing on T (in the usual sense that $t < \tau$ in T implies $s_t \le s_\tau$) whenever $S_.$ is *compact-sublattice valued* (i.e., S_t is a compact sublattice of S for each $t \in T$). Incidentally, this condition implies that a single-valued $S_.$ is "increasing" on T if and only if its unique selection is increasing on T. Second, the set of optimal solutions must be "increasing" in the desired parameter for a broad class of optimization problems.

It turns out that the desired notion of "increasing" is that $S_.$ be *ascending*, i.e., $t < \tau$ in T, $s \in S_t$ and $\sigma \in S_\tau$ imply that the meet $s \wedge \sigma$ and join $s \vee \sigma$ of s and σ exist in S, $s \wedge \sigma \in S_t$ and $s \vee \sigma \in S_\tau$. Figure 6 illustrates this concept. That this definition has the two properties described above is established in Theorems 2 and 8.

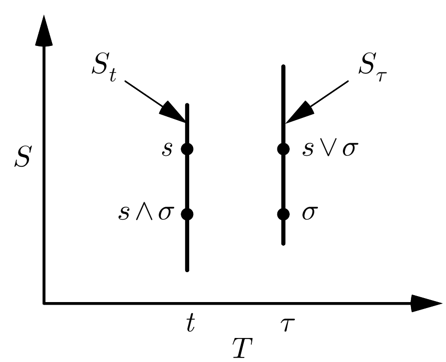


Figure 6. An Ascending Multi-Function

In applications it is desirable that a compact-sublattice-valued ascending multi-function not only have an increasing selection, but also that such a selection be useful and easy to find. One such class of selections can be described with the aid of the following definition.

Linearly order the first n positive integers, and let A be a subset thereof. We say $s \in \Re^n$ is *A-lexicographically smaller than* $\sigma \in \Re^n$, written $s \leq_A \sigma$, if $s = \sigma$ or if the smallest index i (in the linear order) with $s_i \neq \sigma_i$ is such that $s_i < \sigma_i$ if $i \in A$ and $s_i > \sigma_i$ if $i \notin A$. Observe that if the linear order of the first n positive integers is the natural one and if $A = \{1, \ldots, n\}$ (resp., $\emptyset$), the A-lexicographically-smaller-than relation reduces to the ordinary lexicographically-smaller- (resp., larger-) than relation.

THEOREM 2. Increasing Selections from Ascending Multi-Functions. *A compact-sublattice-valued ascending multi-function has increasing least, greatest and A-lexicographically-least selections.*

Proof. Suppose $S_{\cdot} : T \to 2^S$ is ascending on T and is compact-sublattice-valued. Since $S_{\cdot}$ is compact valued, it has an A-lexicographically least selection $s_{\cdot}$, say. Moreover, because $S_{\cdot}$ is also sublattice valued, it has a least (resp., greatest) selection by Proposition 1, and that selection is the A-lexicographically least selection when $A = \{1, \ldots, n\}$ (resp., $\emptyset$). Thus it remains only to show that $s_{\cdot}$ is increasing on T for arbitrary A. To that end, suppose $t < \tau$ in T. Then since $S_{\cdot}$ is ascending, $s_t \wedge s_\tau \in S_t$ and $s_t \vee s_\tau \in S_\tau$. Now if $s_t \not\leq s_\tau$, there is a smallest integer i with $s_{ti} > s_{\tau i}$. If $i \in A$, then $s_t \not\leq_A s_t \wedge s_\tau$, while if $i \notin A$, then $s_\tau \not\leq_A s_t \vee s_\tau$. This contradicts the fact s_t and s_τ are respectively the A-lexicographically least elements of S_t and S_τ. ∎

Ascending multi-functions also arise in another natural way, viz., as sections of sublattices.

LEMMA 3. Sections of Sublattices are Ascending. *If L is a sublattice of $S \times T \subseteq \Re^n$, then the section L_t of L at t is ascending in t on $\pi_T L$.*

Proof. Suppose $t < \tau$ in $\pi_T L$, $s \in L_t$ and $\sigma \in L_\tau$. Since L is a sublattice of $S \times T$, $(s \wedge \sigma, t) = (s,t) \wedge (\sigma, \tau) \in L$ so $s \wedge \sigma \in L_t$. Similarly, $s \vee \sigma \in L_\tau$. ∎

4 ADDITIVE, SUBADDITIVE and SUPERADDITIVE FUNCTIONS

Call a $+\infty$ or real-valued function f on a lattice L in $\Re^n$ *subadditive* if

$$f(r \wedge s) + f(r \vee s) \leq f(r) + f(s) \text{ for all } r, s \in L.$$

Similarly, call f *superadditive* if $-f$ is subadditive. The class of subadditive functions is closed under addition and multiplication by nonnegative numbers, and thus is a convex cone. Moreover, the pointwise limit of any sequence of subadditive functions is subadditive if the limit function doesn't assume the value $-\infty$ anywhere. Also, if g is sublattice preserving from a lattice K to a lattice L and f is subadditive on L, then the composite function $f \circ g$ is subadditive on K.

The *effective domain* of a subadditive function, i.e., the subset on which it is finite, is a sublattice. Thus the problem of minimizing a $+\infty$ or real-valued subadditive function on $\Re^n$ is equivalent to minimizing a real-valued subadditive function on a sublattice of $\Re^n$. This fact leads to the following simple characterization of sublattices of $\Re^n$ in terms of subadditive functions. The *indicator function* of a set L in $\Re^n$, i.e., the function $\delta_L(\,\cdot\,)$ whose value is zero on L and $+\infty$ otherwise, is subadditive on $\Re^n$ if and only if L is a sublattice of $\Re^n$.

Characterization on Finite Products of Chains

Every $+\infty$ or real-valued function on a chain is subadditive. But that is not so for functions on a product of two or more chains. However, there is an important characterization of real-valued subadditive functions on products of $n \ge 2$ chains. We begin with the case $n = 2$.

Suppose S and T are chains and f is real-valued on $S \times T$. Two characterizations of subadditivity of f are immediate from Figure 7. One is in terms of the first differences of f and the other the second differences thereof. In particular, f is subadditive on $S \times T$ if and only if the first difference of f in either variable is decreasing in the other variable, i.e., either

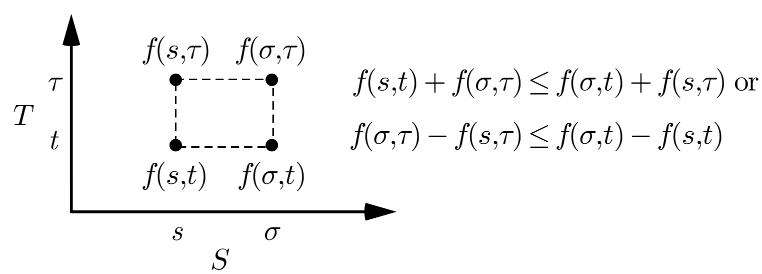


Figure 7. Subadditivity in the Plane

(4a) $\qquad \Delta_1 f(s,t) \equiv f(\sigma,t) - f(s,t)$ is decreasing in t on T for all $s < \sigma$ in S

or

(4b) $\qquad \Delta_2 f(s,t) \equiv f(s,\tau) - f(s,t)$ is decreasing in s on S for all $t < \tau$ in T.

Alternately, f is subadditive on $S \times T$ if and only if the mixed second difference of f is nonpositive on $S \times T$, i.e.,

(5) $\qquad \Delta_{12} f(s,t) \le 0$ for all $s < \sigma$ in S and $t < \tau$ in T

where

$$\Delta_{12} f(s,t) \equiv f(\sigma,\tau) - f(\sigma,t) - f(s,\tau) + f(s,t).$$

The above characterizations take on sharper forms when f is suitably differentiable. In particular, if S is an open interval of real numbers and $f(\,\cdot\,,t)$ is continuously differentiable on S for each $t \in T$, then f is subadditive on $S \times T$ if and only if

(6a) $$D_1 f(s,t) \text{ is decreasing in } t \text{ on } T \text{ for each } s \in S.$$

Symmetrically, if T is an open interval of real numbers and $f(s,\,\cdot\,)$ is continuously differentiable on T for each $s \in S$, then f is subadditive on $S \times T$ if and only if

(6b) $$D_2 f(s,t) \text{ is decreasing in } s \text{ on } S \text{ for each } t \in T.$$

To see that, for example, (4a) and (6a) are equivalent, observe that on $S \times T$,

(7) $$\Delta_1 f(s,t) = \int_s^\sigma D_1 f(u,t)du,$$

so (6a) implies (4a). Conversely, if (6a) does not hold, there is an $s \in S$ and $t < \tau$ in T such that $D_1 f(s,t) < D_1 f(s,\tau)$. Then $D_1 f(\,\cdot\,,t) < D_1 f(\,\cdot\,,\tau)$ on $[s,\sigma]$ for small enough $\sigma > s$. Hence from (7), $\Delta_1 f(s,t) < \Delta_1 f(s,\tau)$, i.e., (4$a$) does not hold, which establishes the equivalence of (4a) and (6a).

Similarly, if S and T are open intervals of real numbers and f is twice continuously differentiable on $S \times T$, then f is subadditive thereon if and only if

(8) $$D_{12} f(s,t) \le 0 \text{ on } S \times T.$$

To see this observe that on $S \times T$,

(9) $$\Delta_{12} f(s,t) = \int_t^\tau \int_s^\sigma D_{12} f(u,v)dudv.$$

In view of (9), (8) implies (5). Conversely, if (8) does not hold for some s,t, then $D_{12}f$ is positive on $[s,\sigma] \times [t,\tau]$ for small enough $\sigma > s$ and $\tau > t$. Then by (9), (5) does not hold, which establishes the equivalence of (5) and (9).

The importance of the above characterizations of real-valued subadditive functions on a product of two chains is that real-valued subadditive functions on a finite product of chains have a characterization in terms of them as the next result shows. To state the result requires a definition. Call a $+\infty$ or real-valued function on a finite product of chains *pairwise subadditive* if it is subadditive on each pair of chains for all fixed values of the other coordinates.

THEOREM 4. Equivalence of Subadditivity and Pairwise Subadditivity. *A real-valued function on a finite product of chains is subadditive if and only if it is pairwise subadditive.*

Proof. It suffices to show that a real-valued pairwise-subadditive function f on a product S of n chains is subadditive. The proof is by induction on n. The result is trivial for $n = 1, 2$. Assume it holds for all positive integers up to $n-1 \geq 2$, and consider n. Suppose $s, \sigma \in S$ are incomparable. By possibly relabeling and permuting variables, we may assume that $s = (a,b,c)$, $\sigma = (\alpha,\beta,\gamma)$, $(a,b) > (\alpha,\beta)$, and $c < \gamma$. Now by the induction hypothesis, $f(\,\cdot\,,b,\,\cdot\,)$ and $f(\alpha,\,\cdot\,,\,\cdot\,)$ are subadditive so

$$f(\alpha,b,c) + f(a,b,\gamma) \leq f(a,b,c) + f(\alpha,b,\gamma)$$

and

$$f(\alpha,\beta,c) + f(\alpha,b,\gamma) \leq f(\alpha,\beta,\gamma) + f(\alpha,b,c).$$

Adding these inequalities and canceling common terms (all finite) yields

$$f(s \wedge \sigma) + f(s \vee \sigma) \leq f(s) + f(\sigma). \blacksquare$$

By combining this result with the equivalence of (4) and (6) for a continuously differentiable function, we obtain the following characterization of subadditivity in terms of its *gradient*, i.e., the vector of partial derivatives of the function.

COROLLARY 5. Continuously Differentiable Subadditive Functions. *A continuously differentiable function on a finite product of open intervals of real numbers is subadditive if and only if its partial derivative in each variable is decreasing in the other variables.*

By combining Theorem 4 with the equivalence of (5) and (8) for a twice continuously differentiable function, we obtain the following characterization of subadditivity in terms of its *Hessian*, i.e., the matrix of mixed partial derivatives of the function.

COROLLARY 6. Twice Continuously Differentiable Subadditive Functions. *A twice continuously differentiable function on a finite product of open intervals of real numbers is subadditive if and only if the off-diagonal elements of its Hessian are nonpositive.*

Additive Functions

Call a real-valued function f on a lattice L in $\Re^n$ *additive* if

$$f(r \wedge s) + f(r \vee s) = f(r) + f(s) \text{ for all } r, s \in L.$$

Evidently, f is additive if and only if it is both subadditive and superadditive on L. The next result is a representation theorem for additive functions on a finite product of chains.

THEOREM 7. Representation of Additive Functions. *A real-valued function f on a product S of n chains $S_1, \ldots, S_n$ is additive if and only if there exist real-valued functions $f_1, \ldots, f_n$ on the n chains for which $f(s) = \sum_1^n f_i(s_i)$ for all $s \in S$.*

Real-Valued Subadditive Functions f on a Product of n Sets of Real Numbers

We now apply these results to give a few examples of subadditive functions most of which will occur frequently in the sequel. The following are important examples of subadditive functions $f(s)$ for $s = (s_i)$ in a product of n sets of real numbers.

- The *negative product* function $f(s) = -s_1 \cdots s_n$ provided either $n = 2$ or $s \geq 0$.
- The *quadratic form* $f(s) = s^{\mathrm{T}} A s$ with A symmetric if and only if the off-diagonal elements of A are nonpositive.
- The *maximum* function $f(s) = s_1 \vee \cdots \vee s_n$.
- The function $f(s_1, s_2) = g(s_1 - s_2)$ if and only if g is convex on $\Re$.
- The function $f(s_1, s_2) = g(s_1, -s_2)$ (resp., $g(-s_1, s_2)$) if and only if g is superadditive.
- The negative of the joint distribution function of n random variables.

Also remember that all of the above examples remain subadditive if we replace each variable s_i by an increasing function thereof.

The easiest way to check for subadditivity of f is to do so under the hypothesis that f is twice continuously differentiable, and then, if warranted, in the general case. For example, by Corollary 6, $f(s) = g(s_1 - s_2)$ is subadditive if and only if $D_{12} f = -D^2 g$ is nonpositive, or equivalently g is convex.

5 MINIMIZING SUBADDITIVE FUNCTIONS ON SUBLATTICES

We now bring together the ideas of sublattices, ascending multi-functions and subadditivity in the qualitative study of optimization problems. To this end, suppose $S \subseteq \Re^n$, $T \subseteq \Re^m$, and f is a $+\infty$ or real-valued function on $S \times T$. Let L be the effective domain of f, g be the *projection* of f defined by

$$g(t) \equiv \inf_{s \in S} f(s,t), t \in \pi_T L,$$

and L_t^o denote the *optimal-reply set at* t, i.e., the set of $s \in S$ for which $g(t) = f(s,t)$. The next result gives conditions on f that assure that the *optimal-reply multi-function* $L_{\cdot}^o$ is ascending on $\pi_T L$ and has increasing (optimal) selections.

THEOREM 8. Increasing Optimal Selections. *If S and T are lattices and f is subadditive on $S \times T$, then the optimal reply set L_t^o is a sublattice and is ascending on the set of $t \in \pi_T L$ for which L_t^o is nonempty. If also for each $t \in \pi_T L$, $f(\,\cdot\,,t)$ is lower-semicontinuous with some level set being nonempty and bounded, then L_t^o is nonempty and compact for each such t, and $L_{\cdot}^o$ has increasing least, greatest and A-lexicographically-least selections.*

Proof. Suppose that $t \le \tau$ in $\pi_T L$, $s \in L_t^o$ and $\sigma \in L_\tau^o$. Since f is subadditive, its effective domain L is a sublattice and so contains (s,t), (σ,τ), $(s \wedge \sigma, t)$, and $(s \vee \sigma, \tau)$. Thus, because f is finite and subadditive on L,

$$0 \ge f(s,t) - f(s \wedge \sigma, t) \ge f(s \vee \sigma, \tau) - f(\sigma,\tau) \ge 0,$$

so equality occurs throughout. Hence $s \wedge \sigma \in L_t^o$ and $s \vee \sigma \in L_\tau^o$. Thus $L_\cdot^o$ is ascending and sublattice valued on the set of t in $\pi_T L$ for which L_t^o is nonempty. Now since $f(\,\cdot\,, t)$ is lower-semicontinuous with some level set being nonempty and bounded for each $t \in \pi_T L$, L_t^o is nonempty and compact for each such t. Thus, by Theorem 2, $L_\cdot^o$ has increasing least, greatest and A-lexicographically-least selections. ∎

The next result is particularly useful in dynamic-programming applications because it assures that subadditivity is preserved under minimization.

THEOREM 9. Projections of Subadditive Functions. *If S and T are lattices, f is subadditive on $S \times T$, and the projection g of f does not equal $-\infty$ anywhere, then g is subadditive on $\pi_T L$.*

Proof. Suppose $t, \tau \in \pi_T L$ and $s, \sigma \in S$. Then since f is subadditive,

$$g(t \wedge \tau) + g(t \vee \tau) \le f(s \wedge \sigma, t \wedge \tau) + f(s \vee \sigma, t \vee \tau) \le f(s,t) + f(\sigma,\tau).$$

Now take infima over $s, \sigma \in S$. ∎

As we shall see, the above results will have myriad applications to many different problems throughout this course. In many of these applications, f is instead real-valued and subadditive on a sublattice L of $S \times T$, and one seeks an s that minimizes $f(s,t)$ over the section L_t of L at $t \in \pi_T L$. This problem is easily reduced to that discussed in the above theorems by extending the definition of f to $S \times T$ by letting f be $+\infty$ on $(S \times T) \setminus L$. Then L is the effective domain of f, f is subadditive on $S \times T$, and one considers instead the equivalent problem of seeking s that minimizes $f(\,\cdot\,, t)$ over S where $t \in T$.

6 APPLICATION TO OPTIMAL DYNAMIC PRODUCTION PLANNING

As an illustration of the application of the above results, consider the problem that a production manager faces in scheduling production of a single product to meet a sales forecast over n periods at minimum total cost. The manager forecasts that the vector of cumulative sales for a single product in the next n periods $1, \ldots, n$ will be $S = (S_1, \ldots, S_n)$, i.e., S_i is the forecast of total sales during the periods $1, \ldots, i$ for $i = 1, \ldots, n$. There is a continuous convex cost $c_i(z)$ of producing (resp., $h_i(z)$ of storing) $z \ge 0$ units of the product in (resp., at the end of) period $i = 1, \dots, n$. There is no initial inventory and none should remain at the end of period n. The manager seeks a vector $X = (X_1, \dots, X_n)$ of cumulative production levels in periods $1, \dots, n$ that minimizes the n-period cost

$$C(X,S) \equiv \sum_{i=1}^{n} [c_i(X_i - X_{i-1}) + h_i(X_i - S_i)] \tag{10}$$

subject to the constraints

$$X_i \ge X_{i-1} \text{ for } i = 1, \dots, n, X_0 = 0, \tag{11a}$$

$$X \ge S \text{ and } X_n = S_n. \tag{11b}$$

Variation of Optimal Cumulative Production with Cumulative Sales

Since the manager is not certain that her sales forecast—particularly the timing thereof—is correct, she is interested in examining how changes in the magnitude and timing of her sales forecast will affect the optimal cumulative production schedule. In this connection, she notes that an increase δ in cumulative sales forecast in a period $i < n$ is equivalent to shifting δ units of the sales forecast from period $i+1$ to period i with no change in the total n-period sales forecast. With this in mind, she poses the following questions.

- Does optimal cumulative production increase with the cumulative sales forecast?
- If so, does optimal cumulative production increase at a slower rate than the cumulative sales forecast in a period?
- Does the incremental cost of optimally satisfying an increase in the cumulative sales forecast in one period fall as the cumulative sales forecast in other periods rise?

Optimal Cumulative Production Rises with the Cumulative Sales Forecast

To answer the first question, observe that the set L of pairs (X, S) satisfying the constraints $(11a)$ and $(11b)$ is a polyhedral sublattice of $\Re^{2n}$ by Example 6. Since the c_i and h_i are convex, a convex function of the difference of two variables is subadditive, and sums of subadditive functions are subadditive, C is subadditive. Moreover, since the c_i and h_i are continuous, so is C. The cumulative sales forecast vector S is feasible if and only if it lies in the projection L' of L on the set of all such vectors, viz., the polyhedral sublattice described by the inequalities $S_n \ge 0$ and $S_n \ge S_i$ for all $1 \le i < n$. Thus by the Increasing-Optimal-Selections Theorem, there is a least $X = X(S)$ that minimizes $C(X, S)$ subject to $(X, S) \in L$, and $X(S)$ is increasing in S on L', i.e., the optimal cumulative production in each period is increasing in the cumulative sales forecast in every period. As one application of this result, observe from the remarks in the preceding paragraph that shifting the sales forecast to earlier periods has the effect of increasing optimal cumulative production.

Optimal Cumulative Production Rises Slower than the Cumulative Sales Forecast

Now consider the second question, viz., is $S_j 1 - X(S)$ increasing in S_j where j is fixed and 1 is here an n-vector of ones? To that end, make the change of variables $Y = S_j 1 - X$ and show that a suitable optimal Y is increasing in S_j. The transformed problem becomes that of choosing $Y = (Y_i)$ to minimize

$$(10)' \qquad C_j(Y, S_j) \equiv \sum_{i=1}^{n} [c_i(Y_{i-1} - Y_i) + h_i(S_j - S_i - Y_i)]$$

subject to

$$(11a)' \qquad Y_i \le Y_{i-1} \text{ for } i = 1, \ldots, n, Y_0 = S_j,$$

$$(11b)' \qquad S_j - Y_i \ge S_i \text{ for } i \ne j \text{ and } Y_i \le 0 \text{ for } i = j,$$

with equality occurring in $(11b)'$ whenever $i = n$. Now observe that the set L_j of pairs (Y, S_j) satisfying $(11a)'$ and $(11b)'$ is a compact polyhedral sublattice with the S_i fixed for all $i \ne j$. Also since the c_i and h_i are continuous and convex, the total cost C_j is continuous and subadditive. Thus by the Increasing-Optimal-Selections Theorem, there is a greatest $Y = Y(S_j)$ minimizing $C_j(\,\cdot\,, S_j)$ subject to $(Y, S_j) \in L_j$, and $Y(S_j)$ is increasing in S_j for $S \in L'$. Incidentally, we chose the greatest $Y = S_j 1 - X$ here because it corresponds to the least X, so $Y(S_j) = S_j 1 - X(S)$.

Subadditivity of Minimum Cost in Cumulative Sales Forecast Vector

The third question has an affirmative answer by observing from the Projections-of Subadditive-Functions Theorem 9 that the minimum $C(S)$ of $C(X, S)$ over the set of X with $(X, S) \in L$ is subadditive in S on L'. Thus by the Equivalence-of-Subadditivity-and-Pairwise-Subadditivity Theorem 4 and (4a), the incremental cost $\Delta_i C(S)$ of optimally satisfying an increase in the cumulative sales forecast in period i, say, falls as the cumulative sales forecast in other periods rise.

This page is intentionally left blank.

3

Noncooperative and Cooperative Games: Competition and Cooperation in Supply Chains

[Na51], [Ta55], [DS63], [To70], [Ve74], [Ow75], [To79], [MR90], [MR92], [De92], [De99]

1 INTRODUCTION

The facilities in a large supply chain are typically owned by many firms with differing interests. This raises questions of how such supply chains might operate. The sequel discusses two contrasting approaches to this problem. One approach is a competitive one in which each firm seeks to maximize its own profits given the strategies of the other firms. This leads to the study of noncooperative games. The second approach is a cooperative one in which the firms collaborate to maximize the aggregate profits of all firms and then allocate the profits among them. This leads to the study of cooperative games. The sequel examines both approaches.

A noncooperative game consists of a (finite) collection of competing firms each of which has a set of available strategies and earns a profit that depends on the *strategy profile*, i.e., the vector of strategies of all firms. An important strategy profile for such games is a *Nash equilibrium*, i.e., the strategy each firm uses in the profile maximizes its profit given the other firms' strategies in the profile. The rationale for supposing that competitive firms will choose strategies that form a Nash equilibrium is that in any other strategy profile at least one of the firms can increase its profits by altering its strategy given that the other firms do not change their strategies.

There are two different hypotheses that assure the existence of a Nash equilibrium. One is that each firm's strategy set is convex and its profit is concave on its strategy set. The second, which §3.2 develops, is that each firm's strategy set is a lattice and the profit the firm earns is superadditive on the product of the strategy set of the firm and each chain of strategies of the other firms. Both hypotheses are useful. The attractive features of the superadditivity hypothesis are similar to those of the lattice-programming problems that §2 discusses and so do not require repetition here.

Although Nash equilibria are firm-by-firm optimal, there are often strategy profiles that simultaneously give each firm higher profits than does any Nash equilibrium. In that event, cooperating to use such a strategy profile would allow all firms to achieve higher profits. This suggests consideration of games that explicitly consider the possibility of cooperation.

A cooperative game consists of a (finite) collection of firms that seek to maximize the aggregate profits of all firms and share the profits among them. The question arises how those profits might be allocated among the firms. One widely recognized criterion for acceptability of a profit allocation (of total profits to the firms) is that it lie in the core. The *core* is the set of profit allocations for which each subset of firms earns at least as much from its profit allocation (the sum of the allocations to each firm in the subset) as it could guarantee by independent operation. Section 3.3 below shows how to find elements in the core of a cooperative linear-programming game by solving a single linear program whose size grows at worst linearly with the number of firms. Finding all elements of the core of such a game appears to require solution of a linear program whose size grows exponentially with the number of firms.

2 NASH EQUILIBRIA of NONCOOPERATIVE SUPERADDITIVE GAMES

Let I be a finite set of firms and $\emptyset \neq T \subseteq \Re^m$ be a set of exogenous *parameters* reflecting the environment in which the firms compete, e.g., costs, technologies, weather, laws, etc. For each firm $i \in I$, let $\emptyset \neq S_i \subseteq \Re^n$ be the set of *strategies* available to i and $S^i \equiv \times_{I\setminus\{i\}} S_j$ be the *strategy profile set* of firms other than i. Let $S \equiv \times_I S_i$ be the *strategy profile set* of all firms. For each firm $i \in I$ and *strategy profile* $s \in S$ of all firms, let $s_i \in S_i$ and $s^i \in S^i$ denote respectively the corresponding strategy of firm i and *strategy profile* of the other firms.

Consider firm $i \in I$, a strategy profile $s \in S$ of all firms, and a parameter $t \in T$. Let $\emptyset \neq S^t_{si} \subseteq S_i$ be firm i's *feasible replies* to s with S^t_{si} being independent of s_i. Let $u_i(s,t)$ be the real-valued *profit* to firm i. Let R^t_{si} be firm i's set of *optimal replies*, i.e., the set of σ_i that maximize $u_i(\sigma,t)$ over $\sigma_i \in S^t_{si}$ given $\sigma^i = s^i$.

Fix a strategy profile $s \in S$ of all firms and a parameter $t \in T$. Let $R^t_s \equiv \times_{i\in I} R^t_{si}$ be the set of *optimal replies* of the firms to s at t. Call s a *Nash equilibrium* at t if $s \in R^t_s$, i.e., s is an optimal reply to itself at t.

THEOREM 1. Existence and Monotonicity of Nash Equilibria with Superadditive Profits. *Suppose that for each firm $i \in I$, the set S^t_{si} of feasible replies to $s \in S$ at $t \in T$ is a nonempty compact sublattice of the compact lattice S_i of the firm's strategies and is ascending in (s,t) on $S \times T$, the profit $u_i(s,t)$ to firm i is real-valued and superadditive in (s,t) on $S_i \times C$ for each chain C in $S^i \times T$, and $u_i(s,t)$ is upper semicontinuous in s_i on S_i for each $(s^i,t) \in S^i \times T$. Then there exist least and greatest Nash equilibria at $t \in T$ and both are increasing in t on T.*

The proof of this result depends on Theorem 2 below which in turn requires some definitions. Suppose $S \subseteq \Re^n$ and $\sigma(\,\cdot\,)$ is a mapping of S into itself. Call $s \in S$ a *deficient*, a *fixed* or an *excessive* point of $\sigma(\,\cdot\,)$ respectively according as $s \le \sigma(s)$, $s = \sigma(s)$ or $s \ge \sigma(s)$. If $\emptyset \ne L \subseteq S$ and S is a compact lattice, then L has a greatest lower (resp., least upper) bound in S, denoted $\wedge L$ (resp., $\vee L$), because the set of all lower (resp., upper) bounds of L is a nonempty compact sublattice of S.

THEOREM 2. Existence and Monotonicity of Fixed-Points of Increasing Mappings. *Suppose $\emptyset \ne S \subseteq \Re^n$ is a compact lattice and $\emptyset \ne T \subseteq \Re^m$. If $\sigma_t(s) \in S$ is increasing in (s,t) on $S \times T$, then $\sigma_t(\,\cdot\,)$ has a common least (resp., greatest) fixed and excessive (resp., deficient) point, and it is increasing in t on T.*

Proof. Since S is a nonempty compact lattice, $\vee S$ exists and is in S. Suppose $t \in T$. Then the set E_t of excessive points of $\sigma_t(\,\cdot\,)$ contains $\vee S$ and so is nonempty. Again since S is a compact lattice, $\wedge E_t$ exists and is in S. Thus, for each $s \in E_t$, $s \ge \sigma_t(s) \ge \sigma_t(\wedge E_t)$ because $\sigma_t(\,\cdot\,)$ is increasing, whence $\wedge E_t \ge \sigma_t(\wedge E_t) \ge \sigma_t^2(\wedge E_t)$. Hence, $\wedge E_t$ and $\sigma_t(\wedge E_t)$ are excessive points of $\sigma_t(\,\cdot\,)$ and $\wedge E_t$ is the least such point, so $\sigma_t(\wedge E_t) \ge \wedge E_t$. Thus $s_t \equiv \wedge E_t$ is a fixed point of $\sigma_t(\,\cdot\,)$ and, since E_t contains all fixed points of $\sigma_t(\,\cdot\,)$, s_t is the least fixed point of $\sigma_t(\,\cdot\,)$.

Now suppose $t \le \tau \in T$. Then $E_t \supseteq E_\tau$ because $s \in E_\tau$ implies $s \ge \sigma_\tau(s) \ge \sigma_t(s)$, the last inequality holding by hypothesis. Thus, $s_t = \wedge E_t \le \wedge E_\tau = s_\tau$ as claimed.

The fact that $\sigma_t(\,\cdot\,)$ has a common greatest fixed and deficient point and that it is increasing in t on T follows *dually*, i.e., by reversing the orderings of S and T (e.g., replace S and T by their negatives), and applying the result just shown. ∎

Proof of Theorem 1. It follows from the hypotheses of Theorem 1 and the Increasing-Optimal Selections Theorem 8 of §2.5 that each firm i has a least optimal reply $\sigma^i_t(s)$ to each $(s,t) \in S \times T$, and $\sigma^i_t(s)$ is increasing in (s,t) on $S \times T$. Thus, $\sigma_t(s) \equiv (\sigma^i_t(s)) \in S$ is increasing in (s,t) on $S \times T$. Hence, by Theorem 2, it follows that $\sigma_t(\,\cdot\,)$ has a least fixed point s_t at $t \in T$, s_t is a Nash equilibrium and s_t is increasing in t on T.

Next we show that s_t is the least Nash equilibrium. To that end, suppose $s \in R^t_s$, i.e., s is a Nash equilibrium at $t \in T$. Then since s is an optimal reply to itself and $\sigma_t(s)$ is the least optimal reply to s at t, it follows that $\sigma_t(s) \le s$. Since s is an excessive point of $\sigma_t(\cdot)$ and s_t is the least such excessive point by Theorem 2, it follows that $s_t \le s$. Thus s_t is the desired least Nash equilibrium. The existence of a greatest Nash equilibrium at $t \in T$ and its monotonicity in t on T follows dually. ∎

Not all equilibria are equally attractive to all firms in multifirm games. However, in the present setting, there are natural sufficient conditions assuring that all firms prefer the greatest (resp., least) Nash equilibria. The next result gives such conditions.

THEOREM 3. Preference for Least and Greatest Nash Equilibria. *Suppose for each firm $i \in I$ and $t \in T$ that $S^t_{si} \subseteq S^t_{\sigma i}$ (resp., $S^t_{si} \supseteq S^t_{\sigma i}$) for each $s, \sigma \in S$ with $s^i \le \sigma^i$ and $u_i(s,t)$ is increasing (resp., decreasing) in s^i on S^i for each $s_i \in S_i$. If $s \le \sigma$ are Nash equilibria at t, then all firms prefer σ (resp., s). In particular, under the additional hypotheses of Theorem 1, all firms prefer the greatest (resp., least) Nash equilibrium for each $t \in T$.*

Proof. It suffices to prove the result reading without parentheses. The other case follows dually. The hypotheses imply that $U_i(s^i,t) = \max\{u_i(s,t) \colon s_i \in S^t_{si}\}$ is increasing in s^i on S^i. Thus, if $s \le \sigma$ are Nash equilibria at t, then $u_i(s,t) = U_i(s^i,t) \le U_i(\sigma^i,t) = u_i(\sigma,t)$. ∎

Application to Multifirm Competitive Price Setting

Suppose that a finite set I of firms sell versions of a product that compete in a market. Suppose also that the unit cost that firm i incurs to manufacture the product is $c_i > 0$, $i \in I$, and $c = (c_i)$. The demand $d_i(p)$ that firm i experiences for the product is positive and depends on the prices $p = (p_i)$ that the firms charge. The profit that each firm $i \in I$ earns is

$$d_i(p)(p_i - c_i). \tag{1}$$

It is reasonable to assume that each firm will consider only prices that at least cover their costs plus a minimum acceptable profit $\pi_i \ge 0$ say. For this reason, assume that

$$c_i + \pi_i \le p_i \le P_i \text{ for } i \in I \tag{2}$$

where $P_i \ge c_i + \pi_i$ is the maximum price that firm $i \in I$ is willing to consider charging.

Now maximizing firm i's profit $d_i(p)(p_i - c_i)$ is equivalent to maximizing its natural log, viz., $u_i(p,c) \equiv \ln d_i(p) + \ln(p_i - c_i)$. Also, $u_i(p,c)$ is superadditive in p if and only if, as we assume in the sequel, $\ln d_i(p)$ has that property because $\ln(p_i - c_i)$ is additive in p. Also, the set of feasible prices p is a compact sublattice, so firm i's feasible prices are a sublattice and are ascending in p^i for all i. Thus it follows from Theorem 1 that there are least and greatest Nash equilibrium prices.

Next consider what effect an increase in the unit costs c of manufacturing has on the Nash equilibrium prices. Now $u_i(p,c)$ is superadditive in (p,c) if and only if each $\ln(p_i - c_i)$ is superadditive. But the last is so because the natural log is concave. Also, the set of feasible vectors (p,c) is a sublattice and so each firm i's set of feasible prices p_i is ascending in c and in the prices p^i of the other firms. Thus, it follows from Theorem 1 that the least and greatest Nash equilibrium prices rise with c. Hence, an increase in unit manufacturing costs at any subset of firms leads to higher Nash equilibrium prices at all firms.

Finally, it is reasonable to suppose that the demand $d_i(p)$ at each firm i is increasing in the prices p^i of the other firms. In that event, $u_i(p,c)$ is also increasing in p^i. Thus, by Theorem 3, all firms prefer the highest Nash equilibrium prices.

The above results remain valid if any of the additional restrictions given below is present. This is because they are all sublattices and so firm i's set of feasible prices ascends and becomes larger (in the sense of set inclusion) as the other firms raise their prices.

- Firm i wishes to limit its prices to a given finite set.
- Firm i insists on being a price leader in a subset J of the firms I, i.e., $p_i \le p_j$ for $j \in J$.
- Firm i wants to be within 20% of the minimum price, i.e., $p_i \le 1.2p_j$ for all $j \in I$.

3 CORE of a COOPERATIVE LINEAR-PROGRAMMING GAME

Consider a finite set I of firms each of whom has operations, e.g., supply chains, that have representations as linear programs. Suppose the linear program representing the operations of firm i in I entails choosing an n-column vector $x \ge 0$ of *activity levels* that maximizes the firm's *profit*

$$cx \tag{3}$$

subject to the constraint that its consumption Ax of resources minorizes its *resource vector* b^i, i.e.,

$$Ax \le b^i. \tag{4}$$

The firms are considering forming an alliance that combines their operations with the aim of increasing overall profits and sharing the benefits. The question arises how the profit of the alliance might be allocated among its members.

An *alliance* (or *coalition*) is a subset of the firms. If an alliance S pools its resource vectors, the linear program that S faces is that of choosing an n-column vector $x \ge 0$ that maximizes the profit (3) that S earns subject to its resource constraint (4) with $b^S \equiv \sum_{i\in S} b^i$ replacing b^i there. Let v^S be the resulting maximum profit of S.

Now consider the *grand alliance* (or *grand coalition*), i.e., the set I of all firms. A *profit allocation* to the firms is an $|I|$-vector $p = (p^i)$ that distributes the grand alliance's maximum profit among the firms, i.e., allocates each firm i the profit p^i from the total $p^I = v^I$ where $p^S \equiv \sum_{i\in S} p^i$ for all $S \subseteq I$. Each alliance S can reasonably insist that an acceptable profit allocation ought to earn S at least as much as S could earn by combining operations of its members, i.e., $p^S \ge v^S$. For if $v^S > p^S$, the firms in S would find it attractive to withdraw from the grand alliance, form an alliance S and improve on p by distributing the extra profit $v^S - p^S > 0$ among its members. The *core* is the set of profit allocations p in which each alliance is as well off as it would be by independent operation, i.e., $p^S \ge v^S$ for all alliances S. The core can have many elements. In that event, the core provides a set of profit allocations over which the firms are likely to negotiate.

Checking whether or not a profit allocation is in the core would seem to require solving $2^{|I|}$ linear programs, one for each alliance, and so rises exponentially with the number $|I|$ of firms. However, for the cooperative linear programs under discussion, it is possible to find elements in the core by solving only a single linear program, viz., that of the grand alliance. In particular one profit allocation in the core entails finding an optimal dual price vector π for the linear program of the grand alliance and then allocating each firm i a profit πb^i equal to the value of the resource vector that i provides at those optimal dual prices.

THEOREM 4. Core of Cooperative Linear Programming Game. *For each optimal dual price vector for the linear program of the grand alliance, allocating each firm the value of its resource vector at those prices yields a profit allocation in the core.*

Proof. Suppose π is optimal for the dual of the linear program of the grand alliance. Let $p = (p^i)$ be the profit allocation that gives each firm $i \in I$ the profit $p^i = \pi b^i$. Then $p^I = \pi b^I = v^I$ by the duality theorem. Also, π is feasible for the dual of the linear program for each alliance S, so $p^S = \pi b^S \ge v^S$ by weak duality. ∎

The above result shows that the set of profit allocations generated by the set of optimal dual price vectors for the linear program of the grand alliance is a subset of the core. Unfortunately, the converse is generally false. However, the converse is true in a larger "r-subsidiary" game provided only that the data A, b and c are rational. The *r-subsidiary game* consists of allowing each firm to subdivide into r, say, subsidiaries of equal size by dividing the resources of the firm into r equal parts. Each of the r subsidiaries of firm i faces a linear program in nonnegative variables like (3)-(4) above except that the resource vector $r^{-1}b^i$ replaces b^i. Each firm then becomes a "holding company" with r identical subsidiaries each of which is free to form alliances with subsidiaries of any firm in I. Of course the linear program for the grand alliance of all subsidiaries coincides with that of the original grand alliance.

THEOREM 5. Core of Cooperative Linear Programming r-Subsidiary Game. *If the data are rational, the following sets of profit allocations to the firms coincide: (i) the set of allocations that give each firm the value of its resource vector at a common optimal dual price vector for the linear program of the grand alliance; (ii) the core of the r-subsidiary game for all large enough positive integers r.*

Differing Activities, Resource Types or Unit Profits. It is possible to reduce problems in which firms have differing activities, resource types and/or unit profits to the problem discussed heretofore in which A and c are not firm dependent. To do this, form the linear program of the grand alliance with upper bounds (possibly $+\infty$) on all activity levels (variables) included. The upper bounds on activity levels are new resource types. Form each firm's linear program from that of the grand alliance as follows. Set the upper bound on an activity level of a firm equal to zero if the activity (or unit profit thereof) is not available to the firm. Set the firm's resources of a given type equal to its portion of the total available to the alliance, e.g., zero if the firm does not have a resource type and its activities do not consume it.

This page is intentionally left blank.

4

Convex-Cost Network Flows & Supply Chains: Substitutes/Complements/Ripples

[Sh61], [IV68,69], [Ve68b,75], [GP81], [GV85], [CGV90a,b]

1 INTRODUCTION

The inventory manager discussed in §2.6 may also be interested in knowing that there exist optimal production levels in each period that are increasing in the actual, as distinguished from the cumulative, sales forecast in each period. It does not appear to be possible to deduce this result directly from the Increasing-Optimal-Selections Theorem because the set of feasible solutions is not a sublattice. However, as we saw in §1.2, the constraints can be expressed as a network-flow problem in which the sales in each period are fixed demands at nodes other than node zero, or equivalently, fixed flows from nodes other than node zero to that node. This suggests that it might be useful to explore the variation of optimal network flows with their parameters. That is the goal of this section.

We study the qualitative variation of minimum-cost network flows and their associated costs with various parameters of the problem, e.g., arc-flow bounds, node demands, cost-function parameters, etc. The aim will be to establish when the optimal flow in one arc is independent of or monotone in the parameter associated with a second arc, to give bounds on the rate of change of the second arc flow with the first arc parameter and to show that the magnitude of change in the second arc flow diminishes the "less biconnected" it is to the first arc. Moreover, we investigate when the minimum cost is additive, subadditive or superadditive in the parameters associated with two arcs. Finally, we apply these results to a multi-period production/inventory problem by discussing the effect of changes in parameters on optimal production, inventories and sales.

Minimum-Cost Network-Flow Problem

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a (*directed*) *graph* with *nodes* $\mathcal{N}$ and *arcs* $\mathcal{A}$. Each arc is directed from one node, called its *tail*, to a different node, called its *head*. Several arcs may have the same head and tail. Denote by $\mathcal{A}_i^+$ (resp., $\mathcal{A}_i^-$) the set of arcs having tail (resp., head) i. There is a given *demand* d_i at each node i. Let x_a be the flow in arc a and t_a be an associated *parameter* (e.g., an upper or lower bound on x_a), the last restricted to a lattice T_a in $\Re^k$ for some k depending on a. There is a *cost* $c_a(x_a, t_a)$ incurred when the flow in arc a is x_a and the associated parameter is t_a. Assume that c_a is a $+\infty$ or real-valued function on $\Re \times T_a$ for each arc a. The problem is to find a vector $x = (x_a)$ that minimizes the total cost

$$C(x,t) \equiv \sum_{a \in \mathcal{A}} c_a(x_a, t_a) \tag{1}$$

associated with a *parameter vector* $t = (t_a)$ subject to the *flow-conservation equations*

$$\sum_{a \in \mathcal{A}_i^-} x_a - \sum_{a \in \mathcal{A}_i^+} x_a = d_i \text{ for } i \in \mathcal{N}. \tag{2}$$

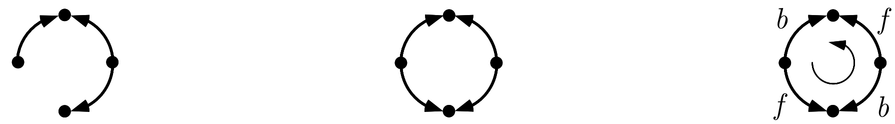
Call a vector x satisfying (2) a *flow* and, if $d_i = 0$ for all i, a *circulation*. The difference of two flows is a circulation. Call a parameter vector t *feasible* if there is a flow x, called *feasible* for t, such that $C(x,t)$ is finite.

Let T be a subset of $\times_a T_a$, $\mathcal{C}(t)$ be the infimum of $C(x,t)$ over all flows x, and $X(t)$ be the set of feasible flows that minimize $C(\cdot, t)$ for $t \in T$. Call $X(\cdot)$ the *optimal-flow multifunction* on the set T^o of $t \in T$ for which $X(t) \neq \emptyset$, and a selection therefrom an *optimal-flow selection*.

It is useful now to explore when changing the parameter in one arc does not affect the optimal flow in another arc. First, recall a few facts about (directed) graphs.

Graphs

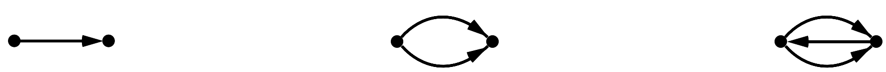
Simple Paths and Cycles. A *simple path* is an alternating finite sequence of distinct nodes and arcs that begins and ends with a node and such that each arc joins the nodes immediately preceding and following it in the sequence. If instead the first and last nodes in the sequence are the same, call the sequence a *simple cycle*. Call a simple cycle *oriented* if a direction of traversing the cycle is specified. In that event, call the arcs that are traversed in their natural order *forward* arcs and call the others *backward* arcs. Call a circulation *simple* if its induced subgraph is a simple cycle.[1] Figure 1 illustrates these definitions. Label the forward (resp., backward) arcs of the oriented simple cycle in Figure 1 by f (resp., b).



Simple Path Simple Cycle Oriented Simple Cycle

Figure 1. Simple Paths and Cycles

Connectivity. Call two distinct nodes in a graph *k-connected* if $k > 0$ and there exist at least k internally node-disjoint simple paths joining the two nodes or if $k = 0$ and there is no simple path joining the two nodes. We use *disconnected*, *connected*, *biconnected*, and *triconnected* interchangeably with 0-connected, 1-connected, 2-connected and 3-connected respectively. Figure 2 illustrates the last three definitions.



Connected Graph Biconnected Graph Triconnected Graph

Figure 2. Some k-Connected Graphs

Call a graph *complete* if every pair of distinct nodes is joined by a single arc. Call a graph with two or more nodes *k-connected* if $k > 0$ and every pair of distinct nodes is k-connected or if $k = 0$ and some pair of distinct nodes is 0-connected. Call a graph with one node 1-*connected*. Here again we use *disconnected, connected, biconnected* and *triconnected* interchangeably with 0-connected, 1-connected, 2-connected and 3-connected respectively. It is not difficult to show that a graph is biconnected if and only if there is a pair of distinct arcs, and each such pair is contained in a simple cycle.

Decomposition into Connected Components. A *subgraph* of a graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ is a graph whose node and arc sets are respectively subsets of $\mathcal{N}$ and $\mathcal{A}$. A *connected component* of a graph is a maximal[2] connected subgraph. It is easy to show that the maximal connected components of a directed graph form a partition thereof. For example, the three graphs in Figure 2 can be considered to be the three connected components of the combined graph consisting of all three. Hence, the set of flows in a graph is a (direct) product of the sets of flows in each connected component thereof. Moreover, there is a flow if and only if the sum of the demands in each connected com-

[1] The subgraph *induced* by a flow is the set of arcs having nonzero flow and the nodes incident thereto.

[2] A subset L of a set S is *maximal* among those with a property P if there is no subset of S with the property P that properly contains L.

ponent is zero, which we assume in the sequel. Thus a minimum-cost network-flow problem can be solved by decomposing its directed graph into its connected components and solving the resulting network-flow problem on each such connected component.

Decomposition into Biconnected Components. The question arises whether it is possible to decompose a network-flow problem on a connected graph into independent subproblems. The answer is that it is, provided that the graph has a *cut node*, i.e., a node ν for which there are two subgraphs whose union is the graph and that have only node ν in common.

In order to see why, consider the graph of Figure 3. Suppose that the demands at the three nodes are, in order from left to right, $1, 2, -3$. Then the node shared by arcs a and e is a cut node whose deletion disconnects the other two nodes in the graph. Now since flow is conserved, the net flow into the cut node from arcs a and b is -1 and that from d and e is 3. Since the demand at the cut node is the sum of these two demands, it follows that a flow in the graph is a (direct) product of two flows, one in the simple cycle containing a, b and the other in the simple cycle containing d, e. Also the cost of a flow is the sum of the costs of the flows in the two simple cycles. Thus one can find a minimum-cost flow by splitting the network-flow problem into two independent subproblems, one on each simple cycle.

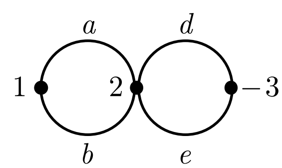


Figure 3. A Connected Graph With a Cut Node

A *biconnected component* of a graph is a maximal subgraph among those that are either biconnected, or comprise a single node or a single arc. It is known that a connected graph can be decomposed into biconnected components. Each cut node belongs to at least two biconnected components; every other node belongs to exactly one biconnected component; and each two distinct biconnected components share at most one (cut) node. Moreover, the (undirected bipartite) graph whose nodes are the cut nodes and the biconnected components, and whose arcs join cut nodes to the biconnected components to which they belong, is a tree, as Figure 4 illustrates. It can be shown from these results that a connected network-flow problem can be decomposed into independent network-flow problems on each biconnected component.

Restriction to Biconnected Network-Flow Problems

It follows from the above discussion that changing the parameter of an arc in one biconnected component of a graph has no effect on the minimum-cost flows in other biconnected components or on their costs. Also, changing the parameter associated with an arc of a biconnected component that consists of a single arc cannot change the flow therein. And a graph consisting of a single node has no arcs. For these reasons, it suffices to consider the effect of changing arc parameters on minimum-cost flows in a biconnected graph.

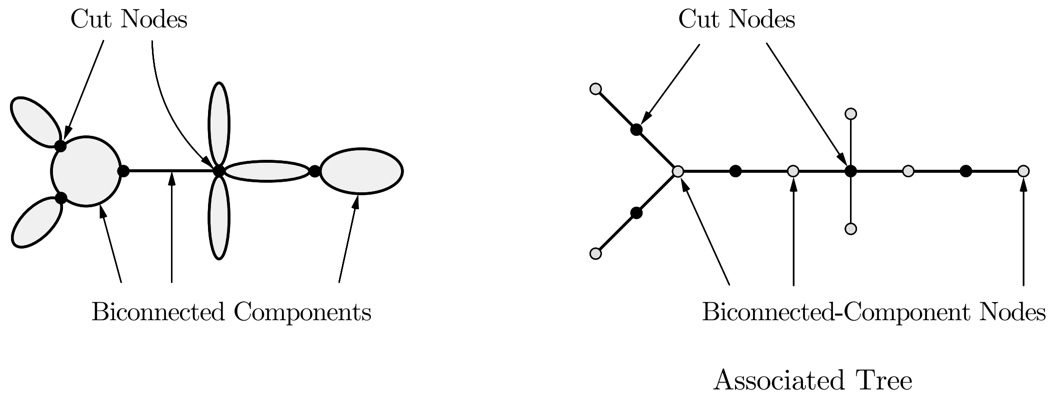


Figure 4. A Tree of Biconnected Components

2 RIPPLE THEOREM

Changing the parameter of an arc b say, has effects that ripple through the network. It is plausible that the magnitude of the resulting change in the optimal flow in an arc a diminishes the "more remote" a is from b. This subsection shows that this is indeed the case if "more remote" means "less-biconnected-to". To see this, it is necessary to give a few definitions and establish a fundamental Circulation-Decomposition Theorem.

Circulation-Decomposition Theorem [Be61], [Be73, p. 91]

Call two vectors (u_i) and (v_i) *conformal* if $u_i v_i \geq 0$ for all i. Call a set of vectors *conformal* if that is so of each pair in the set.

A *simple chain* (resp., *circuit*) is a simple path (resp., cycle) that orients all arcs in the same way. The simple chain *joins* i to j if i and j are each end nodes of the chain, i is the tail of an arc in the chain, and j is the head of an arc therein.

In order to motivate the next result, consider the circulation $x = (x_a, x_b, x_d) = (2\ 4\ 2)$ in the graph with two nodes and three arcs in Figure 5. The question arises whether x can be expressed as a sum of simple circulations. The answer is it can, e.g., $x = x^+ + x^-$ is the sum of the simple circulations $x^+ = (0\ \ 4\ \ 4)$ and $x^- = (2\ \ 0\ {-2})$. Observe that these simple circulations are not

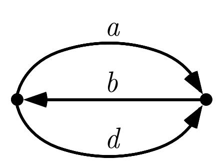


Figure 5. A Graph with Two Nodes and Three Arcs

conformal because $x_d^- x_d^+ = -8 < 0$. On the other hand, we can also express $x = x' + x''$ as the sum of the simple circulations $x' = (2 \ \ 2 \ \ 0)$ and $x'' = (0 \ \ 2 \ \ 2)$ that are conformal since $x'_a x''_a = 0$, $x'_b x''_b = 4 \geq 0$ and $x'_d x''_d = 0$. The next result asserts that it is possible to do this for any circulation.

THEOREM 1. Circulation Decomposition. *Each circulation (resp., integer circulation) with $p \geq 2$ arcs in its induced subgraph is a sum of at most $p-1$ simple conformal circulations (resp., integer circulations), each with distinct induced cycles.*

Proof. Let x be a circulation. We begin by proving the result reading without parentheses for the case $x \geq 0$. The proof is by induction on p. The result is obvious for $p = 2$. Suppose it holds for $p-1$ (≥ 2) or less arcs and consider p.

It is enough to construct a simple circulation $0 \leq y \leq x$ with $y_a = x_a > 0$ for at least one arc a. For then $z \equiv x - y \geq 0$ is a circulation with at most $p-1$ arcs in its induced subgraph, and the result follows from the induction hypothesis.

To construct y, let $\mathbb{M}$ be a maximal simple chain in the subgraph induced by x, and suppose $\mathbb{M}$ joins i to j, say. Since x is a nonnegative circulation, there is an arc (j,k), say, in the induced subgraph. Since $\mathbb{M}$ is maximal, k must be a node of $\mathbb{M}$. Let $\mathbb{C}$ be the set of arcs in the subchain of $\mathbb{M}$ that joins k to j together with the arc (j,k). Evidently the arcs in $\mathbb{C}$ form a simple circuit in the subgraph induced by x. Put $\lambda \equiv \min_{a \in \mathbb{C}} x_a$. The desired simple circulation $y = (y_a)$ is formed by putting $y_a = \lambda$ for $a \in \mathbb{C}$ and $y_a = 0$ otherwise. By construction $y \leq x$ and $y_a = x_a$ for some $a \in \mathbb{C}$, so that a cannot belong to the subgraph induced by $z = x - y$. Finally, observe that if x is integer, the simple circulation y is also integer.

It remains to consider the case $x \not\geq 0$. To that end, let x^* be formed by replacing each arc (i,j) in the set $\mathbb{A}$ of arcs in which the flow is negative by its reverse arc (j,i) and letting the flow in (j,i) be minus the flow in (i,j). Then $x^* \geq 0$ is a circulation in the graph in which the arcs in $\mathbb{A}$ are replaced by the set $\mathbb{A}^*$ of arcs that reverse arcs in $\mathbb{A}$. Now apply the representation established above for nonnegative flows to x^*, reverse the arcs in $\mathbb{A}^*$ to restore the graph to its original form, and replace the flow in each arc in $\mathbb{A}^*$ in the simple circulations in the representation of x^* by their negatives. This preserves the conformality of the simple circulations in the representation and establishes the desired result. ∎

Ripple Theorem

We now apply this result to show that if one changes the parameter of a single arc b say, the change in one optimal flow is a sum of simple conformal circulations each of whose induced simple cycle contains the arc. Thus the simple cycles induced by those simple circulations might appear as in Figure 6. Moreover, the magnitude of the change in the optimal flow in another arc diminishes the "less biconnected" the arc is to arc b.

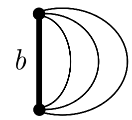
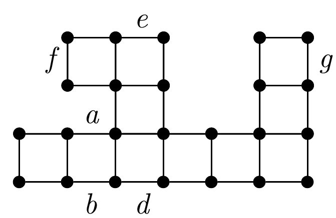


Figure 6. Simple Cycles Induced by Changing Arc b's Parameter

We say that arc a is *less biconnected to* arc b than is an arc d, written $a \le_b d$, if every simple cycle containing a and b also contains d. The relation $\le_b$ is easily seen to be reflexive ($a \le_b a$ for all a) and transitive ($a \le_b d$ and $d \le_b e$ imply that $a \le_b e$) and so quasi-orders the set of arcs. Figure 7 illustrates the relation $\le_b$ in a series-parallel graph.



$$b \le_b a, a \le_b b, d \le_b a, e \le_b d, f \le_b e, g \le_b d, g \not\le_b f$$

Figure 7. Less-Biconnected-to-b Relation

THEOREM 2. Ripple. *In a biconnected graph, if t and $t' \in T$ differ only in component $b \in \mathcal{A}$, $c_a(\,\cdot\,, t_a)$ is convex for each $a \in \mathcal{A} \setminus \{b\}$, x is an element (resp., integer element) of $X(t)$, and $X(t')$ has an element (resp., integer element), then for one such element x',*

1° *$x' - x$ is a sum of simple conformal circulations (resp., integer circulations) each of whose induced cycle contains b and*

2° *$|x'_a - x_a|$ decreases the less biconnected a is to b.*

Proof. To establish 1°, suppose x' is an element (resp., integer element) of $X(t')$. Then $x' - x$ is a sum of simple conformal circulations (resp., integer circulations) by the Circulation-Decomposition Theorem. Let y be the sum of the simple circulations whose induced cycle contains b and let z be the sum of the remaining simple circulations. Then y and z are conformal.

We claim that

$$c_a(x_a + y_a, t'_a) + c_a(x_a + z_a, t_a) \le c_a(x'_a, t'_a) + c_a(x_a, t_a) \text{ for } a \in \mathcal{A}.$$

To see this observe that if $a \neq b$, then $t'_a = t_a$ and the inequality follows from the convexity of $c_a(\,\cdot\,, t_a)$. If instead $a = b$, the inequality holds as an identity on observing that $x_b + y_b = x'_b$ and $z_b = 0$. Summing the above inequalities over all arcs yields

$$C(x+y, t') + C(x+z, t) \le C(x', t') + C(x, t).$$

Hence $x + y \in X(t')$ and $x + z \in X(t)$ because $x' \in X(t')$, $x \in X(t)$ and the two terms on the right-hand side of the above inequality are finite. Moreover, if x and x' are integer, so is $x + y$. Therefore $x + y$ is the desired optimal flow for t'.

It remains to establish 2°. To that end, observe that if $a \le_b d$, d is contained in each of the induced cycles that contains a. Thus since the corresponding simple circulations are conformal, the result follows. ∎

As an application of the second assertion of the Ripple Theorem, observe that if one changes the parameter of arc b in a network-flow problem on the graph of Figure 7, then the absolute change δx_α in arc α's optimal flow satisfies $\delta x_b = \delta x_a \ge \delta x_d \ge \delta x_e \ge \delta x_f$ and $\delta x_d \ge \delta x_g$.

3 CONFORMALITY AND SUBADDITIVITY

When is One Optimal Arc-Flow Monotone in Another Arc's Parameter?

There are two types of assumptions needed to assure that the optimal flow in one arc of a biconnected graph is monotone in the parameter of a second arc. One concerns the nature of the arc costs and the other the relative position of the two arcs in the network.

Assumptions on the Arc Costs

In order to motivate the assumptions needed on the arc costs to assure that the arc flows are monotone in certain parameters, consider first the simple case that Figure 5 illustrates in which there are zero demands at the two nodes and real parameters associated with the three arcs. Then $x_d = x_b - x_a$ in each flow, so the total cost of a flow becomes

$$c_a(x_a, t_a) + c_b(x_b, t_b) + c_d(x_b - x_a, t_d).$$

In order to be assured that the set of optimal x_a, x_b will be ascending in the parameters t_a and t_b, it suffices to require that $c_a(\,\cdot\,,\,\cdot\,)$ and $c_b(\,\cdot\,,\,\cdot\,)$ be subadditive, $c_d(\,\cdot\,, t_d)$ be convex for each t_d, and the set of pairs (x_a, x_b) with minimum total cost be nonempty and compact for each (t_a, t_b) by the Increasing-Optimal-Selections Theorem. For this reason, assume in the sequel that the arc costs are subadditive in the arc-flow parameter-pairs and convex in the arc flows.

The subadditivity and convexity assumptions on the arc costs are quite flexible. They permit, for example, the parameter t_a to be an upper or lower bound on the flow in arc a, a fixed flow in arc a, a parameter of arc a's cost function, or a vector containing several. To illustrate, if one is interested in studying the effect of changes in the upper and lower bounds and in a parameter of the flow cost for arc a, T_a could be the sublattice of $\Re^3$ consisting of triples (u,l,p) for which $l \le u$. The following examples illustrate these possibilities for an arc with flow ξ therein, parameter τ a real number and flow cost $c(\xi,\tau)$. In these examples it will be convenient to denote by δ_+ and δ_0 the indicator functions respectively of the subsets $\Re_+$ and $\{0\}$ of $\Re$.

Example 1. Translation of an Arc's Flow by its Parameter. The most common applications of network flows involve translating an arc flow by a parameter. This is handled by putting $c(\xi,\tau) \equiv c(\xi-\tau)$ where $c(\cdot)$ is $+\infty$ or real-valued. Then $c(\cdot,\cdot)$ is subadditive if and only if $c(\cdot)$ is convex. And in that event, $c(\cdot,\cdot)$ is also convex. We now apply this idea to accommodate upper and lower bounds on arc flows, and fixed flows therein.

Upper and Lower Bounds. The function $c(\cdot,\cdot)$ is subadditive if either

(a) $$c(\xi,\tau) = c(\xi) + \delta_+(\xi-\tau)$$

or

(b) $$c(\xi,\tau) = c(\xi) + \delta_+(\tau-\xi)$$

where $c(\cdot)$ is $+\infty$ or real-valued. This is because $c(\cdot)$ is trivially subadditive and δ_+ is convex so $\delta_+(\xi-\tau)$ is subadditive. Observe that τ can be thought of as a lower (resp., upper) bound on ξ in (a) (resp., (b)) because the cost is infinite when ξ is less (resp., greater) than τ. If also $c(\cdot)$ is convex, then $c(\cdot,\tau)$ will be convex as well. This shows that upper and lower bounds on an arc's flow can be absorbed into the arc's cost function without destroying its subadditivity or convexity.

Fixed Flows. The flow ξ in an arc can be fixed at a value τ by putting $c(\xi,\tau) = c(\xi) + \delta_0(\xi-\tau)$ where $c(\cdot)$ is $+\infty$ or real-valued. Since δ_0 is convex, $c(\cdot,\cdot)$ is subadditive. If $c(\cdot)$ is convex, the same is so of $c(\cdot,\cdot)$. As we shall see subsequently, this technique is useful for studying the effect of changes in the demands at the various nodes.

Example 2. Join of an Arc's Flow and its Parameter. In some situations an arc flow incurs its normal cost if the flow is above a given level and the cost of the given level otherwise. This is encompassed by putting $c(\xi,\tau) \equiv c(\xi \vee \tau)$ where τ is the given flow level. Observe that $c(\cdot,\cdot)$ is subadditive (resp., convex) if and only if $c(\cdot)$ is increasing (resp., increasing and convex). In that event $c(\xi,\tau) = c(\xi) \vee c(\tau)$.

Example 3. Product of an Arc's Flow and its Parameter. In applications to flows with linear cost functions or with gains and losses, interest centers on the effects of multiplying an arc's flow by its parameter. This is treated by putting $c(\xi,\tau) \equiv c(\xi\tau)$ with $c(\,\cdot\,)$ being $+\infty$ or real-valued. Then $c(\,\cdot\,,\tau)$ is convex if that is so of $c(\,\cdot\,)$. Also if $c(\,\cdot\,)$ is finite-valued and twice continuously differentiable, then $D_{12}c(\xi,\tau) = Dc(\xi\tau) + \xi\tau D^2c(\xi\tau)$ so $c(\,\cdot\,,\,\cdot\,)$ is subadditive (resp., superadditive, additive) if $Dc(u) + uD^2c(u)$ is uniformly nonpositive (resp., nonnegative, zero) in u. In particular, $c(\,\cdot\,,\tau)$ is convex and $c(\,\cdot\,,\,\cdot\,)$ is superadditive if $c(u) = u$, i.e., in the case of linear costs with τ being the unit cost. Also, $c(\,\cdot\,,\tau)$ is convex and $c(\,\cdot\,,\,\cdot\,)$ is subadditive (resp, superadditive, additive) if for $u \ge 0$, $c(u)$ is $-u^\alpha$ with $0 < \alpha < 1$ (resp., u^α with $\alpha < 0$ or $1 < \alpha$, $-\ln u$).

Example 4. Product of an Arc's Cost and its Parameter. One is often interested in the impact of increasing or decreasing costs by a given percent. This is encompassed by putting $c(\xi,\tau) \equiv \tau c(\xi)$ with $c(\,\cdot\,)$ being $+\infty$ or real-valued. Then $c(\,\cdot\,,\,\cdot\,)$ is subadditive (resp., superadditive) if $c(\,\cdot\,)$ is decreasing (resp., increasing); and $c(\,\cdot\,,\tau)$ is convex if $c(\,\cdot\,)$ is convex and $\tau \ge 0$. In that event $100(\tau - 1)$ is the (possibly negative) percent increase in the cost function $c(\,\cdot\,)$.

Assumptions on the Arc Positions

In order to see why the *position* of two arcs in a network can prevent the optimal flow in one of the arcs from being monotone in the parameter of the other arc, consider the complete graph K_4 (or wheel W_4) on four nodes in Figure 8. Assume that $c_b(x_b, t_b) = \delta_0(x_b - t_b)$, assuring that $x_b = t_b$. Also assume that all other arc flows lie in the unit interval and have linear costs thereon. In particular, the unit costs are one for arcs d and e, and zero otherwise. Thus the arc costs are each convex in the arc flows and c_b is also subadditive. Finally, assume that all demands are zero.

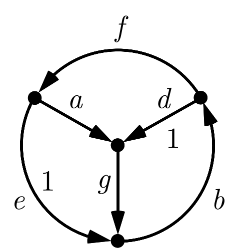


Figure 8. Complete Graph K_4 (or Wheel W_4) on Four Nodes

In this event, for t_b between zero and one, it is optimal to send t_b units along the cycle $b\ f\ a\ g$. However, when $t_b = 1$, arcs f, a and g become saturated so that increasing t_b above one requires sending $t_b - 1$ units along the only unsaturated cycle, viz., $b\ d\ a\ e$, thereby reducing the flow in arc a. Thus the optimal flow in arc a is $t_b \wedge (2 - t_b)$, and so increases with t_b on $[0,1]$ and decreases with t_b on $[1,2]$. Hence, the optimal flow in arc a is not monotone in t_b.

The difficulty in this example is that there are two simple cycles containing both arcs a and b with one cycle orienting the arcs in the same way and the other cycle orienting the arcs in opposite ways. Moreover, depending on the arc costs, increasing t_b by a small amount can cause the flow to increase on either cycle, so the optimal flow in arc a may rise or fall with t_b. This suggests that we shall need to restrict attention to pairs of arcs for which two such simple cycles do not exist. This turns out to be the condition on the position of the arcs required to assure the desired monotonicity of arc flows.

Substitutes and Complements in Biconnected Graphs

Call an arc in a biconnected graph a *complement* (resp., *substitute*) of a second arc if every simple cycle containing both arcs orients them in the same (resp., opposite) way. Call an arc *conformal* with a second arc if the former is either a complement or substitute of the second. No arc is both a complement and a substitute of a second arc. All three relations are symmetric. Here are a few examples.

- Every arc is a complement of itself.
- Two arcs sharing a node that is a head of one and a tail of the other are complements.

- Two distinct arcs with common heads or common tails are substitutes.

Substitutes and Complements in Planar Graphs

Another important example of conformal pairs of arcs arises in a special class of graphs called "planar". Call a graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ *planar* if it can be embedded in the plane in such a way that nodes are points and arcs are simple curves that intersect only at nodes. Examples of planar graphs include those of Figures 5, 7 and 8, as well as the production-planning graph of Figure 1.1. Although the complete graph K_4 on four nodes is planar, that is not so of the complete graph K_5 on five nodes or of the complete bipartite graph K_{33} as Figure 9 illustrates.[3]

Faces and Boundary of a Planar Graph. Each embedding of a planar graph divides the plane into connected regions, called *faces*, viz., the maximal open connected subsets not meeting any node or arc. If also the graph is biconnected, the arcs on the boundary of each of these faces form a simple cycle as all of the examples of planar graphs cited above (each of which is also biconnected) illustrate.

[3]Indeed, a famous result of Kuratowski asserts that a graph is planar if and only if it does not "contain" either of these two graphs.

Figure 9. Two Nonplanar Graphs

Conformal Arcs in Planar Graphs. Two arcs in a planar graph need not be conformal, e.g., a pair of node-disjoint arcs in the wheel W_4. On the other hand, two arcs on the boundary of a common face of a biconnected planar graph are conformal. This may be proved as follows. Suppose that (i, j) and (k, l) are arcs on the boundary of a common face of a biconnected planar graph as Figure 10 illustrates. Since the boundary of the face is a simple cycle, there is no loss of

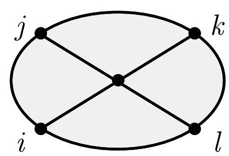


Figure 10. Arcs on the Boundary of a Face of a Biconnected Planar Graph

generality in assuming that i, j, k, l occur in clockwise order around the boundary of the face. If the arcs are not conformal, they must be node-disjoint and there must exist node-disjoint simple paths from j to l and from i to k. But because the graph is planar and the arcs lie on the boundary, those two paths must share a common node as Figure 10 illustrates—a contradiction. Indeed, it is known that two node-disjoint arcs in a triconnected planar graph are conformal if and only if they lie on the boundary of a common face.

Wheels. An important example of a triconnected planar graph is a *wheel* W_p on $p \geq 4$ nodes, i.e., a graph formed from a simple cycle with $p - 1$ (*rim*) *arcs* by appending a *hub node* and *spoke arcs* joining the hub node to each node in the cycle. As we shall see subsequently, the wheel on p nodes arises in the study of cyclic inventory problems with period $p - 1$. Then the spoke arcs are the production arcs and the rim arcs are the storage arcs in each period. Two arcs in a wheel are conformal if and only if they are incident or both lie on the rim of the wheel, and so on the boundary of a common face, viz., the exterior one. In particular, the rim arcs of the wheel W_5 of Figure 11 are complements, the distinct spoke arcs are substitutes, and a rim arc is a complement (resp., substitute) of a spoke arc if and only if the head of the spoke arc is the tail (resp., head) of the rim arc.

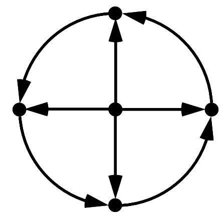


Figure 11. The Wheel W_5

Substitutes and Complements in Series-Parallel Graphs

As the wheel W_5 illustrates, in general some, but by no means all, pairs of arcs in a biconnected graph are conformal. Nevertheless, it is possible to characterize constructively the subclass of biconnected graphs, called *pairwise conformal*, in which every pair of arcs is conformal. In order to give the characterization, it is necessary to give a definition. Call a biconnected graph *series-parallel* if it can be constructed from a simple cycle on two nodes by means of a finite sequence of *series-parallel expansions* each of which involves replacing an arc either by

- two arcs in *series*, i.e., two arcs that respectively join the head and tail of the given arc to a new appended node, or
- two arcs in *parallel*, i.e., two arcs that both join the head and tail of the given arc.

Figure 12 illustrates these expansions. Since both of these expansions preserve biconnectedness and pairwise conformality, it is clear that biconnected series-parallel graphs are pairwise conformal. Indeed, it is known that the converse is also true. The series-parallel expansions also preserve planarity, so series-parallel graphs are planar.

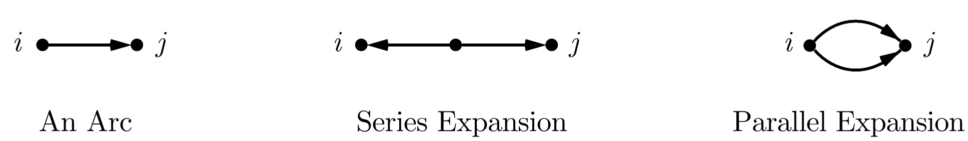


Figure 12. Series-Parallel Expansions

It is easy to check whether a biconnected graph is series-parallel by instead doing *series-parallel contractions*. In particular, if there are three or more arcs, look for a pair of arcs that is in series or in parallel and replace them by a single arc. Repeat these series-parallel contractions until no such pair exists. If the process terminates with two arcs in parallel, the original graph is series-parallel. In the contrary event the graph is not series-parallel.

As an example of series-parallel contractions, consider the production-planning graph of Figure 1.1. Observe that the production arc in the last period is in series with the inventory arc in the preceding period and so may be replaced by a single arc. The arc so formed is in parallel with the production arc in the preceding period and so may be deleted. This leaves a production-planning graph with one fewer periods. Repeating the above construction yields two arcs in parallel, viz., the production arc in the first period and a copy thereof. Thus the production-planning graph is series-parallel.

Another example of a series-parallel graph is given in Figure 7. An example of a planar graph that is not series-parallel is the wheel W_4 in Figure 8. This is because there is no pair of arcs that is in series or in parallel. Indeed, it can be shown that a biconnected graph is series-parallel if and only if it does not "contain" W_4.

Conformal Arcs

To sum up, a pair of arcs of a biconnected graph is conformal if any of the following three conditions holds:

- the arcs share a node,
- the arcs lie on the boundary of a single face of a planar graph or
- the graph is series-parallel.

4 MONOTONICITY OF THE OPTIMAL ARC FLOWS IN ARC PARAMETERS

We now come to our second main result. It asserts that in a network in which the graph is biconnected, the arc cost functions are convex in the arc flows and subadditive in the arc-flow parameter-pairs, and mild regularity hypotheses are satisfied, there is an optimal-flow selection $x(\cdot)$ for which the optimal flow in each arc is increasing (resp., decreasing) in the parameter associated with every arc that is a complement (resp., substitute) thereof. Moreover, $x(\cdot)$ has the *ripple property*, i.e., for each $t, t' \in T^o$ that differ only in a single component $b \in \mathcal{A}$ say, $x(t') - x(t)$ is a sum of simple conformal circulations each of whose induced cycle contains b. As we showed in proving Theorem 1, such a selection necessarily has the property that $|x_a(t') - x_a(t)|$ diminishes the less biconnected a is to b.

THEOREM 3. Monotone Optimal-Flow Selection. *In a biconnected graph, suppose $c_a(\cdot, \tau)$ is convex and lower semicontinuous for each $\tau \in T_a$ and c_a is subadditive for $a \in \mathcal{A}$, and $X(t)$ is nonempty and bounded for each $t \in T$. Then there is an iterated optimal-flow selection $x(\cdot)$ with the ripple property such that $x_a(t)$ is increasing* (*resp., decreasing*) *in t_b whenever a and b are complements* (*resp., substitutes*).

Before proving this result, two remarks are in order. First, the definition of the term "iterated" in the statement of the Theorem is deferred to Appendix 2. Second, observe that subadditivity of the arc flow costs is required only for arcs whose parameters are to be changed. This is because we can take T_a to be a singleton set for every other arc a, in which case the assumption that c_a is subadditive is automatically satisfied.

Since the cost of a flow is subadditive in the flow and the given arc parameter, one might hope to establish the above monotonicity result by applying the Increasing-Optimal-Selections Theorem. Unfortunately the set of flows is not a sublattice because each conservation-of-flow equation may contain two or more variables whose coefficients have the same sign (c.f., Example 6 of §2.2). Thus the above result does not apply directly to the minimum-cost-flow problem. Nevertheless, as we now show, the result can be applied to the projected problem in which one first optimizes over all arc flows other than the arc in question. Incidentally, this technique of optimizing out some of the variables, called *partial optimization*, is also extremely useful in many other applications of lattice programming where the original problem can not be expressed as that of minimizing a subadditive function over a sublattice, but a partially optimized projection of the problem can be expressed in that form.

LEMMA 4. Ascending Optimal-Arc-Flow Multi-Function. *In a biconnected graph, if T^o is a sublattice of T whose elements differ only in component $b \in \mathcal{A}$ and c_b is subadditive on $R \times \pi_b T^o$, then $\pi_b X(t)$ is ascending in t on T^o.*

Proof. Observe that for $t \in T^o$, the minimum cost $C_b(x_b, t_b)$ over all flows with given flow x_b in arc b can be expressed as

$$C_b(x_b, t_b) = c_b(x_b, t_b) + C_b(x_b)$$

where the last term is the minimum cost of flows in arcs other than b. Since $c_b(\,\cdot\,,\,\cdot\,)$, and hence $C_b(\,\cdot\,,\,\cdot\,)$, is subadditive on $\Re \times \pi_b T^o$ and $\mathcal{C} < +\infty$ on T^o, the result follows from the Increasing-Optimal-Selections Theorem of §2.5. ∎

Proof of Theorem 3 With Strict Convexity

We are now able to prove Theorem 3 for the case where the $c_a(\,\cdot\,, t_a)$ are strictly convex. Then $x(t)$ is unique. Now let $t' \in T$ be obtained from $t \in T$ by increasing t_b to t'_b. Then $x(t')$ is unique, and by the Ripple Theorem, $y \equiv x(t') - x(t)$ is a sum of simple conformal circulations $y^1, \ldots, y^k$, say, each of whose induced cycle contains b.

Now by Lemma 4 and the conformality of the simple circulations, their b^{th} components are each positive. Also if a and b are complements (resp., substitutes), then by the conformality of the simple circulations, $y_a^i \geq 0$ (resp., ≤ 0) for $i = 1, \ldots, k$ and so $y_a \geq 0$ (resp., ≤ 0). This is the desired result when the $c_a(\,\cdot\,, t_a)$ are strictly convex. ∎

Method of Proving Theorem 3 Without Strict Convexity

To complete the proof, it is necessary to consider the case where the $c_a(\,\cdot\,,t_a)$ are convex, but not necessarily strictly so. When that is so, there are generally many optimal-flow selections. Moreover, some of them do not have the monotonicity properties given in Theorem 3. For example, that is clearly the case where the arc flow costs are all identically zero because then every feasible flow is optimal for every choice of t.

One solution to this problem is to perturb the arc flow costs to make them strictly convex, let the perturbations converge to zero, and use the limit of the optimal perturbed flows. This is straight forward except for one point. Do the optimal perturbed flows converge. The answer is that in general they do not. However, we show in Appendix 2 how to do the perturbation in such a way that the optimal perturbed flows do indeed converge. The key idea is to perturb each arc flow cost so as to be subadditive in its arc flow and perturbation parameter, and to be strictly convex in its arc flow. This assures that the optimal perturbed flow in each arc is monotone in its perturbation parameter and, by the Ripple Theorem, majorizes changes in the optimal flows in the other arcs, from which facts the claim is established.

Simultaneous Changes in Several Arc Parameters

Observe that the Monotone-Optimal-Flow-Selection Theorem describes the effect of changes in a single arc parameter on the iterated optimal flow in other arcs that are conformal with it. If interest centers instead on simultaneous changes in several arc parameters, it is still possible to apply the Theorem by reducing such changes to a sequence of simple changes, each involving only a single arc parameter. However, then each of the intermediate parameter vectors that one constructs must feasible. Although that is often the case, it is by no mean always so as the discussion below makes clear.

Effect of Changes in Demands

It is often of interest to consider the effect of changing the demands in a subset S of the nodes $\mathcal{N}$. In order to reduce changes of this type to those studied herein, construct the *augmented network* illustrated in Figure 13 as follows. Append a new node τ having demand $\sum_{j\in S} d_j$ there, append arcs (i,τ) from each node $i \in S$ to τ with fixed flow d_i therein by letting d_i be the arc parameter and $\delta_0(\xi - d_i)$ be the arc flow cost, and replace the demand at each node $i \in S$ by zero. Now the sum of the changes in the demands at nodes in S must be zero to preserve feasibility. Thus it follows that if there is a feasible flow and one changes the demand at a single node in the original graph, there does not exist a feasible flow for the altered problem. This means that it is necessary to change the demands simultaneously at two or more nodes if the altered problem is to remain feasible.

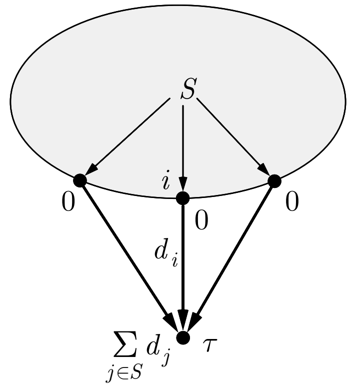


Figure 13. Augmented Network to Allow Changes in Demands at Nodes in S

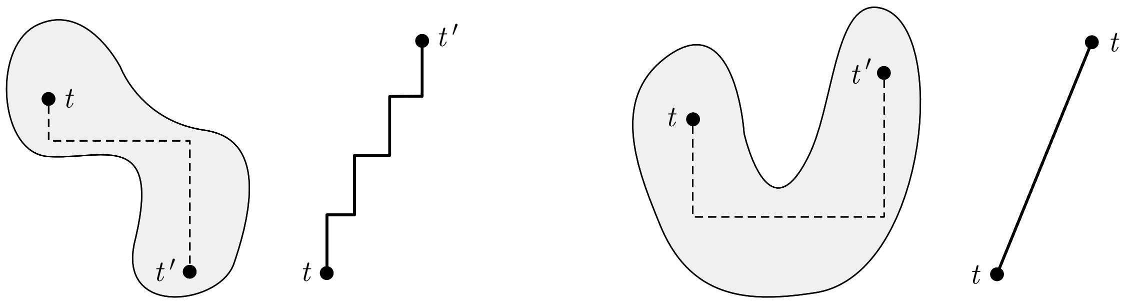
Suppose now that the flow cost in each arc a is convex in its flow and consider changing the vector of demands from one feasible vector $d = (d_i)$ to another $d' = (d_i')$ with $d_i' = d_i$ for all $i \in \mathcal{N} \setminus S$. Then there exist feasible flows x and x' for d and d' respectively. Now by the Circulation-Decomposition Theorem, the circulation $(x', d') - (x, d)$ is a sum $\sum_{k\in K} (x^k, d^k)$ of simple conformal circulations for some finite set K. Since each d^k is a subvector of a simple circulation and the arcs corresponding to all elements of d^k are incident to τ, d^k has exactly two nonzero elements with one being the negative of the other. Also since each arc flow cost is convex in its flow, it follows from the fact that the circulations are conformal that the flow $x + \sum_{k\in J} x^k$ is feasible for the demand vector $d + \sum_{k\in J} d^k$ for every subset J of K. (Observe that conformality assures that the flow in each arc a lies between x_a and x_a'). This reduces the problem of evaluating the effect of changing d to d' to a sequence of changes in which the demand at one node increases and that at another node simultaneously decreases by a like amount.

So far no consideration has been given to simultaneous changes of this type in pairs of arc parameters. However, it is possible to implement increasing the demand at one node i by δ and decreasing the demand at another node j by δ by changing the parameter of only a single arc. To see this, simply append an arc from i to j in the augmented network and fix the flow $d_{ij} = 0$ therein. Then observe that increasing d_i by δ and decreasing d_j by δ has the same effect as increasing d_{ij} by δ. The flow in arc (i, j) can be thought of as the incremental demand δ at i supplied from j. Now repeat a construction of this type for each d^k.

To sum up, it is possible to implement a change in d to d' as a sequence of feasible changes in which only a single arc's parameter is altered at each stage. Moreover, all changes in an arc's parameter are in the same direction because the d^k are conformal. Thus, in the terminology introduced below, two feasible demand vectors are equivalent to two feasible parameter vectors that are "monotonically step-connected" in the set of feasible parameter vectors. Thus the results of Theorems 6 and 7 below also apply to such changes.

Monotonic Step-Connectedness

Call t and t' in T *monotonically step-connected* if there is a finite sequence $t^0 \equiv t, t^1, \ldots, t^k \equiv t'$ in T such that t^i and t^{i-1} differ in at most one component for each i and t^i is *coordinate-wise monotone* in i (i.e., t_a^i is monotone in i for each $a \in \mathcal{A}$). Figure 14 illustrates this concept.



t and t' Monotonically Step-Connected t and t' Not Monotonically Step-Connected

Figure 14. Monotone Step-Connectedness

Weakening the Subadditivity Hypothesis on an Arc's Flow Cost

Observe that if one desires to compare optimal flows associated with two comparable values of an arc's parameter, the strongest form of the Monotone-Optimal-Flow-Selection Theorem is obtained by choosing the parameter set of the arc to be simply those two parameter values. To illustrate, suppose that a parameter associated with an arc is a vector $\tau = (\tau_1, \ldots, \tau_k)$ of real numbers. Then one may examine the effects of increasing several elements of τ by increasing them one at a time. In this event, to apply the Monotone-Optimal-Flow-Selection Theorem, it is not necessary to require the arc's flow cost $c(\xi, \tau)$ be jointly subadditive in ξ, τ, but rather that $c(\xi, \tau)$ merely be subadditive in ξ, τ_j for each j. This last hypothesis is weaker than joint subadditivity. For example, $c(\xi, \tau) = c((\sum_{i \in S} \tau_i) - \xi)$ is subadditive in ξ, τ_j for each $j \in S$ with S an arbitrary subset of the first k positive integers if and only if $c(\,\cdot\,)$ is convex. But $c(\xi, \tau)$ is jointly subadditive in ξ, τ if and only if $c(\,\cdot\,)$ is linear.

5 SMOOTHING THEOREM

Bounds on the Rate of Change of Optimal Arc Flows with Parameters

The Monotone-Optimal-Flow-Selection Theorem asserts that the optimal flow in one arc is monotone in the parameters of certain other arcs. We now explore when the *rate of change* of an optimal arc flow does not exceed that of an arc parameter changed. In order to motivate the result, consider the case of two nodes and two complementary arcs a and b, say, joining them, with $T_a \subseteq \Re$. If $c_a(x_a, t_a) = x_a^2 - 4x_a t_a$ and $c_b = 0$, then the cost of a flow is $c_a(x_a, t_a)$ and the hypotheses of Theorem 3 are satisfied. But $x_a = 2t_a$ is optimal and so increases twice as fast as t_a.

Thus, to assure that the optimal x_a does not increase faster than t_a, it is necessary to impose an additional condition on c_a. In order to see what condition will suffice, observe that the optimal $x_a(t_a)$ does not increase as fast as t_a if and only if $t_a - x_a(t_a)$ is increasing in t_a. This suggests making the change of variables $y_a \equiv t_a - x_a$ and seeking conditions under which y_a is increasing in t_a.

To this end, if c is a $+\infty$ or real-valued function of two real variables, define its *dual* $c^{\#}$ by the rule $c^{\#}(\xi,\tau) = c(\tau-\xi,\tau)$, so $c^{\#\#} = c$. Observe that $c_a^{\#}(y_a,t_a) = c_a(t_a - y_a, t_a) = c_a(x_a,t_a)$. Thus if $c_a^{\#}$ is subadditive, the set of optimal y_a will be ascending in t_a by Lemma 4, and if also c_a is subadditive, the set of optimal x_a will also be ascending in t_a by Lemma 4 again. This suggests the following definition.

Doubly Subadditive Functions

Call a $+\infty$ or real-valued function c of two real variables *doubly subadditive* if c and $c^{\#}$ are both subadditive. The class of doubly subadditive functions is clearly closed under addition, multiplication by nonnegative scalars, $+\infty$ or real-valued pointwise limits and taking duals. Two examples of doubly subadditive functions appear below. A more general doubly subadditive function is formed by taking a sum of four functions, one from each example and a dual of one from each example. Other examples arise by rescaling an arc's parameter by a increasing transformation.

Example 5. Arc Cost a Function Only of its Flow. The function $c(\xi,\tau) = c(\xi)$ and its dual $c^{\#}(\xi,\tau) = c(\tau-\xi)$ are both doubly subadditive if and only if $c(\,\cdot\,)$ is convex as in Example 1. In particular, this situation arises if the parameter is an upper or lower bound on, or a fixed value of, the flow.

Example 6. Arc Cost a Function Only of Join of Arc's Flow and Parameter. The function $c(\xi,\tau) = c(\xi \vee \tau)$ and its dual $c^{\#}(\xi,\tau) = c((\tau-\xi)\vee\tau)$ are both doubly subadditive if and only if $c(\,\cdot\,)$ is increasing and convex as in Example 2.

Observe that the doubly subadditive functions in Examples 5 and 6 are each convex. However, a doubly subadditive function need not be convex in either variable. For example, $c(\xi,\tau)$ is doubly subadditive if either τ is zero-one and $c(\xi,\tau) = c(\xi)$ with $c(\,\cdot\,)$ being periodic with period one or if $c(\xi,\tau) = c(\tau)$.

Smoothing Theorem

LEMMA 5. Ascending Arc-Parameter-Minus-Optimal-Flow Multifunction. *In a biconnected graph, if T^o is a chain whose elements differ only in component $b \in \mathcal{A}$, $T_b \subseteq \Re$, $c_b^{\#}$ is subadditive and $c_a(\,\cdot\,,t_a)$ is convex for $t_a \in \pi_a T^o$ and each $a \in \mathcal{A}\setminus\{b\}$, then $t_b - \pi_b X(t)$ is ascending in t on T^o.*

Proof. Suppose $t \in T^o$ and put $y_b = t_b - x_b$. Then on defining $C_b(\,\cdot\,,\,\cdot\,)$ and $C_b(\,\cdot\,)$ as in the proof of Lemma 4, observe that $C_b(x_b, t_b) = C_b^{\#}(y_b, t_b) = c_b^{\#}(y_b, t_b) + C_b(t_b - y_b)$ is subadditive on $\Re \times \pi_b T^o$ because $c_b^{\#}$ is subadditive and the minimum-cost $C_b(x_b)$ of flows in arcs other than b is convex by the Projection Theorem for convex functions. Now use $\mathcal{C} < +\infty$ on T^o and apply the Monotone-Optimal-Selections Theorem. ∎

Let $\|u\|_1 \equiv \sum_{a\in\mathcal{A}} |u_a|$ and $\|u\|_\infty \equiv \max_{a\in\mathcal{A}} |u_a|$ for $u \in \Re^{|\mathcal{A}|}$.

THEOREM 6. Smoothing. *In a biconnected graph, suppose $T_a \subseteq \Re$, $c_a(\,\cdot\,,\tau)$ is convex and lower semicontinuous for each $\tau \in T_a$ and c_a is doubly subadditive for all $a \in \mathcal{A}$; and $X(t)$ is nonempty and bounded for each $t \in T$. Then there is an iterated optimal-flow selection $x(\,\cdot\,)$ with the ripple property such that $x_a(t)$ and $t_b - x_a(t)$ (resp., $-t_b - x_a(t)$) are increasing (resp., decreasing) in t_b whenever a and b are complements (resp., substitutes). Moreover,*

$$\|x(t') - x(t)\|_\infty \le \|t' - t\|_1$$

for all monotonically step-connected t and t' in T.

Proof. Suppose a and b are complements (resp., substitutes). Then by the Monotone-Optimal-Flow-Selection Theorem, $x_a(t)$ is increasing (resp., decreasing) in t_b. Also by Lemma 3 of Appendix 2 and Lemma 5, $t_b - x_b(t)$ is increasing in t_b. Hence

$$t_b - x_a(t) = [t_b - x_b(t)] + [x_b(t) - x_a(t)]$$

$$(\text{resp., } -t_b - x_a(t) = [-t_b + x_b(t)] + [-x_b(t) - x_a(t)])$$

is increasing (resp., decreasing) in t_b because that is so of each of the two bracketed terms above, the latter by the Ripple Theorem.

Now suppose that t and t' differ in only one coordinate, say the b^{th}. Then since $a \le_b b$ for every arc a,

$$|x_a(t') - x_a(t)| \le |x_b(t') - x_b(t)| \le |t_b' - t_b|,$$

and so $\|x(t') - x(t)\|_\infty \le \|t' - t\|_1$.

Next consider the case where t and t' are monotonically step-connected. Then there exist $t^0 = t, t^1, \ldots, t^k = t'$ such that $t^i - t^{i-1}$ has only one nonzero element for each i and t^i is coordinatewise monotone in i. Thus from what was shown above,

$$\|x(t') - x(t)\|_\infty \le \sum_{i=1}^{k} \|x(t^i) - x(t^{i-1})\|_\infty \le \sum_{i=1}^{k} \|t^i - t^{i-1}\|_1 = \|t' - t\|_1,$$

the last equality following from the monotonic step-connectedness of t and t'. ∎

Application to Vector Parameters

One hypothesis of the Smoothing Theorem is that each arc's parameter is a real number, and so consideration of vector parameters would appear to be precluded. However, that the Theorem applies equally well to the case in which each arc parameter is a vector of real numbers, provided that the vector parameter set for each arc is monotonically step-connected. This is because, as §4.4 discusses, a change in a vector parameter can be expressed as a sequence of simple changes in which only a single element of the vector is changed. The Smoothing Theorem can be applied to each simple change of the parameter vector because that amounts to a change in a single real parameter.

As in §4.4, this approach to vector parameters reveals that each arc cost $c(\xi,\tau)$ as a function of its flow and vector parameter $\tau=(\tau_1,\ldots,\tau_k)$ must only be doubly subadditive in each pair ξ,τ_j for each j. The arc cost need not also be doubly subadditive in ξ,τ as we have assumed heretofore. One example of such a function is

$$c((\textstyle\sum_{i\in S}\tau_i)-\xi)\ (\text{resp., } c(\bigvee_{i\in S}(\xi\vee\tau_i))),$$

and its dual with respect to any τ_j with $1\le j\le k$, S a subset of the first k positive integers, and $c(\,\cdot\,)$ a convex (resp., increasing convex) function. It is easily verified from Example 5 (resp., 6) that the arc-cost function and its dual are doubly subadditive in ξ,τ_j for each j, but not subadditive in ξ,τ.

Rescaling a Parameter

The Smoothing Theorem can also be used to give bounds on the rate of change of optimal arc flows even when that rate exceeds the rate of change of the parameters. This is accomplished by simply *rescaling* the original parameter t_b by means of a strictly increasing function ϕ_b from T_b into itself. On putting $\underline{t}_b\equiv\phi_b(t_b)$, the arc cost can be expressed in terms of the rescaled parameter $\underline{t}_b$ by putting $\underline{c}(x_b,\underline{t}_b)\equiv c_b(x_b,t_b)$. Now since ϕ_b is monotone, $\underline{c}_b$ is subadditive because that is so of c_b. Thus it suffices to choose ϕ_b so that $\underline{c}_b^{\#}$ is also subadditive. Then if t and t' in T differ only in their b^{th} elements, the iterated optimal-flow selection $x(\,\cdot\,)$ satisfies

$$\|x(t')-x(t)\|_\infty\le|\phi_b(t'_b)-\phi_b(t_b)|. \tag{3}$$

Below are two examples of this technique.

Example 7. Arc Cost a Quadratic Form Strictly Convex in its Flow. It is always possible to linearly rescale arc b's parameter $\tau=t_b$ so that the arc flow cost is doubly subadditive whenever it is a quadratic form that is strictly convex in the arc flow ξ. To see this, observe that such a flow cost can be expressed in the form $\alpha c(\xi,\tau)+\beta\tau^2$ with $c(\xi,\tau)=(\xi-\gamma\tau)^2$ for some constants $\alpha > 0$ and β, γ. Moreover, the dual of the flow cost is $\alpha c^{\#}(\xi,\tau) + \beta\tau^2$. Now choose $\underline{\tau} \equiv \phi_b(\tau) \equiv \gamma\tau$. Then $\underline{c}(\xi,\underline{\tau}) = (\xi - \underline{\tau})^2$ is doubly subadditive in $\xi, \underline{\tau}$. Thus conclude from the Smoothing Theorem that if $\gamma > 0$ (resp., $\gamma < 0$), then the iterated optimal flow $x_b(t)$ in arc b does not increase (resp., decrease) faster than $|\gamma|$ times the rate of change of t_b.

Example 8. Product of Arc's Parameter and Exponential Function of Flow. If the flow cost for an arc b can be expressed in the form $c(\xi,\tau) = \tau e^{-\xi}$ for $\tau > 0$ and all real ξ, then c is subadditive. On putting $\underline{\tau} = \ln\tau$, we have that $\underline{c}(\xi,\underline{\tau}) = e^{-(\xi-\underline{\tau})}$ which is doubly subadditive because it is a convex function of the difference of the flow and the transformed parameter as Example 5 discusses. Thus it follows from the Smoothing Theorem that the iterated optimal flow $x_b(t)$ in arc b does not change faster than $\ln t_b$ does. As a particular illustration of this result, observe that the perturbed convex flow cost for each arc b given in (1) of Appendix 2 is the sum of a convex function of the arc flow and a function of the form discussed here with perturbation parameter ϵ_b. The sum of these two functions is doubly subadditive in x_b, $\ln\epsilon_b$. One implication of this result is that the iterated optimal perturbed flow in arc b may grow as fast as $\ln\epsilon_b$.

6 UNIT PARAMETER CHANGES

Call a $+\infty$ or real-valued function on the real line *affine between integers* if it is affine on each closed interval whose end-points are successive integers having finite function values, and is $+\infty$ otherwise. Figure 15 illustrates of this concept. When each arc flow cost is affine between integers in its arc flow, it is possible to refine the Ripple, Monotone-Optimal-Flow-Selection, and Smoothing Theorems.

THEOREM 7. Unit Parameter Changes. *In a biconnected graph, suppose $T_a \subseteq \Re$ for each $a \in \mathcal{A}$; t, t' are integer elements of T that differ only in component $b \in \mathcal{A}$ with $t'_b = t_b + 1$ (resp., $t'_b = t_b - 1$); $c_a(\,\cdot\,,\tau)$ is affine between integers for each $\tau \in \{t_a, t'_a\}$ and $a \in \mathcal{A}$, and is convex for each $a \in \mathcal{A} \setminus \{b\}$; c_b is doubly subadditive; $x \in X(t)$ is integer and $X(t')$ is nonempty. Then either $x \in X(t')$ or there is an $x' \in X(t')$ such that $x' - x$ is a simple circulation whose induced cycle contains b and (c.f., Figure 16)*

$$x'_b = x_b + 1 \text{ (resp., } x'_b = x_b - 1). \tag{4}$$

Proof. Since x is integer, the demands are integers. We claim that there is an integer element of $X(t')$. To see this, observe that by hypothesis there is an $x'' \in X(t')$. If x'' is not integer, then fix the integer elements of x'' and allow the others to vary only between the integers obtained by rounding the fractional flows up and down. The restricted problem is one of finding a minimum-linear-cost flow with integer demands and integer upper and lower bounds on each arc flow. Hence there is an integer element of $X(t')$ as asserted.

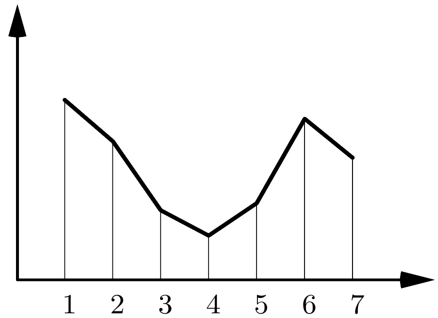


Figure 15. A Function that is Affine Between Integers

Thus, by the Ripple Theorem, there is an integer $x' \in X(t')$ such that $x' - x$ is a sum of simple conformal integer circulations each of whose induced cycle contains b. Hence it suffices to show that either $x \in X(t')$ or (4) holds. To that end observe that either $x'_b > x_b$ (resp., $x'_b < x_b$), or by Lemma 4, $x_b \in \pi_b X(t')$, whence $x \in X(t')$. Similarly, either $x'_b \le x_b + 1$ (resp., $x'_b \ge x_b - 1$) or, by Lemma 5, $x_b \in \pi_b X(t')$, whence $x \in X(t')$, because $t_b - x_b \le$ (resp., $\ge$) $t'_b - x'_b$ Thus, because x' is integer, either $x \in X(t')$ or (4) holds. ∎

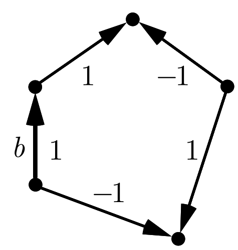


Figure 16. Simple Circulation $x' - x$ Induced by the Change $t'_b = t_b + 1$

Remark. The hypotheses on the flow cost in arc b are satisfied if that flow cost is a sum of functions $c(\,\cdot\,,\,\cdot\,)$ of the types in Examples 5 and 6 in which the associated functions $c(\,\cdot\,)$ are themselves affine between integers.

Minimum-Cost Cycles

Suppose now that $b = (k,l)$ is an arc, t, t' in T^o, $\pm(t' - t)$ is the b^{th} unit vector and $x \in X(t)$ is integer. Then Theorem 4 reduces the problem of finding an integer $x' \in X(t')$ to that of finding a minimum-cost simple cycle containing arc b where the arc costs are defined below. Such a cycle can easily be found using standard methods for finding a minimum-cost simple path (excluding b) from l to k. To be specific, let F_i be the set of nodes that are heads of arcs in $\mathcal{A}_i^+$ or tails of arcs in $\mathcal{A}_i^-$ and

$$c_a^{\pm} = \begin{cases} c_a(x_a \pm 1, t_a) - c_a(x_a, t_a),\ a \ne b \\ \\ c_b(x_b \pm 1, t'_b) - c_b(x_b, t_b)\,,\ a = b. \end{cases}$$

Let c_{ij}^{+} (resp., c_{ij}^{-}) be the smallest of the arc costs c_a^{+}(resp., c_a^{-}) among all arcs $a \neq b$ with tail i and head j. If $t'_b = t_b + 1$ (resp., $t'_b = t_b - 1$), then put

$$c_{ij} = c_{ij}^{+} \wedge c_{ji}^{-} \text{ (resp., } c_{ij} = c_{ij}^{-} \wedge c_{ji}^{+} \text{).}$$

Since x is optimal for t, there cannot exist a cycle not containing b with negative cycle cost. Denote by C_i the minimum cost among all simple paths (excluding b) from i to k, so $C_k = 0$. The remaining C_i may be determined by solving Bellman's equation

$$C_i = \min_{j \in F_i} [c_{ij} + C_j] \text{ for } i \in \mathcal{N} \setminus \{k\}. \tag{5}$$

One way of doing this is to let C_i^m be the minimum-cost among all simple paths (excluding b) from i to k with m arcs or less. Then

$$C_i^m = \min_{j \in F_i} [c_{ij} + C_j^{m-1}] \text{ for } i \in \mathcal{N} \setminus \{k\} \tag{6}$$

for $m = 1, \ldots, n-1$ where $n \equiv |\mathcal{N}|$, $C_k^0 = 0$ and $C_j^0 = \infty$ otherwise. Moreover, $C_i^{n-1} = C_i$ is the desired minimum cost from i to k. If $c_b^{+} + C_l \geq 0$ (resp., $c_b^{-} + C_l \geq 0$), then one should choose $x' = x$. In the contrary event, a simple path from k to l is found from (6). If the arc b is appended to this path, the result is the simple cycle induced by the simple circulation $x' - x$.

A Parametric Algorithm

Theorem 7 suggests a simple parametric algorithm for finding an optimal flow when all arc-flow costs are doubly subadditive, and are convex and affine between integers in their flows. Let $t^0, \ldots, t^p = t$ be a sequence of integer parameter vectors in T^o with the property that $\pm(t^i - t^{i-1})$ is a unit vector for $i = 1, \ldots, p$. Also suppose t^0 is such that there is an "obvious" choice of an integer optimal flow x. Then given an integer optimal flow x^{i-1} for t^{i-1}, one constructs an integer optimal flow x^i for t^i by applying Theorem 7 and the algorithm for finding minimum-cost cycles discussed above. This is possible because $\mathbb{C}(t^i - t^{i-1})$ is a unit vector. With this algorithm $p = \|t - t^0\|_1$ if t^i is coordinatewise monotone in i. Incidentally, this algorithm is essentially a specialized refinement of the "out-of-kilter" method [Fu61].

7 SOME TIPS FOR APPLICATIONS

In order to efficiently obtain the most qualitative information about the effect of changes in the parameters of a minimum-cost network flow problem on the optimal arc flows by using the theory of substitutes, complements and ripples is often useful to take advantage of a few tips that simplify the task of applying the theory.

Omit Unnecessary Arcs. Always formulate a problem with as few arcs as possible. For example, if the flow cost associated with an arc is identically zero, shrink the arc so that its end nodes coalesce, thereby eliminating one arc and one node. The reason for doing this, whenever possible, is that the smaller the set of arcs in a graph, the larger the set of pairs of arcs that are conformal (resp., comparable under the less-biconnected-to relation).

Series-Parallel Contractions. In order to simplify the task of determining which pairs of arcs in a biconnected graph $\mathcal{G}$ are conformal, it is a good idea to repeatedly replace each pair of arcs that is in series or in parallel with a single arc until no such pair of arcs remains. This leaves a *series-parallel-free* graph $\mathbb{G}$. The reason for doing this is that in practice $\mathbb{G}$ is frequently much smaller than $\mathcal{G}$ and so it is easier to identify conformal pairs of arcs in $\mathbb{G}$ than in $\mathcal{G}$. Moreover, by construction, each arc a in $\mathbb{G}$ is a *series-parallel-contraction* of a subgraph $\mathcal{G}_a$ of $\mathcal{G}$. Also two arcs in $\mathcal{G}$ are conformal if and only if their series-parallel-contractions are conformal in $\mathbb{G}$. In particular, two arcs in $\mathcal{G}$ whose series-parallel-contractions coincide are certainly conformal.

Use Equivalent Networks in Different Variables and Parameters. It is often the case that there are several equivalent formulations of a problem in terms of network flows in different variables and parameters. When this is so, the theory of substitutes, complements and ripples should be applied to each equivalent formulation to obtain the most information about the qualitative impact of parameter changes. As we shall see in §4.8, this technique is useful in the study of dynamic inventory problems.

Changing an Upper or Lower Bound on an Arc Flow. A lower (resp., upper) bound on the flow in an arc is *inactive* if it is strictly below (resp., above) the optimal flow (in an optimal flow selection) in the arc. Otherwise the bound is *active*. *Tightening* (resp., *loosening*) a bound on an arc flow means moving the bound in a direction that reduces (resp., enlarges) the set of feasible flows. For example, raising a lower bound tightens the bound while lowering the bound loosens it. Tightening an inactive bound does not change the optimal flow until the bound becomes active, at which point further tightening maintains the active status of the bound. Dually, loosening an active bound maintains the active status of the bound until it becomes inactive, at which point further loosening maintains the inactive status of the bound.

These facts may easily be seen as follows. Let $C_b(x_b)$ be the minimum-cost over all flows with given flow x_b (ignoring the lower or upper bound on x_b) in arc b and let t_b be a lower (resp., upper) bound on x_b. By the convexity and lower semi-continuity of the flow costs, the boundedness of the set of optimal flows, and the Projection Theorem for convex functions, $C_b(\,\cdot\,)$ is convex and lower semi-continuous. Now extend $C_b(\,\cdot\,)$ to the extended real line by putting $C_b(\pm\infty) \equiv \lim_{z\to\pm\infty} C_b(z)$. Then $C_b(\,\cdot\,)$ has a (possibly infinite) global minimum $\underline{x}_b^o$ say. Also the global minimum x_b^o of $C_b(x_b)$ subject to the lower- (resp., upper-) bound constraint $x_b \ge$ (resp., $\le$) t_b is evidently given by $x_b^o = \underline{x}_b^o \vee t_b$ (resp., $\underline{x}_b^o \wedge t_b$), from which the above claims are immediate.

Changing Parameters that Enter the Flow Costs of Several Arcs. In practice, a parameter may enter the flow cost of several arcs. One way to deal with changes in such parameters is to view them as sequences of separate changes in the parameters of each arc on which the flow cost depends. For example, a change in a common upper bound on several arcs can be viewed as a sequence of changes in separate upper bounds on those arcs which restores equality of the upper bounds on the arcs only after all changes are completed.

Costs Depending on Net Sums of Arc Flows in Arcs Incident to a Node. It is often the case that there are costs or bounds associated with the net sum of the arc flows in a specified subset $\mathcal{S}$ of the arcs that are incident to a common node i say. By the net sum we mean the sum of the flows in arcs in $\mathcal{S}$ whose tail is i minus the sum of the flows in arcs in $\mathcal{S}$ whose head is i. Such a non-arc-additive arc-flow cost network-flow problem is easily reduced to an arc-additive one by appending a new node τ, an arc joining i and τ with flow therein equaling the desired net sum of arc flows, and replacing the end node i of each arc in $\mathcal{S}$ by τ as Figure 17 illustrates.

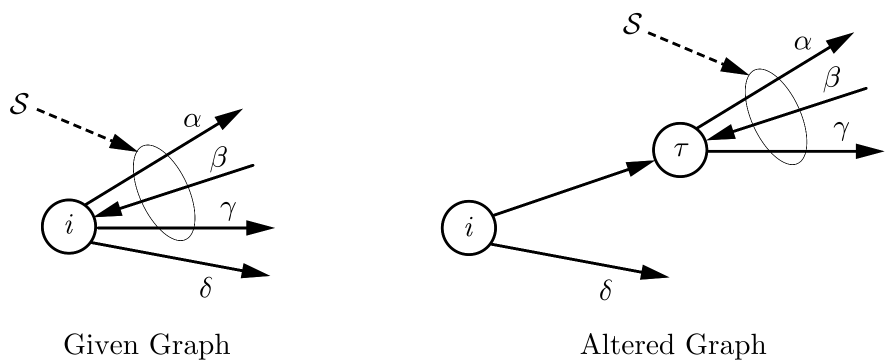


Figure 17. Cost of Net Sum of Arc Flows Incident to a Node

8 APPLICATION TO OPTIMAL DYNAMIC PRODUCTION PLANNING [Jo57], [KV57a], [Dr57], [Da59], [Ve64], [Ve66b], [HMMS60], [CGV90b]

The purpose of this subsection is to formulate a fundamental single-item multi-period inventory model as a minimum-convex-cost network-flow problem and to study its qualitative properties with the aid of the theory of substitutes, complements, and ripples developed above. To this end, let x_i, s_i and y_i denote respectively the amounts produced (or ordered) in, sold in, and stored at the end of period i, $1 \le i \le n$. Negative values of x_i, s_i and y_i signify respectively disposal of stock, return of sales, and backlogged demand. Because there is conservation of stock in each period,

$$x_i - s_i + y_{i-1} - y_i = 0,\ 1 \le i \le n. \tag{7}$$

Assume without loss of generality that $y_0 = y_n = 0$. This is a network-flow model with zero demands at each node. Figure 18 illustrates the associated graph for $n = 4$.

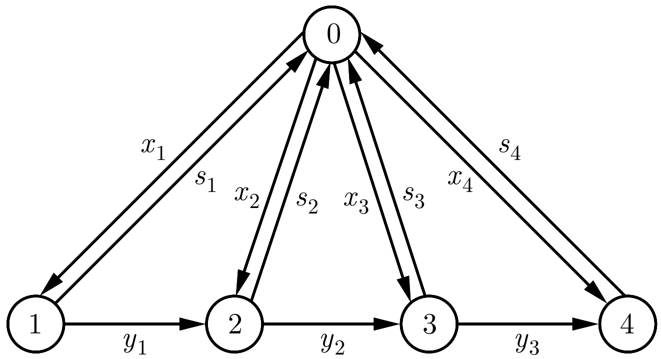


Figure 18. Single-Item Production/Inventory Network

Now suppose there is a cost $c_i(x_i, t_i)$ of producing x_i units in period i when the production parameter is t_i in that period. The parameter t_i may represent an upper or lower bound on production in period i, or a parameter of the cost function. Similarly, there is a cost $h_i(y_i, u_i)$ of storing y_i units in period i when u_i is the storage parameter in that period. Finally, there is revenue $-r_i(s_i, v_i)$ resulting from the sale of s_i units in period i when the sales parameter is v_i in that period. In practice decision makers often control sales only indirectly by setting price. As Figure 19 illustrates, sales is usually a strictly decreasing function of price—especially in monopolies or oligopolies. Then the price $\pi_i(s_i)$ needed to assure a given level s_i of sales is a strictly decreasing function of sales s_i. In this event, $-r_i(s_i, v_i) = \pi_i(s_i)s_i$. Hence it is convenient (and equiv-

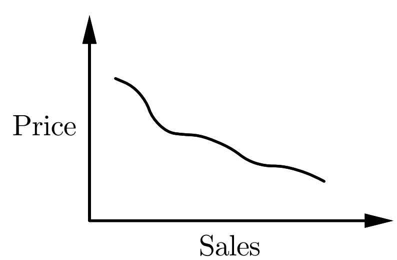


Figure 19. Price-Sales Graph

alent) to think in terms of controlling sales with the understanding that the price is chosen to assure the desired level of sales. However, it should be recognized that the assumption below that $r_i(\,\cdot\,, v_i)$ is convex has implications for the price elasticity of sales in this setting. The total cost is thus

$$\sum_{i=1}^{n} [c_i(x_i, t_i) + h_i(y_i, u_i) + r_i(s_i, v_i)]. \tag{8}$$

The problem is to find a schedule of production, storage and sales that minimizes the total cost (8) subject to the stock-conservation constraint (7).

Assume now that c_i, h_i, and r_i are $+\infty$ or real-valued, subadditive, and convex and lower semicontinuous in the first variable. This encompasses cases where the t_i, u_i, and v_i are upper or lower bounds, or are fixed values of the variables. For example, if the sales in period i is known to be v_i, put $r_i(s_i, v_i) = \delta_0(s_i - v_i)$. Observe that the motive for carrying inventories in this model is that there may be a temporal increase in the marginal cost of supplying demand.

Effect of Changes in Parameters on Optimal Production, Inventories and Sales

The graph in Figure 18 is series-parallel so that, as Table 1 summarizes, each two arcs are either substitutes (indicated by $\mathbb{S}$) or complements (indicated by $\mathbb{C}$). This fact and the Monotone-Optimal-Flow-Selection Theorem determine the direction of change of optimal production, inventory, and sales in each period as a result of changes in the production, storage and sales parameters. Table 2 summarizes the results with the arrows indicating the direction in which one set of optimal variables change as the parameters increase. (Of course $\uparrow$ means increasing and $\downarrow$

TABLE 1. Substitutes and Complements in Production, Inventories and Sales

	x_j	y_j	s_j
x_i	$\mathbb{S}\,(i \neq j)$ $\mathbb{C}\,(i = j)$	$\mathbb{S}\,(i > j)$ $\mathbb{C}\,(i \leq j)$	$\mathbb{C}$
y_i	$\mathbb{S}\,(i < j)$ $\mathbb{C}\,(i \geq j)$	$\mathbb{C}$	$\mathbb{S}\,(i \geq j)$ $\mathbb{C}\,(i < j)$
s_i	$\mathbb{C}$	$\mathbb{S}\,(i \leq j)$ $\mathbb{C}\,(i > j)$	$\mathbb{S}\,(i \neq j)$ $\mathbb{C}\,(i = j)$

TABLE 2. Monotonicity of Optimal Production, Inventories and Sales in Parameters

	$t_j \uparrow$	$u_j \uparrow$	$v_j \uparrow$
x_i	$\downarrow (i \neq j)$ $\uparrow (i = j)$	$\downarrow (i > j)$ $\uparrow (i \leq j)$	$\uparrow$
y_i	$\downarrow (i < j)$ $\uparrow (i \geq j)$	$\uparrow$	$\downarrow (i \geq j)$ $\uparrow (i < j)$
s_i	$\uparrow$	$\downarrow (i \leq j)$ $\uparrow (i > j)$	$\downarrow (i \neq j)$ $\uparrow (i = j)$

means decreasing.) For example, x_i is increasing in u_j for $i \le j$ and decreasing in u_j for $i > j$. Now apply these results and the Ripple Theorem to explore their implications in several practical problems.

Example 9. Increase in Wage Rates and Technological Improvements. Suppose there is a temporary fixed percent increase in wage rates, and the same percent increase in other production costs, in period j, say. How should optimal production, storage and sales decisions respond? To answer this question, put

$$c_j(x_j,t_j) = \begin{cases} t_j c_j(x_j), x_j \ge 0 \\ \\ c_j(x_j),\ x_j < 0 \end{cases}$$

where t_j is the index of wage rates in period j. Current wage rates correspond to $t_j = 1$, and an increase therein corresponds to $t_j \ge 1$. Then $c_j(\,\cdot\,,\cdot\,)$ is superadditive and $c_j(\,\cdot\,,t_j)$ is convex provided that $c_j(\,\cdot\,)$ is convex, $c_j(\,\cdot\,)$ is increasing on $[0,\infty)$ and $c_j(0) = 0$. In that event, increasing wage rates in period j, i.e., increasing t_j, has the following effects on optimal production, storage and sales decisions.

- Sales fall (i.e., prices rise) in all periods and total production falls by the amount that total sales falls.
- Production falls in period j and rises in all other periods.
- Inventories rise before and fall after period j, and, by the Ripple Theorem, the absolute change in inventories in period i is quasiconcave in i with maximum at $i-1$ or i.
- The change in production in period j exceeds all other changes.

If the increase in wage rates in period j is *permanent*, i.e., it is applicable in period j and thereafter, then the conclusions given above remain valid before period j, but not in period j or thereafter, except that sales do fall in every period in this case as well. Also total production in periods $j,\dots,n$ falls by at least as much as total sales in those periods falls.

The effect of technological improvements is usually to reduce production costs. If the percent reduction is the same at all production levels, the impact can be studied with the above model by instead reducing the index t_j to $0 \le t_j \le 1$. Then $c_j(\,\cdot\,,\,\cdot\,)$ remains superadditive. However, in order to assure that $c_j(\,\cdot\,,t_j)$ remains convex, it is necessary to assume also that $c_j(\,\cdot\,)$ is decreasing on $(-\infty,0]$. Then the qualitative impact of temporary or permanent technological improvements in period j is precisely opposite to that of higher wage rates.

Example 10. Increase in Market Price. Suppose the product is sold in a competitive market with the price in period i being v_i. Then $r_i(s_i,v_i) = -s_iv_i$, which is subadditive. Thus an increase in market price in period j has the following effects on optimal production, storage and sales decisions.

- Production rises in every period.
- Sales rise in period j and fall in all other periods.
- Total sales rise by the amount total production rises.
- Inventories behave as in Example 9 above.
- The change in sales in period j exceeds all other changes.

Example 11. Increase in Fixed Sales. Suppose there are fixed levels v_i of sales that must be met in each period i. Then $r_i(s_i, v_i) = \delta_0(s_i - v_i)$ is doubly subadditive. Thus the effect of increasing sales in period j is like a price increase in that period as Example 10 discusses, except, of course, the fixed sales in other periods remain unchanged. Also, if s and s' are two sales schedules, x, y is optimal for s, and there is an optimal schedule for s', then one such pair x', y' is such that $\|(x,y) - (x',y')\|_\infty \le \|s - s'\|_1$.

Effect of Parameter Changes on Optimal Cumulative Production

We can obtain additional qualitative properties of optimal schedules by making a change of variables. To this end let $X_i \equiv \sum_{j \le i} x_j$ and $S_i \equiv \sum_{j \le i} s_j$ be respectively the cumulative production and (fixed) sales in periods $1, \ldots, i$. Then, under the assumptions of Example 11 above (i.e., fixed sales in each period) with $h_i(y_i, u_i) = h_i(y_i)$ for all i and $h_n(\,\cdot\,) = \delta_0(\,\cdot\,)$, (7) and (8) can be rewritten as

$$(7)' \qquad x_i + X_{i-1} - X_i = 0,\ 1 \le i \le n$$

and

$$(8)' \qquad \sum_{i=1}^{n} [c_i(x_i, t_i) + h_i(X_i - S_i)]$$

where $X_0 \equiv 0$. The restrictions $(7)'$ describe a set of circulations as Figure 20 illustrates for $n = 4$. Notice that since $h_i(\,\cdot\,)$ is convex, $h_i(X_i - S_i)$ is doubly subadditive.

Observe that the graph in Figure 20 is also series-parallel, so that all pairs of arcs are either substitutes or complements as Table 3 indicates. Thus one optimal cumulative production schedule has the following properties.

- Cumulative production increases with cumulative sales.
- Cumulative production increases slower than cumulative sales.

Incidentally, recall that §2.6 obtains both of these results by another method, viz., directly from the Increasing-Optimal-Selections Theorem of Lattice Programming. However, with the present

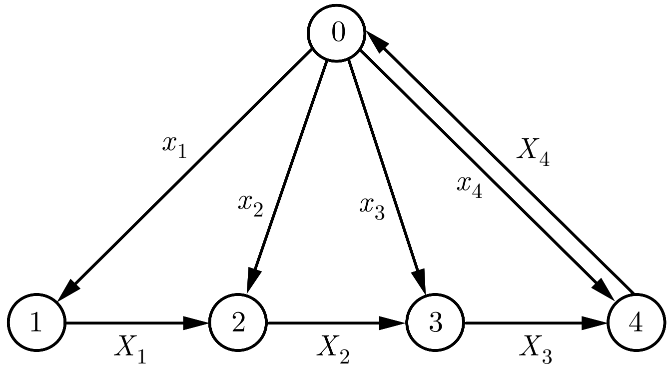


Figure 20. Cumulative-Production Network

TABLE 3. Substitutes and Complements in Production

	x_j	X_j
x_i	$\mathbb{S}$ $(i \neq j)$	$\mathbb{S}$ $(i > j)$
	$\mathbb{C}$ $(i = j)$	$\mathbb{C}$ $(i \leq j)$
X_i	$\mathbb{S}$ $(i < j)$	$\mathbb{C}$
	$\mathbb{C}$ $(i \geq j)$	

method we can go farther and show that an increase in cumulative sales in period j has the following effects on optimal production.

- Production rises before and in period j, and falls thereafter.
- The change in cumulative sales in period j exceeds all other changes.

One implication of the above results is that increasing sales in a given period increases cumulative production in every period and actual production before or in the given period, but the size of the increase in production diminishes the further the given period is in the future.

All of the above results can also be obtained by a direct analysis of the original network as Figure 18 illustrates. This is because increasing S_j by $\bar{\sigma} > 0$, say, can be represented in that network by appending an arc $\alpha = (j, j{+}1)$ from node j to node $j{+}1$ carrying flow $\sigma = \bar{\sigma}$ as discussed in §4.4 on the effect of changes in demand. Now by the Ripple Theorem, increasing the flow σ in α from 0 to $\bar{\sigma}$ increases the flow in one or more simple conformal circulations containing α, with the sum of the flows in each equaling $\bar{\sigma}$. From the structure of the network, each such simple circulation carrying flow $\epsilon > 0$, say, carries flow ϵ in arc α and either entails (i) storing ϵ less in period j or (ii) producing ϵ more in some period $i \leq j$, producing ϵ less in some period $k > j$ and storing ϵ more in periods $i, \ldots, j{-}1, j{+}1, \ldots, k{-}1$. The claimed results then follow readily from these facts.

Algorithms

The qualitative results given for each of the above network-flow formulations can be used to give parametric algorithms for finding optimal integer schedules as discussed following the Unit-Parameter-Changes Theorem. Alternately, this problem may be solved by any standard algorithm for finding minimum-convex-cost flows.

5

Convex-Cost Network Flows and Supply Chains: Invariance

[Ve71, Ve87]

1 INTRODUCTION

This section develops a different class of results about minimum-convex-cost network-flow problems in which there are upper and lower bounds on the flow in each arc and interest centers on the "invariance" of optimal solutions of the primal problem and its (Lagrangian) dual. The primal form of this result asserts that if there is a flow, then one such flow simultaneously minimizes every "d-additive-convex" function of the flows in the arcs emanating from a single node of the graph. Thus the optimal flow is invariant over the class of d-additive-convex functions. The corresponding result for the dual of the problem states that the "order" of the optimal dual variables—but not their values—remains invariant as the "conjugate" of the primal objective function ranges over the d-additive-convex functions.

In the special case in which the graph has the form of the production-planning graph, the problem and its dual admit a graphical method of solution. The method entails threading a string between two rigid curves in the plane and pulling the string taut. This "taut-string" solution can be computed in linear time. One example of a primal problem that can be solved by this method is that of optimal production planning in a single- or multi-facility supply chain with time-dependent additive homogeneous convex production costs, capital charges associated with investment in inventories, and upper and lower bounds on inventories. An example of a dual problem that can be solved by the taut-string method is that of choosing the minimum-expected-cost amounts of a product to stock at each facility of a serial supply chain with random demands at the retailer. Another entails choosing the maximum-profit times to buy and sell shares of a single stock when there are transaction costs and the schedule of buy and sell prices is known.

In order to develop these ideas, it is first necessary to discuss the notions of conjugate functions and subgradients.

2 CONJUGATES AND SUBGRADIENTS [Ro70]

Call a $+\infty$ or real-valued function f on $\Re^n$ *proper* if f is finite somewhere and is bounded below by an affine function. If f is proper, so is its *conjugate* f^* defined by

$$f^*(x^*) \equiv \sup_x [\langle x, x^* \rangle - f(x)] \text{ for } x^* \in \Re^n \tag{1}$$

where $\langle \cdot , \cdot \rangle$ is an "inner product" on $\Re^n$. To see this, observe that since f is finite at some x, say, then f^* is bounded below by the affine function $\langle x, \cdot \rangle - f(x)$. Similarly, since f is bounded below by an affine function $\langle \cdot , x^* \rangle - u$, say, $\langle x, x^* \rangle - f(x) \le u$ for all x so $f^*(x^*) \le u < \infty$ as claimed.

Observe that f^* is convex because it is a supremum of affine functions, even if f is not convex. The geometric interpretation of the conjugate is given in Figure 21.

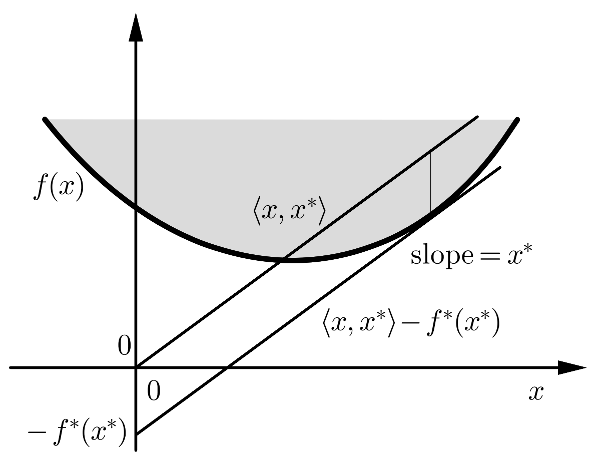
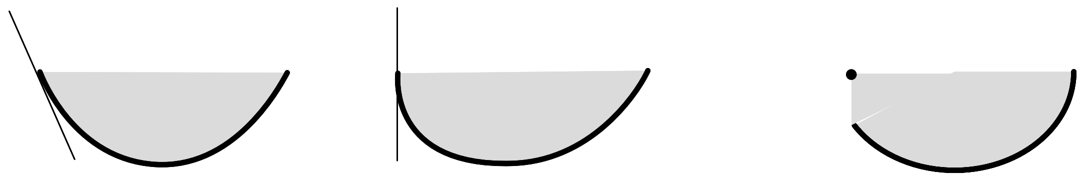


Figure 1. Conjugate Functions

If f is a proper function, define its *closure* $\operatorname{cl} f$ by

$$(\operatorname{cl} f)(x) \equiv \sup_{f(\cdot) \ge \langle \cdot, x^* \rangle - u^*} [\langle x, x^* \rangle - u^*] \tag{2}$$

i.e., $\operatorname{cl} f$ is the supremum of all affine functions lying below f. Call a proper function *closed* if $\operatorname{cl} f = f$. Closed functions are convex, but the converse need not be so as Figure 2 illustrates.



Sup Attained Sup Not Attained

Closed Nonclosed

Figure 2. Closed and Nonclosed Convex Functions

Also f^* is closed for every proper function f. Now from (1) and (2) (since $u^* = f^*(x^*)$ in (2))

$$f^{**}(x) = \sup_{x^*} [\langle x, x^* \rangle - f^*(x^*)] = \operatorname{cl} f(x). \tag{3}$$

(These facts should be checked geometrically in Figure 1.) Thus, if f is a closed proper (convex) function, then $f^{**} = f$.

If f is a proper convex function on $\Re^n$ and $f(x)$ is finite, denote by $\partial f(x)$ the set of *subgradients* of f at x, i.e., the set of vectors x^* such that

$$f(y) \geq f(x) + \langle y - x, x^* \rangle \text{ for all } y \in \Re^n, \tag{4}$$

or equivalently, as Figure 3 illustrates, for some $u^* \in \Re$,

$$\begin{aligned} &f(y) \geq \langle y, x^* \rangle - u^* \text{ for all } y \in \Re^n \text{ and} \\ &f(x) = \langle x, x^* \rangle - u^*, \end{aligned} \tag{4$'$}$$

or equivalently

$$x \text{ maximizes } \langle \,\cdot\,, x^* \rangle - f(\,\cdot\,) \text{ over } \Re^n. \tag{4$''$}$$

Clearly $u^* = f^*(x^*)$ in $(4)'$.

If f is a closed proper convex function, then

$$x^* \in \partial f(x) \text{ if and only if } x \in \partial f^*(x^*), \tag{5}$$

i.e., ∂f and ∂f^* are inverse mappings. To establish (5), observe from $(4)'$ that $x^* \in \partial f(x)$ if and only if

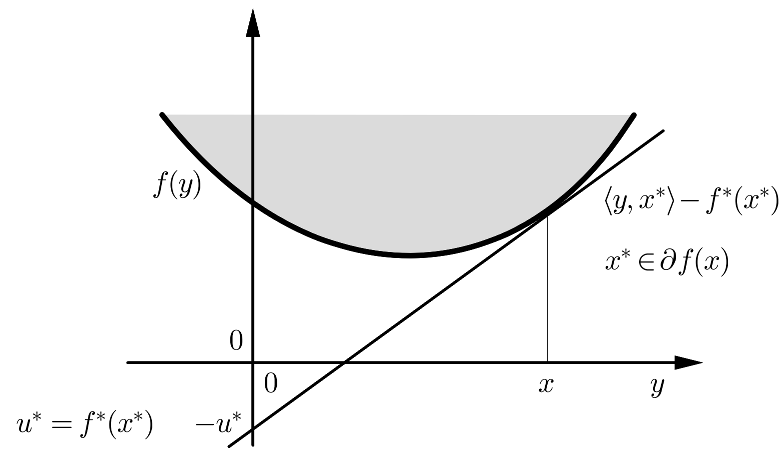


Figure 3. Subgradient

$$f(x) + f^*(x^*) = \langle x, x^* \rangle. \tag{6}$$

On applying this equivalence to f^* instead of f and using the fact that $f^{**} = f$, it follows that $x \in \partial f^*(x^*)$ if and only if (6) holds, which establishes (5).

3 d-ADDITIVE CONVEX FUNCTIONS

In the sequel, we shall make extensive use of an important class of functions called d-additive convex. Let $d = (d_i) \in \Re^n$ be a given positive vector. Call f *d-additive convex* if

$$f(x) = \sum_{i=1}^{n} d_i \widehat{f}\,(\frac{x_i}{d_i}) \tag{7}$$

where $\widehat{f}\,(\,\cdot\,)$ is a proper convex function on $\Re$. Observe that the class of d-additive convex functions whose effective domains contain a given vector is a convex cone. Also, $\sum_{i=1}^{n} d_i \widehat{f}\,(\frac{x_i}{d_i})$ is subadditive in (x, d) for $x \ge 0$ and $d \gg 0$ provided that $d_i \widehat{f}\,(\frac{x_i}{d_i})$ is subadditive in (x_i, d_i) for $x_i \ge 0$ and $d_i > 0$. The last is so because the right-hand derivative of $d_i \widehat{f}\,(\frac{x_i}{d_i})$ with respect to x_i, viz., $\mathrm{D}^+ \widehat{f}\,(\frac{x_i}{d_i})$, diminishes with d_i. Here are a few examples of functions f that are d-additive convex.

Examples of d-Additive Convex Functions

1° $f(x) = \sum_{i=1}^{n} d_i^{-q} |x_i|^{q+1}$ where $q \ge 0$, $\widehat{f}\,(z) = |z|^{q+1}$.

2° $f(x) = \sum_{i=1}^{n} \max_{1 \le j \le k} (d_i \alpha_j + \beta_j x_i)$ where $\widehat{f}\,(z) = \max_{1 \le j \le k} (\alpha_j + \beta_j z)$.

3° $f(x) = \sum_{i=1}^{n} \delta_{[d_i L,\, d_i U]}(x_i)$ where $\widehat{f}\,(z) = \delta_{[L,\, U]}(z)$ and $L \le U$.

Interpretation of d-Additive Convex Functions

There are a number of useful economic interpretations of d-additive convex functions (7). In many of these, there are n *activities* $1,\ldots,n$, an amount d_i of a *resource* is allocated to activity i, x_i is the output of activity i, and $\widehat{f}(\xi)$ is the cost per unit of resource assigned to operate activity i when ξ is the output per unit of resource assigned. Then $d_i\widehat{f}(\frac{x_i}{d_i})$ is the total cost incurred by activity i when allocated d_i units of the resource to generate the output x_i, and $f(x)$ is the total cost over all n activities to generate the output vector x when the resource vector is d.

For example, suppose the i^{th} activity consists of production in period $i=1,\ldots,n$. The resource is working days and d_i is the number of working days in period i exclusive of holidays, weekends, vacations, downtime, etc. Moreover, suppose x_i is the total production during period i, the daily production rate $\frac{x_i}{d_i}$ during the period is constant and the cost of producing ξ in a day is $\widehat{f}(\xi)$. Then the total production cost during period i is $d_i\widehat{f}(\frac{x_i}{d_i})$, so $f(x)$ is the total cost of production over n periods.

Incidentally, it is not necessary to *assume* that production is spread uniformly over the period. *That schedule is actually optimal.* To see this, suppose for simplicity that time is measured in small enough subperiods, e.g., days of a month, so that each d_i is integer. Also suppose $\widehat{f}(\xi)$ is the cost of producing ξ in a subperiod of period i and one wishes to produce x_i during period i. Then the problem of finding a minimum-cost production schedule during period i is that of choosing (y_j) to minimize $\sum_{j=1}^{d_i}\widehat{f}(y_j)$ subject to $\sum_{j=1}^{d_i}y_j=x_i$. As we saw in §1.2 (and indeed will show in the sequel), one optimal solution to this problem is to put $y_j=\frac{x_i}{d_i}$ for $j=1,\ldots,d_i$ with attendant minimum cost $d_i\widehat{f}(\frac{x_i}{d_i})$ during period i.

Conjugates of d-Additive Convex Functions

The question arises whether it is possible to choose the inner product so that if f is d-additive convex, then so is its conjugate f^*. The answer is that it is, provided that the inner product is $\langle x,y\rangle\equiv x^{\mathrm{T}}D^{-1}y$ for $x,y\in\Re^n$ where $D\equiv\operatorname{diag}(d_1,\ldots,d_n)$. To see this, observe that

$$f^*(x^*)=\sup_x\Big[\sum_i x_i(\frac{x_i^*}{d_i})-\sum_i d_i\widehat{f}(\frac{x_i}{d_i})\Big]=\sup_y\Big[\sum_i d_i\{y_i(\frac{x_i^*}{d_i})-\widehat{f}(y_i)\}\Big]=\sum_{i=1}^n d_i\widehat{f}^*(\frac{x_i^*}{d_i})$$

where $\widehat{f}^*\equiv(\widehat{f})^*$ and

$$\widehat{f}^*(z^*)=\sup_z\,[zz^*-\widehat{f}(z)].$$

d-Quadratic Functions

A d-*quadratic* function is a d-additive convex function f for which $\widehat{f}(z)=\frac{z^2}{2}$. In that event, $\widehat{f}^*=\widehat{f}$ so $f^*=f$, i.e., d-quadratic functions are *self-conjugate*.

Actually, up to a constant, every additive strictly-convex quadratic function can be made d-quadratic by making a change of variables. To see this, suppose that

$$g(x) = \sum_1^n \alpha_i(\frac{1}{2}x_i^2 + \beta_i x_i)$$

where $\alpha_i > 0$ for all i. Setting $d_i = \dfrac{1}{\alpha_i}$, completing the square and making the change of variables $y_i = x_i + \beta_i$, shows that

$$g(x) = f(y) - f(\beta)$$

where $y = (y_i)$, $\beta = (\beta_i)$ and f is d-quadratic, establishing the claim.

4 INVARIANCE THEOREM FOR NETWORK FLOWS [Ve71]

Recall that if f is convex, then one optimal solution of the problem of choosing $x_1, \ldots, x_n$ that minimize

$$\sum_{i=1}^n f(x_i)$$

subject to

$$\sum_{i=1}^n x_i = R$$

is to put $x_i = \dfrac{R}{n}$ for $i = 1, \ldots, n$. Thus the optimal solution to this problem depends on the convexity of f, but not on its precise form. Hence the optimal solution is "invariant" for the class of all convex functions f.

The goal of this section is to develop a far-reaching generalization of this result and apply it to a number of problems in inventory control. In order to see how to do this, observe that the constraints in the above problem can be expressed as the network-flow problem with zero demands as Figure 4.20 illustrates for $n = 4$ and $S_4 \equiv R$. Thus the above result can be expressed by saying that there is a feasible flow that simultaneously minimizes every 1-additive convex function of the flows emanating from node zero.

The purpose of this section is to show that for every feasible network-flow problem with arbitrary upper and lower bounds on the flows in each arc, there is a flow that simultaneously minimizes every d-additive convex function of the flows emanating from a single node 0, say. A dual of this result will also be obtained. These results have applications to a broad class of inventory problems. Since the argument is a bit involved, it is useful to list the six steps in the development.

- Form the primal capacitated network-flow problem $\mathcal{P}$ in the flow x.
- Form the Stöer dual $\mathcal{P}^*$ of choosing $\pi \equiv (t, u, v^+, v^-)$ to maximize $\inf_x L(x, \pi)$ subject to $v^\pm \geq 0$ where L is the Lagrangian and show that $(v^+)^{\mathrm{T}} v^- = 0$ without loss of generality.
- Characterize saddle points of Lagrangian as equilibrium conditions.
- Show that t contains u.
- Characterize equilibrium conditions in d-quadratic case.
- Prove Invariance Theorem.

Primal Network-Flow Problem

Consider a (directed) graph $(\mathcal{N},\mathcal{A})$ with nodes $\mathcal{N}=\{0,\ldots,n\}$ and arcs $\mathcal{A}\equiv\{(i,j)\in\mathcal{N}^2 : i<j\}$. Let x_{ij} be the flow from node i to node j, $(i,j)\in\mathcal{A}$. Although we do not allow arcs (i,j) with $i>j$, one can interpret a negative flow from i to j as a positive flow from j to i. The problem is to choose a preflow $x=(x_{ij})$ that

$$\text{minimizes } f(x_0) \tag{8}$$

subject to

$$\sum_{j=0}^{i-1} x_{ji} - \sum_{j=i+1}^{n} x_{ij} = c_i,\ 1\le i\le n \tag{9}$$

and

$$a \le x \le b \tag{10}$$

where $x_0=(x_{0j})$, $j\ge 1$, is the vector of flows emanating from node zero to the other n nodes, $c=(c_i)$, $i\ge 1$, is the given vector of flows from those other nodes to node 0; $a=(a_{ij})$ and $b=(b_{ij})$ are respectively given extended-real-valued lower and upper bounds on the flow x with $a_{ij}<\infty$ and $-\infty<b_{ij}$ for all $(i,j)\in\mathcal{A}$; and $f(x_0)$ is a $+\infty$ or real-valued cost associated with the flows emanating from node 0. It is important to note that the cost does not depend on the flows emanating from any node other than 0. Call this problem $\mathcal{P}$. Figure 4 illustrates the problem for $n=4$. Our aim is to show that if (9), (10) is feasible, then there is a flow x that simultaneously minimizes every d-additive convex $f(x_0)=\sum_1^n d_i\widehat{f}\,(\frac{x_{0i}}{d_i})$ of x_0 for a fixed positive vector d. The flow found does not depend on $\widehat{f}$, though it does depend on d.

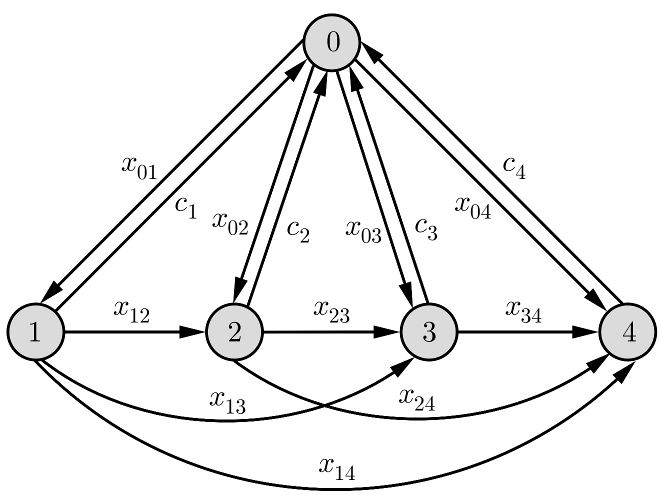


Figure 4. A Network

Stöer (or Lagrangian) Dual $\mathcal{P}^*$ of $\mathcal{P}$

To analyze the problem $\mathcal{P}$, it is convenient first to construct the dual $\mathcal{P}^*$ of $\mathcal{P}$. To this end, associate multipliers (prices) $u=(u_i)$ with the equations (9), $v^-=(v^-_{ij})\ge 0$ with the left-hand inequality in (10), and $v^+ = (v^+_{ij}) \geq 0$ with the right-hand inequality in (10). For notational convenience, put $v_0^- \equiv (v_{0i}^-)$, $v_0^+ \equiv (v_{0i}^+)$, and $t \equiv (t_i) \equiv u - v_0^+ + v_0^-$. Then on letting $\pi = (t, u, v^+, v^-)$, the Lagrangian becomes

$$L(x,\pi) = f(x_0) + \sum_{i=1}^{n} (c_i - \sum_{j=0}^{i-1} x_{ji} + \sum_{j=i+1}^{n} x_{ij}) u_i - (x-a)^{\mathrm{T}} v^- - (b-x)^{\mathrm{T}} v^+. \tag{11}$$

Now the *Stöer dual* $\mathcal{P}^*$ of $\mathcal{P}$ is that of choosing π that

$$\text{maximizes } \inf_x L(x,\pi) \tag{12}$$

subject to $v^+ \geq 0$, $v^- \geq 0$. Observe that the u_i are unconstrained because (9) consists of equations and the $v^\pm$ are nonnegative because (10) consists of inequalities.

In order to make this problem more explicit, it is necessary to compute $\inf_x L(x,\pi)$. To that end, put $\langle p, q \rangle \equiv p^{\mathrm{T}} D^{-1} q$ for all $p,q \in \Re^n$ where $D \equiv \operatorname{diag}(d_1, \ldots, d_n)$. Rewrite (11) by collecting terms that depend on x_0 and x_{ij} $(i \geq 1)$ yielding

$$L(x,\pi) = a^{\mathrm{T}} v^- - b^{\mathrm{T}} v^+ + c^{\mathrm{T}} u - [\langle Dt, x_0 \rangle - f(x_0)] + \sum_{1 \leq i < j \leq n} (u_i - u_j + v_{ij}^+ - v_{ij}^-) x_{ij}. \tag{13}$$

Evaluating the infimum of $L(x,\pi)$ over x_0 requires computation of the supremum over x_0 of the bracketed term above which becomes $f^*(Dt)$. Since the infimum of $L(x,\pi)$ over x_{ij} is $-\infty$ if the coefficient of x_{ij} $(i \geq 1)$ is not zero, we can assume that each such coefficient is zero. Putting these facts together with (12), the Stöer dual $\mathcal{P}^*$ becomes that of choosing π that

$$\text{maximizes } a^{\mathrm{T}} v^- - b^{\mathrm{T}} v^+ + c^{\mathrm{T}} u - f^*(Dt) \tag{14}$$

subject to

$$v_0^+ - v_0^- - u + t = 0, \tag{15}$$

$$v_{ij}^+ - v_{ij}^- - u_j + u_i = 0,\ 1 \leq i < j \leq n, \tag{16}$$

$$\begin{gathered} v^+ \geq 0,\ v^- \geq 0, \\ v_{ij}^+ = 0 \text{ if } b_{ij} = \infty \text{ and } v_{ij}^- = 0 \text{ if } a_{ij} = -\infty \end{gathered} \tag{17}$$

for $0 \leq i < j \leq n$. (Note that (15) is simply the equation defining t.)

It is useful to observe that (15)-(17) is homogeneous and so is always trivially feasible because all the variables may be set equal to zero. More important is the fact that for any fixed (t, u), it is easy to determine the (v^+, v^-) that maximize (14). They are

$$v_0^\pm = (u-t)^\pm \text{ and } v_{ij}^\pm = (u_j - u_i)^\pm,\ 1 \leq i < j \leq n, \tag{18$'$}$$

where z^- is the negative part of z. Since $z = z^+ - z^-$, these $v_{ij}^{\pm}$ satisfy (15)-(17). They are also uniquely determined by (15)-(17) and $v_{ij}^+ v_{ij}^- = 0$ for all i, j, or equivalently (15)-(17) and

$$(v^+)^{\mathrm{T}} v^- = 0. \tag{18}$$

In order to see why (v^+, v^-) satisfying (15)-(18) maximizes (14) subject to (15)-(17) with (t, u) fixed, suppose $v_{ij}^+ v_{ij}^- > 0$ for some i, j and (15)-(17) hold. Put $\epsilon = v_{ij}^+ \wedge v_{ij}^-$. Then reducing v_{ij}^+ and v_{ij}^- by $\epsilon > 0$ preserves the conditions (15)-(17) and increases (14) by $(b_{ij} - a_{ij})\epsilon$, which is nonnegative because $a_{ij} \le b_{ij}$ and $\epsilon > 0$.

Characterization of Saddle Points as Equilibrium Conditions

The argument given above shows that the difficult part of the dual problem is that of determining the optimal (t, u). In order to proceed further, we need to give the equilibrium conditions associated with optimality. To that end, assume in the remainder of this section that f is a closed proper d-additive convex function, whence that is so of f^* as well.

Now recall from the Stöer duality theory for nonlinear programs that if (x, π) is a saddle point of the Lagrangian $L(\,\cdot\,,\cdot\,)$, then x is optimal for $\mathcal{P}$ and π is optimal for $\mathcal{P}^*$. It is easy to write down conditions characterizing saddle points. First, π must maximize $L(x,\,\cdot\,)$ subject to $v^+, v^- \ge 0$. In particular, observe that a necessary and sufficient condition that u_i maximize $L(x, \pi)$ is that the coefficient of u_i in parentheses in (11) is zero, i.e., (9) holds. Similarly, from (11) again, (v^+, v^-) maximizes $L(x, \pi)$ subject to $v^+, v^- \ge 0$ if and only if (10) holds and

$$(b - x)^{\mathrm{T}} v^+ = 0,\ (x - a)^{\mathrm{T}} v^- = 0. \tag{19}$$

To see this, observe that if $x_{ij} < a_{ij}$ for some $i < j$, then $L(x, \pi) \uparrow \infty$ as $v_{ij}^- \uparrow \infty$, which contradicts the hypothesis that $L(x,\,\cdot\,)$ assumes its maximum. That is why the left-hand inequality in (10) must hold. The justification for the right-hand inequality in (10) is similar. Having established that (10) must hold, observe from (11) again that (19) is necessary and sufficient for (v^+, v^-) to maximize $L(x, \pi)$ because the last two terms in (11) are nonpositive always, and they assume their maxima when they are zero, which would be the case if, for example, $(v^+, v^-) = 0$.

To complete the characterization of saddle points (x, π), it is necessary to characterize when x minimizes $L(\,\cdot\,, \pi)$. To do this, observe from (13) and the d-additivity of f that $x_i \equiv x_{0i}$ minimizes $L(x, \pi)$ if and only if x_i maximizes $d_i[t_i \frac{x_i}{d_i} - \widehat{f}(\frac{x_i}{d_i})]$, or equivalently by $(4)''$,

$$t_i \in \partial \widehat{f}(\frac{x_i}{d_i}),\ 1 \le i \le n. \tag{20}$$

Finally, observe from (13) again that x_{ij} minimizes $L(x, \pi)$ if and only if (from constructing the dual problem $\mathcal{P}^*$) (16) holds. Since (15) is simply the definition of t, (15) holds as well. Thus

(x,π) is a saddle point of $L(\cdot,\cdot)$ if and only if (9), (10), (15)-(17), (19) and (20) hold. Call these the *equilibrium conditions*.

t Contains u

The next step is to explore the relationship between the optimal t and u in the dual problem $\mathcal{P}^*$, or more precisely in the equilibrium conditions. To this end say t *contains* u if each element of u is contained among the elements of t. If $w \in \Re^n$, say that $w' \in \Re^n$ *preserves the order* (resp., *zeroes*) of w if $w_i \le w_j$ (resp., $w_i = 0$) implies $w'_i \le w'_j$ (resp., $w'_i = 0$) for all i, j.

LEMMA 1. t contains u. *If (x,t) satisfies the equilibrium conditions for some (u,v^+,v^-), then it does so for some u contained in t.*

Proof. By hypothesis (x,t,u,v^+,v^-) satisfies the equilibrium conditions. There is no loss in generality in assuming that (18) holds. (If not, choosing the $v^\pm$ as in (18)$'$ produces the desired result.) Now relabel the nodes so that $t_1 \le \cdots \le t_n < t_{n+1} \equiv \infty$. For each $1 \le i \le n$, let

$$u'_i = \begin{cases} t_1, \text{ if } u_i < t_1 \\ t_j, \text{ if } t_j \le u_i < t_{j+1},\ 1 \le j \le n, \end{cases}$$

as Figure 5 illustrates for $n=3$. Thus, given (t,u'), choose v'^+, v'^- as in (18)$'$. This assures that (15)-(18) hold. By construction (t,u') preserves the order of (t,u), and so (v'^+,v'^-) preserves the zeroes of (v^+,v^-). Consequently, since (v^+,v^-) satisfies (19), so does (v'^+,v'^-). Also since (x,t) is unchanged, the conditions (9), (10), and (20) still hold. Thus (x,t,u',v'^+,v'^-) satisfies the equilibrium conditions. Also, t contains u' by construction. ∎

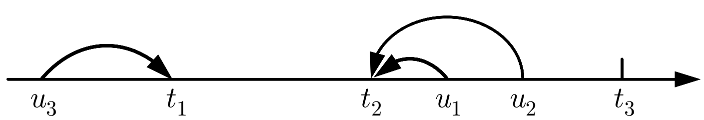


Figure 5. The Contains Relation

Remark. The importance of this Lemma is that once we have found a (t,u) satisfying the equilibrium conditions, we can assume u is one of up to n^n vectors in $\Re^n$ each of whose elements is an element of t. Thus the most important part of the dual problem $\mathcal{P}^*$ is to find t.

The d-Quadratic Case

As will be seen shortly, the d-quadratic case, i.e., where f, or equivalently f^*, is d-quadratic, plays an important role in the general theory. The reason is that a solution to that case can be transformed easily into a solution for any d-additive convex f. For this reason it is useful to explore a few implications of the duality theory for convex quadratic programs in the d-quadratic case. To begin with, since $\mathcal{P}^*$ is always feasible, it follows that $\mathcal{P}$ is feasible if and only if $\mathcal{P}^*$ has an optimal solution. And in either case, $\mathcal{P}$ and $\mathcal{P}^*$ have optimal solutions x and π satisfying the equilibrium conditions where (20) specializes to

$$t_i = \frac{x_i}{d_i},\ 1 \le i \le n, \tag{20$'$}$$

because when f is d-quadratic, $\partial \widehat{f}(z) = \{z\}$.

Invariance Theorem

We can now state and prove our main result.

THEOREM 2. Invariance.

1° **Invariance of Optimal Flow.** *If $\mathcal{P}$ is feasible, there is a flow x that simultaneously minimizes every d-additive convex f.*

2° **Invariance of Order of Optimal Dual Variables.** *If there is a t that is optimal for $\mathcal{P}^*$ in the d-quadratic case, there is an associated optimal u contained in t. If also f^* is a closed d-additive convex function and $t_i \in \partial \widehat{f}^*(\Re)$ for $1 \le i \le n$, there is a t' preserving the order of t for which $t_i \in \partial \widehat{f}^*(t'_i)$, $1 \le i \le n$, and a unique u' such that (t', u') preserves the order of (t, u). In addition (t', u') is optimal for $\mathcal{P}^*$ with f^*.*

Remark 1. Contrasting Invariance in the Primal and Dual. Part 1° asserts invariance of the optimal solution of $\mathcal{P}$ as f ranges over the class of d-additive convex functions. By contrast, 2° asserts instead invariance of the order of the optimal (t, u) as f^* ranges over most closed d-additive convex functions.

Remark 2. Solving $\mathcal{P}$ and $\mathcal{P}^*$. One can solve both $\mathcal{P}$ and $\mathcal{P}^*$ by first finding optimal x and π with u contained in t in the d-quadratic case, perhaps by standard quadratic network-flow algorithms. Then x is optimal for $\mathcal{P}$. To find a π' optimal for $\mathcal{P}^*$, proceeds as follows. First, given t, by hypothesis there is a t'_i so that $t_i \in \partial \widehat{f}^*(t'_i)$, or equivalently by (4)$''$, t'_i maximizes $t_i t'_i - \widehat{f}^*(t'_i)$. Since this function is superadditive in (t_i, t'_i), one can choose the maximizer t'_i to be increasing in t_i by the Increasing-Optimal-Selections Theorem. Thus t' preserves the order of t. Now let $u'_i = t'_j$ if $u_i = t_j$. This rule defines the unique $u' = (u'_i)$ such that (t', u') preserves the order of (t, u). To sum up, once an optimal (t, u) is found for $\mathcal{P}^*$ in the d-quadratic case, the optimal t' is found by solving n one-dimensional optimization problems indexed by the parameters t'_i.

Proof of Theorem 2. First, prove 2°. Since there exists an optimal t for $\mathcal{P}^*$ in the d-quadratic case, there exist x and $\pi = (t, u, v^+, v^-)$ that satisfy the equilibrium conditions by the duality theorem for quadratic programming. Without loss of generality, assume that (18) holds, and by Lemma 1, that t contains u as well. Now the t' that Remark 2 above constructs preserves the order of t. Also by $(20)'$, $\frac{x_i}{d_i} = t_i \in \partial \widehat{f}^*(t'_i)$, or equivalently by (5), $t'_i \in \partial \widehat{f}(\frac{x_i}{d_i})$, $1 \le i \le n$, since $\widehat{f}^*$ is closed, proper and convex. Moreover, as Remark 2 discusses, there is a unique u' such that (t', u') preserves the order of (t, u). Consequently, the unique (v'^+, v'^-) determined by (15)-(18) (with (t', u') replacing (t, u)) preserves the zeroes of (v^+, v^-). Thus x and $\pi' \equiv (t', u', v'^+, v'^-)$ satisfy the equilibrium conditions with f and so are optimal for $\mathcal{P}$ and $\mathcal{P}^*$ respectively, which proves 2°. This also proves that if f is d-additive convex, closed and satisfies $t_i \in \partial \widehat{f}^*(\Re)$, or equivalently by (5), $\partial \widehat{f}(t_i) \ne \emptyset$ for $1 \le i \le n$, then x is optimal for $\mathcal{P}$ with f. To complete the proof, suppose f is d-additive convex, so $\widehat{f}$ is proper and convex. Then there is a sequence of closed proper convex $\widehat{f}_m$ with $\partial \widehat{f}_m(t_i) \ne \emptyset$ for all $1 \le i \le n$ and $m \ge 1$ such that $\widehat{f}_m$ converges pointwise to $\widehat{f}$. Let $f_m(x_0) \equiv \sum_i d_i \widehat{f}_m(\frac{x_i}{d_i})$. Thus from what was shown above, x is optimal for $\mathcal{P}$ with f_m for each m, and so also for $\mathcal{P}$ with $\lim f_m = f$. ▮

5 TAUT-STRING SOLUTION

Specialization $\mathbb{P}$ of $\mathcal{P}$

There is an important instance of the Invariance Theorem that can be solved graphically! Let $\mathbb{P}$ be the specialization of $\mathcal{P}$ in which $-a_{0i} = b_{0i} = \infty$ for $1 \le i \le n$, $a_{ij} = b_{ij} = 0$ for $1 \le i < j \le n$ with $j \ne i+1$, and $c_i = 0$ for $1 \le i < n$. Now let $X_i \equiv x_{i,i+1}$, $E_i \equiv a_{i,i+1}$ and $F_i \equiv b_{i,i+1}$ for $1 \le i < n$; also $X_n \equiv X_{n-1} + x_n$ and $E_n \equiv F_n \equiv c_n$. Figure 4.20 illustrates the network for $n = 4$ where $S_4 \equiv c_4$. Observe that $x_i = X_i - X_{i-1}$, so $\mathbb{P}$ can be expressed as choosing x_0 to

$$\text{minimize} \sum_{i=1}^{n} d_i \widehat{f}(\frac{x_i}{d_i}) = f(x_0) \tag{21}$$

subject to

$$E \le X \le F \tag{22}$$

where $E \equiv (E_i)$, $X \equiv (X_i)$, and $F \equiv (F_i)$.

To solve $\mathbb{P}$, specialize to $\widehat{f}(z) \equiv \sqrt{z^2+1}$, which is strictly convex. Then (21) becomes the d-additive convex function

$$\sum_{i=1}^{n} d_i \sqrt{\left(\frac{x_i}{d_i}\right)^2 + 1} = \sum_{i=1}^{n} \sqrt{x_i^2 + d_i^2}. \tag{21$'$}$$

Now plot E_i, X_i and F_i versus $D_i \equiv \sum_{j=1}^{i} d_j$ on the plane for $i = 0, \ldots, n$ as illustrated in Figure 6 for $n = 5$. Then draw the polygonal path joining (D_{i-1}, X_{i-1}) and (D_i, X_i) for $0 < i \le n$ where $D_0 \equiv E_0 \equiv F_0 \equiv X_0 \equiv 0$. Observe that $(21)'$ gives the length of that polygonal path because $\sqrt{x_i^2 + d_i^2}$ is the distance between the points (D_{i-1}, X_{i-1}) and $(D_i, X_i) = (D_{i-1}, X_{i-1}) + (d_i, x_i)$.

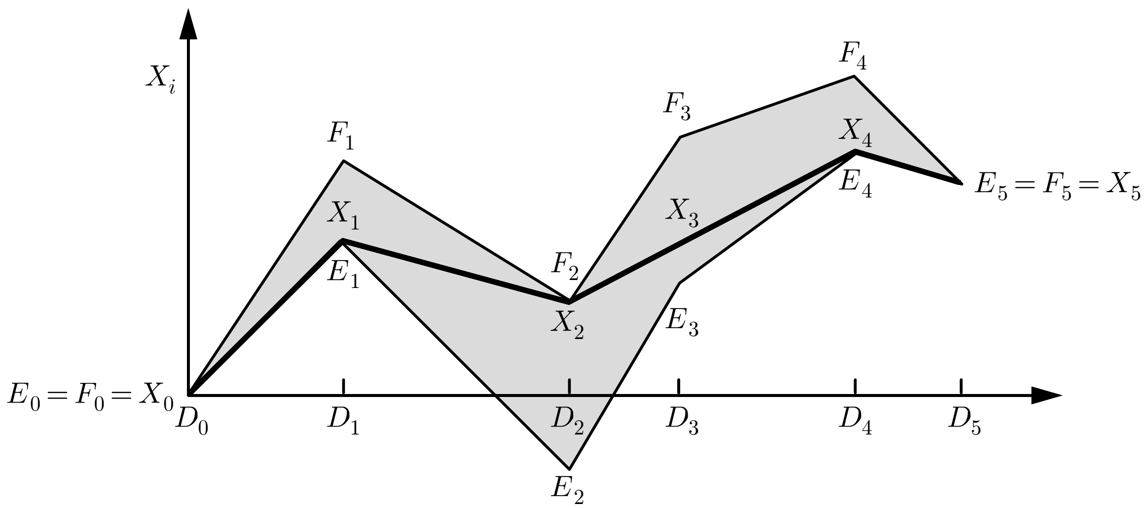


Figure 6. Taut-String Solution of $\mathbb{P}$

Thus the problem of choosing X that minimizes $(21)'$ subject to (22) has a geometric interpretation, viz., find the shortest path in the plane from the origin to (D_n, E_n) among those paths that lie between (D_i, E_i) and (D_i, F_i) for $0 < i < n$! This path can be found as follows. Put pins in the plane at (D_i, E_i) and (D_i, F_i) for $0 \le i \le n$. Then tie a string to the pin at the origin, thread it between the pins at (D_i, E_i) and (D_i, F_i) for $1 \le i \le n$, and pull the string taut at (D_n, E_n). The taut string traces out the desired shortest path. This taut-string solution can be computed in $O(n)$ time.

The importance of this construction is that by the Invariance Theorem 2, the x_i determined by the taut-string simultaneously solve $\mathbb{P}$ for all d-additive convex f because $(21)'$ is strictly convex and d-additive.

Dual $\mathbb{P}^*$ of $\mathbb{P}$

Now specialize $\mathcal{P}^*$ to give the dual $\mathbb{P}^*$ of $\mathbb{P}$. To begin with, since $-a_{0i} = b_{0i} = \infty$, it follows from (17) that $v_{0i}^+ = v_{0i}^- = 0$, so by (15), $u = t$. Thus there is no loss in generality in eliminating u from the problem. Also, by $(18)'$, assume $v_{i,i+1}^+ = (t_{i+1} - t_i)^+$ and $v_{i,i+1}^- = (t_{i+1} - t_i)^-$ for $1 \le i < n$, thereby eliminating the restrictions (15)-(17) of the dual problem. On putting $t_{n+1} \equiv 0$, the dual $\mathbb{P}^*$ of $\mathbb{P}$ thus becomes that of choosing $t = (t_i) \in \Re^n$ to maximize (14), or equivalently,

$$\sum_{i=1}^{n} [E_i(t_{i+1} - t_i)^- - F_i(t_{i+1} - t_i)^+ - d_i \widehat{f}^*(t_i)] \tag{23}$$

because $E_n(t_{n+1} - t_n)^- - F_n(t_{n+1} - t_n)^+ = c_n[(-t_n)^- - (-t_n)^+] = c_n u_n$.

It is often the case that one has an inequality of the form $t_i \ge t_{i+1}$ where $1 \le i < n$. This can be assured by setting $F_i = \infty$. In particular, if one has the system of inequalities $t_1 \ge t_2 \ge \cdots \ge t_n$, it suffices to set $F_i = \infty$ for $i = 1, \ldots, n-1$. Similarly, if one has an inequality of the form $t_i \le t_{i+1}$

where $1 \le i < n$, this can be assured by setting $E_i = -\infty$. In particular, if one has the system of inequalities $t_1 \le t_2 \le \cdots \le t_n$, it suffices to set $E_i = -\infty$ for $i = 1, \ldots, n-1$.

The general dual problem $\mathbb{P}^*$ can be solved with a closed d-additive convex $\widehat{f}^*$ by finding the taut-string solution x_0 to the dual $\mathbb{P}$ of $\mathbb{P}^*$ and then finding the optimal t_i, preserving the order of the $\frac{x_i}{d_i}$, from (20). (Observe that $\frac{x_i}{d_i}$ is the slope of the taut string on the interval (D_{i-1}, D_i).) By (5), this is equivalent to choosing t_i so $\frac{x_i}{d_i} \in \partial \widehat{f}^*(t_i)$, or what is the same thing by $(4)'$, choosing t_i to maximize $\frac{x_i}{d_i} t_i - \widehat{f}^*(t_i)$ for $1 \le i \le n$. It is important to note that given x_i, the optimal t_i depends on d_j only for $j = i$.

6 APPLICATION TO OPTIMAL DYNAMIC PRODUCTION PLANNING [MH55], [Ve71], [Ve87]

Application of $\mathbb{P}$ to Production Planning

Observe that $\mathbb{P}$ can be interpreted as the instance of the production planning problem in Figure 4.20 where $c_i(x_i) \equiv d_i \widehat{f}(\frac{x_i}{d_i})$, $h_i(y_i) \equiv \delta_+(y_i - A_i) + \delta_+(B_i - y_i)$ for all i, $A \equiv (A_i)$, $B \equiv (B_i)$, $S \equiv (S_i)$, $E \equiv S + A$ and $F \equiv S + B$. Thus this model encompasses nonstationary production costs by means of the parameters d_i. Direct storage costs are not allowed, unless it is possible to absorb them into the production costs by an appropriate choice of the *scale parameters* d_i, e.g., as we show below for capital costs associated with investment in inventories, a major component of storage costs. However, lower and upper bounds A and B are permitted on inventories. The motive for carrying inventories in this model is that there may be a temporal increase in the marginal cost of supplying demand.

Observe that the optimal production schedule is independent of $\widehat{f}$ provided only that $\widehat{f}$ is proper and convex. This means one does not need to know $\widehat{f}$ in order to schedule optimally! Also during any interval I of periods in which inventories are not equal to their upper or lower bounds, it follows from the taut-string solution that $\frac{x_i}{d_i} = \lambda$, say, or equivalently, $x_i = \lambda d_i$, for all $i \in I$, i.e., optimal production in each period in I is proportional to the scale parameter. Thus, optimal production rises or fall according as the scale parameter rises or falls over the interval I.

Positively Homogeneous Convex Production Cost Function. As a particular example, suppose that the present value of the production cost in period i is $c_i(z) = \beta^i |z|^{q+1} c$ for some discount factor $\beta > 0$ and unit cost $c > 0$, so $c_i(\,\cdot\,)$ is positively homogeneous of degree $q + 1 \ge 1$. This formulation accounts for the cost of capital invested in inventories. Then the production cost is d-additive convex with $\widehat{f}(z) = |z|^{q+1} c$ and scale parameters

$$d_i = (1+\rho)^{\frac{i}{q}},\ i = 1, \ldots, n$$

where $\beta \equiv \frac{1}{1+\rho}$ and $100\rho\%$ is the interest rate. Observe that if $\rho = 0$, then $d_i = 1$ for all i. If instead $\rho > 0$ (resp., $\rho < 0$), then d_i expands (resp., contracts) geometrically with the precise rate being greater than, equal to or less than $|\rho|$ according as $0 < q < 1$, $q = 1$ or $1 < q$. In this event, during intervals in which inventories are not equal to their upper or lower bounds, optimal production rises or falls geometrically according as $\rho > 0$ or $\rho < 0$.

Application of $\mathbb{P}^*$ to Production Planning

It is of interest to note that $\mathbb{P}^*$, like $\mathbb{P}$, can be interpreted as a special case of the production-planning problem in Figure 4.20. In particular, letting $X_i \equiv t_{i+1}$, $S_i = 0$, $c_i(z) \equiv F_i z^+ - E_i z^-$, and $h_i(z) = d_{i+1} f^{*\prime}(z)$ for $0 \le i \le n$ where $d_{n+1} \equiv 0$ produces a special case of the problem in Figure 4.20 except in the present case X_0 is a variable to be chosen by the decision maker and so one adds the cost $h_0(X_0) = d_1 \widehat{f}^*(X_0)$ to $(4.8)'$. In this event, F_i is the unit production cost and E_i is the unit disposal revenue in period i.

Planning Horizons

In the production planning interpretations of $\mathbb{P}$ and $\mathbb{P}^*$, call period k a *planning horizon* if the optimal choice of x_1 (or t_1) depends only on (D_i, E_i, F_i) for $i \le k$ for all parameter sequences D, E, F in a specified set under consideration. Thus a planning horizon can be thought of as the number of periods ahead one must be able to forecast correctly in order to act optimally initially. Planning horizons are easy to determine graphically. As an illustration, $k = 2$ is a planning horizon for the example in Figure 6.

In seasonal (i.e., periodic) problems, planning horizons typically occur prior to one complete cycle and then at periodic intervals thereafter. To illustrate, suppose $D_i = i$ for all i, $E = S$ and $F_i = \infty$ for all $1 < i < \infty$ and that the sales schedule is periodic, i.e., for some positive integer p, $s_i = s_{i+p}$ for all $i \ge 1$ where $s_i = S_i - S_{i-1}$ for $i \ge 1$ and $S_0 = 0$. Then for the example in Figure 7 with $p = 12$ months,

$$\frac{S_i}{i} \le \frac{S_6}{6} \text{ for } 1 \le i < 6 \text{ and } \frac{S_i - S_6}{i - 6} \le \frac{S_{18} - S_6}{12} \text{ for } 6 < i < 18,$$

the first planning horizon occurs at six months, and subsequent planning horizons occur every 12

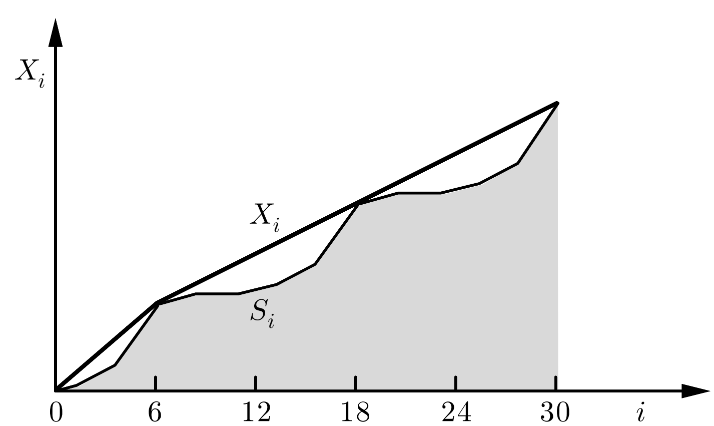


Figure 7. Production Planning Horizon

months thereafter. In this event, the optimal monthly production rate is constant during the initial six months and then falls to a lower rate[1] that remains constant thereafter. Observe that the planning horizons occur in periods i of zero inventories and falling sales, i.e., $s_i \geq s_{i+1}$. Also, if two successive planning horizons do not occur in subsequent periods, then they are separated by at least one period j of strictly rising sales, i.e., $s_j < s_{j+1}$.

[1]In the present case this occurs immediately. But in general the decline in production may occur over several months.

6

Concave-Cost Network Flows and Supply Chains

[WW58], [Ma58], [Za69], [Ve69], [Lo72], [Kal72], [EMV79,87], [Ro85,86], [WVK89], [FT91], [AP93]

1 INTRODUCTION

So far attention has been focused largely on inventory problems that could be formulated as minimum-convex-cost network-flow problems. These models exhibit diseconomies-of-scale. Also optimal production schedules tend to smooth out fluctuations in sales and other parameters, c.f., the Smoothing Theorem. In practice it often happens that the cost functions instead exhibit economies-of-scale, e.g., they are concave. Under these conditions, optimal production schedules tend to amplify fluctuations in sales and other parameters—sharply contrasting with the situation of convex costs.

From a mathematical viewpoint, the problem of finding minimum-cost production schedules in the presence of scale economies often reduces to the problem of finding a vector x in a polyhedral convex set X that minimizes a concave function c on X. For this reason, §6.2 characterize when such a minimum is attained. That section also shows, under mild conditions, that if the minimum is attained, it is attained at an extreme point of X.

It is often the case that the polyhedral sets X arising in the study of inventory problems are sets of feasible flows in a network. For this reason, it is useful to specialize the general theory of §6.2 to characterize extreme flows in networks in §6.3. That section also shows how to reduce uncapacitated network flow problems with arbitrary additive-concave-cost functions to equivalent problems in which the arc costs are nonnegative. This fact proves useful in the development of efficient algorithms for solving such problems.

Once the extreme points of a convex polyhedral set have been characterized, there remains the important problem of developing an efficient algorithm for searching them to find one that is optimal. Examining all of them is generally impractical since the number of extreme points typically rises exponentially with the size of the problem. This difficulty can be overcome in the simplest and most prominent instance of this problem, viz., linear programming. This is because linear objective functions enjoy the twin properties of concavity and convexity, the former assuring that the minimum is attained at an extreme point and the latter assuring that a local minimum is a global minimum. Consequently, in searching for an improvement of an extreme point of a linear program, it is enough to examine the "adjacent" extreme points. Unfortunately, that is not so for nonlinear concave functions since they are not convex.

Nevertheless, it is possible to develop efficient algorithms for globally minimizing additive concave functions on uncapacitated networks in important classes of network flow problems that arise in the study of inventory systems. (It suffices to study the uncapacitated problem since, as §6.6 shows, it is possible to reduce the capacitated to the uncapacitated problem.) Section 6.4 develops a "send-and-split" dynamic programming method for doing this. The running time of the method is polynomial in the numbers of nodes and arcs of the graph, but exponential in the number of *demand nodes*, i.e., nodes at which there is a nonzero demand. Section 6.5 shows that in the case of planar graphs the running time is also polynomial in the number of demand nodes, though it is exponential in the number of faces of the planar graph containing the demand nodes. The importance of this for inventory problems is that, as §6.7 shows, the graphs of single and tandem facility n-period inventory problems are planar with all demand nodes lying in a common face. Thus, the send-and-split algorithm solves such problems in polynomial time.

Unfortunately, the networks arising from the study of multi-retailer distribution systems are not planar, so the send-and-split method is not efficient for them. Nevertheless, §6.8 shows that if at each facility of a one-warehouse N-retailer inventory system, the ordering costs are of set-up cost type, the storage costs are linear, and the costs and demand rates are stationary, it is possible to find a schedule that is guaranteed to be within 6% of the minimum average cost per unit time in $O(N \log N)$ time.

2 MINIMIZING CONCAVE FUNCTIONS ON POLYHEDRAL CONVEX SETS [HH61], [Ro70], [Ve85]

The first question that arises in attempting to minimize a concave function on a convex polyhedral set is whether the minimum is attained? In order to motivate the answer to this question, consider the simplest situation, viz., $X \subseteq \Re$. There are three cases to examine.

- $X = \Re$. In this event, c attains its minimum if and only if c is constant on $\Re$. Thus this case is uninteresting.

- $X = [a, b]$. In this event c attains its minimum at an end point of X, i.e., at a or b (at b in Figure 1).

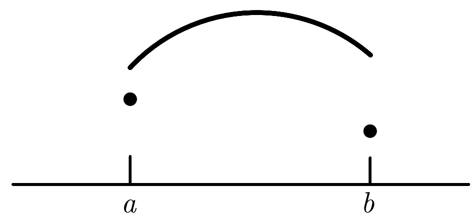


Figure 1. A Concave Function on a Bounded Interval

- $X = [a, \infty)$. In this event c attains its minimum on X if and only if c is bounded below on X. And when that is so, the minimum is attained at a because c must be increasing on X. For if c is not increasing, then $c(x) \to -\infty$ as $x \to \infty$, in which case c cannot attain its minimum on X. There is, of course, one other case to consider, viz., $X = (-\infty, a]$. But this is reduced to the case just considered by reversing the axis and so need not be discussed.

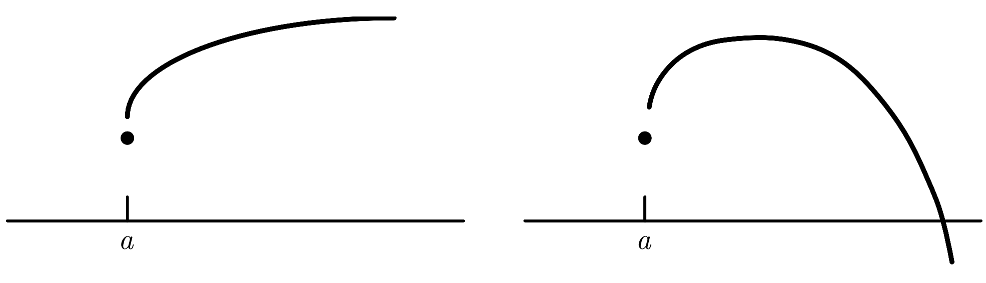


Figure 2. Concave Functions on Semi-Infinite Intervals

The first of the above cases does not arise if X does not contain a line. The last two cases are unified by asserting that if c is bounded below on each "half-line" in X, then c assumes its minimum on X at an extreme point of X.

It is now appropriate to consider the general case. To that end, suppose X is a polyhedral convex subset of $\Re^n$. Call an element e of X an *extreme point* (of X) if e is not one-half the sum of two distinct elements of X. A *half-line* in X *emanating* from $x \in \Re^n$ in the *direction* (of recession) $d \in \Re^n$, $d \neq 0$, is a subset of X of the form $\{x + \lambda d : \lambda \in \Re_+\}$. Call a half-line H in X *extreme* if no element of H can be expressed as one-half the sum of two points in $X \setminus H$. In this event, call the associated directions of H *extreme*. It is known [Ro70, pp. 162-72] that if X contains no lines, then $X = \operatorname{conv} E + \operatorname{cone} D$ where E is the nonempty finite set of extreme points of X, D is the set of extreme directions of X, "conv" means "convex hull of" and "cone" means "convex cone hull of." Note that $\operatorname{cone} D$, called the *recession cone* of X, is the set of directions.

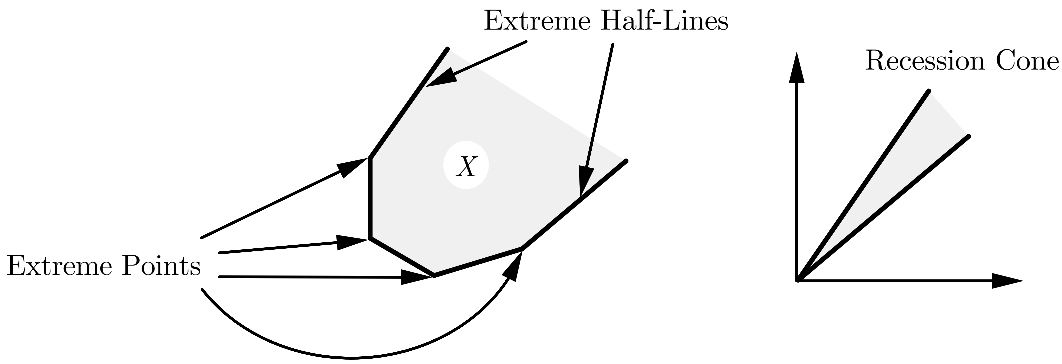


Figure 3. Extreme Points and Half-lines; Recession Cone

A necessary condition that c attain its minimum on a polyhedral convex set X is that c be bounded below on the half-lines emanating from a single element of X in every extreme direction thereof. It is known from the theory of linear programming that this necessary condition is also sufficient if c is linear. The question arises whether the condition remains sufficient if c is merely concave. The answer is "no" as the following example illustrates.

Example 1. A Concave Function that is Bounded Below in Extreme Directions from Some Points, but Not Others. Let $X = \{(v,w) \in \Re^2_+ : w \le 1\}$, $c(v,w) = 0$ for $0 \le v$ and $0 \le w < 1$, and $c(v,w) = -v$ for $0 \le v$ and $w = 1$. Then c is concave on X and is bounded below on the half-line emanating in the (unique) extreme direction $(1,0)$ from $(0,0)$, but not from $(0,1)$.

Concave Functions With Bounded Jumps

This example shows that another necessary condition on c is required. To that end, let C_X be the class of real-valued concave functions on a convex set X in $\Re^n$ and B_X be the subclass of functions $c \in C_X$ with *bounded jumps*, i.e., for which $\lim_{\lambda \downarrow 0} [c(x) - c(x + \lambda(y-x))]$ is bounded below uniformly in $x, y \in X$. The function c in the above example does not have bounded jumps.

The class B_X is closed under addition and nonnegative scalar multiplication, and so is a convex cone. The continuous (which include the linear) functions in C_X are in B_X because they have no jumps. When $n = 1$, the elements of C_X have at most two jumps, so $B_X = C_X$. Thus, for $n \ge 1$, B_X contains the additive (and hence the fixed-charge) functions $c \in C_X$, i.e., those for which $c(x) =$

$\sum_{i=1}^{n} c_i(x_i)$. Finally, the bounded functions in C_X are in B_X. Many other concave functions that arise in practice are also in B_X because they can be generated from a set of real-valued concave functions, each of which is continuous, additive, or bounded, by alternately taking pointwise minima and nonnegative linear combinations of members of the set.

A necessary condition for c to attain its minimum on X is that $c^- \equiv c \wedge 0$ have bounded jumps. Of course, $c \in B_X$ implies $c^- \in B_X$. But $c \in B_X$ is not a necessary condition for c to attain its minimum, as the function $c(v,w)+v$ illustrates where c is as defined in Example 1.

Existence and Characterization of Minima of Concave Functions

THEOREM 1. Existence and Characterization of Minima of Concave Functions. *If c is a real-valued concave function on a nonempty polyhedral convex set X in $\Re^n$ that contains no lines, the following are equivalent.*

1° c attains its minimum on X at an extreme point thereof.

2° $c^- = c \wedge 0$ has bounded jumps and c is bounded below on the half-lines emanating from a single element of X in every extreme direction therefrom.

3° c is bounded below on the half-lines emanating from each extreme point of X in each extreme direction therefrom.

Proof. Let D be the set of extreme directions of X. For each $x \in X$ and $d \in D$, put $x^\theta \equiv x + \theta d$. Since c is concave, $c(x^\theta)$ is bounded below in $\theta \ge 0$ if and only if $c(x^\theta) \ge c(x)$ for all $\theta \ge 0$.

Clearly 1° implies 2°. Now assume that 2° holds, so there is a $y \in X$ such that $c(y^\theta) \ge c(y)$ for each $\theta \ge 0$ and $d \in D$. Suppose $x \in X$ and put $m \equiv c(x) \wedge c(y) \wedge 0$. We show that $c(x^\theta)$ is bounded below in $\theta \ge 0$. Since c^- is concave, we have for each $0 \le \theta$ and $0 < \lambda \le 1$ that

$$c^-(x^\theta + \lambda(y^\theta - x^\theta)) = c^-(\lambda y^{\theta/\lambda} + \underline{\lambda} x) \ge \lambda c^-(y^{\theta/\lambda}) + \underline{\lambda} c^-(x) \ge m$$

where $\underline{\lambda} \equiv 1 - \lambda$. Now since $c \ge c^-$ and the jumps of c^- are bounded below by a finite number l, it follows from the above inequality that

$$c(x^\theta) \ge \lim_{\lambda \downarrow 0} [c^-(x^\theta) - c^-(x^\theta + \lambda(y^\theta - x^\theta))] + \lim_{\lambda \downarrow 0} c^-(x^\theta + \lambda(y^\theta - x^\theta)) \ge l + m$$

for each $\theta \ge 0$, so 3° holds.

Next, suppose that 3° holds and $x \in X$. Then since $X = \operatorname{conv} E + \operatorname{cone} D$,

$$x = \sum_{i=1}^{p} \alpha_i e^i + \sum_{j=1}^{q} \beta_j d^j = \sum_{i,j} \alpha_i \beta_j [e^i + d^j]$$

for some $e^i \in E$, $d^j \in D \cup \{0\}$, $\alpha_i > 0$, $\sum_k \alpha_k = 1$, $\beta_j > 0$ and $\sum_l \beta_l = 1$, so $\sum_{i,j} \alpha_i \beta_j = 1$ where E is the set of extreme points of X. Thus since c is concave and $c(e^i + d^j) \ge c(e^i)$,

$$c(x) \ge \sum_{i,j} \alpha_i \beta_j c(e^i + d^j) \ge \sum_{i,j} \alpha_i \beta_j c(e^i) \ge \min_i c(e^i). \blacksquare$$

In order to make use of the above result, it is necessary to characterize the extreme points of polyhedral sets arising in inventory systems. This is done in the sequel for an important case arising in practice, viz., uncapacitated network flows, especially those that are planar.

Once the extreme flows have been characterized, there remains the formidable problem of searching them to find one with minimum cost. There is a dynamic-programming method, called the send-and-split method, for doing this for uncapacitated networks. This algorithm, not surprisingly, runs in exponential time in general. However, the algorithm can be refined to run in polynomial time when either the number of *demand nodes*, i.e., nodes at which there are nonzero demands, is bounded or the graph is planar and all but possibly one of the demand nodes lie in the same face. The importance of this in inventory control is that the networks arising in inventory systems often have one of these properties.

3 MINIMUM-ADDITIVE-CONCAVE-COST UNCAPACITATED NETWORK FLOWS [Za68], [EMV79, 87]

Formulation

Consider a (directed) graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ consisting of a set $\mathcal{N}$ of n *nodes* together with a set $\mathcal{A}$ of a ordered pairs of distinct nodes called *arcs*. There is a *demand* r_i at each node i, and $r = (r_i)$ is the *demand* vector. Negative demand at a node is, of course, a *supply*. Let $\mathcal{D}$ be the collection of $d+1$, say, nodes, called *demand nodes*, with nonzero demands.

Let x_{ij} be the number of units of a commodity flowing from node i to node j along arc (i,j). A *flow* is a vector $x = (x_{ij})$ that satisfies the *conservation-of-flow equations*

$$\sum_{(j,i)\in\mathcal{A}} x_{ji} - \sum_{(i,k)\in\mathcal{A}} x_{ik} = r_i \text{ for } i \in \mathcal{N}.$$

A flow is *feasible* if it is nonnegative, and *extreme* if it is an extreme point of the polyhedral set of feasible flows. The directions are the nonnegative nonnull circulations.

There is a (real-valued) additive concave *flow-cost function* $c(x) = \sum_{(i,j)\in\mathcal{A}} c_{ij}(x_{ij})$ defined on the set of nonnegative vectors $x = (x_{ij}) \ge 0$ with each $c_{ij}(\,\cdot\,)$ being concave on the nonnegative real line. We can and do assume in the sequel without loss of generality and without further mention that $c_{ij}(0) = 0$ for all i, j, so $c(\,\cdot\,)$ is *vector subadditive*, i.e., $c(u+v) \le c(u) + c(v)$ for all $u, v \ge 0$, and $c(0) = 0$. The objective is to find a *minimum-cost flow*, i.e., a feasible flow with minimum cost.

Since $c(\cdot)$ is real-valued, additive and concave on the set of feasible flows, it follows from Theorem 1 that $c(\cdot)$ assumes its minimum thereon if and only if there is a feasible flow and $c(\cdot)$ is bounded below on each half-line emanating from some feasible flow in each extreme direction of the set of feasible flows. And in that event the minimum is attained at an extreme flow. For these reasons, it is necessary to characterize the extreme flows and directions.

Extreme Flows and Directions

The characterization of the extreme flows and directions requires a few definitions. Call a graph a *tree* if each pair of nodes is connected by exactly one simple path. Call a node in a tree an *end-node*, or a *leaf*, if it is incident to only one arc. The *union of two graphs* $(\mathcal{N},\mathcal{A})$ and $(\mathcal{N}',\mathcal{A}')$ is the graph $(\mathcal{N}\cup\mathcal{N}',\ \mathcal{A}\cup\mathcal{A}')$. A *forest* is a graph that is a union of node-disjoint trees, or equivalently, that contains no simple cycles. As an illustration, the graph in Figure 4 is a forest with the leaves of its two trees being the nodes $\gamma,\delta,\epsilon,\zeta$ and η,κ,λ respectively.

THEOREM 2. Characterization of Extreme Flows and Directions. *The polyhedral set of feasible flows has the following properties.*

1° *A feasible flow is extreme if and only if its induced subgraph is a forest. Moreover, the leaves of the trees in the forest are demand nodes.*

2° *A direction is extreme if and only if its induced subgraph is a simple circuit.*

Proof. Consider 1° first. To that end, suppose x is an extreme flow and induces a subgraph that is not a forest. Then the induced subgraph contains a simple cycle γ. Let y^ϵ be a simple circulation whose induced cycle is γ and whose flow around γ is ϵ. Then for small enough $\epsilon>0$, $x+y^\epsilon$ and $x-y^\epsilon$ are both feasible flows and x is one-half their sum, contradicting the fact that x is extreme. Conversely, suppose that the subgraph induced by x is a forest and $x=\frac{1}{2}(x'+x'')$ for some feasible flows x' and x''. We must show that $x_\alpha=x'_\alpha=x''_\alpha$ for all arcs α. This is so for arcs not in the forest because the elements of x on arcs not in the forest are zero. Also, by flow conservation, the flow in each arc (i,j) in the forest is unique since it is the sum of the demands at all nodes k for which there is a unique simple path in the forest joining i and k that contains j. Thus 1° holds.

Next consider 2°. To that end, suppose that d is an extreme direction. Then d is a nonnegative circulation. By the Circulation-Decomposition Theorem, d is a sum of distinct simple nonnegative circulations. Thus $d=d'+d''$ where d' is one of the simple circulations and d'' is the sum of the rest. If $d''=0$, the assertion is proved since the subgraph induced by d' is a simple circuit. If not, $d'\neq d''$ and $d=\frac{1}{2}(2d'+2d'')$, which contradicts the fact that d is an extreme direction. Conversely, suppose that d is a nonnegative nonnull simple circulation and $d=d'+d''$ for some nonproportional nonnegative circulations d' and d''. Then $d'_\alpha=-d''_\alpha$ for each arc α not in the simple circuit induced by d. Now if $d'_\alpha \neq 0$, then $d'_\alpha d''_\alpha < 0$, whence d' and d'' cannot both be nonnegative, which is impossible. Thus $d'_\alpha = d''_\alpha = 0$ for each arc α not in the simple circuit. Hence d' and d'' are proportional, which is impossible. ∎

Example 2. Forest Induced by an Extreme Flow. The forest induced by an extreme flow is given in Figure 4. The nodes are labeled by letters. The number to the left of a node (resp., an arc) is the demand there (resp., flow therein). The letters to the right of each arc are the demand nodes that lie below the arc. Observe that the leaves of the two trees are indeed demand nodes, and both positive and negative demands are possible. Also, a node that is not a leaf of a tree need not be a demand node, e.g., node θ.

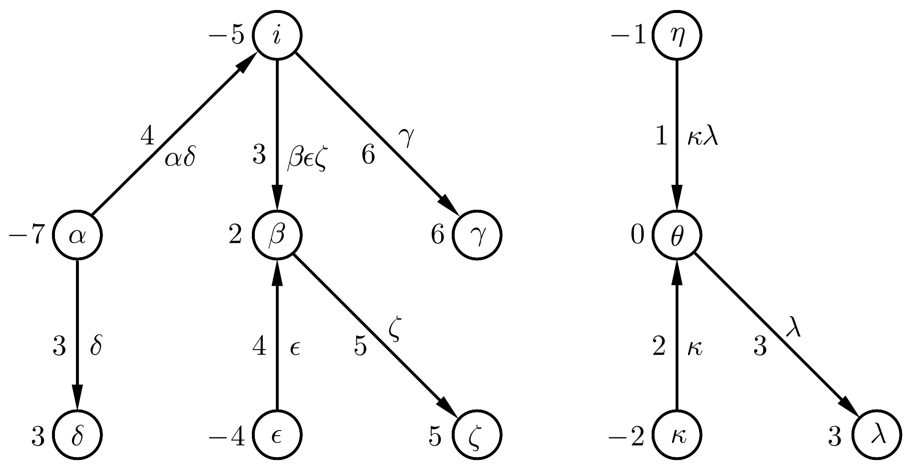


Figure 4. A Forest Induced by an Extreme Flow

Strong Connectedness and Strong Components of a Graph

In order to proceed further, it is helpful to introduce a few concepts about (directed) graphs. One node is *strongly connected* to a second node if there is a simple chain from the first node to the second. A graph is *strongly connected* if each node is strongly connected to each other node. Strongly-connected graphs are connected, but not necessarily conversely. For example, wheels are connected and the wheel of Figure 4.8 is strongly connected, but the wheel of Figure 4.11 is not.

Any graph can be decomposed into maximal strongly-connected subgraphs called the *strong components* of the graph. The node sets of the strong components of a graph partition the nodes of the graph. If we contract the nodes and arcs in each strong component of a directed graph into a single strong-component node and delete copies of other arcs, the resulting *strong-component graph* has no simple circuits. For example, the wheel of Figure 4.11 has two strong components, viz., the hub node and the set of rim nodes. The strong-component graph is then the single arc directed from the hub node to the rim-node set. As a second example, each node of the production-planning network of Figure 1.1 is a strong component of the graph, so the strong-component graph coincides with the graph itself.

Existence of Minimum-Cost Flows and Reduction to Nonnegative Arc Costs

The additivity of $c(\cdot)$ permits the characterization of the existence of minimum-cost flows obtained from Theorem 1 to be improved. The improvement, given in Theorem 3, is to require that $c(\cdot)$ be bounded below on each half-line emanating in each extreme direction only from the origin instead of from some flow. This last condition has the advantage of being independent of the demand vector.

There is another useful characterization of the existence of minimum-cost flows. To describe it, we first require a definition. A *path* is an alternating sequence of nodes and arcs that begins and ends with a node and for which each arc joins the nodes immediately preceding and following it in the sequence. A *chain* is a path in which all arcs are oriented in the same way. Paths and chains generalize simple paths and chains by allowing repetition of nodes and arcs.

Let $\mathbb{G}$ be the augmented graph in which one appends the node ν to $\mathcal{N}$ and appends an arc (i,ν) to $\mathcal{A}$ for exactly one node i in each strong component of $\mathcal{G} = (\mathcal{N},\mathcal{A})$ that is not strongly connected to a distinct strong component of $\mathcal{G}$. As illustrated in Figure 5, this assures that each

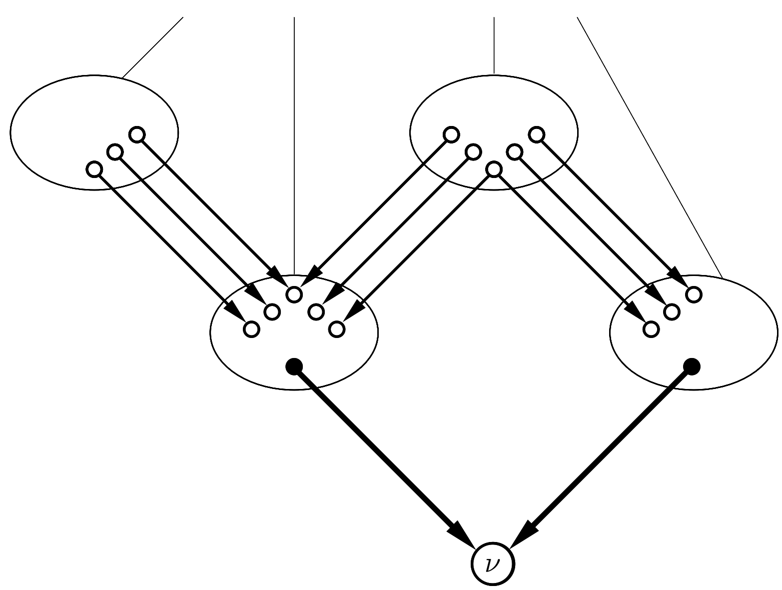


Figure 5. Augmented Graph $\mathbb{G}$

node in $\mathbb{G}$ is strongly connected to ν. Let $\dot{c}_{ij}(\infty) = \lim_{\lambda\to\infty}\dot{c}_{ij}(\lambda)$ where $\dot{c}_{ij}(\cdot)$ denotes the right-hand derivative of $c_{ij}(\cdot)$. Let $\underline{c}_{ij} \equiv \dot{c}_{ij}(\infty)$ for each arc (i,j) in a strong component of $\mathcal{G}$ (and hence $\mathbb{G}$) and let $\underline{c}_{ij}$ be arbitrary, though finite, for every other arc (i,j) in $\mathbb{G}$. Let $\underline{\pi}_i$ be the infimum of the costs of the chains in $\mathbb{G}$ from node $i \neq \nu$ to node ν where the $\underline{c}_{ij}$ are the arc costs in $\mathbb{G}$. Theorem 3 asserts that if there is a flow, there is a minimum-cost flow if and only if $\underline{\pi} = (\underline{\pi}_i)$ is finite. Moreover, $\underline{\pi}$ is finite if and only if the cost of traversing each simple circuit in $\mathcal{G}$ (with the arc costs $\underline{c}_{ij}$) is nonnegative.

If a minimum-cost flow exists, so $\underline{\pi}$ is finite, then Theorem 3 asserts that it is possible to reduce the minimum-cost-flow problem to an equivalent problem with nonnegative arc-flow costs. This is very useful because the running time of the send-and-split algorithm (which will be described shortly) for finding minimum-cost flows can be significantly reduced when the flow costs are nonnegative.

To construct the equivalent problem, let z be an upper bound on the feasible flows in arcs joining distinct strong components of $\mathcal{G}$. For example, $z = \sum_i r_i^+$ will do. The equivalent minimum-cost-flow problem is formed by setting $\pi = (\pi_i) = \underline{\pi}$, $c_{ij}^{\pi}(y) \equiv c_{ij}(y) - (\pi_i - \pi_j)y$ and $c^{\pi}(x) \equiv \sum_{i,j} c_{ij}^{\pi}(x_{ij})$. Observe that for each flow x, $c^{\pi}(x) = c(x) + \sum_i \pi_i r_i$, so that the set of feasible flows that minimize $c^{\pi}(\,\cdot\,)$ is independent of π. Also, since $\pi_i - \pi_j \leq \underline{c}_{ij}$ for all i, j, the altered arc costs $c_{ij}^{\pi}(\,\cdot\,)$ are nonnegative for each feasible flow as illustrated in Figures 6a and 6b. Thus the problem of finding a minimum-cost flow with the altered flow-cost function $c^{\pi}(\,\cdot\,)$ has nonnegative arc-flow costs and is equivalent to the original problem.

THEOREM 3. Existence of Minimum-Cost Flows and Reduction to Nonnegative Arc Costs. *In networks with graph $\mathcal{G}$, the following are equivalent.*

1° *There is a minimum-cost flow for some demand vector.*

2° *The null circulation is a minimum-cost nonnegative circulation.*

3° *If there is a feasible flow, there is a minimum-cost flow that is extreme.*

4° *Each nonnegative simple circulation has nonnegative flow cost.*

5° $\sum_{a \in \gamma} \underline{c}_a \geq 0$ *for all simple circuits with arc-set γ in $\mathcal{G}$.*

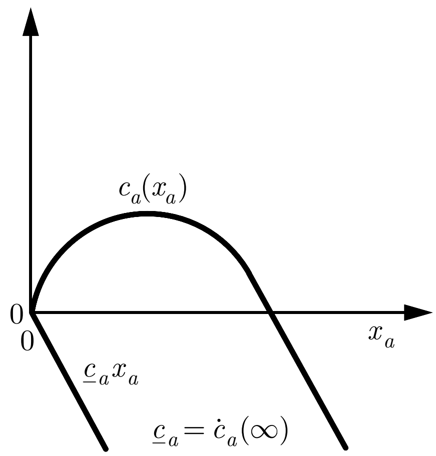
6° *There is a simple minimum-cost chain from each node in $\mathbb{G}$ to ν where the arc-cost vector is $\underline{c}$.*

If also $\underline{c}_a z \leq c_a(z)$ for all arcs a joining distinct strong components of $\mathcal{G}$ where $z \equiv \sum_i r_i^+$ and r is a demand vector, the above are equivalent to:

7° *There is a π for which $c_a^{\pi}(x) \geq 0$ for all arcs a in $\mathcal{G}$ and feasible flows x. Moreover, the vector $\underline{\pi}$ of minimum simple-chain costs in $\mathbb{G}$ is one such π.*

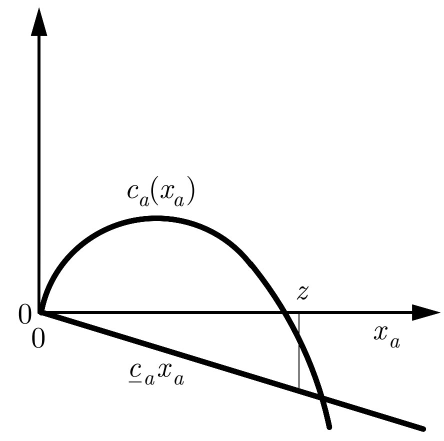
Proof. Let x be a feasible flow and y be an extreme direction of the corresponding set of feasible flows. Then by Theorem 2, the subgraph of $\mathcal{G}$ induced by y is a simple circuit with arcs γ, say. Since $c(\,\cdot\,)$ is concave and subadditive,

$$\frac{1}{2}c(2x) + \frac{1}{2}c(2\theta y) \leq c(x + \theta y) \leq c(x) + c(\theta y).$$



$c_a^{\pi}(x_a) \geq 0$ for $0 \leq x_a$

Figure 6a. Arc a in Strong Component of $\mathcal{G}$



$c_a^{\pi}(x_a) \geq 0$ for $0 \leq x_a \leq z$

Figure 6b. Arc a Not in Strong Component of $\mathcal{G}$

Thus $c(x+\theta y)$ is bounded below in $\theta \geq 0$ if and only if that is so of $c(\theta y)$, or equivalently, $c(\theta y) \geq 0$ for all $\theta \geq 0$. This implies the equivalence of 1°, 3° and 4° by Theorems 1 and 2. Also, 4° and 5° are equivalent because $c(\theta y) \geq 0$ for all $\theta \geq 0$ is equivalent to the inequality $\sum_{a\in\gamma} \dot{c}_a(\infty) \geq 0$. Further, 2° trivially implies 1° and, because the null circulation is the unique extreme point when the demand vector is null, 3° implies 2°.

Clearly 5° implies 6°. To show that 6° implies 7°, observe that the flow in each arc not in a strong component of $\mathcal{G}$ cannot exceed z. Thus, by definition of $\underline{c}$ and the facts that the c_a are concave and vanish at the origin, $c_a(x_a) \geq \underline{c}_a x_a$ for all flows x. Also, $\underline{\pi}_i - \underline{\pi}_j \leq \underline{c}_{ij}$ for all arcs (i,j) in $\mathcal{G}$. Hence, 7° holds with $\pi = \underline{\pi}$. Finally, 7° implies 4° because for each nonnegative simple circulation x, $c(x) = c^{\underline{\pi}}(x) \geq 0$. ∎

It follows from 7° that we can put $\underline{c}_a = \dot{c}_a(\infty)$ for all a for which $\dot{c}_a(\infty)$ is finite. In particular, if c is linear, i.e., $c(x) = cx$, we can put $\underline{c} = c$. Then the Theorem implies the familiar result that a minimum-linear-cost flow exists if and only if the cost of traversing every simple circuit is nonnegative. However, if $\dot{c}_a(\infty) = -\infty$, which occurs if, for example, $c_a(y) = -y^2$ for some arc a in $\mathcal{G}$, but not in a strong component thereof, then we cannot take $\underline{c}_a = \dot{c}_a(\infty)$ because then $\underline{\pi}_i = -\infty$, whence $c^{\underline{\pi}}$ would not be well-defined and finite on its domain.

Computations

To check for the existence of a minimum-cost flow, and if so, find π to make the arc costs nonnegative, proceed as follows. First find the strong components of $\mathcal{G}$ and put $\underline{c}_a \equiv \dot{c}_a(\infty)$ for all arcs a therein. For each remaining arc a in $\mathcal{G}$, choose $\underline{c}_a$ so $\underline{c}_a z \leq c_a(z)$ where $z \equiv \sum_i r_i^+$. Finally, set $\underline{c}_{i\nu} = 0$ for all (i,ν) in $\mathbb{G}$. Then use a suitable minimum-cost-chain algorithm to compute $\pi = \underline{\pi}$ with arc costs $\underline{c} = (\underline{c}_a)$. One such method, successive approximations, entails computation of minimum-cost-k-or-less-arc simple chains to ν for $k = 1, \ldots, n$ (e.g., as in (4.6)) and requires up to na operations (i.e., additions and comparisons) to find π.

4 SEND-AND-SPLIT METHOD IN GENERAL NETWORKS [EMV79, 87]

Subproblems

We now develop the send-and-split method for finding a minimum-cost flow, assuming that such a flow exists. The idea of the algorithm is to embed the problem in a family of subproblems that differ from the given problem only in their demand vectors and to solve the subproblems by induction on the cardinality of their demand-node sets. A typical *subproblem*, denoted $i \to I$, takes the form of finding the minimum-cost C_{iI} of satisfying the *demands at a subset* I of the demand nodes $\mathcal{D}$ from a supply $r_I \equiv \sum_{j \in I} r_j$ available at node i. More precisely, C_{iI} is the minimum cost among all feasible flows for the subproblem $i \to I$ in which one first replaces r_j by zero for all $j \in \mathcal{N} \setminus I$ and then subtracts r_I from the resulting demand at node i. By Theorem 3, either a minimum-cost flow exists for the subproblem $i \to I$, in which case C_{iI} is finite, or no feasible flow exists for the subproblem, in which case one sets $C_{iI} = +\infty$. On putting $I_i \equiv I \setminus \{i\}$, notice that the desired minimum cost over all feasible flows for the original problem is $C_{i\mathcal{D}} = C_{j\mathcal{D}_j}$ for all $i \in \mathcal{N}$ and $j \in \mathcal{D}$ because $r_{\mathcal{D}} = 0$, so the subproblems $i \to \mathcal{D}$ and $j \to \mathcal{D}_j$ coincide for all such i, j. But other C_{iI} will also often be of interest for sensitivity-analysis studies, for example, if the possibility of satisfying the demands only at a subset of the demand nodes is contemplated.

Dynamic-Programming Equations

We begin by showing that the C_{iI} satisfy the dynamic-programming equations (1) and (2) below where the B_{iI} are defined by (3). If $r_I = 0$, which is so when $I = \mathcal{D}$, the subproblem $i \to I$ coincides with the subproblems $j \to I_j$ for all $j \in I$, so

$$(1) \qquad C_{iI} = C_{jI_j} \text{ for all } j \in I.$$

This possibility is illustrated in Figure 4 where $I = \{\eta, \kappa, \lambda\}$, $r_I = 0$ and $i \notin I$.

In order to discuss the case $r_I \neq 0$, it is convenient to introduce a few definitions that will allow us to treat the situations in which r_I is positive or negative in a unified way. For that reason, when we speak about *sending a negative flow* $z < 0$ along a chain from node i to node j in the sequel, we mean sending the positive flow $-z$ along the *reverse chain* (formed by the reverse of each arc in the original chain) from j to i. Also, the cost $c_{ij}(z)$ of sending the negative flow z through arc (i, j) is the cost $c_{ji}(-z)$ of sending the positive flow $-z$ through the reverse arc (j, i). In order to exploit these definitions, for each set $\mathcal{S}$ of arcs, it is convenient to put $\mathcal{S}_I \equiv \mathcal{S}$ if $r_I > 0$

and $\mathcal{S}_I \equiv \{(i,j) : (j,i) \in \mathcal{S}\}$ if $r_I < 0$. In this notation, if $r_I < 0$, then $c_{ij}(r_I) = c_{ji}(-r_I)$ for each $(i,j) \in \mathcal{A}_I$.

Now if $r_I \neq 0$, in which case $\emptyset \subset I \subset \mathcal{D}$, then as we discuss below and prove in Theorem 4, C_{iI} satisfies

$$C_{iI} = \min_{(i,j)\in\mathcal{A}_I} [c_{ij}(r_I) + C_{jI}] \wedge B_{iI} \tag{2}$$

where for $|I| > 1$,

$$B_{iI} \equiv \min_{\emptyset\subset J\subset I} [C_{iJ} + C_{i,I\setminus J}], \tag{3}$$

and for $|I| = 1$, $B_{iI} \equiv 0$ if $I = \{i\}$ and $B_{iI} \equiv +\infty$ otherwise. In order to understand why these equations hold, observe that if the subproblem $i \to I$ has a feasible flow, then one minimum-cost flow is extreme and the graph induced thereby is a forest, for example as might be illustrated in Figure 4.

Now there are two possibilities. One is that it is optimal to *send* the entire supply r_I at node i through a single arc (i,j) to some node j and then optimally satisfy the demands at I from j. Because of the vector subadditivity of the flow costs, the minimum cost of so doing with fixed j cannot exceed the sum $c_{ij}(r_I) + C_{jI}$ of the cost of sending r_I through (i,j) and the minimum cost for the subproblem $j \to I$. Moreover, if j is chosen to minimize that sum and the arc (i,j) is not in the subgraph induced by the minimum-cost flow for $j \to I$, both of which are so with an extreme minimum-cost flow for $i \to I$, then the minimum cost of sending r_I through a single arc incident to i is $\min_j [c_{ij}(r_I) + C_{jI}]$. For example, if $I = \{\beta, \epsilon, \zeta\}$ in Figure 4, then $r_I = 3$ and it is optimal to send the supply $r_I = 3$ at i through arc (i, β) to β and then to satisfy optimally the demands at nodes β, ϵ, ζ from β, which costs $c_{i\beta}(3) + C_{\beta I}$.

The second possibility is that it is optimal to send the supply r_I at i out through two or more arcs. In that event, it is optimal to *split* the subproblem $i \to I$ into two subproblems $i \to J$ and $i \to (I \setminus J)$ with $\emptyset \subset J \subset I$, and solve the two subproblems. Because of the vector subadditivity of the flow cost, the cost of the sum of the two minimum-cost flows for these two subproblems, which is a feasible flow for the subproblem $i \to I$, is at most $C_{iJ} + C_{i,I\setminus J}$. Moreover, equality obtains if the two minimum-cost flows induce subgraphs that are arc-disjoint, as is the case with an extreme minimum-cost flow. Thus, in this event, $C_{iI} = B_{iI}$ where the minimum cost B_{iI} when the problem $i \to I$ is split into two subproblems is given by (3). This possibility is illustrated in Figure 4. For example, suppose that $I = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \kappa, \lambda\}$. Then $C_{iI} = B_{iI}$ and it is optimal to split $i \to I$ into the two subproblems $i \to J$ and $i \to (I \setminus J)$. The subproblem $i \to I$ can be split optimally in several ways, e.g., by setting J equal to $\{\alpha, \delta\}$, $\{\beta, \epsilon, \zeta\}$, $\{\gamma\}$, or $\{\eta, \kappa, \lambda\}$, or a union of up to three of these four sets. Each of the first three of these sets J has the property that $|r_J|$ is the flow in an arc incident to i in the extreme flow of Figure 4. The fourth set J is the set of demand nodes in the tree in the forest that does not contain i. Any other choice of J is not optimal if the extreme flow depicted in Figure 4 is the unique optimal flow.

To sum up for each subproblem $i \to I$ with $r_I \neq 0$, it is optimal either to

- send $|r_I|$ through a single arc incident to i or
- split the subproblem $i \to I$ into two subproblems $i \to J$ and $i \to (I \setminus J)$.

Also, the minimum cost C_{iI} is the smaller of the costs of sending and of splitting at i and so is given by (2) with B_{iI} given by (3).

Send-and-Split Method

The *send-and-split method* uses (1)-(3) to compute the C_{iI} by induction on the cardinality of the demand-node sets $\emptyset \subset I \subseteq \mathcal{D}$. In particular, suppose that the C_{iI} have been found for all i and $|I| < k$, and consider $|I| = k$. If $r_I = 0$, then C_{iI} can be computed from (1) because $|I_j| < k$ for each $j \in I$. If instead $r_I \neq 0$, the B_{iI} can be computed from the *set-splitting equations* (3) because $|J| \vee |I \setminus J| < k$. Then, one can find the C_{iI} for each $i \in \mathcal{N}$ by solving the equations (2).

To solve the last system, it is convenient to reduce (2) to an equivalent system for finding minimum-cost chains. To that end, construct a graph $\mathcal{G}'_I = (\mathcal{N}', \mathcal{A}'_I)$ by appending a new node ν to $(\mathcal{N}, \mathcal{A}_I)$ so that $\mathcal{N}' \equiv \mathcal{N} \cup \{\nu\}$ and $\mathcal{A}'_I \equiv \mathcal{A}_I \cup (\mathcal{N} \times \{\nu\})$. The cost c_{ij} associated with each arc $(i,j) \in \mathcal{A}'_I$ is $c_{ij}(r_I)$ if $j \neq \nu$ and B_{iI} if $j = \nu$ as illustrated in Figure 7. We suppress the dependence of the arc costs on I for simplicity. Then (2) can be rewritten as Bellman's equations for finding the minimum costs $C_i = C_{iI}$ among all chains in $\mathcal{G}'_I$ from each node $i \in \mathcal{N}$ to ν, viz.,

$$(2)' \qquad C_i = \min_{(i,j) \in \mathcal{A}'_I} [c_{ij} + C_j],\ i \in \mathcal{N},$$

where $C_\nu \equiv 0$. Observe that C_i can be thought of as the minimum cost of sending r_I from i through the graph to another node k at which it is optimal to split the subproblem $k \to I$. By Theorem 3, the cost of traversing each simple circuit in $\mathcal{G}$, and hence in $\mathcal{G}'_I$, is nonnegative. Thus, as is well known, the *minimum-cost-chain equations* $(2)'$ have a greatest $+\infty$ or real-valued solution, viz., the desired C_i.

Existence and Uniqueness of Solution to Dynamic-Programming Equations

Unfortunately, $C = (C_{iI})$ is not always the unique solution of (1) and (2) where (3) defines the B_{iI}. For example, if $c_{ij}(\,\cdot\,) = 0$ for all arcs (i,j), $r_I \neq 0$ for all nonempty proper subsets I of $\mathcal{D}$, and $\mathcal{G}$ is strongly connected, then $C = 0$ and $C' = (C'_{iI})$ also satisfies (1) and (2) where

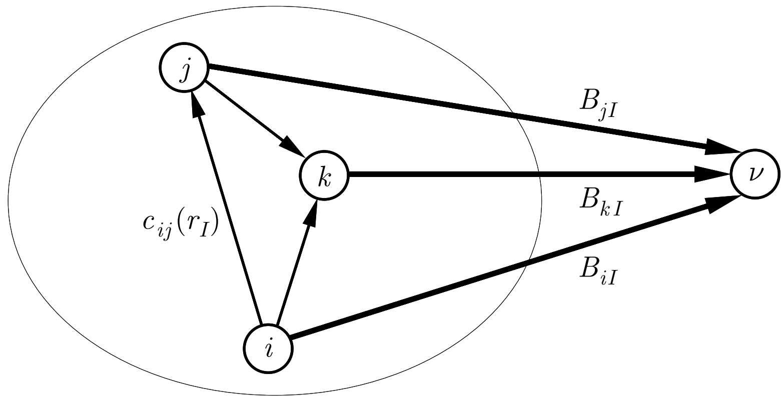


Figure 7. The Arc Costs on the Graph $\mathcal{G}'_I$

$C'_{iI} = -(|I| \wedge d)$ for $i \in \mathcal{N}$ and $\emptyset \subset I \subseteq \mathcal{D}$. However, C is the greatest solution of (1) and (2) as the next result shows. Moreover, C is the only such solution if condition 4° of Theorem 3 is strengthened to require that each nonnegative simple circulation has positive, rather than merely nonnegative, cost. The proof of the next Theorem appears in §4 of the Appendix.

THEOREM 4. Dynamic-Programming Equations. *If there is a minimum-cost flow, then C is the greatest $+\infty$ or real-valued solution of* (1) *and* (2). *If also each nonnegative simple circulation has positive cost, then C is the only such solution.*

Reduction of Minimum-Cost-Chain Problems to Ones with Nonnegative Arc Costs

By choosing $\underline{c}$, z and π as discussed following Theorem 3, one sees that $|r_I| \le z$ for all $I \subseteq \mathcal{D}$. Also, for each arc (i, j), either $c^{\pi}_{ij}(r_I) \ge 0$ or $c^{\pi}_{ij}(-r_I) \ge 0$ according as r_I is nonnegative or nonpositive. Thus we can take all the arc costs in the minimum-cost-chain equations $(2)'$ to be nonnegative. Hence, $(2)'$ can be solved by successive approximations, e.g., as in (4.6), or more efficiently by Dijkstra's method.

Dijkstra's Method For Finding Minimum-Nonnegative-Cost Chains [Di59]

Dijkstra's method can be used to find minimum-cost simple chains from each node to node ν when the arc costs c_{ij} are nonnegative. At each stage of this inductive method, one has at hand a subset S of the nodes that includes ν, the minimum cost C_i over all simple chains from i to ν for $i \in S$, and the minimum cost C_i over all simple chains from i to ν that contain some arc (i, j) with $j \in S$ for $i \in \mathcal{N} \setminus S$ (if there is no such arc (i, j), then $C_i = +\infty$). (Initially, one has $S = \{\nu\}$, $C_\nu = 0$, and, for $i \in \mathcal{N}$, $C_i = c_{i\nu}$ if (i, ν) is an arc and $C_i = \infty$ otherwise.) One first finds an $i = k$ that minimizes C_i over $i \in \mathcal{N} \setminus S$. Then one replaces S by $S' \equiv S \cup \{k\}$ and C_i by C'_i for $i \in \mathcal{N}$ where

$$C_i' \equiv \begin{cases} C_i & , i \in S' \\ C_i \wedge (c_{ik} + C_k) & , i \in \mathcal{N} \setminus S'. \end{cases}$$

The process terminates when $\mathcal{N} \setminus S = \emptyset$, in which case the C_i are then the desired minimum costs over all simple chains from each node i to ν. Incidentally, at each stage, it follows by construction that the minimum-cost simple chains to ν from nodes in S lie entirely in S.

To justify this method, it suffices to show that C_k is indeed the minimum cost over all simple chains from k to ν. To that end, consider any simple chain from k to ν that exits $\mathcal{N} \setminus S$ at a node $j \in \mathcal{N} \setminus S$. Since the arc costs are nonnegative, the cost of that chain is at least C_j by definition thereof. On the other hand, $C_k \le C_j$ by definition of k, so C_k is the desired minimum cost over all simple chains from k to ν.

Finding a Minimum-Cost Flow

We now discuss how to find a minimum-cost flow by doing a little extra bookkeeping while solving (1), (2)$'$ and (3). For each $i \in \mathcal{N}$ and $\emptyset \subset I \subset \mathcal{D}$ with $r_I \neq 0$, record a $j = j_{iI}$ attaining the minimum in (2)$'$. Do this in such a way that the arcs (i, j_{iI}) form a tree T_I spanning $\mathcal{G}_I'$ with all arcs directed towards ν. (The spanning tree T_I will automatically be constructed in the course of solving (2)$'$ iteratively by standard methods without extra computation provided that care is taken not to change an arc used in the tree from one iteration to the next unless this strictly reduces the appropriate cost.) If also $|I| > 1$, record a $J = J_{iI}$ achieving the minimum in (3).

We now show how to construct a minimum-cost flow inductively from the trees T_I and sets J_{iI}. To that end, let k_{iI} be the node adjacent to ν in the unique chain in T_I (Initially, there is only one unsolved subproblem, viz., $i \to \mathcal{D}$ for any node i.) Choose an unsolved subproblem $i \to I$. If $r_I = 0$, replace $i \to I$ in the set of unsolved subproblems by $j \to I_j$ for any $j \in I$. Now suppose that $r_I \neq 0$. If also $I = \{i\}$, delete $i \to I$ from the set of unsolved subproblems. Thus suppose instead that $I \neq \{i\}$. Then if $i \neq k_{iI}$, construct a "chain preflow" by sending r_I units along the unique chain in T_I from i to k_{iI} (remembering to reverse the direction of the flow in the chain if $r_I < 0$), and replace $i \to I$ in the set of unsolved subproblems by $k_{iI} \to I$. However, if instead $i = k_{iI}$, then split I into two subsets J_{iI} and $I \setminus J_{iI}$, and replace $i \to I$ in the set of unsolved subproblems by $i \to J_{iI}$ and $i \to (I \setminus J_{iI})$. Repeat the above construction until the set of unsolved subproblems is empty. Then let Y be the set of chain preflows so constructed and x be their sum. Now by the vector subadditivity of $c(\,\cdot\,)$, $C_{i\mathcal{D}} = \sum_{y \in Y} c(y) \ge c(x) \ge C_{i\mathcal{D}}$, so equality occurs throughout. Hence, x is the desired minimum-cost flow. Since at most d set splits can occur, no more than nd extra additions and table-lookups are required to find a minimum-cost flow once the trees T_I and sets J_{iI} are available.

Finding a Minimum-Cost Flow that is Extreme

The minimum-cost flow x found by the send-and-split method is not generally extreme. However, a minimum-cost flow that is extreme is readily available, e.g., any extreme flow in the graph induced by x.

Running Time in General Networks: $\frac{n}{2}3^d + n^2 2^d$ Operations

The number of operations, i.e., additions and comparisons, needed to execute the send-and-split method is the sum of two dominant terms, viz., that needed to do the minimum-cost chain and the set-splitting computations. The total number of operations required is at most $\frac{n}{2}3^d + n^2 2^d$. As we show below, the first term accounts for the number of operations to do the set splitting and the second for the number to find the minimum-cost chains.

Minimum Cost Chains. It follows from (1) that we can fix any node $\sigma \in \mathcal{D}$ and split only subsets of $\mathcal{D}_\sigma$. There are 2^d subsets I of $\mathcal{D}_\sigma$, and for each (nonempty) one of these, the minimum-cost simple chains must be found. The minimum-cost simple chains for a fixed subset I can be found using Dijkstra's method with at most n^2 operations. To see this, observe that for each subset S, minimizing C_i over $\mathcal{N} \setminus S$ requires $|\mathcal{N} \setminus S| - 1$ comparisons. Since this must be done for only one set S for which $|\mathcal{N} \setminus S| = j$ for each $1 \le j \le n$, the total number of comparisons is at most $\sum_1^n (j-1) = \frac{n}{2}(n-1) \le \frac{n^2}{2}$. The C_i must also be updated for each subset S. This requires one operation for each arc (i,k) encountered. Since each such arc is in $\mathcal{A}$, the number of such arcs does not exceed $n(n-1)$. Also, one-half of the possible arcs in the graph are never considered because if arc (i,k) is encountered, then (k,i) is not. Thus the number of operations required to do the updating is at most $\frac{n}{2}(n-1) \le \frac{n^2}{2}$. Hence, if Dijkstra's method is used, at most $n^2 2^d$ operations are needed to find all the minimum-cost chains.

Set Splitting. Observe that for each fixed node $i \in \mathcal{N}$, set splitting requires one operation for each pair of sets I and J satisfying $J \subset I \subset \mathcal{D}_\sigma$. Also, each choice of I and J amounts to a partition of $\mathcal{D}_\sigma$ into the three subsets J, $I \setminus J$ and $\mathcal{D}_\sigma \setminus I$. The number of partitions of a set with d elements into three subsets is at most 3^d because there are three subsets into which each element of the set $\mathcal{D}_\sigma$ can independently be assigned. This number can be cut in half by noting that if we split any subset I into two proper subsets J and $I \setminus J$, there is no loss in generality in restricting attention to the J that contain any given node in I, say i_I, since then $I \setminus J$ does not contain i_I. Thus the total number of pairs (I,J) that need to be considered for any given node i is at most $\frac{1}{2}3^d$. Since this must be done for each of the n nodes, the number of operations required to do the set splitting is at most $\frac{n}{2}3^d$.

Biconnected and Strong Components

In many applications, including several of those in §6.6, it is necessary to exploit the connectivity structure of $\mathcal{G}$ in order to make the send-and-split algorithm efficient. There is no loss in generality in assuming that $\mathcal{G}$ is biconnected. For if not, one can split the problem into independent subproblems on each biconnected component thereof as discussed in §4.1. Even under this assumption, the set-splitting and minimum-cost-chain calculations can both be reduced further by constructing the (circuitless) *strong-component graph* $\widehat{\mathcal{G}}$ formed by contracting each strong component of $\mathcal{G}$ into a node and deleting duplicates of arcs in $\widehat{\mathcal{G}}$. For each subproblem $i \to I$, the *demand at a strong component* is the sum of the subproblem's demands therein. A necessary and sufficient condition for the subproblem to be infeasible, in which case $C_{iI} = \infty$ and there is no need to split I at node i, is that the restriction of the subproblem to the strong-component graph is infeasible, a fact that can be checked by solving a maximum flow problem on that graph. A sufficient condition for infeasibility is that some strong component with positive (resp., negative) demand is not preceded (resp., followed) by a strong component with negative (resp., positive) demand.

The minimum-cost-chain problems for I can be solved inductively by solving the problems on each strong component of $\mathcal{G}'_I$ that does not precede another strong component of $\mathcal{G}'_I$ in the strong-component graph $\widehat{\mathcal{G}}'_I$ for which the minimum-cost-chain problems have not yet been solved. (The graph $\widehat{\mathcal{G}}'_I$ is formed from $\widehat{\mathcal{G}}$ by reversing the arcs in $\widehat{\mathcal{G}}$ if $r_I < 0$, and appending a node ν and arcs from each node in $\widehat{\mathcal{G}}$ to ν.) Moreover, the strong component can be decomposed into a tree T of biconnected components, each of which is necessarily strongly connected. The minimum-cost-chains from each node of the strong component to ν can then be found inductively by the following *two-pass* divide-and-conquer method.

First, let C be a biconnected component that is a leaf of the tree T, and denote by $T \setminus C$ the subtree of remaining biconnected components other than C. Let i be the cut node that separates C and $T \setminus C$. Now find the minimum-cost chains from each node of C to ν through arcs in C, and let π_i be the resulting (restricted) minimum cost from i to ν. Next replace the cost associated with arc (i, ν) by π_i. Now find the minimum-cost chains from each node in $T \setminus C$ to ν through arcs in $T \setminus C$, and denote by π'_i the resulting (restricted) minimum-cost from i to ν. To execute this last step when $T \setminus C$ consists of two or more biconnected components, one applies the procedure being described recursively. Then replace the cost π_i associated with arc (i, ν) by π'_i. Finally, find the minimum-cost chains from each node in C to ν through arcs in C. (Of course, if $\pi_i = \pi'_i$, these chains are the ones previously found, so no new computations are required.) These last chains are the desired minimum-cost chains from each node of C to ν through all nodes of the strong component.

The above procedure entails finding minimum-cost chains from each node in each biconnected component to ν through arcs in that component, and for some, but not all, of these biconnected components, those chains must be found twice. Of course, one can and should arrange that the component in which minimum-cost chains are definitely found only once is the component that requires the most computation. This two-pass method has the effect of significantly reducing the number of operations required to find the minimum-cost chains. For example, if the number of nodes in each biconnected component of each strong component is bounded above as n grows, then the minimum-cost chains can all be found in linear time. This fact is used often in §6.5.

5 SEND-AND-SPLIT METHOD IN 1-PLANAR AND NEARLY 1-PLANAR NETWORKS

[EMV79, 87]

Call a network *planar* if its graph is planar. Call the set of demand nodes in a single face of a planar embedding of the graph a *face set*.

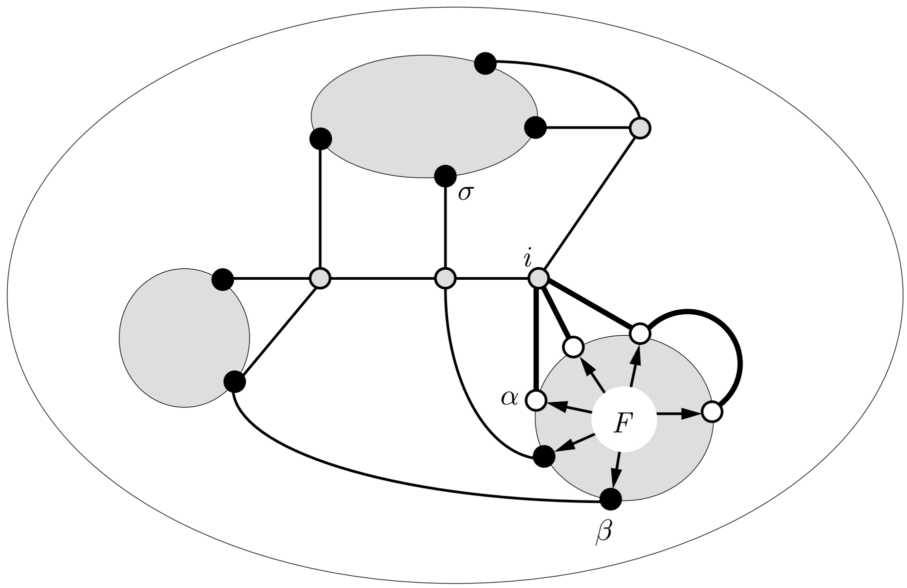
It turns out that the send-and-split method can be refined to run in polynomial time in planar networks in which the set of demand nodes can be covered by a uniformly bounded number of face sets. Here we shall be content to establish this fact for 1-planar and nearly 1-planar networks. Call a network 1-*planar* (resp., *nearly* 1-*planar*) if it is planar and all (resp., all but possibly one) of the demand nodes lie in the boundary of a single face set. For example, the production-planning network of Figure 1.1 is 1-planar, while wheel networks in which every node is a demand node are nearly 1-planar, but not 1-planar.

Extreme Flows in Planar Networks

If F is a face set of a planar network $(\mathcal{G}, r)$ with $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, denote by $\mathcal{G}'' = (\mathcal{N}'', \mathcal{A}'')$ the augmented graph formed by appending a node t to the node set $\mathcal{N}$ and arcs (i, t) from each node i in F to t. The augmented graph evidently inherits the planarity of $\mathcal{G}$. Then it is possible to cyclically order the nodes of F clockwise in $\mathcal{G}''$ around t.

The extreme flows in such a network have a special form that allows attention to be restricted to nonempty subsets I of $\mathcal{D}$ such that $I \cap F$ is a subinterval of F. A *subinterval* of F is a subset $[\alpha, \beta)$ of F that equals F when $\alpha = \beta \in F$ and that, when $\alpha \neq \beta$, is the set of nodes in F that lie between α and β in the clockwise order in $\mathcal{G}''$ around t, including α but not β.

If the graph is connected, there is a minimal tree containing a forest. Since each extreme flow induces a forest, call a minimal tree that contains the forest an *induced tree*. Denote by $i : j$ the unique simple path in the tree that joins nodes i and j therein. Say that a node k in the tree *separates* nodes i and j in the tree if k lies on the path $i : j$. The next result, which Figure 8 illustrates, characterizes the tree that an extreme flow induces in a connected planar network.



○ Demand Nodes in Face Set F Separated from σ by i

● Other Demand Nodes

◎ Nodes in Induced Tree Other than Demand Nodes

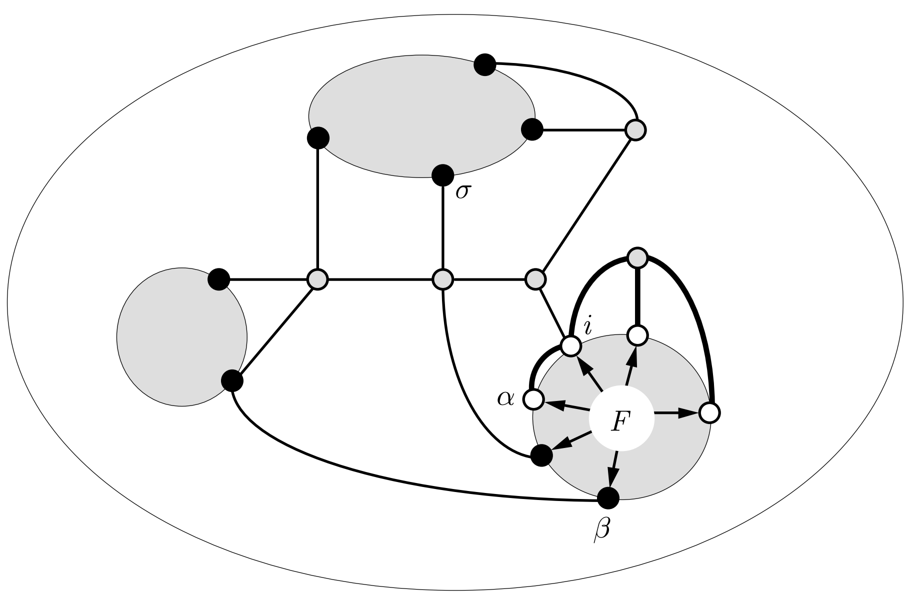


Figure 8. Tree Induced by an Extreme Flow in a Connected Planar Graph

THEOREM 5. Extreme Flows in Planar Networks. *If a connected network* $(\mathcal{G}, r)$ *has a planar embedding, each extreme flow induces a tree with the following properties for each demand node* σ*, node* i *in the tree and face set* F*.*

$1°$ *The subset of* F *that* i *separates from* σ *is a subinterval* $[\alpha, \beta)$ *of* F*.*

$2°$ *If also* i *is an internal node of* $[\alpha, \beta) \subset F$*, then* i *separates* $[\alpha, i)$ *from* $[i, \beta)$*.*

Proof. Let x be an extreme flow and T be a tree that x induces. Let $\mathcal{G}''$ be the augmented graph formed by appending a node t and arcs from each node of F to t. Then $\mathcal{G}''$ is planar. Let I be the subset of F that is separated from σ by i. If $i = \sigma$, then $I = F$, so $1°$ holds trivially, and the hypothesis of $2°$ is not satisfied. Thus it suffices to consider the case $i \neq \sigma$.

If $1°$ does not hold, there exist distinct $j, k, l, m \in F$ in clockwise order in $\mathcal{G}''$ around t such that I contains k and m, but not j and l, and $j = \sigma$ if $\sigma \in F$. Since $\mathcal{G}''$ is planar, the cycle consisting of the paths and arcs $i : k$, (k, t), (m, t), $i : m$, divides the plane into two regions, one containing σ and the other containing either j or l. Thus neither the path $\sigma : j$ nor $\sigma : l$ contains i, but one of them does meet one of the paths $i : k$ or $i : m$ at a node $p \neq i, \sigma$ as Figure 9 illustrates. Hence, T contains the cycle formed by the paths $\sigma : p, p : i, \sigma : i$, contradicting the fact that T is a tree.

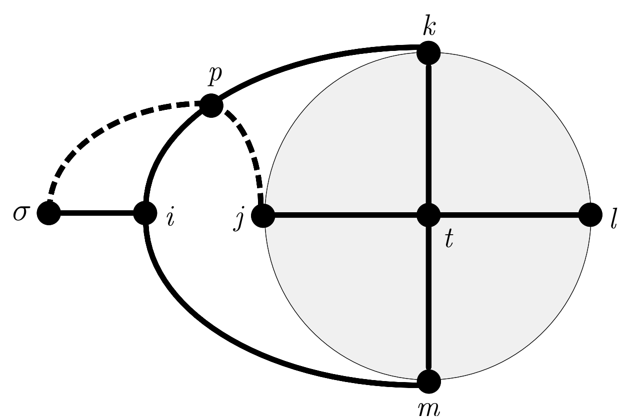


Figure 9. Illustration of Proof of Theorem 5

It remains to prove $2°$. To that end, suppose $l \in [\alpha, i)$ and $j \in [i, \beta)$. We must show that $i : l$ and $i : j$ meet only at i. This is trivially so when $j = i$ since then $i : j = \{i\}$. Thus suppose $j \neq i$. Since $[\alpha, \beta) \subset F$, there is a $k \in F \setminus [\alpha, \beta)$. Also i separates k from j and l, so $i : j$ and $i : l$ meet $i : k$ only at i. Now since $\mathcal{G}''$ is planar, the simple cycle consisting of the path and arcs $k : i$, (i, t), (k, t) in $\mathcal{G}''$ divides the plane into two regions R_j and R_l, say, with $j \in R_j$ and $l \in R_l$. Hence the internal nodes of the path $i : j$ (resp., $i : l$) lie in the interior of R_j (resp., R_l); for if not, $i : j$ (resp., $i : l$) meets $i : k$ at a node other than i, which is a contradiction. ∎

The value of the above characterization of extreme flows in planar networks is that it reduces the number of subsets I that must be considered in executing the send-and-split method provided that some face set has four or more elements. In the sequel, we estimate the running time of the send-and-split method for nearly 1-planar and 1-planar networks.

Running Time in Nearly 1-Planar Networks: $\frac{n}{2}d^3 + n^2d^2$ Operations

Consider a nearly 1-planar network that is not 1-planar, so there is a demand node σ for which $\mathcal{D}_\sigma$ is a face set of the network. By Theorem 5, for each node i, it suffices to consider only subsets I that are subintervals $[\alpha,\gamma)$ of $\mathcal{D}_\sigma$ in solving (1) and (2)$'$. If also, $|I| > 1$ and $r_I = 0$, it suffices to choose $j = \alpha$ in (1), whence I_j is a subinterval of $[\alpha,\gamma)$. Similarly, if instead $r_I \neq 0$, it suffices to consider only those subsets J in (3) for which J and $I \setminus J$ are both subintervals of I. The resulting number of operations to execute the send-and-split method is at most $\frac{n}{2}d^3 + n^2d^2$. As we show below, the first term accounts for the set-splitting and the second term for the minimum-cost chains.

Minimum-Cost Chains. Since each choice of a subinterval I of $\mathcal{D}_\sigma$ is determined by its end points and there are d possible choices of each end point of I, there are at most d^2 subintervals of $\mathcal{D}_\sigma$. Hence, if Dijkstra's method is used, the number of operations required to solve the minimum-cost-chain problems is at most n^2d^2.

Set Splitting. The subintervals $\mathcal{D}_\sigma \setminus I$, $I \setminus J$ and J comprise a partition $\mathcal{D}_\sigma$. Such a partition into three subintervals is determined by the choice of three points in $\mathcal{D}_\sigma$. Since there are at most d ways to choose each point, the number of such partitions is at most d^3. However, it suffices to consider one-half of these partitions, just as in the general case. Hence the number of operations to execute set splitting is at most $\frac{n}{2}d^3$.

Running Time in 1-Planar Networks: $\frac{n}{6}d^3 + \frac{1}{2}n^2d^2$ Operations

The running time of the send-and-split method in 1-planar networks can be reduced by a factor of two to three over that in nearly 1-planar networks. In a 1-planar network with face set $\mathcal{D}$, it is convenient to cyclically label the demand nodes around $\mathcal{D}$ by $1,\ldots,d{+}1$ with distinguished node $\sigma \equiv d+1$. Then each subinterval $[\alpha,\gamma)$ of $\mathcal{D}_\sigma$ consists of the integers $\alpha,\ldots,\gamma{-}1$ where $1 \le \alpha < \gamma \le d+1$. The number of operations required by the send-and-split method is $\frac{n}{6}d^3 + \frac{1}{2}n^2d^2$ as we now show.

Minimum-Cost Chains. Evidently the number of subintervals $[\alpha,\beta)$ of $\mathcal{D}_\sigma$ is at most $\frac{1}{2}d^2$. Thus, if Dijkstra's method is used, the number of operations required to solve the minimum-cost-chain problems is at most $\frac{1}{2}n^2d^2$.

Set Splitting. The number of pairs of subintervals $I = [\alpha, \gamma)$ of $\mathcal{D}_\sigma$ and proper subintervals $J = [\alpha, \beta)$ of I is the number of integer triples (α, β, γ) that satisfy $1 \le \alpha < \beta < \gamma \le d + 1$. In order to compute the number of such triples, it is convenient to recall that the binomial coefficients satisfy the important identity

$$\sum_{i=0}^{j} \binom{i+k}{k} = \binom{j+k+1}{k+1},\ j, k = 0, 1, \dots,$$

as is readily verified by induction. Now since $\binom{i}{0} = 1$ for $i \ge 0$, it follows that the number of triples (α, β, γ) is, on substituting $i = \alpha - 1$, $j = \beta - 2$ and $k = \gamma - 3$,

$$\sum_{0 \le i \le j \le k \le d-2} \binom{i}{0} = \sum_{0 \le j \le k \le d-2} \binom{j+1}{1} = \sum_{0 \le k \le d-2} \binom{\kappa+2}{2} = \binom{d+1}{3} \le \frac{1}{6} d^3.$$

Hence, the total number of operations to execute set splitting is at most $\frac{n}{6} d^3$.

6 REDUCTION OF CAPACITATED TO UNCAPACITATED NETWORK FLOWS [Wa59]

The send-and-split method can be used to solve capacitated network-flow problems with nonnegative flows because they can be reduced to uncapacitated ones with nonnegative flows in the following simple way. Consider a network-flow problem in which there is an arc (i, j) with nonnegative flow x that cannot exceed an upper bound u. This may be expressed by writing the capacity constraint as $x + y = u$ where $y \ge 0$ is the excess capacity in the arc. This capacitated problem can be reduced to an uncapacitated one by subtracting the capacity constraint from the original flow-conservation constraint for node j and replacing the latter by the constraint so formed. The reduced problem has one less arc with upper-bounded flow, one additional node ν, say, (associated with the capacity constraint) and demand u at ν. Also, the arc (i, j) is replaced by arcs (i, ν) and (j, ν) carrying nonnegative flows x and y respectively, and the demand η, say, at node j is replaced by $\eta - u$. Figure 10 illustrates this reduction.

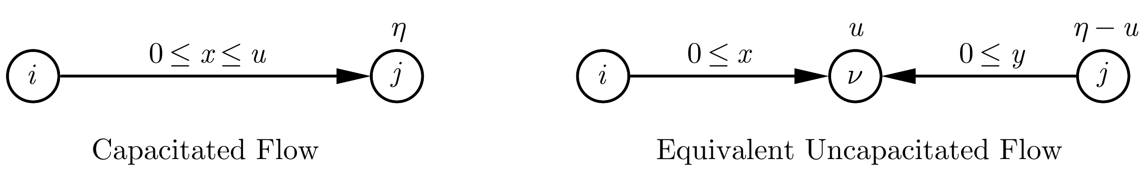


Figure 10. Reduction of Capacitated Arc Flows to Uncapacitated Ones

If the above construction is repeated for every arc in the network having upper-bounded flows, the capacitated problem is reduced to an uncapacitated one. If the original flow-cost function is additive and concave, then the send-and-split method can thus be applied to the reduced problem. On the other hand, since the reduction for each capacitated arc adds a node and an arc, and up to two demand nodes, the reduction for all capacitated arcs may significantly increase the computational effort required to solve the problem.

Observe that the reduction of a planar network is also planar. Thus, if the upper-bounded arc flows all occur on arcs in the boundary of the face of a nearly 1-planar (resp., 1-planar) network that contains all but one (resp., all) of the demand nodes, the near 1-planarity (resp., 1-planarity) of the network is inherited by the reduced network. Thus, the send-and-split method solves such minimum additive-concave-cost problems in polynomial time.

7 APPLICATION TO DYNAMIC LOT-SIZING AND NETWORK DESIGN [WW58], [Ma58], [Za69], [Ve69], [DW71], [Ko73], [JM74], [EMV79, 87], [WVK89], [FT91], [AP93]

Many inventory and network-design problems exhibit scale economies and can be formulated as minimum-concave-cost network-flow problems. The purpose of this section is to show how our theory permits many such problems to be solved in an efficient and unified way. These problems include the dynamic single-facility economic-order-and-sale-interval problem, the dynamic tandem-facilities economic-order-interval problem, and the Steiner-tree problem in graphs.

Dynamic Single-Facility Economic-Order-and-Sale-Interval Problem [WW58], [Ma58], [Za69], [Ve69], [Ko73], [EMV79, 87], [WVK89], [FT91], [AP93]

An inventory manager seeks a minimum-cost plan to order, sell, store and backorder a single product over n periods, labeled $1,\dots,n$. The cost is an additive concave function of the amounts ordered, sold, stored and backordered in the n periods, with each of the functions of one variable vanishing at the origin. Let x_k and s_k be respectively the nonnegative (variable) amounts ordered and sold in period $1 \le k \le n$, and let y_k and z_k be respectively the amounts stored and backordered at the end of period $0 \le k \le n$. The initial and final inventories and backorders are fixed. In addition to the (variable) amounts ordered and sold in a period, there is a fixed (possibly negative) *demand* r_k in period k, $1 \le k \le n$, which, when $k = 1$ (resp., $k = n$), also includes the fixed initial net backorders $z_0 - y_0$ (resp., final net inventory $y_n - z_n$). The network for this problem is 1-planar as illustrated in Figure 11 with $n = 4$.

Let $\sigma = 0$ be the demand node that is incident to all the order-and-sale-quantity arcs. Then the subsets of demand nodes that must be considered are the subintervals of $\mathcal{D}_0$. We can take $\mathcal{D}_0$ to be the set $\{1,\dots,n\}$, though this may entail including nodes with zero demand in that set.

Cubic Running Time. Since there are $n+1$ nodes and each can be a demand node, straightforward application of the send-and-split method to this problem will run in $O(n^4)$ time. In fact it is possible to implement the send-and-split method (strictly speaking, a variant thereof) in $O(n^3)$ time by cutting the number of subintervals of demand nodes that must be considered

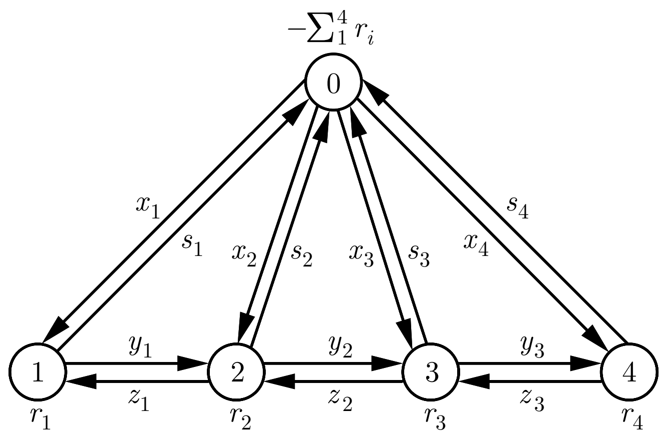


Figure 11. 1-Planar Network for Four-Period Single-Facility Economic-Order-and-Sale-Interval Problem

from $\binom{n+1}{2}$ to n. Indeed, the only subintervals of $\mathcal{D}_0$ that need to be considered take the form $(k] \equiv \{1,\ldots,k\}$ for some $1 \le k \le n$. To see this, observe that it is enough to restrict the nodes of the subtrees induced by the extreme flows for the subproblem $0 \to (k]$ to the set $\{0\} \cup (k]$. Moreover, in executing the set-splitting routine (3) at node 0 for the interval $(k]$, it suffices to choose the subinterval $J \equiv \{i{+}1,\ldots,k\}$ so that it will not be split again at node 0 and that $|r_J|$ will be the flow in some arc joining node 0 and some node $j \in J$. If $r_J > 0$ (resp., < 0), the flow $|r_J|$ is in arc $(0,j)$ (resp., $(j,0)$), and so is the amount ordered (resp., sold) in period j. If a tree induced by an extreme flow contains arc $(0,j)$ (resp., $(j,0)$), then the subtree of the tree that is separated from 0 by j is a path that spans J and no other nodes. Hence, the cost c_{ik}^j of satisfying the demands at nodes in J from j is uniquely defined. If there is no way of satisfying those demands, that cost is, of course, ∞. Moreover, on setting $C_k \equiv C_{0,(k]}$ and

$$c_{ik} \equiv \min_{i<j\le k} c_{ik}^j,$$

we see that the send-and-split dynamic-programming equations can be written as

$$C_k = \min_{0\le i<k} [C_i + c_{ik}],\ k = 1,\ldots,n,$$

where $C_0 \equiv 0$. We show below that the c_{ik} can all be calculated with up to $\frac{1}{3}n^3 + O(n^2)$ additions and $\frac{1}{6}n^3 + O(n^2)$ comparisons and function evaluations. Once the c_{ik} are found, the C_k can be calculated with up to $\frac{1}{2}n^2 + O(n)$ additions and comparisons because at most one addition and comparison is required for each $0 \le i < k \le n$.

To see how to compute the c_{ik} in the claimed running time, let $R_j \equiv \sum_1^j r_i$ and $R_{ij} \equiv R_j - R_i$ for $1 \le i < j$. Let $h_j(w)$ be the cost of storing $w \ge 0$ (resp., backordering $-w \ge 0$) units of the product at the end of period j. Let $c_j(w)$ be the cost of ordering $w \ge 0$ (resp., selling $-w \ge 0$)

units of the product in period j. We can and do assume that $c_j(0) \equiv h_j(0) \equiv 0$ without loss of generality. Now since extreme flows do not entail both production and sales or both inventories and backorders in the same period, the c^j_{ik} can be calculated from

$$c^j_{ik} = B_{ij} + c_j(R_{ik}) + H_{jk}$$

where

$$B_{ij} \equiv \sum_{i<l<j} h_l(-R_{il}) \text{ and } H_{jk} \equiv \sum_{j\le l<k} h_l(R_{lk})$$

for all $0 \le i < j \le k \le n$. Observe that the R_j can all be computed in $O(n)$ time and so the R_{ij} can all be computed in $O(n^2)$ time. Hence, since the B_{ij} and H_{jk} can be computed recursively respectively from

$$B_{i,j+1} = B_{ij} + h_j(-R_{ij}) \text{ and } H_{jk} = H_{j+1,k} + h_j(R_{jk}),$$

it follows that the B_{ij} and H_{jk} can also all be computed in $O(n^2)$ time.

Thus, after the above computation, each c^j_{ik} can be computed with two additions and one function evaluation. And the c_{ik} can all be computed with at most one comparison for each c^j_{ik}. Since the number of c^j_{ik} is the number of triples (i, j, k) of integers that satisfy $0 \le i < j \le k \le n$, it is easy to see, by an argument like that used to estimate the running time of set splitting in 1-planar networks, that the number of such triples is at most $\frac{1}{6}n^3 + O(n^2)$. This verifies our claim that the c_{ik} can all be computed with at most $\frac{1}{6}n^3 + O(n^2)$ comparisons and function evaluations and twice that many additions.

Upper Bounds on Storage or Backorders. If storage is (resp., backorders are) permitted in a period, but not backorders (resp., storage), and if there is an upper bound on storage (resp., backorders) in the period, then that capacitated arc may be replaced by two tandem uncapacitated arcs as discussed in §6.5. (Actually, this method can be extended to handle upper bounds on both storage and backorders in any period, as well as on orders and sales in periods 1 and n, but we omit a discussion of this case for brevity.) The transformed problem is an uncapacitated, $(2n-1)$-period, single-facility, economic-order-and-sale-interval problem in which ordering and selling are prohibited in even-numbered periods, storage or backorders are allowed alternately in each period, and demands are both positive and negative. The transformed network is 1-planar and the above implementation of the send-and-split method requires $\frac{4}{3}n^3 + O(n^2)$ additions and $\frac{2}{3}n^3 + O(n^2)$ comparisons.

Quadratic Running Time. The running time can be reduced to $O(n^2)$ whenever the c_{ik} can all be computed in that time. This is possible if, for example, the demands are nonnegative and there are no sales, backorders or upper bounds on storage. Then we can take $j = i + 1$ in the definition of the c_{ik}. This implementation of the send-and-split method requires up to $\frac{3}{2}n^2 + O(n)$ additions and $\frac{1}{2}n^2 + O(n)$ comparisons.

Dynamic Serial-Facilities Economic-Order-Interval Problem [Za69], [Ko73], [JM74], [EMV79, 87]

A significant generalization of the dynamic economic-order-interval problem is to the case of N serial facilities, labeled $1, \ldots, N$. One seeks a minimum-cost plan to order, store and backorder a single product at each facility over n periods, labeled $1, \ldots, n$. Facility i orders only from facility $i-1$ for $2 \le i \le N$, and facility 1 orders from a supplier. Each facility i orders and stores nonnegative amounts in each period with x_k^i being the amount ordered in period $1 \le k \le n$ and y_k^i being the amount stored at the end of period $0 \le k \le n$. In addition, facility N backorders $z_k \ge 0$ units at the end of period $0 \le k \le n$. The initial and final inventories and backorders are zero. There is a demand r_k at facility N for the product sold there in each period $1 \le k \le n$. The demand in period one includes the initial net backorders $z_0 - y_0^N$ and the demand in period n includes the final net inventory $y_n^N - z_n$. The cost is an additive concave function of the amounts ordered, stored and backordered in each period, with each function of one variable vanishing at the origin. The network associated with this problem is biconnected and 1-planar as Figure 12 illustrates for $N = 3$ and $n = 4$.

Running Time. The send-and-split algorithm solves the problem in at most $\frac{1}{6}(N+1)n^4 + O(Nn^3)$ operations. To see this, let σ be the (black) node at the top of the graph $\mathcal{G}$ and L be the set of n (black) nodes at the bottom of the graph corresponding to facility N. Then the subgraph $\mathcal{G}_L$ induced by L is a strong component of $\mathcal{G}$, as is each other single node in the graph. Also, $\mathcal{G}_L$ has $n-1$ biconnected components, each a *bicircuit*, i.e., simple circuit on two nodes. Thus, for each $i \in L$ and subinterval I of $\mathcal{D}_\sigma \subseteq L$, the subproblem $i \to I$ can be solved by decomposing it into $n-1$ independent network-flow problems, each of whose graphs is a bicircuit.

Since each of the bicircuit subproblems at facility N has at most one arc entering (resp., leaving) each node, set splitting is not required at nodes in L. By applying the running-time analysis for 1-planar networks, it is easy to see that the set splitting at the remaining $(N-1)n+1$ nodes entails at most $\frac{1}{6}(N-1)n^4 + O(n^3)$ operations.

The minimum-cost-chain computations can all be carried out with $O(Nn^3)$ operations using the two-pass divide-and-conquer method given at the end of §6.4. This follows from the fact that each biconnected component of a strong component of $\mathcal{G}$ has at most two nodes, there are $O(Nn)$ arcs in the augmented graph and the minimum-cost-chain problems must be solved for $O(n^2)$ subintervals of demand nodes.

In the special case in which the demands at facility N are nonnegative in each period and there are no backorders (resp., is no storage) there, the running time of the send-and-split method

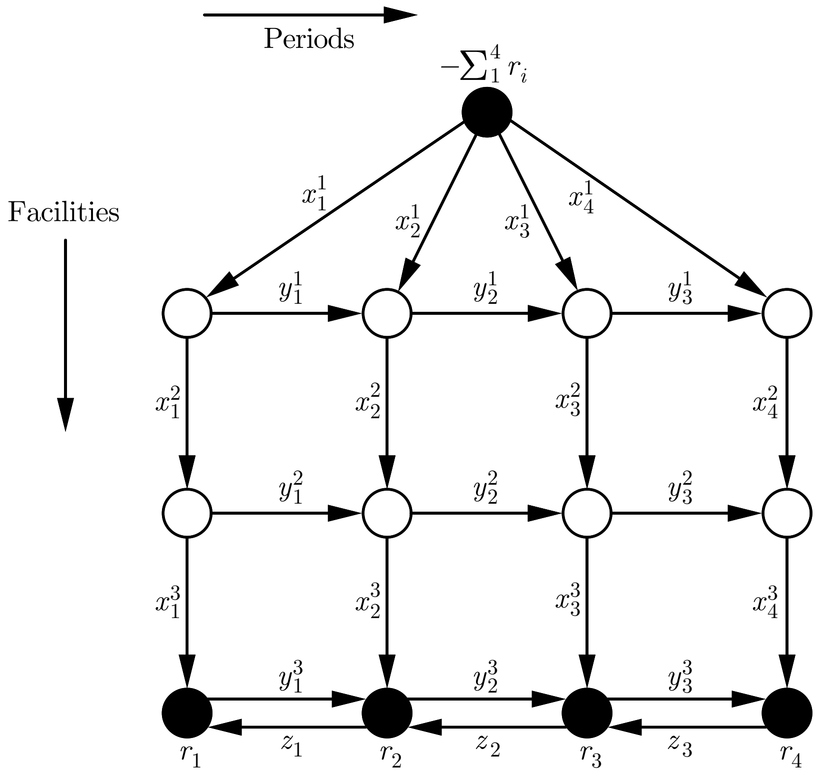


Figure 12. 1-Planar Network for Four-Period Three-Serial-Facility Economic-Order-Interval Problem

improves to $\frac{1}{24}(N-1)n^4 + O(n^3)$ operations because that is so of the set splitting. To see this, observe that the graph is circuitless. Also, for each node l and subinterval I of $\mathcal{D}_\sigma$, the subproblem $l \to I$ is feasible only if $l = \sigma$ or l is above and to the left (resp., right) of the left (resp., right) end node of I in the graph. Thus, the number of operations required to carry out the needed set splitting is

$$(N-1)\sum_{1\le i<j-1\le n} i(j-i-1) \le \frac{1}{24}(N-1)n^4.$$

Network Design: Minimum-Cost Forest [EMV79,87], [DW71]

Consider the problem of designing a minimum-cost network, e.g., of roads, pipelines, railways, transmission lines, etc., to meet given demands for service when there are scale-economies in building arc capacities. To formulate the problem precisely, we require a few definitions. A subgraph of the graph *spans* the $d+1$ demand nodes if they all lie in the subgraph. A subgraph is *feasible* if there is a flow in the network that induces a subgraph of the given subgraph. Of course, a feasible subgraph necessarily spans the set of demand nodes. Moreover, there is a unique flow in a feasible forest, and the flow in each arc is then called the arc's *capacity*. The *cost* of a feasible forest is the sum of the costs of its arc capacities, with the cost of an arc's capacity being concave therein and vanishing at the origin. The network-design problem is that of finding a minimum-cost feasible forest. The problem is equivalent to one of finding a minimum-cost network flow, and so can be solved by the send-and-split method with the usual running times. The case in which the network is 1-planar arises in this context if, for example, the network is located on a land mass that is bounded by water with the demand nodes all lying on the coast.

The undirected version of the above problem arises if an arc is in the graph only if its reverse arc is also in the graph, and both arcs have the same capacity costs. In the undirected problem, a forest is feasible if and only if it spans the demand nodes and the sum of the demands at the nodes in each tree in the forest is zero.

Minimum-Cost Arborescence. In the special case in which there is a single demand node with negative demand, called the *source*, the feasible forests are precisely the arborescences that are rooted at the source and that span the demand nodes. (An *arborescence* is a tree in which all arcs are directed away from a distinguished node called the *root*.) In this event the minimum-cost-forest problem becomes the minimum-cost-arborescence problem. The send-and-split method solves this problem in polynomial time in $(d+1)$-demand-node networks and in nearly 1-planar networks. Observe that when the arc capacity costs are setup costs, the cost of an arborescence with given source node depends only on the set of demand nodes and not on the size of the demands at those nodes. Thus, in that case, we can assume without loss of generality that the demand is $-d$ at the source and 1 at the other d demand nodes.

Minimum-Cost Chain. The minimum-cost-chain problem is the special case of the minimum-cost arborescence problem in which $d = 1$, because then the feasible arborescences are simply the chains from the source to the other demand node, and the flow in each such chain is the same. For this problem, the send-and-split method reduces to solving the minimum-cost-chain problem, and so is as efficient as any method for so doing.

Steiner-Tree-Problem in Graphs. The Steiner-tree problem in an undirected graph is that of finding a minimum-cost tree that spans a given set of $d+1$ nodes of the graph where the arc costs are setup costs. The problem is equivalent to the undirected minimum-cost-arborescence problem in which the capacity costs are setup costs. The send-and-split method solves this problem in polynomial time in $(d+1)$-demand-node and nearly-1-planar networks.

Minimum Spanning Tree. The minimum-spanning-tree problem is the special case of the Steiner-tree problem in graphs in which $d+1 = n$. This problem was solved by Kruskal [Kr56] in $O(n^2)$ time and by Prim [Pr57] in $O(a \ln a)$ time, both of which are much faster than the send-and-split method for this problem.

8 94%-EFFECTIVE LOT-SIZING: ONE-WAREHOUSE MULTI-RETAILER SUPPLY CHAINS

[Ro85]

One of the most interesting problems arising in supply chains occurs where there is a single warehouse supplying N retailers as Figure 13 illustrates. Unfortunately, the only known dynamic-

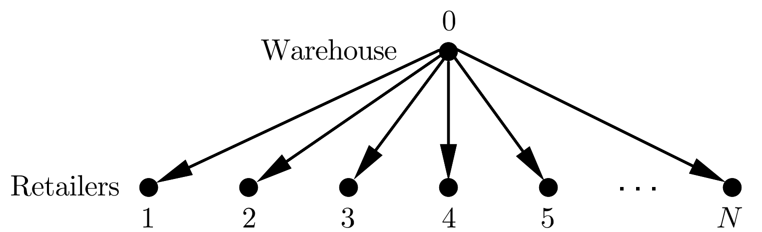


Figure 13. One-Warehouse N-Retailer System

programming algorithms for solving this problem run in exponential time. For example, it is possible to generalize the algorithm that the preceding sections discuss for n-period serial supply chains to the one-warehouse N-retailer problem. Although the running time is polynomial in n, it is exponential in N. Similarly, for the case of linear storage-cost rates and setup ordering costs, an alternate algorithm is available whose running time is linear in N, but exponential in n. This raises the possibility of seeking instead a polynomial-running-time algorithm for finding a schedule that, although not necessarily optimal, is nearly so. That is the goal of this section.

To that end, consider an infinite-horizon continuous-time version of the one-warehouse N-retailer problem in which the costs and demand rates are stationary and facility dependent, the storage costs are linear and the production costs are of setup type. The sequel shows how to find, in $O(N \log N)$ time, a schedule that has long-run average cost within 6% of the minimum possible!

For purposes of discussion, it is convenient to think of the *system inventory*, i.e., the sum of all inventories that the warehouse and retailers hold, as consisting of N different products with each retailer n stocking product n and no others. The warehouse holds inventories of all products.

Sales Rates

Sales occur at each retailer at a constant deterministic rate. Without loss of generality choose the unit of each product so the sales rate therefor, i.e., the *sales rate* per unit time at each retailer, is two. This formulation simplifies notation in the sequel and allows the sales rates measured in common units to be retailer dependent. Sales must be met as they occur over an infinite horizon without shortages or backlogging. Orders by retailers generate immediate sales at the warehouse and deliveries of orders by the warehouse are instantaneous.

Costs

In the sequel, it will sometimes be convenient to refer to the retailers and the warehouse as *facilities*, with the warehouse being facility zero and retailer $n > 0$ being facility n. There is a *setup cost* $K_n > 0$ for placing each order at facility n.

There is a *unit holding cost rate* $h'_n > 0$ per unit time for storing product n at retailer n and a *unit holding cost rate* $h^n > 0$ per unit time for storing product n at the warehouse. Because of the choice of units, the holding cost rates at the warehouse may be product dependent. The *incremental holding cost rate* at retailer n is $h_n \equiv h'_n - h^n > 0$. The assumption that $h_n > 0$ is not essential, but generally so and simplifies the exposition.

Effectiveness

The goal is to find a policy with minimum or near-minimum long-run average cost per unit time. Since there is no known way of finding an optimal policy for this problem, and optimal policies are usually very complex in any case, we are led to seek policies with high guaranteed effectiveness. The *effectiveness* of a policy is 100% times the ratio of the infimum of the average cost over all policies to the average cost of the policy in question. Optimal policies are those with 100% effectiveness. Occasionally, it is also convenient to say that a policy with $100e\%$ effectiveness has *effectiveness* e.

The sequel introduces the class of "integer-ratio policies" and shows, for any data set, that there is a policy in the class with effectiveness at least 94%! Moreover, this development shows how to find such a policy in $O(N \log N)$ time.

Integer-Ratio Policies

Let R be an infinite subset of the positive integers and their reciprocals that has no least or greatest element. For each $r \in R$, let r_- (resp., r_+) be the greatest (resp., least) element of R that is less (resp., greater) than r. Assume that $1 \in R$ and $\frac{r}{r_-} \le 2$ for all $r \in R$. Thus $\frac{1}{2}$ and 2 are also in R. Two examples of such sets R are the positive integers and their reciprocals, and the integer powers of two.

An *integer-ratio* policy is a sequence $\mathcal{T} = (T_0, \ldots, T_N)$ of positive numbers such that each facility n places an order once every $T_n > 0$ units of time for an amount equal to the demand at the facility during the interval until the next order, and $\frac{T_n}{T_0} \in R$ for each retailer n. The term integer-ratio reflects the fact that either $\frac{T_n}{T_0}$ or $\frac{T_0}{T_n}$ is an integer. In the sequel it will often be convenient to denote the order interval T_0 at the warehouse by T.

If $T_n \le T$ for all n, the order quantities at all facilities are stationary. However, if $T_n > T$ for some n, the order quantities at the warehouse are periodic. For example suppose, as Figure 14 illustrates, that there are two retailers, $T = 1$, $T_1 = \frac{1}{2}$ and $T_2 = 2$. The order quantities of product two at the warehouse are $4, 0, 4, 0, \ldots$. The order quantities at the retailers are stationary.

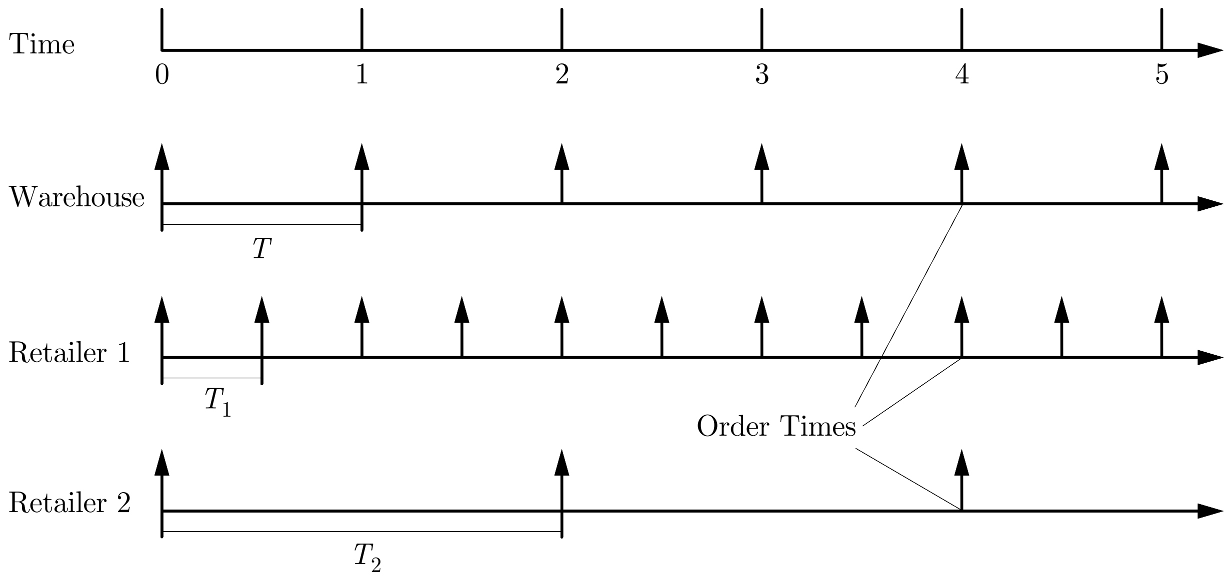


Figure 14. Timing of Orders in a Simple Integer-Ratio Policy

Average Cost of Supplying Retailer n

In the sequel it will be necessary to find the average cost $c_n(T, T_n)$ per unit time of supplying the demand for product n with an integer-ratio policy $\mathcal{T} = (T, T_1, \ldots, T_N)$. The average cost c_n includes the setup costs and holding costs at retailer n, and the cost of holding product n at the warehouse. The c_n's do not include the setup costs at the warehouse.

Case 1. Retailer Order Interval Majorizes that at Warehouse. When $T_n \ge T$, the warehouse orders whenever retailer n orders. Therefore the warehouse holds no inventory of product n, and the only costs to consider are those that the retailer incurs. Thus

$$c_n(T, T_n) = \frac{K_n}{T_n} + h'_n T_n.$$

This is the average-cost function that leads to the Harris square-root formula in the one-facility model.

Case 2. Retailer Order Interval Minorizes that at Warehouse. Suppose instead that $T_n \le T$. The *system inventory* of product n is the sum of the inventory of product n at the warehouse and the inventory at retailer n as Figure 15 illustrates. The average holding cost of product n is the product of the average system inventory of product n and its holding cost rate at the warehouse, plus the product of the average inventory at retailer n and the incremental holding cost rate there. The average cost is thus

$$c_n(T,T_n) = \frac{K_n}{T_n} + h_n T_n + h^n T.$$

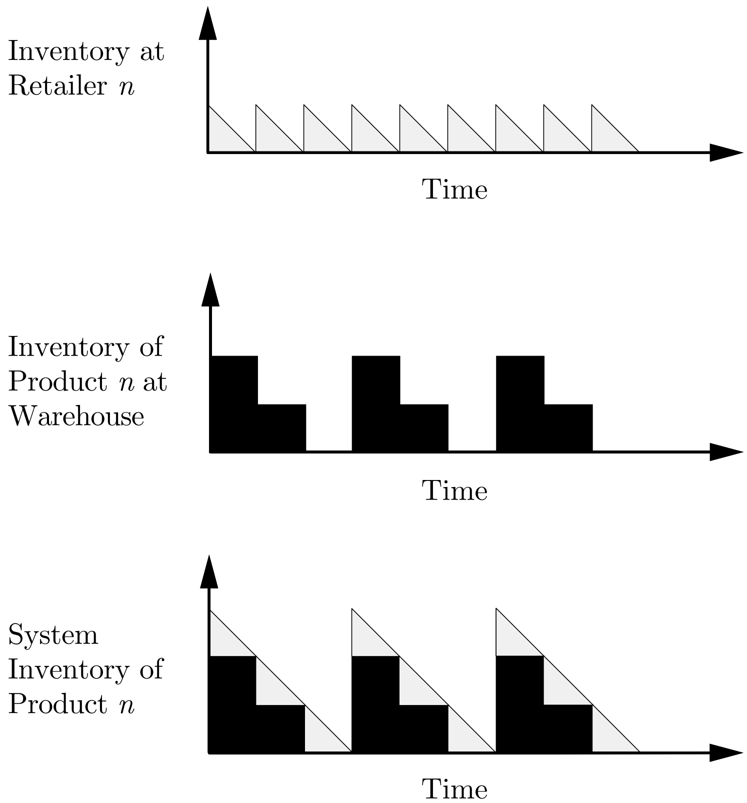


Figure 15. Inventory Patterns when $T_n \le T$

Average Cost of an Integer-Ratio Policy

Since $h'_n = h_n + h^n$, it is possible to combine the above formulas for c_n as

$$c_n(T,T_n) = \frac{K_n}{T_n} + h_n T_n + h^n(T \vee T_n). \tag{4}$$

Note that c_n is convex on the positive orthant. The *average cost* of the integer-ratio policy $\mathcal{T}$ is $c(\mathcal{T}) \equiv \frac{K_0}{T} + \sum_{n\ge 1} c_n(T,T_n)$. Clearly c is strictly convex on the positive orthant.

Lower Bound on the Average Cost of all Policies

Now drop the integer-ratio constraints $\frac{T_n}{T} \in R$, use (4) to extend the definition of c to the entire positive orthant and consider the relaxed problem of finding an *optimal relaxed order-interval policy* $\mathcal{T} = \mathcal{T}^*$ that minimizes $c(\mathcal{T})$ over all $\mathcal{T} \gg 0$. Evidently, $\mathcal{T}^*$ exists and is unique.

The reason for considering this relaxation is that the minimum $\mathbb{B} \equiv c(\mathcal{T}^*)$, which is clearly a lower bound on the average cost of all integer-ratio policies, is in fact a lower bound on the average cost of all policies! Moreover, although the relaxation $\mathcal{T}^*$ need not be an integer-ratio policy, there is a nearby integer-ratio policy that always has 94% effectiveness.

Sign-Preserving Integer-Ratio Policies

In seeking such a nearby integer-ratio policy $\mathcal{T}$, it will prove useful to restrict attention to those that are "sign preserving". To define this concept, let T_n^* be the order interval for retailer n and T^* be that at the warehouse when $\mathcal{T}^*$ is used. Partition the retailers into the three sets $G \equiv \{n : T_n^* > T^*\}$, $E \equiv \{n : T_n^* = T^*\}$, and $L \equiv \{n : T_n^* < T^*\}$ corresponding respectively to the retailers whose optimal relaxed order intervals are greater than, equal to, or less than that at the warehouse. Call the vector $\mathcal{T}$ *sign preserving* if $T = T^*$ and if $T_n - T^*$ preserves the sign of $T_n^* - T^*$ for each retailer n. Thus, $\mathcal{T}$ is sign preserving if and only if $T = T^*$, $T_n \ge T^*$ for $n \in G$, $T_n = T^*$ for $n \in E$, and $T_n \le T^*$ for $n \in L$.

Effectiveness of a Sign-Preserving Integer-Ratio Policy

The next step is to show that the average cost of a sign-preserving integer-ratio policy is a sum of average costs of the single-facility lot-sizing type, and that the lower bound $\mathbb{B}$ is the sum of the minima of these functions. This fact facilitates estimation of the effectiveness of a sign-preserving integer-ratio policy by comparison of each single-facility average cost with its minimum.

To establish these facts, observe from (4) that, for each sign-preserving integer-ratio policy $\mathcal{T}$, it is possible to rewrite $c(\mathcal{T})$ as a sum

$$(5) \qquad c(\mathcal{T}) = (\frac{K}{T} + HT) + \sum_{n \in E^c} (\frac{K_n}{T_n} + H_n T_n)$$

of average-cost functions of the single-facility lot-sizing type where $K \equiv K_0 + \sum_{n \in E} K_n$, $H \equiv \sum_{n \in E} h_n' + \sum_{n \in L} h^n$, $H_n \equiv h_n'$ for $n \in G$ and $H_n \equiv h_n$ for $n \in L$. It is useful to think of K and H respectively as the aggregate setup cost and average holding cost per unit time associated with the warehouse and those retailers whose order intervals coincide with that at the warehouse. Also, H_n is the average holding cost per unit time associated with retailer $n \in E^c$.

Since $\mathcal{T}^*$ minimizes c on the positive orthant and (5) holds for T, T_n close enough to T^*, T_n^* for $n \in E^c$, $\mathcal{T}^*$ is a local minimum of the right-hand side of (5). Hence since the right-hand side of (5) is convex, it follows that T^* and $T_n^* > 0$, $n \in E^c$, respectively minimize the terms in parentheses in (5) on the positive half line. Thus

$$(6) \qquad \mathbb{B} = c(\mathcal{T}^*) = M + \sum_{n \in E^c} M_n$$

where $M \equiv \frac{K}{T^*} + HT^*$ and $M_n \equiv \frac{K_n}{T_n^*} + H_n T_n^*$ for $n \in E^c$.

In order to find a lower bound on the effectiveness of a sign-preserving integer-ratio policy $\mathcal{T}$, it is convenient to write (5) in an alternate form. To that end, let $e(\alpha) \equiv \dfrac{2}{\alpha+\alpha^{-1}}$ for $\alpha > 0$. Then since $\dfrac{K}{T^*} = HT^* = \dfrac{M}{2}$ and $\dfrac{K_n}{T_n^*} = H_n T_n^* = \dfrac{M_n}{2}$ for retailers $n \in E^c$, it follows from (5) that

$$(7) \qquad c(\mathcal{T}) = M + \sum_{n\in E^c} \frac{M_n}{e(q_n)}$$

where $q_n \equiv \dfrac{T_n}{T_n^*}$ since $\dfrac{M_n}{e(q_n)} = \dfrac{M_n}{2}(\dfrac{T_n^*}{T_n} + \dfrac{T_n}{T_n^*}) = \dfrac{K_n}{T_n} + H_n T_n$. Observe that if $\mathcal{T} = \mathcal{T}^*$, then (7) reduces to (6) because $e(1) = 1$. Also note that if $\mathcal{T}$ is a sign-preserving integer-ratio policy, then $e(q_n)$ is the effectiveness of $\mathcal{T}$ at retailer $n \in E^c$. Thus the effectiveness of $\mathcal{T}$ at facility n depends only on the quotient q_n of the order intervals that $\mathcal{T}$ and $\mathcal{T}^*$ use there, and is otherwise independent of the cost and sales data. For this reason call the q_n *effectiveness quotients*.

It follows from (7) and the fact $\mathbb{B}$ is a lower bound on the minimum average cost that the effectiveness of a sign-preserving integer-ratio policy $\mathcal{T}$ is at least

$$(8) \qquad \frac{c(\mathcal{T}^*)}{c(\mathcal{T})} = \frac{1}{\dfrac{M}{\mathbb{B}} + \displaystyle\sum_{n\in E^c} \frac{M_n}{\mathbb{B}} \frac{1}{e(q_n)}} \ge \min_{n\in E^c} e(q_n).$$

Thus a sufficient condition for $\mathcal{T}$ to have effectiveness at least ϵ, say, is that $\mathcal{T}$ have at least that effectiveness at each retailer in E^c.

Lower-Bound Theorem

It is now possible to begin the proof of the Lower-Bound Theorem, i.e., to show that the minimum relaxed average cost $\mathbb{B}$ is a lower bound for the average cost of all feasible policies. The proof exploits the idea that it is possible to reallocate the holding cost rates among the facilities in such a way that

- the average cost that any policy incurs majorizes the average reallocated cost thereof and
- the sum of the minimum average reallocated costs at each facility, when considered in isolation from the other facilities, is the desired lower bound $\mathbb{B}$.

Define the *reallocated holding cost rate* H_n for facility n as follows. For $n \in E^c$, define H_n as in (5). For $n \in W \equiv E \cup \{0\}$, define H_n so that $\dfrac{K_n}{T^*} = H_n T^*$. Let M_n be the minimum of $\dfrac{K_n}{T_n} + H_n T_n$ over all $T_n > 0$ for each facility n, which is in agreement with the prior definition of M_n for $n \in E^c$.

LEMMA 6. Reallocated Holding Cost Rates. *One has* $H = \sum_{n\in W} H_n$ *and* $\mathbb{B} = \sum_{n\ge 0} M_n$. *Also* $h_n \le H_n \le h'_n$ *for each retailer* n. *Finally,* $H_0 = \sum_{n\ge 1} H^n$ *where* H^n *is defined so* $H_n + H^n \equiv h'_n$ *for each retailer* n.

Proof. Since $\frac{K}{T^*} = HT^*$ and $\frac{K_n}{T^*} = H_nT^*$ for $n \in W$, $\frac{K}{T^*} = T^*(\sum_{n\in W} H_n)$. Hence $H = \sum_{n\in W} H_n$. Now recall that $M = \frac{K}{T^*} + HT^*$ and $M_n = \frac{K_n}{T^*} + H_nT^*$ for $n \in W$, so $M = \sum_{n\in W} M_n$. This fact and (6) establish $\mathbb{B} = \sum_{n\geq 0} M_n$. For the second assertion of the Lemma, observe that $H_n = h_n'$ for $n \in G$ and $H_n = h_n$ for $n \in L$. Thus suppose $n \in E$. Since $T^* > 0$ minimizes $c_n(T^*, \cdot)$, it follows from (4) and $H_n = \frac{K_n}{T^{*2}}$ that

$$-H_n + h_n = D_2^- c_n(T^*, T^*) \leq 0 \leq D_2^+ c_n(T^*, T^*) = -H_n + h_n',$$

and so $h_n \leq H_n \leq h_n'$ as claimed. The final assertion of the Lemma follows from the fact that $H_0 = H - \sum_{n\in E} H_n = \sum_{n\in E} (h_n' - H_n) + \sum_{n\in L} h^n = \sum_{n\geq 1} H^n$. ∎

It is now possible to prove the Lower-Bound Theorem.

THEOREM 7. Lower-Bound. *The minimum relaxed average cost $\mathbb{B}$ is a lower bound on the average cost of all feasible policies for every finite horizon.*

Proof. Consider an arbitrary policy over the infinite horizon. Let $C(t')$ be the average cost that this policy incurs over the interval $[0, t')$. It suffices to show that $C(t') \geq \mathbb{B}$ for all $t' > 0$.

Let J_n be the number of orders facility n places in $[0, t')$, I_n^t be the inventory at retailer n at time t, $S_n^t \geq I_n^t$ be the system inventory of product n at time t, and $I_0^t \equiv \sum_{n\geq 1} \frac{H^n}{H_0} S_n^t$ be the average value over all products of the system inventory at time t. Observe that I_n^t is right-continuous in t, has jumps (upward) at the times at which facility $n \geq 0$ orders, and decreases linearly in t with slope -2 otherwise.

By Lemma 6, it is possible to obtain a lower bound on the total holding cost incurred in $[0, t')$ as follows:

$$\sum_{n\geq 1} \int_0^{t'} (h_n I_n^t + h^n S_n^t)\, dt \geq \sum_{n\geq 1} \int_0^{t'} (H_n I_n^t + H^n S_n^t)\, dt = \sum_{n\geq 0} \int_0^{t'} H_n I_n^t\, dt.$$

Now the n^{th} term in the sum on the right-hand side of the above equality is the total holding cost incurred in $[0, t')$ in a single-item economic-lot-size problem in which there are J_n orders in $[0, t')$, the demand rate per unit time is two, and the unit holding cost per unit time is H_n. The sequel shows that the minimum-cost policy for this problem among those with J_n orders in the interval $[0, t')$ entails ordering every $T_n \equiv \frac{t'}{J_n}$ units of time with the resulting total holding cost $t'H_nT_n$. Thus

$$C(t') \geq \sum_{n\geq 0} [K_nJ_n + \int_0^{t'} H_n I_n^t\, dt]\frac{1}{t'} \geq \sum_{n\geq 0} [\frac{K_n}{T_n} + H_nT_n] \geq \mathbb{B}.$$

It remains to show that ordering at equally-spaced points in time is optimal for the single-item problem in which the number of orders in $[0, t')$ is J_n, the setup cost is K_n, the unit storage cost rate per unit time is H_n and the demand rate per unit time is two. To that end, let t_i be the time between the i^{th} and $(i+1)^{th}$ orders, $i = 1, \ldots, J_n-1$. From the results of §6.7, there is no loss in generality in assuming that orders occur only when stock runs out. Now the total setup costs during $[0, t')$ is the constant $K_n J_n$ and the total storage cost during that interval is $H_n f(t)$ where $f(t) \equiv \sum_{i=1}^{J_n} t_i^2$ and $t = (t_i)$. The latter cost attains its minimum at the value of $t \gg 0$ that minimizes $f(t)$ subject to $\sum_{i=1}^{J_n} t_i = t'$. Since $f(t)$ is 1-quadratic, it follows from the Invariance Theorem that the optimal choice of t is $t_1 = t_2 = \cdots = t_{J_n}$ as claimed. ∎

94%-Effective Integer-Ratio Lot-Sizing

The next step is to show how to find an integer-ratio policy whose effectiveness is at least $e(\sqrt{2}) \approx .94$. To that end, let $r_n^* \equiv \frac{T_n^*}{T^*}$ be the *optimal relaxed order-interval ratio* at retailer n. Also let $r_n \equiv \frac{T_n}{T^*}$ be the *order-interval ratio* at retailer n for some integer-ratio policy $\mathcal{T}$ for which $T = T^*$. Note that T^* and the r_n $(\in R)$, $1 \le n \le N$, uniquely determine an integer-ratio policy $\mathcal{T}$. Also, $\mathcal{T}$ is sign preserving if and only if $T = T^*$, $r_n \ge 1$ whenever $r_n^* \ge 1$, and $r_n \le 1$ whenever $r_n^* \le 1$ for each retailer n.

THEOREM 8. 94% Effectiveness. *There is an integer-ratio policy with effectiveness at least* $e(\sqrt{2}) = \frac{1}{3}\sqrt{8} > .94$.

Proof. Construct the desired integer-ratio policy $\mathcal{T}$ as follows. Set $T = T^*$. For each $n \in E$, put $T_n = T^*$, so $r_n = 1 \in R$. And for each $n \in E^c$, choose $r \in R$ so that $r_n^* \in (r_-, r]$, put

$$r_n = \begin{cases} r_-, \text{ if } r_n^* \le \sqrt{rr_-} \\ r \;\;, \text{ if } r_n^* > \sqrt{rr_-}, \end{cases}$$

and choose T_n so that $\frac{T_n}{T^*} = r_n$ $(\in R)$. Then $\mathcal{T}$ is an integer-ratio policy. Also $\mathcal{T}$ is sign preserving because $r_n^* \le 1$ implies that $r_n \le r \le 1$, $r_n^* > 1$ implies that $r_n \ge r_- \ge 1$, and $r_n^* = 1$ implies that $r_n = r = 1$.

Thus, by the Lower-Bound Theorem, it follows from (8) that the effectiveness of $\mathcal{T}$ is at least $e(\sqrt{2})$ if $e(q_n) \ge e(\sqrt{2})$ for each $n \in E^c$. The last inequality will hold if and only if $q_n \in [\frac{1}{\sqrt{2}}, \sqrt{2}]$ because $e(\alpha)$ is strictly quasiconcave, achieves its maximum at $\alpha = 1$, and satisfies $e(\alpha) = e(\frac{1}{\alpha})$ for all $\alpha > 0$. Since $q_n = \frac{r_n}{r_n^*}$, it suffices to show that $\frac{r_n}{r_n^*} \in [\frac{1}{\sqrt{2}}, \sqrt{2}]$. To that end, observe that if $r_n^* \le \sqrt{rr_-}$, then

$$1 \ge \frac{r_n}{r_n^*} \ge \frac{r_-}{\sqrt{rr_-}} = \sqrt{\frac{r_-}{r}} \ge \frac{1}{\sqrt{2}},$$

while if $r_n^* > \sqrt{rr_-}$, then

$$1 \le \frac{r_n}{r_n^*} \le \frac{r}{\sqrt{rr_-}} = \sqrt{\frac{r}{r_-}} \le \sqrt{2}. \;∎$$

Remark 1. Geometric Mean. Observe that $\sqrt{rr_-}$ is the geometric mean of r_- and r. Thus the Theorem calls for rounding the optimal relaxed order-interval ratio $r_n^* \in (r_-, r]$ at retailer n down to r_- or up to r respectively according as r_n^* is less than or greater than the geometric mean of r_- and r. Figure 16 illustrates the result.

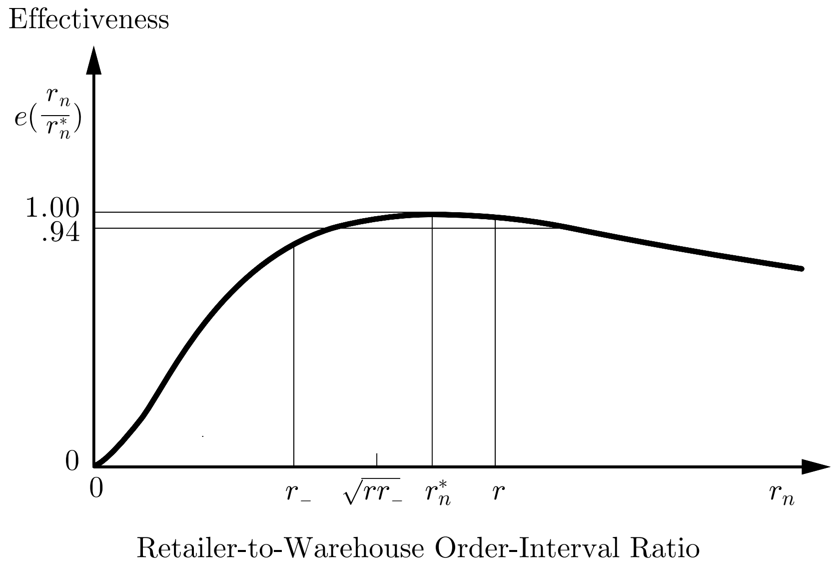


Figure 16. Rounding a Retailer's Optimal Relaxed Order-Interval Ratio

Remark 2. Flatness of Effectiveness Function. The above result depends on the fact that the effectiveness function $e(\alpha)$ is rather flat near its maximum $e(1) = 1$. As one indication of this fact, observe that in order to reduce the effectiveness by 6% from its maximum value, it is necessary either to increase α by over 40% (to $\sqrt{2}$) or decrease α by nearly 30% (to $\frac{1}{\sqrt{2}}$)—a wide range indeed!

Remark 3. Ex Post Estimate of Effectiveness. Observe that once an integer-ratio policy is at hand, a higher lower bound on its effectiveness is available from the equality in (8).

Remark 4. Integer Powers of Two. It is notable that the Theorem is valid when R is the set of integer powers of two. In that event it suffices to restrict the order interval at each retailer to integer-power-of-two multiples of the optimal relaxed order interval at the warehouse. Since then $\frac{r}{r_-} = 2$ is constant for every $r \in R$, it follows that the percentage by which one rounds a retailer's optimal relaxed order interval is independent of its magnitude. In particular, one rounds large optimal relaxed order intervals by large amounts.

Remark 5. Optimal Rounding. It is also possible to show that the rounding procedure in the proof of Theorem 8 produces a policy with maximum effectiveness among all integer-ratio policies with $T = T^*$.

Minimizing the Relaxed Average Cost

The next step is to develop an algorithm to find $\mathcal{T}^*$. To that end, first minimize $c_n(T, T_n)$ over all $T_n > 0$ for fixed $T > 0$. It is easy to verify that

$$(9) \qquad b_n(T) \equiv \inf_{T_n > 0} c_n(T, T_n) = \begin{cases} 2\sqrt{K_n h_n'} & , \; T < \tau_n' \\ \dfrac{K_n}{T} + h_n' T & , \; \tau_n' \le T \le \tau_n \\ 2\sqrt{K_n h_n} + h^n T, & \tau_n < T \end{cases}$$

where $\tau_n' \equiv \sqrt{\dfrac{K_n}{h_n'}}$ and $\tau_n \equiv \sqrt{\dfrac{K_n}{h_n}}$. The value of T_n that minimizes $c_n(T, T_n)$ is τ_n' if $T < \tau_n'$, T if $\tau_n' \le T \le \tau_n$, and τ_n if $\tau_n < T$. Note that τ_n' (resp., τ_n) is the order interval at retailer n given by the Harris square-root formula for the single-facility problem with setup cost K_n, demand rate two, and holding cost rate h_n' (resp., h_n). Also note that $b_n(T)$ is convex and continuously differentiable.

Let $B(T) \equiv \dfrac{K_0}{T} + \sum_{n \ge 1} b_n(T)$. Then the order interval T^* at the warehouse that the optimal relaxed-order-interval policy $\mathcal{T}^*$ uses evidently minimizes B on the positive half-line, so $\mathbb{B} = B(T^*)$. Since B is strictly convex and continuously differentiable, and $B(T) \to \infty$ as $T, T^{-1} \to \infty$, T^* is the unique positive solution to $B'(T^*) = 0$.

The main work in finding the 94%-effective integer-ratio lot-sizing rule is in minimizing the relaxed average cost $B(T)$. It is possible to do this efficiently as follows. By (5), $B(T)$ has the form $\dfrac{K(T)}{T} + M(T) + H(T)T$ where $K(T)$, $M(T)$ and $H(T)$ are piecewise-constant functions of T. The functions $K(T)$ and $H(T)$ are similar to K and H in (5) where $E = E(T)$ (resp., $L = L(T)$, $G = G(T)$) is instead $\{n : \tau_n' \le T \le \tau_n\}$ (resp., $\{n : \tau_n < T\}$, $\{n : T < \tau_n'\}$). The values of $K(T)$, $M(T)$ and $H(T)$ change only when T crosses a τ_n' or a τ_n. If T moves from right to left across τ_n (resp., τ_n'), this has the effect of shifting retailer n from L to E (resp., from E to G). These $2N$ "breakpoints" give rise to $2N + 1$ "pieces" inside of which $K(T)$, $H(T)$ and $M(T)$ are constant. Since $B(T)$ is strictly convex and continuously differentiable, and since $B(T) \to \infty$ as $T, T^{-1} \to \infty$, B attains its minimum at the unique positive number T^* satisfying $B'(T^*) = 0$. Therefore $T = T^*$ if and only if $T = \sqrt{K(T)/H(T)}$. The minimum-relaxed-average-cost algorithm begins with the right-most piece, the one in which T is larger than the largest breakpoint. It moves left from piece to piece until it finds the one containing T^*. Figure 17 illustrates the procedure. For convenience denote $K(T)$ by K and $H(T)$ by H.

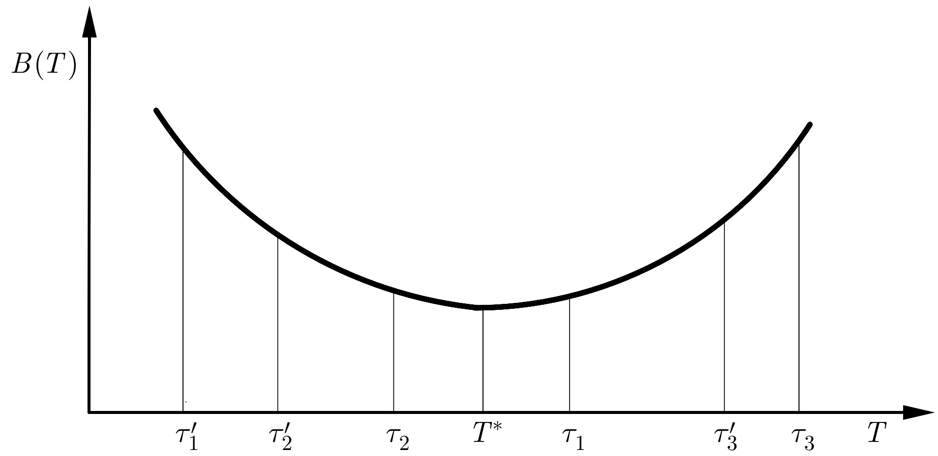


Figure 17. Breakpoints and Pieces of $B(T)$

Minimum-Relaxed-Average-Cost Algorithm

Step 1. Calculate and Sort the Breakpoints.

Calculate the breakpoints $\tau_n' = \sqrt{K_n/h_n'}$ and $\tau_n = \sqrt{K_n/h_n}$, and sort them to form an increasing sequence of $2N$ numbers. Label each breakpoint with the value of n and with an indicator showing whether it is the left breakpoint τ_n' or the right breakpoint τ_n.

Step 2. Initialize E, G, L, K, and H.

Set $E = G = \emptyset$, $L = \{1, \ldots, N\}$, $K = K_0$, and $H = \sum_{n\ge 1} h^n$.

Step 3. Cross the Largest Uncrossed Breakpoint.

Let τ be the largest previously uncrossed breakpoint. If $\tau^2 \ge K/H$ and $\tau = \tau_n$ is a right breakpoint, cross τ and update E, L, K and H by $E \leftarrow E \cup \{n\}$, $L \leftarrow L \setminus \{n\}$, $K \leftarrow K + K_n$ and $H \leftarrow H + h_n$. Then go to Step 3. If $\tau^2 > K/H$ and $\tau = \tau_n'$ is a left breakpoint, cross τ and update E, G, K and H by $E \leftarrow E \setminus \{n\}$, $G \leftarrow G \cup \{n\}$, $H \leftarrow H - h_n'$ and $K \leftarrow K - K_n$. Then go to Step 3. Otherwise T^* is in the current piece. Go to Step 4.

Step 4. Calculate T^* and $\mathbb{B}$.

Set $T^* = \sqrt{K/H}$, $T_n^* = \tau_n'$ for $n \in G$, $T_n^* = T^*$ for $n \in E$, $T_n^* = \tau_n$ for $n \in L$, and $\mathbb{B} = c(\mathcal{T}^*)$.

It remains to show that Step 4 occurs before crossing the last breakpoint. Otherwise, $H = 0$ and $E \cup L = \emptyset$. If τ is the only uncrossed breakpoint, it is a left breakpoint τ_n'. Then in Step 3, $K = K_0 + K_n$ and $H = h_n'$, so $\dfrac{K}{H} > \tau^2$. Therefore Step 4 occurs and the algorithm terminates.

Running Time of the Minimum-Relaxed-Average-Cost Algorithm

Sorting in Step 1 requires at most $2N \log N$ comparisons. The number of operations to execute all other phases of the algorithm is linear in N. Also, the algorithm requires at most N square roots if Steps 1 and 3 use $\tau_n'^2$ and τ_n^2 instead of τ_n' and τ_n. Hence the entire algorithm runs in $O(N \log N)$ time.

7

Stochastic Order, Subadditivity Preservation, Total Positivity and Supply Chains

1 INTRODUCTION

Distribution Depending on a Parameter

It is often the case that the demand D for a single product in a period is a random variable with known distribution $F(\,\cdot\,|\,t)$ depending on a real parameter t. For example, t might be

- the level of advertising expenditures,
- an index of business conditions,
- a (decreasing) function of price,
- the size of the total potential market for the product, or
- a sufficient statistic for the posterior demand distribution where the demand distribution depends on an unknown parameter having known prior distribution.

Monotonicity of Optimal Starting Stock in Parameter

Suppose the inventory manager must choose the *starting stock* y of the product before observing the demand D. The resulting cost is $g(y, D)$. The conditional expected cost $G(y, t)$ given the starting stock y and demand parameter t is

$$G(y,t) = \int g(y,u)\,dF(u \mid t). \tag{1}$$

In that event it is natural to choose the starting stock y to minimize $G(y,t)$. This raises questions like the following. Is it optimal to increase the starting stock y when there is an increase in advertising, the index of business conditions, the potential market size, or the sufficient statistic, or when the price drops? Each of these questions involves determining whether the optimal y is increasing in t.

The Increasing-Optimal-Selections Theorem provides a simple answer to these questions. It is that (apart from compactness hypotheses) if G is subadditive, then one optimal y is increasing in t.

Conditions Assuring that G is Subadditive

This raises the following question. What conditions on g and F assure that G is subadditive? We shall explore this problem in the special case where g is subadditive. (This is so if, for example, $g(y,D) = \hat{g}(y-D)$ with $\hat{g}(\,\cdot\,)$ being convex. In that event $\hat{g}(z)$ is the end-of-period cost of storing or backlogging $|z|$ units according as $z \ge 0$ or $z \le 0$). Now suppress y and $\epsilon > 0$, and put $h(u) \equiv -g(y+\epsilon,u) + g(y,u)$ and $H(t) \equiv -G(y+\epsilon,t) + G(y,t)$. Then from (1)

$$H(t) = \int h(u)\,dF(u \mid t). \tag{2}$$

Evidently, subadditivity of g is equivalent to h being increasing for each $\epsilon > 0$ and y. A corresponding relationship exists between G and H. Thus, F has the property that G is subadditive whenever g is subadditive and the integral (1) exists if and only if H is increasing whenever h is increasing and the integral (2) exists. Therefore, it suffices to consider the latter problem in the sequel.

If it is true of F that H is increasing for every increasing function h for which the integral (2) exists, then that must be so when h is the step function

$$h(u) = \begin{cases} 0, & u \le v \\ 1, & u > v. \end{cases}$$

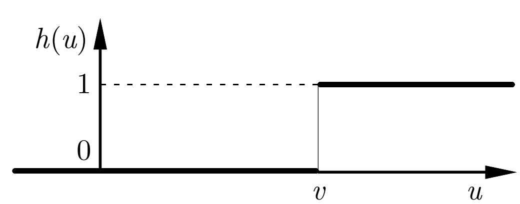


Figure 1. Step Function

This is because h is increasing and bounded for each fixed v, so the integral (2) exists. Thus

$$H(t) = 1 - F(v \mid t) = \Pr(D > v \mid t) \equiv \overline{F}(v \mid t),$$

so $\overline{F}(v \mid t)$ is increasing in t for each v. This says that the probability that D exceeds each fixed number v is increasing in t. It turns out, as we shall see below, that this necessary condition is also sufficient to assure that F carries the class of increasing functions into itself.

2 STOCHASTIC AND POINTWISE ORDER [Le59, p 73], [Ve65c], [St65]

If F is a distribution function, denote by $\overline{F}$ its *complement*, i.e., $\overline{F}(u) \equiv 1 - F(u)$ for all u. If X and Y are random variables with distributions F and G, then X (resp., F) is *stochastically smaller than* Y (resp., G), written $X \subseteq Y$ (resp., $F \subseteq G$), if $\Pr(X > z) \le \Pr(Y > z)$ (resp., $\overline{F}(z) \le \overline{G}(z)$) for all z. In this terminology, the condition of the preceding paragraph is that $F(\,\cdot \mid t)$ is stochastically increasing in t.

Pointwise Order

An important sufficient condition for $X \subseteq Y$ to hold is that $X \le Y$ *pointwise* because then $\Pr(X > z) = \Pr(Y \ge X > z) \le \Pr(Y > z)$.

Example 1. Scalar Multiplication. If $X \ge 0$ and $\alpha \ge 1$, then $X \subseteq \alpha X$ because $X \le \alpha X$ pointwise.

Example 2. Mixtures. If $X \le Y$, then the *mixture* $Z_\alpha \equiv (1-\alpha)X + \alpha Y$ is stochastically increasing in α since $Z_\alpha = X + \alpha(Y - X)$ is increasing in α.

Example 3. Nonnegative Addition. If $Y \ge 0$, then $X \subseteq X + Y$ since $X \le X + Y$ pointwise. Below are three instances of this idea.

- **Normal Distribution.** Observe that $X \subseteq X + \mu$ for each random variable X and nonnegative number μ. This implies that the normal distribution is stochastically increasing in its mean.

- **Poisson Distribution.** Suppose X and Y are Poisson random variables with $\mathrm{E}X = \lambda$ and $\mathrm{E}Y = \mu > \lambda$. Now Y has the same distribution as $X + W$ where $W \ge 0$ is independent of X and Poisson with $\mathrm{E}W = \mu - \lambda$. Since $X \le X + W$, $X \subseteq X + W$, so $X \subseteq Y$, i.e., the Poisson distribution is stochastically increasing in its mean.

- **Binomial Distribution.** Suppose X and Y are binomial random variables with parameters (m, p) and (n, p) with $n > m$. Now Y has the same distribution as $X + W$ where $W \ge 0$ is independent of X and is binomial with parameters $(n-m, p)$. Since $X \le X + W$, $X \subseteq X + W$, so $X \subseteq Y$. Thus if X_n is a binomial random variable with parameters (n, p), then X_n is stochastically increasing in n. Also since $n - X_n$ is binomial with parameters $(n, 1-p)$, $n - X_n$ is stochastically increasing in n as well. Thus the binomial distribution with parameters (n, p) increases stochastically in n, but not as fast as n does.

It is natural to ask whether the binomial random variable X_n is also stochastically increasing in p. In order to investigate this, observe that X_n has the same distribution as $\sum_1^n B_i$ where the B_i are independent *Bernoulli*, i.e., 0-1 valued, random variables with common probability p of assuming the value one. It is immediate that each B_i is stochastically increasing in p. The question arises whether this implies that their sum $\sum_1^n B_i$, and so X_n, is likewise stochastically increasing in p. This is an instance of a more general question. Namely, is every increasing (Borel) function $h(B_1,\ldots,B_n)$ stochastically increasing in p? (The special case above is where $h(B_1,\ldots,B_n)=\sum_1^n B_i$.) The answer is it is. Indeed, Theorem 2 below shows that even more is true, viz., the function $h(Y_1,\ldots,Y_n)$ increases stochastically whenever that is so of each Y_i provided that $Y_1,\ldots,Y_n$ are independent random variables and h is any increasing (Borel) function of them.

Representation of Random Variables with Uniform Random Variables

In order to establish this fact, it is necessary to establish a preliminary result which will shed further light on the notion of stochastic order. To this end, suppose in the sequel that distribution functions are right continuous. If F is a distribution function, let

$$F^{-1}(u) \equiv \min\{v : F(v) \geq u\} \text{ for all } 0<u<1 \tag{3}$$

be the (smallest) $100u^{th}$ *percentile* of F. For example, if $u=.75$, $F^{-1}(.75)$ is the 75^{th} percentile of F. Observe that because $F(-\infty)=0$, $F(\infty)=1$, and F is increasing and right continuous, the minimum in (3) is always attained. It might seem more natural to require that $F(v)=u$ in (3), but this is impossible in general unless F is continuous as Figure 2 below illustrates. It is immediate from the definition of F^{-1} that

$$F(F^{-1}(u)-) \leq u \leq F(F^{-1}(u)) \text{ for all } 0<u<1. \tag{4}$$

Moreover, equality occurs throughout if $F(\,\cdot\,)$ is left-continuous at $F^{-1}(u)$.

The importance of the function F^{-1} for our purposes is that it permits us to reduce questions about distributions of arbitrary random variables to questions about uniform random variables as the following result shows.

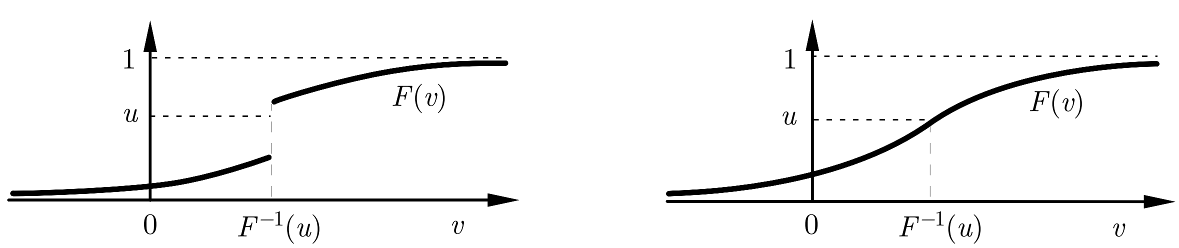


Figure 2. Distribution Function

LEMMA 1. *If U is a uniformly distributed random variable on $(0,1)$ and F is a distribution function, then $F^{-1}(U)$ has the distribution function F.*

Proof. First observe that for $0 < u < 1$,

$$F^{-1}(u) \le v \text{ if and only if } u \le F(v). \tag{5}$$

The "if" part follows from the fact $F^{-1}(u)$ is the least v satisfying $u \le F(v)$. The "only if" part follows by observing that $u > F(v)$ implies that $F^{-1}(u) > v$. Then from (5) and the fact that U is a uniform random variable on $(0,1)$, $\Pr(F^{-1}(U) \le v) = \Pr(U \le F(v)) = F(v)$. ∎

Equivalence of Stochastic and Pointwise Monotonicity Using Common Random Variables

The next result establishes an equivalence between stochastic and pointwise order.

THEOREM 2. Equivalence of Stochastic and Pointwise Monotonicity. *Suppose $X_t = \{X_{it} : i \in I\}$ is an indexed family of independent random variables for each $t \in T \subseteq \Re^k$ with F_{it} denoting the distribution of X_{it} for all i, t. Then the following are equivalent.*

$1°$ *X_{it} is stochastically increasing in t on T for each $i \in I$.*

$2°$ *$F_{it}^{-1}(u)$ is increasing in t on T for all $0 < u < 1$ and $i \in I$.*

$3°$ *There is an indexed family of random variables $X'_t = \{X'_{it} : i \in I\}$ for each $t \in T$ with X'_t and X_t having the same distributions for all $t \in T$ and with X'_t being increasing in t on T.*

$4°$ *$\mathrm{E}h(X_t)$ is increasing in t on T for every increasing real-valued bounded Borel function h (resp., every increasing real-valued Borel function h for which the expectations exist).*

$5°$ *$h(X_t)$ is stochastically increasing in t on T for every increasing real-valued Borel function h.*

Proof. $1° \Rightarrow 2°$. Evidently, the fact that X_{it} is stochastically increasing in t implies that $F_{it}(v)$ is decreasing in t, which in turn implies $F_{it}^{-1}(u)$ is increasing in t for all $0 < u < 1$.

$2° \Rightarrow 3°$. Let $\{U_i : i \in I\}$ be an indexed family of independent uniformly-distributed random variables on $(0,1)$ and $X'_t \equiv \{F_{it}^{-1}(U_i) : i \in I\}$. By Lemma 1, X'_t has the same distribution as X_t. Also, $2°$ implies that X'_t is increasing in t.

$3° \Rightarrow 4°$. Since X'_t has the same distribution as X_t, $h(X'_t)$ has the same distribution as $h(X_t)$. Also since $h(\,\cdot\,)$ and $X'_{(\cdot)}$ are increasing, $h(X'_t)$ is increasing in t on T, so the same is so of $\mathrm{E}h(X_t) = \mathrm{E}h(X'_t)$.

$4° \Rightarrow 5°$. Let $g(z) = 0$ for $z \le v$ and $g(z) = 1$ for $z > v$. Then g is increasing and bounded, so the same is so of $g(h(\,\cdot\,))$. Thus by $4°$, $\Pr(h(X_t) > v) = \mathrm{E}g(h(X_t))$ is increasing in t on T.

$5° \Rightarrow 1°$. On choosing $h(X_t) = X_{it}$, $5°$ implies that X_{it} is stochastically increasing in t on T. ∎

The next result is immediate from the Equivalence-of-Stochastic-and-Pointwise-Monotonicity Theorem on setting $h(x) = x$ in $4°$.

COROLLARY 3. Monotonicity of Expected Values. *If X_t is a random variable that is stochastically increasing in $t \in T \subseteq \Re^k$, then $\mathrm{E}X_t$ is increasing in t on T provided that the expectations exist.*

Remark. [St65], [KKO77] Actually the equivalence of 3°-5° remains valid even where the components of X_t are not independent. However, the proof is much more complex. It is an elegant application of the duality theory for the (continuous) transportation problem with linear side conditions.

As another application of the Equivalence Theorem, let $P = (p_{ij})$ be the (possibly countable) transition matrix of a Markov chain on a state space $I \subseteq \Re^k$. Call P *stochastically increasing* if the i^{th} row p_i of P is stochastically increasing in i, i.e., $p_i h$ is increasing in i for each increasing function $h = (h_j)$ for which $p_i h \equiv \sum_j p_{ij} h_j$ exists.

COROLLARY 4. Closure of Stochastically-Increasing Markov Matrices Under Multiplication. *The class of stochastically-increasing Markov matrices on a countable state space $I \subseteq \Re^k$ is closed under multiplication. In particular, the positive integer powers of a stochastically-increasing Markov matrix on I are stochastically increasing on I.*

Proof. Suppose P and Q are stochastically increasing and h is increasing and bounded on I. Then $g = Qh$ is increasing and bounded on I, whence the same is so of $PQh = Pg$. ∎

Theorem 2 is a powerful tool for assessing the direction of the effect of changes in the distributions of random variables governing the evolution of a stochastic system on measures of the system's performance. One arena of application is in showing the monotonicity of the power of statistical tests in their test parameters. Another is in studying the effect of changes of arrival and service rates on the performance of queueing systems. To illustrate the latter, consider the following example concerning nonstationary $GI/GI/1$ queues.

Example 4. Monotonicity of Nonstationary $GI/GI/1$ Queue Size in Arrival and Service Rates. The question arises whether or not the number N_t (resp., average number $\bar{N}_t$) of customers in a nonstationary $GI/GI/1$ queue—including the customer being served—at time t (resp., during the interval $[0,t]$) increases or decreases stochastically as a function of the interarrival and service-time distributions. To that end, let a_i be the time between the arrival of the $(i-1)^{th}$ and i^{th} customers and s_i be the service time of the i^{th} customer (once service begins), $i = 1, 2, \ldots$. Assume that $a_1, a_2, \ldots$ and $s_1, s_2, \ldots$ are independent positive random variables and that customers are served in order of arrival. Let $A_j \equiv a_1 + \cdots + a_j$ be the arrival time of customer j,

$S_{ij} \equiv s_i + \cdots + s_j$ be the total service times of customers $i, \ldots, j$, and D_j be the departure time of customer j. Then

$$D_j = \max_{1 \le i \le j} (A_i + S_{ij}),\ j = 1, 2, \ldots. \tag{6}$$

In order to justify (6), observe that $D_j \ge A_i + S_{ij}$ because customer j cannot depart until customer i arrives and customers $i, \ldots, j$ are subsequently served. On the other hand, $D_j = A_i + S_{ij}$ for the first customer i that arrives for which $N_t > 0$ for all $A_i \le t < D_j$, i.e., since the customer arrived to start the latest busy period.

Monotonicity of Queue Size in Service Rate. Let $A(t)$ and $D(t)$ be the numbers of customers that respectively arrive and depart in the interval $[0, t]$. Of course, $A(t) = \max\{j : A_j \le t\}$ and $D(t) = \max\{j : D_j \le t\}$. Now since $N_t = A(t) - D(t)$, $A(t)$ is independent of the service times $s = (s_1, s_2, \ldots)$, and $D(t)$ is decreasing in s by (6), it follows that N_t is increasing in s, and so also stochastically increasing therein.

Monotonicity of Queue Size in Arrival Rate. In general, N_t is not stochastically decreasing in the interarrival times as may be seen by considering the case in which $a_1 + s_1 < t < a_1 + s_1 + \epsilon$ and $t < A_i$ for all $1 < i$. Then increasing a_1 to $a_1 + \epsilon$ will increase N_t from zero to one. However, $\bar{N}_t$ is stochastically decreasing in the interarrival times $a = (a_1, a_2, \ldots)$. To see this, observe that customer j spends

$$W_j^t = (D_j - A_j) \wedge (t - A_j)^+,\ j = 1, 2, \ldots, \tag{7}$$

units of time waiting in the queue before time t. Thus $W \equiv \sum_{j=1}^{\infty} W_j^t$ is the total waiting time spent in the queue by all customers up to time t. Hence, the average number of customers in the queue up to time t is $\bar{N}_t = \dfrac{W}{t}$. Now it follows from (6) and (7) that W_j^t is decreasing in a. Thus W and $\bar{N}_t$ are decreasing in a, so $\bar{N}_t$ is stochastically decreasing in a.

3 SUBADDITIVITY-PRESERVING TRANSFORMATIONS AND SUPPLY CHAINS

COROLLARY 5. Subadditivity Preservation by Stochastic Monotonicity. *If $Y \subseteq \Re^n$ is a lattice, $T \subseteq \Re$, D is a random m-vector with range a product $\mathcal{D} \subseteq \Re^m$ of chains and with conditional distribution $F(\cdot \,|\, t)$ for each $t \in T$, D has independent components with each component being stochastically increasing in t on T, g is a real-valued (Borel) function on $Y \times \mathcal{D} \times T$ that is subadditive on $Y \times C \times T$ for each chain C in $\mathcal{D}$, and $G(y, t) \equiv \int g(y, u, t)\, dF(u \,|\, t)$ exists and is finite for each $(y, t) \in Y \times T$, then G is subadditive on $Y \times T$.*

Proof. Suppose $y, y' \in Y$ and $t, t' \in T$. Without loss of generality, assume that $t \le t'$. By Theorem 2, there exist random m-vectors $D \le D'$ having distributions $F(\,\cdot\,|\,t)$ and $F(\,\cdot\,|\,t')$ respectively. Then by the subadditivity of g,

$$G(y,t) + G(y',t') = \mathrm{E}[g(y,D,t) + g(y',D',t')]$$

$$\ge \mathrm{E}[g(y \wedge y', D, t) + g(y \vee y', D', t')] = G(y \wedge y', t) + G(y \vee y', t'),$$

so G is subadditive. ▮

Below are several applications of the above result.

Example 5. Monotonicity of Optimal Starting Stock in Demand Distribution. Consider again the problem discussed at the beginning of this section, viz., that of determining the variation of the optimal starting stock y of a single product as a function of a real parameter t of the distribution of demand D for the product during a single period. To make the discussion concrete, assume that the holding and penalty cost $g(z)$ when z is the net stock on hand at the end of the period is convex with $g(z) \to \infty$ as $|z| \to \infty$. Then $G(y,t) \equiv \mathrm{E}(g(y-D) \mid t)$ is the expected one-period cost, which we take to exist and be finite for all pairs (y,t). Thus $G(y,t) \to \infty$ as $|y| \to \infty$. Now it follows from Corollary 5 and the fact that $g(y-u)$ is subadditive in (y,u) that $G(y,t)$ is subadditive in (y,t) if D is stochastically increasing in t, which we assume in the sequel. Hence, from the Increasing-Optimal-Selections Theorem, the least optimal starting stock $y = y_t$ is increasing in t. And, of course, this is so even if the product can be ordered only in restricted quantities, e.g., in boxes, crates, etc.

One common situation in which D is stochastically increasing in t is where $D = Ct$ with C being a random variable not depending on t, and C and t being nonnegative. For example, this may be so when t is an index of business conditions, or when t is the level of advertising for the product in the period, or when demands in successive periods are positively correlated and t is the demand in the preceding period or the cumulative demand in all earlier periods. In each of these cases, y_t is increasing in t.

As a second example in which D is stochastically increasing in t, suppose that t is the number of independent potential customers for the product, and each potential customer buys the product with probability $p > 0$. Then D is binomially distributed with parameters (t,p), and so increases stochastically with t. Actually, $G(y,t)$ is doubly subadditive in this example because $t - D$ is also binomially distributed with parameters $(t, 1{-}p)$. Thus y_t and $t - y_t$ are both increasing in t.

Example 6. Stochastic Price Elasticity/Inelasticity of Demand Implies Price Elasticity/Inelasticity of Optimal Production. Suppose that the demand D for a product during a period is a random variable whose distribution depends on the price $p \ge 0$ that the firm charges for the product. The firm must decide on the amount $y \ge 0$ of the product to produce in the period before observing the demand for the product. The cost of producing $y \ge 0$ units of the product in the period is $c(y)$. Then the one-period net cost incurred when $y \ge 0$ units of the product are produced during the period is $c(y) - p(y \wedge D)$. The expected value of this function is not generally subadditive or superadditive in (y, p) when, as in usually the case, D is stochastically decreasing in p. Thus, one cannot generally expect the optimal production y_p to be monotone in the price p.

However, it is possible to obtain useful information about the price elasticity of optimal production by making the change of variables $Y \equiv py$ and $R \equiv pD$. Observe that Y (resp., R) is the total revenue if the entire production is sold (resp., entire demand is satisfied). Then the one-period net cost is $g(Y, R, p) \equiv c(\frac{Y}{p}) - (Y \wedge R)$, which is subadditive in (Y, p, R) (resp., superadditive in $(Y, p, -R)$) provided that $c(\frac{Y}{p})$ is subadditive (resp., superadditive) in (Y, p) (see Example 4.3 for a discussion of when this last condition is satisfied). Thus, if R (resp., $-R$) is stochastically increasing in p, then $G(Y, p) \equiv \mathrm{E}g(Y, R, p)$ is subadditive (resp., superadditive) in (Y, p) by Corollary 5. The interpretation of the condition that R (resp., $-R$) is stochastically increasing in p is that a given percent increase in the price p causes the demand D to fall (stochastically) by a smaller (resp., larger) percentage, i.e., demand is *stochastically price inelastic* (resp., *elastic*). Also, $G(Y, p) \to \infty$ as $y \to \infty$. Thus, by the Increasing-Optimal-Selections Theorem, the least optimal $Y = Y_p$ is increasing (resp., decreasing) in p. This means that a given percent increase in the price p of the product entails reducing the optimal production $y_p = \dfrac{Y_p}{p}$ by a smaller (resp., larger) percentage. In short, stochastic price inelasticity (resp., elasticity) of demand implies price inelasticity (resp., elasticity) of optimal production. (Actually, price inelasticity of optimal production really means that the price elasticity of optimal production is at least -1, and so optimal production may rise with price.) By combining the assumptions of both of the above cases, it follows that $c(\frac{Y}{p})$ is additive in (Y, p) and R is stochastically independent of p. Thus $G(Y, p)$ is additive in (Y, p) and so the optimal choice of Y is independent of p. Hence, $y_p = \dfrac{Y}{p}$ and $D = \dfrac{R}{p}$.

Example 7. Monotonicity of Optimal Starting Stocks of Complementary Products in Demand Distribution. A manufacturer of circuit boards must produce in anticipation of uncertain aggregate demand D for boards from n independent customers, with the demand from each customer being nonnegative. Each board consists of N different chips labeled $1, \ldots, N$, with each being of a different type. If the manufacturer makes y_i chips of type i, the number of usable chips of that type is $f_i y_i$ where the yields $0 < f_i < 1$ are independent random variables. The number of boards the manufacturer can assemble from the vector $y = (y_i)$ of chips manufactured is $\wedge_i f_i y_i$ because a board can be made only from usable chips. It costs the manufacturer $c_i > 0$ to make each chip of type i. The revenue from selling a board is $s > \sum_{i=1}^N \frac{c_i}{\mathrm{E} f_i}$. If the manufacturer makes the vector $y = (y_i)$ of chips, has yield vector $f = (f_i)$, and has demand D for boards, the net cost is

$$g(y, D, f) = \sum_{i=1}^N c_i y_i - s[(\bigwedge_{i=1}^N f_i y_i) \wedge D].$$

Observe that $g(\,\cdot\,,\,\cdot\,, f)$ is subadditive for each f, and D is stochastically increasing in n, so by Corollary 5, the conditional expected net cost $G(y, n) = \mathrm{E}(g(y, D, f) \mid n)$ is continuous and subadditive in (y, n), and approaches infinity as $\|y\| \to \infty$. Hence, the least $y \geq 0$ minimizing $G(y, n)$ is increasing in the number of independent customers by the Increasing-Optimal-Selections Theorem.

8

Dynamic Supply Policy with Stochastic Demands

1 INTRODUCTION

The models of supply chains developed so far provide only two of the motives for carrying inventories, viz., a temporal increase in the marginal cost of supplying demand and scale economies in supply. This section and §9 examine a third motive—both separately and in combination with the other two motives—for carrying inventories, viz., uncertainty in demand.

In practice supply-chain managers cannot usually forecast with certainty future demands for the products whose inventories they manage. Nevertheless, it is often reasonable to assume that the demands in each period are random variables whose values are observed at the ends of the periods in which they occur. Initially assume those distributions are known. Moreover, there are typically costs, e.g., of ordering, storage, shortage, etc., in each period. In such circumstances the aim is usually to choose the inventory levels sequentially over time in the light of actual demands observed so as to minimize the total expected costs over a suitable time interval. This section examines problems of this type.

The approach is to formulate a suitable dynamic-programming recursion relating the minimum expected cost in each period to that in the following period. Then use the recursion to determine properties, e.g., convexity, K-convexity, subadditivity, etc., that the minimum expected cost function in each period inherits from the corresponding function in the following period by an appropriate projection theorem. Finally, exploit this information to study the dependence of the form of the optimal policy on the underlying assumptions about the given cost functions in each period.

Although the focus in this section is only on the single-product case, some of the techniques carry over to the multi-product case. These include, most notably, the use of dynamic-programming recursions to characterize and compute optimal policies, the use of lattice programming to characterize their qualitative structure, and the use of myopic policies. In addition the general approach is applicable in many other fields, e.g., optimal control of queues, optimal maintenance and repair policy selection, and portfolio management among others.

Dynamics

To make these ideas concrete, consider an inventory manager who seeks to manage inventories of a single product over n periods so as to minimize the expected n-period costs. At the beginning of each period i, $1 \le i \le n$, the manager observes the *initial stock* x_i of the product, i.e., the stock on hand less backorders before ordering in the period, and then orders a nonnegative amount of stock with immediate delivery. Call the sum y_i of the initial stock and the order delivered the *starting stock* in period i. Of course $y_i \ge x_i$. Also there is a given distribution $\Phi_i(\,\cdot\,|y_i)$ of initial stock x_{i+1} in period $i+1$ given the starting stock y_i in period i and given all other past history. The point is that the distribution of x_{i+1} depends on the past only through y_i. This formulation encompasses uncertain demands, random deterioration of stock in storage, and rules for handling demands in excess of stock available. For example, if D_i is the demand in period i, then $x_{i+1} = y_i - D_i$ if unsatisfied demands are *backordered* while $x_{i+1} = (y_i - D_i)^+$ if excess demands $(y_i - D_i)^-$ are either *lost* or satisfied by *special* means. More generally, if a fraction $1 - F_i$ of the stock on hand at the end of period i deteriorates, then $x_{i+1} = F_i(y_i - D_i)^+$.

Costs

There is a *cost* $c_i(z)$ *of ordering* $z \ge 0$ units in period i and a (*conditional*) *expected storage and shortage cost* $G_i(y_i)$ in period i given the starting stock is y_i in period i and given the past history. For example, if D_i is the demand in period i and $g_i(z)$ is the *storage and shortage cost* in period i where the starting stock less demand in the period is z, then $G_i(y_i) = \mathrm{E}[g_i(y_i - D_i)|y_i]$.

Dynamic-Programming Recursion

To find an optimal policy in this setting, let $C_i(x)$ be the (*conditional*) *minimum expected cost* in periods $i, \dots, n$ given that the initial inventory is x in period i. Then, under suitable regularity conditions,

$$C_i(x) = \min_{y \ge x} \{c_i(y - x) + G_i(y) + \mathrm{E}[C_{i+1}(x_{i+1})|y]\} \tag{1}$$

for $i = 1, \ldots, n$ where $C_{n+1} \equiv 0$.

The manager computes the C_i recursively for each i and finds a $y = y_i(x) \ge x$ that achieves the minimum in the right-hand-side of (1). Then $y_i(x)$ is the *optimal starting stock* in period i. The manager is interested both in computing $y_i(x)$ and studying how it varies with x. The behavior of $y_i(\,\cdot\,)$ varies markedly according as the ordering cost is convex or concave, just as was the case with known demands.

2 CONVEX ORDERING COSTS [BGG55], [AKS58, Chap. 9], [Ve65-69]

Suppose first that the ordering cost c_i is convex. This provides two motives for carrying inventories, viz., uncertainty in demand and a temporal increase in the marginal cost of supplying demand. Let

$$c(y,x) \equiv c_i(y-x) + \delta_+(y-x) + G_i(y) + \mathrm{E}[C_{i+1}(x_{i+1})|y]$$

The first two terms in this sum are convex functions of the difference of y and x, and so are subadditive in (y,x). The last two terms depend only on y and not x, and so are trivially subadditive. Thus $c(y,x)$ is subadditive. Also $C_i(x) = \min_y c(y,x)$, so by the Increasing-Optimal-Selections Theorem (under suitable regularity hypotheses), the least $y = y_i(x)$ achieving the minimum above is increasing in x.

If the c_j and G_j are convex for all $j \ge i$, unsatisfied demands are backordered, i.e., $x_{i+1} = y_i - D_i$, and the demands D_i are independent, then $y_i(x)$ does not increase as fast as x, or more precisely, $x - y_i(x)$ is increasing in x. To see this, show first by induction on i that C_i is convex. This is trivially so for $i = n+1$. Suppose it is so for $i+1$. Then $\mathrm{E}C_{i+1}(y - D_i)$ is convex, whence the same is so of $c(y,x)$. Thus by the Projection Theorem for convex functions, C_i is convex. Now the dual $c^\#$ of c is subadditive, so $x - y_i(x)$ is increasing in x as claimed. Hence, the larger the initial inventory, the larger the optimal starting stock and the smaller the optimal order (c.f., Figure 1). This proves the next result.

THEOREM 1. Convex Ordering Costs. *If the production cost c_i in period i is convex, the least optimal starting stock $y_i(x)$ in period i is increasing in the initial inventory x in that period. If also the production cost c_j and expected storage and shortage cost G_j are convex in each period $j \ge i$, unsatisfied demands are backordered, and demands are independent, then $x - y_i(x)$ is increasing in x and the minimum expected cost C_i in periods $i, \ldots, n$ is convex.*

Linear Ordering Cost and Limited Production Capacity. As an example of this result, suppose that in each period i the cost of production is linear and there is an upper bound u_i on production. Then $c_i(z) = c_i z + \delta_+(u_i - z)$ and the other hypotheses of Theorem 1 hold. Then $y_i(x) - x = (y_i^* - x)^+ \wedge u_i$ where $y = y_i^*$ is the least minimizer of $c_i y + G_i(y) + \mathrm{E}C_{i+1}(y - D_i)$.

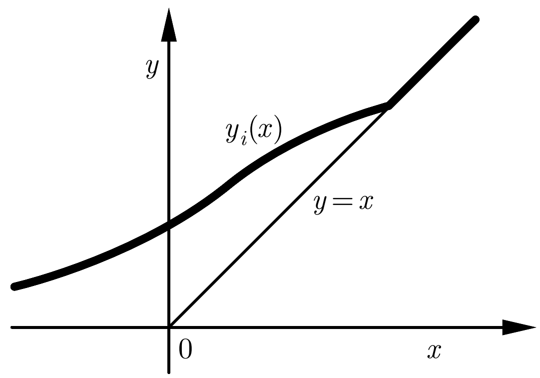


Figure 1. Optimal Supply Policy with Convex Ordering Costs

3 SETUP ORDERING COST AND (s, S) OPTIMAL SUPPLY POLICIES [Sc60], [Ve66a], [Po71], [Sc76]

Now consider instead ordering costs for which $c_i(0) = 0$ and $c_i(z) = K_i \ge 0$ for $z > 0$, i.e., there is a *setup cost* K_i *for ordering* in period i. This is an important instance of the more general case of concave ordering costs. This model combines two motives for carrying inventories, viz., uncertainty in demand and scale economies in supply. In most practical cases, it is valid to assume, as the sequel does, that G_i is quasiconvex and continuous, and that $G_i(y) \to \infty$ as $|y| \to \infty$ for each i.

In order for a policy having a particular form to be optimal in a multi-period sequential decision problem, it invariably must be optimal for the single-period problem. Since the latter problem is generally much easier to analyze, it is useful to begin the study of optimal policies for a multi-period problem with the single-period case.

Single-Period $(\underline{s}, \underline{S})$ Optimal Supply Policy

While discussing the single-period problem, it is convenient to temporarily drop the time-period subscripts. Then the goal is to find a y that minimizes $c(y - x) + G(y)$ subject to $y \ge x$. The above hypotheses on G assure there is an $\underline{S}$ minimizing G and an $\underline{s} \le \underline{S}$ with $G(\underline{s}) = K + G(\underline{S})$. Figure 2 illustrates these definitions.

Now observe that if $x \ge \underline{S}$, it is optimal not to order, i.e., $y = x$, because by so doing one simultaneously minimizes the ordering and the expected storage and shortage costs. If $\underline{s} \le x < \underline{S}$, it is also optimal not to order because the cost is $G(x)$ in that event and, if the contrary event, the cost majorizes $K + G(\underline{S})$ and so $G(x)$. Finally, if $x < \underline{s}$, it is optimal to order to $y = \underline{S}$ because $G(x)$ is the cost if one doesn't order, $K + G(\underline{S})$ is the minimum cost if one does order, and $G(x) \ge K + G(\underline{S})$. To sum up, the optimal one-period ordering policy $y(x)$ has the form

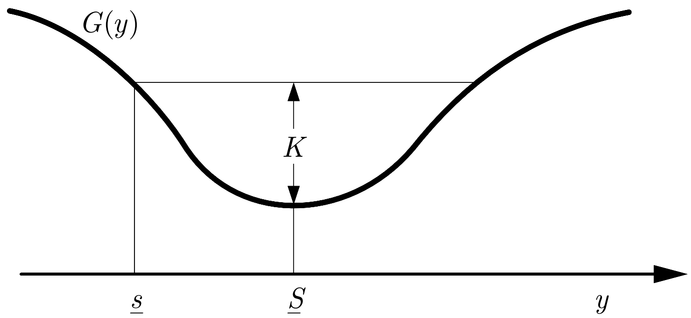


Figure 2. Single-Period $(\underline{s}, \underline{S})$ Optimal Policy

$$y(x) = \begin{cases} \underline{S},\ x < \underline{s} \\ x,\ x \ge \underline{s}. \end{cases}$$

This policy assures that ordering occurs when the initial inventory is x only if the setup cost is less than the reduction $G(x) - G(y(x))$ in expected storage and shortage costs.

(s, S) Policies

This suggests the following definition. An (s,S) *policy*, $s \le S$, is an ordering policy $y(\cdot)$ of the above form where the *reorder point* s and *reorder level* S replace $\underline{s}$ and $\underline{S}$ respectively. In particular the $(\underline{s}, \underline{S})$ policy is optimal for the single-period problem. Figure 3 illustrates an (s,S) policy.

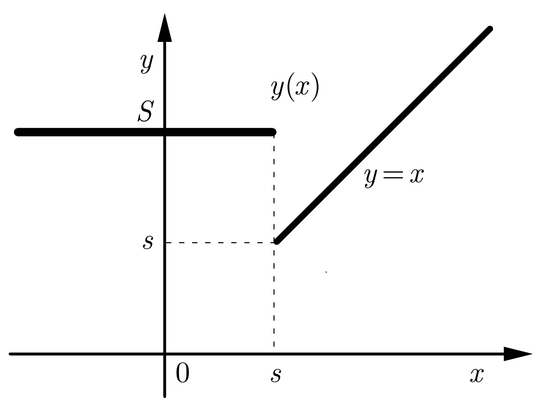


Figure 3. (s, S) Supply Policy

Quantity Discounts, Concave Ordering Cost and Generalized (s, S) Optimal Policies

In practice, there are often quantity discounts in procurement that simple setup costs do not reflect. For example, if the marginal cost of ordering declines with the size of an order, the ordering-cost function is concave. When that is so, (s,S) policies need not be optimal and it is necessary to generalize them to achieve optimality.

Call a policy $y(\cdot)$ *generalized* (s, S) if $s \le S$, $y(x) = x$ for $x \ge s$, and $y(\cdot)$ is decreasing on the interval $(-\infty, s)$ and has infimum S thereon, as Figure 4 illustrates. Observe that the special case in which $y(\cdot)$ is constant on $(-\infty, s)$ is an ordinary (s, S) policy.

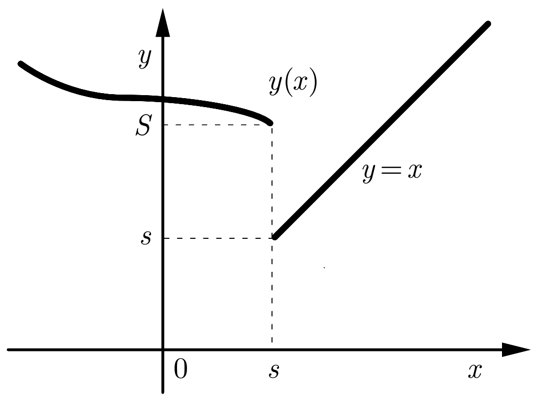


Figure 4. Generalized (s, S) Supply Policy

Why does the presence of general concave ordering costs requires the use of generalized (s, S) policies to achieve optimality? The answer is that if it is optimal to order to a level y when the initial stock is $x < y$, then reducing the initial stock reduces the marginal (not the total) cost of ordering enough to bring the starting stock to any fixed level exceeding the new initial stock. Hence the new optimal starting stock will be at least as large as y, and often strictly larger. By contrast, in the setup-cost case, the above marginal cost remains constant and so the new optimal starting stock remains equal to y.

THEOREM 2. Generalized (s, S) Optimal Single-Period Policies. *If the ordering cost c is nonnegative and concave on $[0, \infty)$ with $c(0) = 0$, the expected storage and shortage cost G is convex, $G(y) \to \infty$ as $|y| \to \infty$, and $G(x) > c(y - x) + G(y)$ for some $x < y$, there is a generalized (s, S) optimal policy $y(\cdot)$ such that $s \le S \le y(x) \le \underline{S}$ for all $x < s$ for the single-period model.*

Proof. Evidently $\hat{c}(y, x) \equiv c(y - x) + G(y)$ is lower semicontinuous and approaches ∞ as $y \to \infty$, so there is a least $y = y(x)$ minimizing $\hat{c}(y, x)$ subject to $y \ge x$ for each x. Since the dual $\hat{c}^{\#}$ of $\hat{c}$ is subadditive, $y(x) - x$ is decreasing in x. Now $y(x) = x$ for $x \ge \underline{S}$ because $\hat{c}(x, x) \le \hat{c}(y, x)$ for $\underline{S} \le x \le y$. Thus since $y(x) > x$ for some x because $\hat{c}(x, x) > \hat{c}(y, x)$ for some $x < y$ by an hypothesis and since $y(\cdot)$ is evidently lower semicontinuous, there is a least s such that $y(s) = s$. Moreover, since $y(x) - x$ is decreasing, $y(x) = x$ for $x \ge s$ and $s \le \underline{S}$. Also, since $\hat{c}(y, x)$ is increasing in $y \ge \underline{S} \vee x$ for all x, it follows that $x < y(x) \le \underline{S}$ for all $x < s$.

Next show that $y(x) \ge s$ for all $x < s$. If not, there is an $x < s$ such that $x < y(x) < s$. Put $v \equiv y(x)$ and $w \equiv y(v)$. By definition of s, $v < w$. Now $\hat{c}(v, x) = c(v - x) + \hat{c}(v, v) > c(v - x)$

$+\,\hat{c}(w,v) \ge \hat{c}(w,x)$ by definition of w, the concavity of c and the fact $c(0)=0$, contradicting the fact that v minimizes $\hat{c}(v,x)$ subject to $v \ge x$. This justifies the claim that $y(x) \ge s$.

It remains to establish that $y(\,\cdot\,)$ is decreasing on $(-\infty, s)$. To this end observe from what was just shown that for $x < s$, the least y minimizing $\hat{c}(y,x)$ subject to $y \ge x$ is the same as the least y minimizing $\hat{c}(y,x)$ subject to $y \ge s$. Thus since c is concave, $\hat{c}(y,x)$ is subadditive in $(y,-x)$ on the sublattice $-x \ge -s$, $y \ge s$, so by the Increasing-Optimal-Selections Theorem, $y(\,\cdot\,)$ is decreasing on $(-\infty, s)$. ∎

Remark. See [Po71] for a multiperiod version of Theorem 2 in which the demands in each period are sums of positive translates of independent exponential random variables.

Quasiconvexity of the Expected Storage and Shortage Cost

If the expected storage and shortage cost function G is bounded below, but is not quasiconvex, then as Figure 5 illustrates, it is possible to choose a small enough setup cost $K > 0$ so that no (s,S) policy is optimal. In that example, it is optimal to order to $\underline{S}$ for initial stocks in the interval (s^2, s^3) and not to order for initial stocks in the interval (s^1, s^2), so the optimal policy cannot be (s,S).

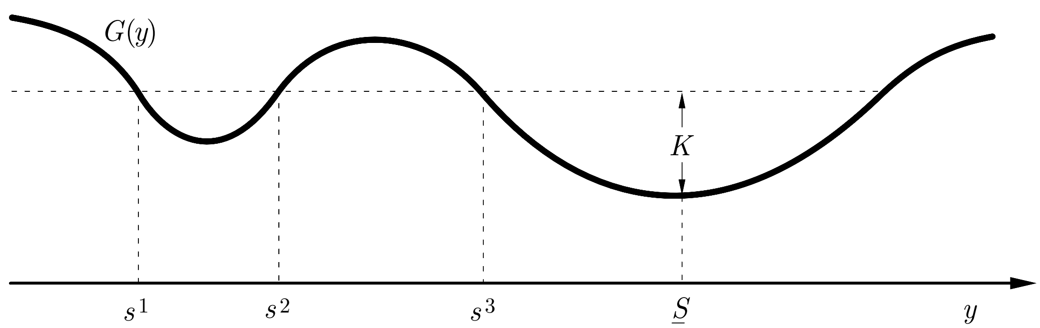


Figure 5. No (s,S) Optimal Policy

The discussion at the beginning of this section shows that (s,S) policies are optimal in the single-period case when the ordering cost is of setup-cost type and the expected storage/shortage cost function is quasiconvex. For this reason, it is reasonable to hope that such policies remain optimal in the multi-period case, though of course they may vary over time. The principal purpose of the remainder of this section is to show that this is indeed the case with a few additional hypotheses concerning the transition function and the way in which the parameters vary over time. These assumptions appear in 2°, 4°, and 5° below together with the assumptions already discussed.

Assumptions Implying the Optimality of Dynamic (s, S) Policies

1° **Setup Ordering Costs.** $c_i(0) = 0$ and $c_i(z) = K_i \geq 0$ for $z > 0$ and $1 \leq i \leq n$.

2° **Declining Setup Costs.** $K_i \geq K_{i+1}$ $(K_{n+1} \equiv 0)$ for $1 \leq i \leq n$.

3° **Quasiconvex Expected Storage and Shortage Costs.** $G_i(y)$ is continuous, quasiconvex and converges to ∞ as $|y| \to \infty$ for $1 \leq i \leq n$.

4° **Transition Functions.** The following hold for each $1 \leq i \leq n$.

(a) **Stochastic Monotonicity.** $\Phi_i(\,\cdot\,|y)$ is stochastically increasing in y.

(b) **Stochastic Boundedness.** For each y there is a b for which $\Phi_i(b|y) = 1$.

(c) **Stochastic Continuity.** At every point (x, y) where $\Phi_i(x|y)$ is continuous in x, it is continuous in y.

Now 3° implies there is an $\underline{S}_i$ minimizing G_i. Call $\underline{S}_i$ the *myopic order level* in period i since it is the optimal starting stock for that period considered by itself provided that the initial stock in that period is below that level and it is optimal to order. The final assumption imposes conditions on the time variation of the costs and transition laws.

5° **Stochastic Accessibility of Myopic Order Levels.** $\Phi_i(\underline{S}_{i+1}|\underline{S}_i) = 1$ for $1 \leq i < n$.

When Assumptions on the Costs and Transition Functions are Satisfied

Before proceeding further, it is useful to discuss briefly when each of the above assumptions holds. The first requires no discussion.

Declining Setup Costs. This hypothesis is satisfied in the stationary case or if the rate of interest exceeds the rate of inflation of the setup costs.

Quasiconvex Expected Storage and Shortage Costs. This assumption is satisfied if $G_i(y) = \mathrm{E}g_i(y - D_i)$ and if either

- g_i is convex, in which case G_i is also convex, or

- g_i is quasiconvex and the density of D_i is "PF_2", i.e., the log of the density is concave, in which case G_i is quasiconvex, but not necessarily convex.

Transition Functions. The stochastic-monotonicity condition normally holds, e.g., that is so if unsatisfied demands are backordered or lost (in the latter event the starting stocks must be nonnegative) and the demands are independent. The remaining assumptions of 4° are merely regularity conditions.

Stochastic Accessibility of Myopic Order Levels. This is so if both

- the initial stock in a period does not exceed the starting stock in the previous period, so $\Phi_i(\underline{S}_i|\underline{S}_i) = 1$, and
- $\underline{S}_i \le \underline{S}_{i+1}$.

Of course the first condition usually holds. For example, that is so if demands are nonnegative and unsatisfied demands are backordered. And the second condition holds in the stationary case or, for example, if demands increase stochastically over time and $G_i(y) = \mathrm{E}g(y, D_i)$ with g being subadditive. For then $G_i(y)$ is subadditive in (i, y) by Corollary 5 of §7 and $\underline{S}_i$ is increasing in i by the Increasing-Optimal-Selections Theorem.

Why the Additional Assumptions 2°, 4° and 5° are Necessary to Assure the Optimality of (s, S) Policies in the Multi-Period Problem

The discussion earlier in this section shows that assumptions 1° and 3°, viz., that the ordering cost is a setup cost and that the expected storage and shortage cost is quasiconvex, are necessary to assure the optimality of an (s, S) policy in the single-period problem. However, these conditions do not assure the optimality of (s, S) policies in the multi-period problem. Indeed, the additional hypotheses 2°, 4° and 5° on the nature of the variation of the costs and transition functions over time are necessary to guarantee the optimality of (s, S) policies in the dynamic problem. Some examples will illustrate why this is so. For this purpose and because it will prove useful in the sequel, put

$$H_i(y) \equiv \mathrm{E}[C_{i+1}(x_{i+1})|y]$$

and

$$J_i(y) \equiv G_i(y) + H_i(y).$$

Need for Declining Setup Costs. If $n = 2$, $0 \le K_1 < K_2$ (so 2° fails to hold), $y_1 = x_2$ (so $\Phi_1(y|y) = 1$ for all y), and G_1 and G_2 are as in Figure 6, then $J_1(y) = G_1(y) + C_2(y)$ and no (s, S) policy is optimal in period one. Indeed, in that period it is optimal to order to $\underline{S}_1$ from x and to $\underline{S}_2$ from x'.

Need for Stochastically-Increasing Transition Functions. If $n = 2$, $K_1 = K_2 = 0$, $G_1(y) = G_2(y) = |y|$, and $x_2(y_1) = y_1$ for $y_1 < 1$ and $x_2(y_1) = 0$ for $y_1 \ge 1$ (so the transition function in period one is not stochastically increasing), then $J_1(y) = G_1(y) + C_2(x_2(y))$ and no (s, S) policy is optimal in period one. Indeed, the optimal policy in period one is to order to 0 from $x < 0$ and to 1 from the $0 < x' < 1$ as Figure 7 illustrates.

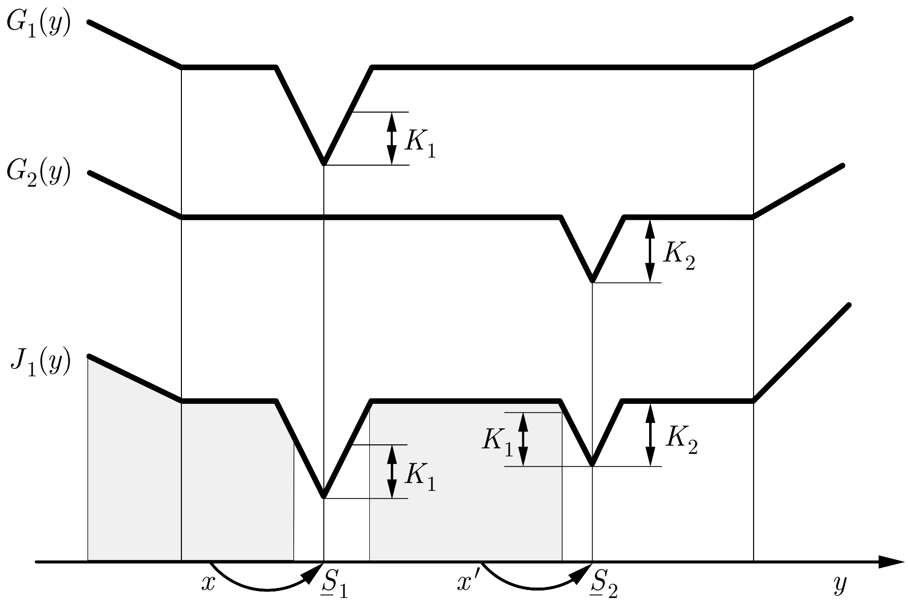


Figure 6. No (s,S) Optimal Policy with Rising Setup Costs

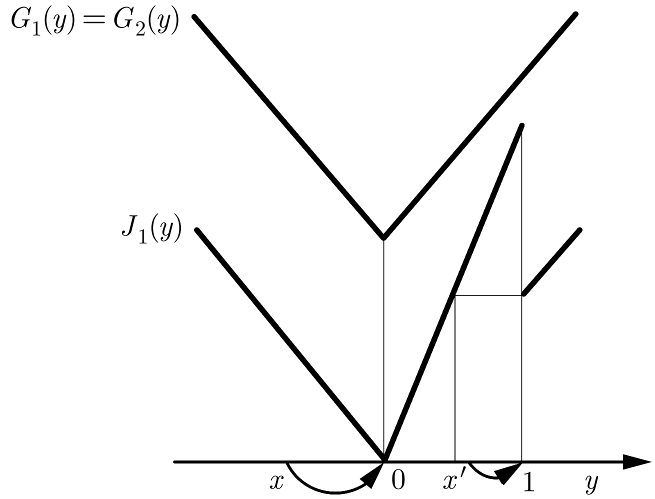


Figure 7. No (s,S) Optimal Policy with Stochastically-Nonmonotone Transition Function

Need for Stochastically-Accessible Myopic Order Levels. If $n = 2$, $K_1 = K_2 = 0$, $y_1 = x_2$, and G_1 and G_2 are as in Figure 8, then $J_1(y) = G_1(y) + C_2(y)$, $\underline{S}_1 > \underline{S}_2$ and $\Phi_1(\underline{S}_2|\underline{S}_1) = 0$, so the myopic order levels are stochastically inaccessible and no (s,S) policy is optimal in period one. Indeed, one optimal policy in that period is to order from x to $\underline{S}_2$ and from x' $(> \underline{S}_2)$ to $\underline{S}_1$.

Relations Between Single- and Multi-Period (s,S) Optimal Policies

Before proving that there is an (s_i,S_i) optimal policy in the i^{th} period of the multi-period problem, it is useful to explore briefly the relation of that policy to the $(\underline{s}_i,\underline{S}_i)$ optimal policy for

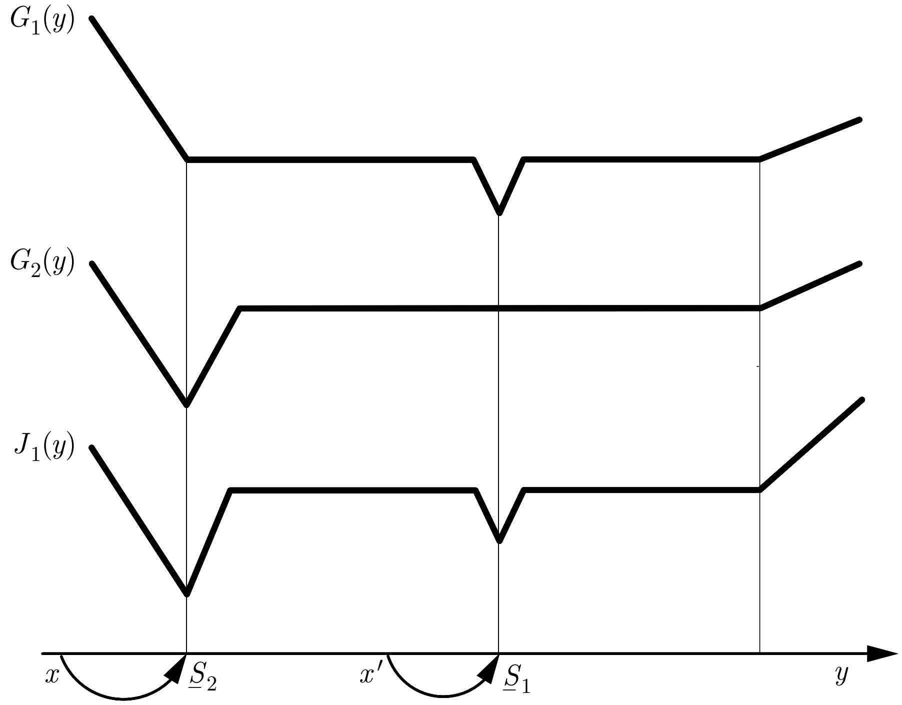


Figure 8. No (s, S) Optimal Policy with Stochastically-Inaccessible Myopic Order Levels

period i considered by itself. It might be hoped that the two policies would be the same. But this is not normally the case. In order to see why, consider the stationary case in which the demands in each period are independent identically-distributed nonnegative random variables, unsatisfied demands are backordered, G is linear between $\underline{s}$ and $\underline{S}$, and G is symmetric about $\underline{S}$. Then using $(\underline{s}, \underline{S})$ in each period assures that the starting stocks in each period would (eventually) all lie between $\underline{s}$ and $\underline{S}$. Moreover, if the demands were small in comparison with $\underline{S} - \underline{s}$, those starting stocks would be roughly uniformly distributed between $\underline{s}$ and $\underline{S}$. Furthermore, the long-run average expected storage/shortage cost per period would be about $G(\underline{S}) + \frac{K}{2}$ and the long-run average expected ordering cost per period would be K divided by the expected number of periods between orders, i.e., the mean number of periods required for the cumulative demand to exceed $\underline{S} - \underline{s}$. Thus if one increases both $\underline{s}$ and $\underline{S}$ by $\frac{\underline{S} - \underline{s}}{2}$, there is no change in the long-run average expected ordering cost per period. However, the long-run average expected storage and shortage cost per period falls to about $G(\underline{S}) + \frac{K}{4}$. This suggests that it is best to choose $s \le S$ so that $\underline{S}$ is about midway between them because then the starting stocks would fluctuate around the global minimum of G.

Indeed, the sequel shows that in general the (s_i, S_i) optimal policy in period i satisfies $\underline{s}_i \le s_i \le \underline{S}_i \le S_i$ where $\underline{s}_i$ is the greatest number not exceeding $\underline{S}_i$ for which $G_i(\underline{s}_i) = K_i + G_i(\underline{S}_i)$. Also $G_i(s_i) \ge G_i(\underline{S}_i)$. Figure 9 depicts these relations.

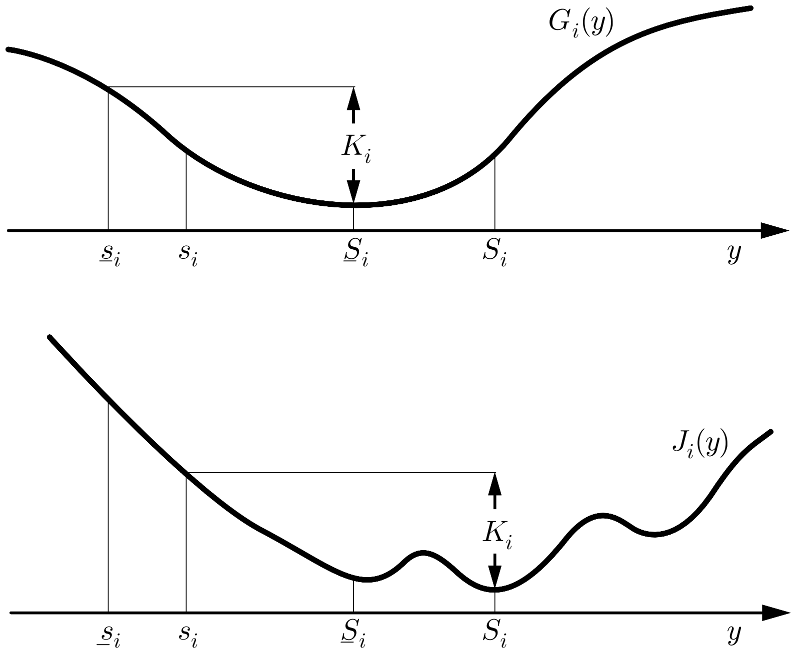


Figure 9. Comparison of Single- and Multi-Period (s, S) Optimal Policies

In order to motivate the proof of optimality of (s, S) policies, consider what property of J_i would assure that such a policy is optimal. For ease of exposition, assume that $J_i(y)$ is continuous and approaches ∞ as $|y| \to \infty$. If J_i were always quasiconvex, then of course the result would follow from what was already established for the single-period case. Unfortunately, however, J_i is not normally quasiconvex. For this reason, it is necessary to discover alternate properties of J_i that both assure an (s_i, S_i) policy is optimal in period i with $s_i \le \underline{S}_i \le S_i$ and that it is possible to inherit.

One sufficient condition is that J_i be, in the terminology introduced below, "$(K_i, \underline{S}_i)$-quasiconvex", i.e., J_i has the following two properties. First, $J_i(x) \le K_i + J_i(y)$ for all $\underline{S}_i \le x \le y$, so it is optimal not to order in period i when the initial inventory is at least $\underline{S}_i$. Second, J_i is decreasing on $(-\infty, \underline{S}_i]$, so J_i attains its global minimum on the real line at an $S_i \ge \underline{S}_i$. These facts imply that there is an $s_i \le \underline{S}_i$ such that $J_i(s_i) = K_i + J_i(S_i)$. Then $J_i(x) \ge K_i + J_i(S_i)$ for $x < s_i$, in which case it is optimal to order to S_i from x; and $J_i(x) \le K_i + J_i(y)$ for $s_i \le x \le y$, in which case it is optimal not to order from x. Thus, the policy (s_i, S_i) is optimal as Figure 9 illustrates.

K-Increasing and (K, S)-Quasiconvex Functions

It is time now to formalize the above ideas and introduce two useful classes of functions of one real variable. (In the sequel the symbols $[-\infty, \infty)$ and $(-\infty, \infty]$ both mean $(-\infty, \infty)$, and $[\infty, \infty) \equiv (-\infty, -\infty] \equiv \emptyset$). Suppose K and S are respectively nonnegative and extended real numbers. Call a real-valued function g on $[S, \infty)$ K-*increasing* if $g(x) \le K + g(y)$ for all $x \le y$ in $[S, \infty)$. The geometric interpretation is that $g(y)$ never falls below $g(x)$ by more than K for $x \le y$. The second concept generalizes the first. Call a real-valued function g on the real line (K,S)-*quasiconvex* if it is decreasing on $(-\infty,S]$ and K-increasing on $[S,\infty)$. Observe that $(K,-\infty)$-quasiconvexity is the same as K-increasing on $\Re$. The following properties of (K,S)-quasiconvex and K-increasing functions are immediate from the definitions.

$1'$ Quasiconvexity (resp., increasing) is equivalent to $(0,S)$-quasiconvexity for some S (resp., 0-increasing).

$2'$ If g is K-increasing and f is increasing on the real line, then the composite function $g(f(\cdot))$ is K-increasing on the real line. If also g is decreasing on $(-\infty,S']$ and $f(S)\le S'$, then $g(f(\cdot))$ is (K,S)-quasiconvex.

$3'$ If $g(y,u)$ is $(K(u),S)$-quasiconvex (resp., $K(u)$-increasing) in y on $\Re$ for each $u\in\Re$ and if Ψ is real-valued and increasing on $\Re$, then $\int g(y,u)d\Psi(u)$ is (K,S)-quasiconvex (resp., K-increasing) in y on $\Re$ where $K\equiv\int K(u)d\Psi(u)$.

$4'$ If g is (K,S)-quasiconvex (resp., K-increasing on $\Re$) and $K\le L$, then g is (L,S)-quasiconvex (resp., L-increasing on $\Re$).

$5'$ If g is (K,S)-quasiconvex and S is real, g is bounded below (by $g(S)-K$).

Optimality of (s,S) Policies: $(K_i,\underline{S}_i)$-Quasiconvexity [Ve66, Po71]

These concepts are useful for establishing the existence of an (s,S) optimal policy for the n-period problem.

THEOREM 3. (s,S) Optimal Policies. *If $1°$-$5°$ hold and $C_{n+1}=0$, then in each period $1\le i\le n$ there is an (s_i,S_i) optimal policy satisfying $\underline{s}_i\le s_i\le \underline{S}_i\le S_i$ and $G_i(s_i)\ge G_i(S_i)$ with equality holding throughout whenever $K_i=0$; moreover, C_i and J_i are $(K_i,\underline{S}_i)$-quasiconvex, and C_i is K_i-increasing on $\Re$.*

Proof. The proof is by induction on i. For notational simplicity in the proof, drop the subscript i on all symbols and replace subscripts $i+1$ thereon by superscript primes, e.g., C_i and C_{i+1} become respectively C and C'.

The first step is to show that that if C' is $(K',\underline{S}')$-quasiconvex and K'-increasing, then H is $(K',\underline{S})$-quasiconvex and K'-increasing. This is clearly so for $i=n$ because $H=C'=0$ and $K'=0$. Thus suppose $i<n$. Observe that since $\Phi(\cdot|y)$ is stochastically increasing in y by $4°(a)$, there is an indexed family of random variables $x'(y)$ such that $x'(y)$ has the distribution function $\Phi(\cdot|y)$ and is increasing in y on $\Re$ by the Equivalence-of-Stochastic-and-Pointwise-Monotonicity Theorem. Also, it follows from $5°$ that $x'(\underline{S})\le\underline{S}'$ almost surely since $\Pr(x'(\underline{S})\le\underline{S}')=\Phi(\underline{S}'|\underline{S})=1$. Thus since C' is $(K',\underline{S}')$-quasiconvex and K'-increasing, $C'(x'(y))$ is almost surely $(K',\underline{S})$-quasiconvex and K'-increasing in y by $2'$. Hence, $H(y)=\mathrm{E}C'(x'(y))$ is $(K',\underline{S})$-quasiconvex and K'-increasing in y on $\Re$ by $3'$.

Now suppose $1 \le i \le n$ and C' is $(K', \underline{S}')$-quasiconvex and K'-increasing. Since $K \ge K'$ by 2°, H is $(K, \underline{S})$-quasiconvex and K-increasing on $\Re$ by what was shown above and 4′. In addition, G is $(0, \underline{S})$-quasiconvex by 3° and 1′, so $J = G + H$ is $(K, \underline{S})$-quasiconvex by 3′. Since H is bounded below by 5′ and $G(y) \to \infty$ as $|y| \to \infty$ by 3°, $J(y) \to \infty$ as $|y| \to \infty$. Moreover, as may be seen by induction using 3°, 4°b and 4°c, J is continuous. Thus since J is decreasing on $(-\infty, \underline{S}]$, J assumes its minimum on $\Re$ at some least $S \ge \underline{S}$. Also, since J is K-increasing on $[\underline{S}, \infty)$, $J(x) \le K + J(y)$ for $\underline{S} \le x \le y$, whence $y(x) = x$ for $\underline{S} \le x$. Moreover, since $J(\underline{S}) \le K + J(S)$, J is decreasing on $(-\infty, \underline{S}]$ and $J(y) \to \infty$ as $y \to \infty$, there is a greatest $s \le \underline{S}$ such that $J(s) = K + J(S)$. Then $J(x) \ge J(s)$ for $x < s$ and $J(x) < J(s)$ for $s < x < \underline{S}$, so $y(x) = S$ for $x < s$ and $y(x) = x$ for $s \le x < \underline{S}$. Hence the (s, S) policy is optimal.

Evidently, $C(x) = J(x \vee s)$. Thus since J is decreasing on $(-\infty, \underline{S}]$, the same is so of C. And for $x \le y$, $C(x) = J(x \vee s) \le K + J(y \vee s) = K + C(y)$. Hence C is $(K, \underline{S})$-quasiconvex and K-increasing on $\Re$.

The next step is to show that $\underline{s} \le s$ and $G(s) \ge G(S)$. Since H is decreasing on $(-\infty, \underline{S}]$, for each x in the interval $[s, \underline{S}]$ it follows that

$$\begin{aligned} 0 &\le K + J(S) - J(x) \le K + J(\underline{S}) - J(x) \\ &= K + G(\underline{S}) - G(x) + H(\underline{S}) - H(x) \le K + G(\underline{S}) - G(x). \end{aligned}$$

Since s is the largest $x \le \underline{S}$ for which the first inequality is an equality, it follows that $\underline{s} \le s$ as claimed. Also since H is K-increasing,

$$0 = K + J(S) - J(s) = K + G(S) - G(s) + H(S) - H(s) \ge G(S) - G(s),$$

so $G(s) \ge G(S)$.

Finally, note that if $K = 0$, $\underline{s} = s = \underline{S} = S$. ∎

Optimality of Myopic Base-Stock Policy

Call an ordering policy in a period *base-stock* if for some base stock S the policy entails ordering $(S - x)^+$ in the period when the initial inventory therein is x. In short, order up to S if $x < S$ and don't order otherwise. If $K_i = 0$, then the $(\underline{S}_i, \underline{S}_i)$ policy in period i is *myopic,* i.e., is optimal for period i alone, and is base-stock with $S = \underline{S}_i$. One specialization of Theorem 3 asserts that if the setup costs are zero in each period, then the myopic base-stock policy is optimal for the multi-period problem.

Optimality of (s, S) Policies with Stochastic Inaccessibility of Myopic Order Levels

Theorem 3 gives the best available results about the optimality of (s, S) policies for the case of stationary cost data and demand distributions, and in other cases as well. A specialization of that result also gives conditions assuring that myopic policies are optimal.

However, Theorem 3 has the important limitation that it is necessary to require stochastic accessibility of the myopic order levels, viz., 5° holds. This condition generally fails during periods in which the demand distributions fall stochastically. The question arises whether (s,S) policies are optimal without any conditions on the variation of the demand distributions. The answer is that they are. The idea is that it is possible to eliminate 5° by strengthening 3° to require convexity of the G_i and 4° to require backordering of unsatisfied demands, both of which are often reasonable assumptions. (In fact, it is possible to weaken the backorders assumptions to allow lost sales and other possibilities). In particular, the new assumptions are:

3^* **Convexity of Expected Storage/Shortage Costs.** $G_i(y)$ is convex and converges to ∞ as $|y| \to \infty$ for $1 \le i \le n$.

4^* **Transition Functions.** The demands $D_1, \ldots, D_n$ in periods $1, \ldots, n$ are independent random variables and unsatisfied demands are backordered so $x_{i+1} = y_i - D_i$ for $i = 1, \ldots, n$.

K-Convex Functions

In order to establish the optimality of (s,S) policies under the above hypotheses, it is necessary to introduce another class of functions. Given a number $K \ge 0$, call a function g on the real line *K-convex* if for each $x < y$ and $b > 0$,

$$g(y) \ge g(x) + (y-x)\frac{[g(x) - g(x-b)]}{b} - K.$$

The geometric interpretation is that the straight line passing through the two points $(x-b, g(x-b))$ and $(x, g(x))$ on the graph of g never lies above the graph of g to the right of x by more than K. Figure 10 illustrates the definition. Note that 0-convexity is equivalent to ordinary convexity.

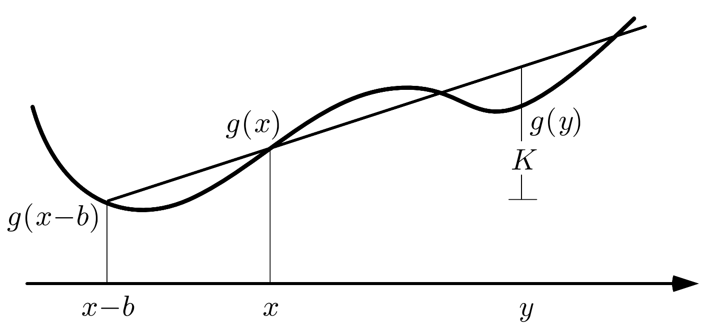


Figure 10. K-Convex Function g

Optimality of (s,S) Policies: K-Convexity [Sc60]

The next result gives conditions assuring that (s,S) policies are optimal for the n-period problem and that the functions C_i and J_i are K_i-convex for each $i = 1, \ldots, n$. A homework problem develops the inductive proof of this result.

THEOREM 4. (s, S) Optimal Policies. *If 1°, 2°, 3^*, 4^* hold and $C_{n+1} = 0$, then in each period $1 \le i \le n$ there is an (s_i, S_i) optimal policy. Moreover, C_i and J_i are K_i-convex on $\Re$ for $i = 1, \ldots, n$.*

Optimality of Base-Stock Policies [BGG55, AKS58]

When $K_i = 0$ for $i = 1, \ldots, n$, then the C_i and J_i are convex for all i in Theorem 4 and $s_i = S_i$ so the optimal policy in each period i is *base-stock* with base stock S_i. Note that by contrast with the application of Theorem 3 to the case of zero setup costs, the base-stock policy resulting from specializing Theorem 4 need not be myopic. Incidentally, a simpler way to establish the optimality of base-stock policies is to specialize Theorem 1 to the case where the ordering cost vanishes.

Computation of (s, S) Optimal Policies

Under the hypotheses of either of the two (s, S)-Optimal-Policies Theorems, the computation of an optimal policy is far simpler than in general dynamic programs. This is because it is possible to exploit the (s_i, S_i) form of an optimal policy in carrying out the computations in the following steps for each i.

- Tabulate J_i.
- Find the global minimum S_i of J_i.
- Find an $s_i \le S_i$ satisfying $K_i + J_i(S_i) = J_i(s_i)$.

Thus only one global minimum and one root must be found. By contrast, in a general inventory problem it is necessary to find an optimal y for each x.

4 POSITIVE LEAD TIMES

Reducing Positive to Zero Lead Times with Backorders

So far orders have been assumed to be delivered instantaneously. Now relax this hypothesis by assuming that there is a positive integer delivery lead time $L > 0$ between placement and delivery of orders. In that event the recursion (1) is no longer valid. The reason is that the state of the system at the beginning of period i must then include not only the initial inventory in that period, but also the orders scheduled for delivery at the beginnings of periods $i+1, \ldots, i+L-1$. The result is that it is necessary to replace (1) in general by a recursion involving a function of L state variables. This means that it is possible to carry out the computations in such problems only where L is small.

However, it is possible to simplify the computations dramatically in one important case. This is where the demands $D_1, D_2, \ldots$ for the product in each period are independent and unsatisfied demands are backordered. Under these hypotheses it is possible to reduce the L-state-variable problem to a one-state-variable problem!

To see this, it is convenient to associate costs arising in period $i+L$ with period i as Figure 11 illustrates for $L=3$. In particular, let $g_i(z)$ be the storage and shortage cost in period $i+L$ when z is the stock on hand less backorders at the end of that period. Also let $c_i(z)$ be the cost—paid on delivery in period $i+L$—of an order for z units placed in period i.

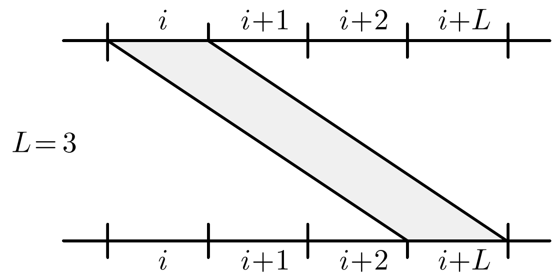


Figure 11. Association of Costs Incurred L Periods in Future with Present

The key idea is that the storage and shortage cost in period $i+L$ is a function only of the difference between the total stock y_i on hand and on order less backorders after ordering at the beginning of period i and the total demand in periods $i,\ldots,i{+}L$. To see this, observe that the stock on hand less backorders at the end of period $i+L$ is $y_i-\sum_{j=i}^{i+L}D_j$. Thus the storage and shortage cost in period $i+L$ is $g_i(y_i-\sum_{j=i}^{i+L}D_j)$. Hence

$$G_i(y_i)\equiv \mathrm{E}[g_i(y_i-\sum_{j=i}^{i+L}D_j)|y_i]$$

is the (conditional) expected storage and shortage cost in period $i+L$ given y_i.

Let $C_i(x)$ be the minimum expected cost in periods $i{+}L,\ldots,n{+}L$ where x is the total stock on hand and on order less backorders before ordering at the beginning of period i. Then C_i can be expressed in terms of C_{i+1} by the dynamic-programming recursion (1) in §8.1. This means that the theory in §8.2 and §8.3, which characterizes the forms of optimal ordering policies for a zero lead time, applies at once to the case of a positive lead time provide that unsatisfied demands are backordered. The only real difference is that when there is positive lead time, $y_i(x)$ is the optimal stock on hand *and on order* less backorders after ordering but before demands in period i.

The assumption that unsatisfied demands are backordered is vital to the above argument. For without it, the stock on hand less backorders at the end of period $i+L$ would depend individually on the stock on hand less backorders before ordering in period i, the outstanding orders after ordering in period i to be delivered in periods $i{+}1,\ldots,i{+}L$, and the demands in periods $i,\ldots,i{+}L$.

Reducing Lead Time Reduces Minimum Expected Cost

When the cost of ordering a given amount is paid on delivery, the effect of reducing lead times is to reduce the minimum expected cost. The reason is that if $0\le K<L$ are two integer lead times,

placing an order in a period with an L-period lead time can be simulated by waiting $L-K$ periods and then placing the order with a K-period lead time. Since the delivery times of all orders are the same with both systems, the ordering and storage/shortage costs are also the same. Thus, to each policy for a system with L-period lead times, there corresponds a policy for the modified system with K-period lead times that has the same expected costs.

Of course, one would not want to use an optimal policy for the L-period-lead-time problem in the manner described above where K-period lead times were in fact available. For this would amount to choosing the order quantities $L-K$ periods earlier than is necessary, thereby ignoring the information acquired in the interim about the actual demands experienced in the subsequent $L-K$ periods. Indeed, such a policy would impede one's ability to adjust order quantities to compensate for actual demands and to keep inventories closer to optimal levels. Thus the amount by which the minimum expected cost with L-period lead times exceeds that for K-period lead times can be thought of as the value of the information acquired by having the opportunity to delay placement of orders $L-K$ periods without affecting the delivery times.

5 SERIAL SUPPLY CHAINS: ADDITIVITY OF MINIMUM EXPECTED COST [CS60], [Ve66b]

Consider a serial supply chain for a single product in which there are N facilities. Label them $1,\ldots,N$. Each facility may carry stocks of the product. In each period, it is possible to order a nonnegative amount of product at any facility j from its immediate predecessor facility $j-1$ up to the amount available there if $j>1$ and from a supplier in any amount if $j=1$. Let x^j be the *initial echelon stock* at facility j in period i, i.e., the sum of the amounts on hand at facilities $j,\ldots,N$ at the beginning of period i. Call $x=(x^1,\ldots,x^N)$ the *initial echelon stock*. After observing the initial echelon stock in period $1\le i\le n$, the supply-chain manager places orders at each facility for nonnegative amounts of stock not exceeding available supplies at facilities supplying them. Initially, the lead time to deliver product from one facility to its successor will be one-period. Subsequently that case will be shown to include that of general positive integer lead times. The amounts of echelon stock on hand and on order after ordering at the facilities is the *starting echelon stock* $y=(y^1,\ldots,y^N)$. Demands arise only at facility N and D_i is the nonnegative demand there in period $i=1,\ldots,n{+}1$. The manager satisfies demands at facility N from stock on hand there as far as possible and backorders the excess demand. Thus, if the starting echelon stock in period i is y, the initial echelon stock in period $i+1$ is $y-D_i1$ where 1 is here an N-vector of ones. Evidently, the initial and starting echelon stocks satisfy $y^1\ge x^1\ge y^2\ge x^2\ge\cdots\ge x^{N-1}\ge y^N\ge x^N$.

Ordering costs are paid on delivery. Let $c_i(z)$ be the cost of orders $z=(z^1,\ldots,z^N)$ at the N facilities in period i for delivery in period $i+1$. Let $g_i(z)$ be the holding and shortage cost in period $i+1$ where z is the echelon stock on hand at the end of that period. Given the starting echelon stock y in period i, the (conditional) expected echelon holding and shortage cost in period $i+1$

is $G_i(y) \equiv \mathrm{E}g_i(y-(D_i+D_{i+1})1)$. Let $C_i(x)$ be the minimum expected cost in periods $i{+}1,\dots,n{+}1$ when the initial echelon stock in period i is x. (The expected cost in period i is excluded because decisions in period i do not affect the costs in that period, i.e., those costs are *sunk.*) Assume that all expectations exist and are finite, and that $G_i(y) \to \infty$ as $|y| \to \infty$ for all i. Then

$$C_i(x) = \min_{\substack{x^{j-1} \ge y^j \ge x^j \\ 1 \le j \le N}} \{c_i(y-x) + G_i(y) + \mathrm{E}C_{i+1}(y - D_i 1)\} \tag{2}$$

for $i = 1,\dots,n$ and $x^1 \ge \cdots \ge x^N$ where $x^0 \equiv \infty$ and $C_{n+1} \equiv 0$.

THEOREM 5. Serial Supply Chains. *Consider a serial supply chain with facilities* $1,\dots,N$, *the demands at facility* N *in each period being independent random variables, the delivery lead time at each facility being one period, and the ordering and the holding and shortage cost functions being additive. If at each facility after the first, the ordering cost functions are linear and the holding and shortage cost functions are convex, then the minimum expected cost* $C_i(x)$ *in periods* $i{+}1,\dots,n{+}1$ *is additive in the initial echelon stock* x *and convex in* $(x_2,\dots,x_N)$ *for* $x^1 \ge \cdots \ge x^N$. *Also there exist numbers* y_i^{j*} *such that one optimal starting echelon stock at facility* j *in period* i *is* $x^{j-1} \wedge (y_i^{j*} \vee x^j)$ *for* $j > 1$ *and all* i.

Proof. By hypothesis, there exist real-valued functions $c_i^j(\,\cdot\,)$, $g_i^j(\,\cdot\,)$ and constants c_i^j such that $c_i(z) = \sum_{j=1}^N c_i^j(z^j)$ with $c_i^j(z^j) \equiv c_i^j z^j$ for $j > 1$ and $g_i(z) = \sum_{j=1}^N g_i^j(z^j)$ with $g_i^j(\,\cdot\,)$ being convex for $j > 1$. Thus, $G_i(y) = \sum_{j=1}^N G_i^j(y^j)$ where $G_i^j(y^j) \equiv \mathrm{E}g_i^j(y^j - D_i - D_{i+1})$ for each j. Next show by backward induction on i that

$$C_i(x) = \sum_{j=1}^N C_i^j(x^j) \tag{3}$$

for some real-valued functions $C_i^j(\,\cdot\,)$ that are convex for each $j > 1$ and i. This is certainly true for $i = n{+}1$ on taking $C_{n+1}^j \equiv 0$ for $1 \le j \le N$. Suppose (3) holds for $1 < i{+}1 \le n{+}1$ and consider i. Then from the additivity of $c_i(\,\cdot\,)$, $G_i(\,\cdot\,)$ and $C_{i+1}(\,\cdot\,)$, (2) becomes

$$C_i(x) = \sum_{j=1}^N \min_{x^{j-1} \ge y^j \ge x^j} \{c_i^j(y^j - x^j) + G_i^j(y^j) + \mathrm{E}C_{i+1}^j(y^j - D_i)\}. \tag{4}$$

Suppose first that $j > 1$. Let

$$J_i^j(y^j) \equiv c_i^j y^j + G_i^j(y^j) + \mathrm{E}C_{i+1}^j(y^j - D_i) \tag{5}$$

and y_i^{j*} be the least minimizer of $J_i^j(\,\cdot\,)$. Since $g_i^j(\,\cdot\,)$ is convex, so is $G_i^j(\,\cdot\,)$. Also, $C_{i+1}^j(\,\cdot\,)$ is convex by the induction hypothesis, so $J_i^j(\,\cdot\,)$ is convex as well. Thus, since the minimum of a convex function over an interval is additive and convex in the endpoints of the interval by Problem 4 of Homework 2,

$$\min_{x^{j-1}\ge y^j \ge x^j} J_i^j(y^j) = \underline{J}_i^j(x^j) + \bar{J}_i^j(x^{j-1}) \tag{6}$$

on defining the convex functions $\underline{J}_i^j$ and $\bar{J}_i^j$ by

$$\underline{J}_i^j(x^j) \equiv J_i^j(y_i^{j*} \vee x^j) - \frac{1}{2}J_i^j(y_i^{j*}) \text{ and } \bar{J}_i^j(x^{j-1}) \equiv J_i^j(y_i^{j*} \wedge x^{j-1}) - \frac{1}{2}J_i^j(y_i^{j*}). \tag{7}$$

Moreover, one optimal choice of the starting echelon stock $y_i^j(x)$ at facility j in period i is

$$y_i^j(x) = (y_i^{j*} \vee x^j) \wedge x^{j-1}. \tag{8}$$

Consequently, it follows from (4)-(6) that (3) holds for i with

$$C_i^j(x^j) \equiv [\underline{J}_i^j(x^j) - c_i^j x^j] + \bar{J}_i^{j+1}(x^j) \tag{9}$$

being convex for $j > 1$, $\bar{J}_i^{N+1} \equiv 0$ and for all $j \ge 1$,

$$C_i^j(x^j) \equiv \min_{y^j \ge x^j}\{c_i^j(y^j - x^j) + G_i^j(y^j) + \mathrm{E}C_{i+1}^j(y^j - D_i)\} + \bar{J}_i^{j+1}(x^j). \;\blacksquare \tag{10}$$

Observe from (10) that it is possible to calculate the C_i^j recursively by period in the order $i = n,\dots,1$. Also (10) is the single-facility inventory equation for facility j in period i augmented by the added initial shortage cost function $\bar{J}_i^{j+1}(x^j)$. This function depends on the costs at facilities $j{+}1,\dots,N$ and, from (7), is decreasing convex. *Thus it is possible to solve the N-facility serial-supply-chain problem as a sequence of single-facility problems in the facility order $N,\dots,1$. The ordering costs at each facility other then the first are linear.*

Equation (10) for facility $j = 1$ is, apart from the convex function $\bar{J}_i^2(x^1)$, that for a single facility problem. Consequently, the results describing the form of the optimal policy in Theorem 1 (resp., 4) of §8.3 apply with minor modification to facilty one provided that the $g_i^1(\,\cdot\,)$ are convex and $c_i^1(z)$ is convex (resp., a setup-cost function with K_i being decreasing and nonnegative). The modification is that the first term on the right-hand side of (10) is convex (resp., K_i-convex) in x^1 and so adding $\bar{J}_i^2(x^1)$ preserves that property for $C_i^j(x^j)$.

Linear Holding and Shortage Costs. The most natural situation in which the echelon holding the shortage costs are additive and convex is where $g_i^j(z) = h_i^j z^+ + p_i^j z^-$ for each i, j. The interpretation is that in period j, h_i^j is the amount by which the unit storage cost at facility j exceeds that at $j-1$ and p_i^j is the amount by which the unit shortage cost at facility j exceeds that at facility $j+1$.

Extension to Arbitrary Lead Times. It is easy to extend the above results to the arbitrary positive-integer-lead-time problem by reducing that problem to a minor variant of the one-period lead-time problem in an augmented supply chain. Form the *augmented supply chain* by inserting $N(\bar{L}{-}1)$ in-transit facilities into the original supply chain where $\bar{L}$ is the average lead time over all N facilities in the original supply chain. To describe how to do this, call each facility in the original supply chain a *storage* facility and its immediate predecessor (its exogenous supplier if it is the first facility in the supply chain) its *supplier*. If there is an $L > 1$ period lead time to deliver orders placed by storage facility j, say, in the original supply chain, insert $L-1$ serial *in-transit* facilities between facility j and its supplier, each with a one-period lead time. An order in a period at j is shipped from j's supplier to the first of the in-transit (to j) facilities, then in order to the $2^{nd}, 3^{rd}, \ldots, L-1^{th}$ of these in-transit facilities, and finally from the $L-1^{th}$ in-transit facility to facility j.

Since each in-transit facility in the augmented supply chain requires one-period to deliver stock to its *customer* (the next facility, whether an in-transit or storage facility), the lead time to deliver an order placed by a storage facility with an L period lead time from its supplier in the original supply chain is at least L. To assure that the actual lead time is L, it is necessary to assure that in-transit facilities immediately ship whatever they receive; only storage facilities may hold stock for a period or more. To assure this, whenever facility j's immediate predecessor $j-1$ is an in-transit facility in the augmented supply chain, replace each instance of the inequality $x^{j-1} \ge y^j$ by the equation $x^{j-1} = y^j$. For such facilities j, replace (7) and (8) by

$$\underline{J}_i^j(x^j) \equiv 0 \text{ and } \bar{J}_i^j(x^{j-1}) \equiv J_i^j(x^{j-1}) \tag{7$'$}$$

and

$$y_i^j(x) = x^{j-1} \tag{8$'$}$$

respectively. Equation (9) remains valid as is. Equation (10) holds only at facilities j in the augmented supply chain that have no predecessor (and so order from an exogenous supplier) or for which facility $j-1$ is not an in-transit facility.

Incidentally, in practice one would expect $G_i^j = 0$ for all i at each in-transit facility j since storage and shortage costs are normally incurred only at the storage facilities. Also, since the ordering cost is paid on delivery, then it is natural to expect that $c^j = 0$ at j.

6 SUPPLY CHAINS WITH COMPLEX DEMAND PROCESSES [Sc59], [Ka60], [IK62], [BCL70], [Ha73], [GO01]

The above development makes two basic assumptions about the demand process. One is that the demands in successive periods are known and independent. Another is that demands for a product in a period are for immediate delivery. This section examines how the optimal supply policy changes with various relaxations of these assumptions.

Markov-Modulated Demand [IK62]**.** It is often reasonable to suppose that the demand for a product in a period given depends on the state of an underlying (vector) Markov process in that period. More precisely, the conditional distribution of demand in period i given the current and past states of the Markov process as well as the past demands depends only on the state m, say, of the

Markov process in period i. Before giving some examples of this idea, consider how one would find an optimal ordering policy under this assumption about the demand process for the general single-facility supply problem with backorders and no delivery lead times that §8.1 discusses.

Let $C_i(x,m)$ be the minimum expected cost in periods $i,\dots,n$ given that in period i the initial stock on hand is x and the index of the Markov process is m. Let m_i be the state of the Markov process at the end of period i. Then the analog of (1) becomes

$$C_i(x,m) = \min_{y\ge x}\{c_i(y-x) + G_i(y,m) + \mathrm{E}[C_{i+1}(y-D_i,m_i)|m]\} \tag{11}$$

for $i=1,\dots,n$ where $G_i(y,m)\equiv \mathrm{E}[g_i(y,D_i)|m]$ for all i and $C_{n+1}\equiv 0$. In this event, one optimal starting stock $y_i(x,m)$ in period i depends on both x and m. Moreover the demands and the underlying Markov process are independent of the supply manager's decisions. *It is easy to see from this fact that $y_i(x,m)$ varies with x for the case of convex* (*resp., setup*) *ordering costs precisely as Theorem* 1 (*resp.,* 4) *describes. The only difference is that the optimal policy in period i depends on m.*

Below are three examples to which the above results apply.

Example 1. Index of Business Conditions. In many situations it is natural to suppose that the demand for a product in a period depends on an index of (general or specific) business conditions, e.g., gross national product or total sales of automobiles. In such circumstances, it is reasonable to suppose that the index of business conditions is itself a Markov process whose distribution is known and that the distribution of demand in a period given the index of business conditions and past demands depends only on the index. Then the above results apply.

Example 2. Demand Distributions with Unknown Parameter: a Bayesian Perspective [Sc59], [Ka60]**.** In practice the distributions of demands in successive periods are rarely known with certainty. Instead, one may use expert judgment and/or learn about them from experience. Consider a simple example of this situation in which the demands $D_1,\dots,D_n$ in periods $1,\dots,n$ are independent random variables with a common distribution $\Psi(\,\cdot\,|\,\Lambda)$ that depends on an unknown parameter Λ. Though the parameter Λ is unknown and not directly observable, it is reasonable to suppose that Λ is a random variable and an expert provides its known prior distribution at the beginning of the first period. As one observes subsequent demands, one updates the distribution of Λ to reflect the new information. In particular, in each period i one calculates the posterior distribution of Λ given the demands $D_1,\dots,D_{i-1}$ observed before period i and uses this information to calculate the conditional distribution of the demand D_i given $D_1,\dots,D_{i-1}$. In general this is complex to do because both conditional distributions depends on the $i-1$ previous demands and so are of high dimensionality.

For this reason, it is reasonable to restrict attention to natural distributions $\Psi(\,\cdot\,|\,\Lambda)$ for which the posterior distribution of Λ at the beginning of period i depends on a sufficient statistic of low dimension. One such class of distributions is the exponential family. This family includes the gamma, Poisson and negative-binomial distributions.

If a distribution in the *exponential family* is continuous with density $\psi(\,\cdot\,|\,\Lambda)$, that density assumes the form $\psi(u\,|\,\Lambda) = \beta(\Lambda)e^{\Lambda u}r(u)$ where $r(u) = 0$ for $u < 0$ and $\beta(\Lambda)$ is a normalizing constant chosen so the integral of the density is one for each Λ. Let $p(\lambda)$ be the prior density of Λ. Then a sufficient statistic for the posterior density of Λ in period $i+1$ is the sum $m_i = D_1 + \cdots + D_i$ of the demands in prior periods. Also, the posterior density of Λ at the beginning of period $i+1$ given m_i is

$$p_i(\lambda\,|\,m_i) = \beta(\lambda)^i p(\lambda) e^{\lambda m_i}\theta_i(m_i)$$

where $\theta_i(m_i)$ is a normalizing constant chosen so the integral of the density is one for each m_i. Thus, the posterior density is in the exponential family, the sufficient statistics $m_1, m_2, \ldots, m_n$ is a Markov process, and the posterior density of D_i given the past demands depends only on the sufficient statistic m_{i-1}. Consequently, the above results apply. In this case, $p_i(\lambda\,|\,m_i)$ is also TP_2 since $\ln p_i(\lambda\,|\,m_i)$ is superadditive.

Example 3. Percent-Done Estimating: Style Goods Inventory Management [Ha73]. Percent-done estimating is used by some large firms to predict demands for style goods for a season whose length is n periods. To describe what this means, let $d_1, \ldots, d_n$ be independent positive random variables with known distributions. Assume that the *cumulative demand* m_i for the product in periods $1, \ldots, i$ takes the form $m_i \equiv \prod_{j=1}^{i}(d_j+1)$, so that $m_i = (d_i+1)m_{i-1}$ for $i = 1, \ldots, n$ where $m_0 \equiv 1$. Thus the sequence $m_1, m_2, \ldots, m_n$ is a Markov process and the actual demand in period i is $D_i = d_i m_{i-1}$. Therefore, the conditional expected demand in each period given the cumulative demand through the preceding period is a fixed percentage of the latter with the percentage depending on the period in the season. This is the source of the name "percent-done estimating". Thus if a style produces high cumulative demand part way through a season, the result is high expected demand after that. In short, a product that is initially "hot", is likely to stay "hot". In practice, the expected cumulative demand rises slowly at the beginning of a season, then more rapidly in mid season, and finally more slowly near the end of the season. In this case (11) holds and the above results apply.

Positive Homogeneity. However, there is an interesting situation in which it is possible to reduce the number of state-variables from two to one. To see this, suppose that in each period the ordering and the holding and shortage costs are positively homogeneous[1]. Then it is easy to see

---

[1]Call a real-valued function $f(\,\cdot\,)$ of several real variables *positively homogeneous* if $f(\lambda w) = \lambda f(w)$ for all $\lambda \geq 0$ and all w.

that $C_i(\,\cdot\,,\,\cdot\,)$ and $y_i(\,\cdot\,,\,\cdot\,)$ are positively homogeneous. Thus, on setting $C_i(x) \equiv C_i(x,1)$, $y_i(x) \equiv y_i(x,1)$ and $G_i(y) \equiv G_i(y,1)$, it follows by dividing (11) by m, replacing x/m and y/m by x and y respectively, and factoring out d_i+1 from the last term in (11) that the latter simplifies to

$$(12) \qquad C_i(x) = \min_{y \ge x} \{c_i(y-x) + G_i(y) + \mathrm{E}(d_i+1)C_{i+1}((y-d_i)/(d_i+1))\}$$

with $y = y_i(x)$ being a minimizer of the right-hand side of (12). Thus, $y_i(x,m) = y_i(\frac{x}{m})m$. This completes the reduction of state variables from two to one. Incidentally, positive homogeneity leads to a similar reduction in the number of state variables in other problems as well.

Advance Bookings [BCL71], [GO01]. In practice customers often book orders for delivery of a product in a specific subsequent period. This situation is especially easy to analyze where the bookings are for delivery during the lead time $L \ge 0$ and unsatisfied demands are backordered. Denote by D_{ij} the *bookings* in period i for delivery in period j, $i \le j \le i{+}L$. Assume that the D_{ij} are independent. Let y_i be the stock on hand and on order less backorders and prior bookings after ordering at the beginning of period i. Then $y_i - \sum_{j=i}^{i+L}\sum_{k=j}^{i+L} D_{jk}$ is the stock on hand less backorders at the end of period $i+L$.

Let $g_i(z)$ be the storage and shortage cost in period $i+L$ when z is the stock on hand less backorders at the end of that period. Thus the (conditional) expected storage and shortage cost in period $i+L$ given y_i is

$$G_i(y_i) \equiv \mathrm{E}[g_i(y_i - \sum_{j=i}^{i+L}\sum_{k=j}^{i+L} D_{jk})|y_i].$$

Also let $c_i(z)$ be the cost of placing an order for z units of stock in period i for delivery in period $i+L$.

The goal is the minimize the expected ordering, storage and shortage costs in periods $1{+}L,\ldots,$ $n{+}L$. To that end, let $C_i(x)$ be the minimum expected cost in periods $i{+}L,\ldots,n{+}L$ given that in period i the stock on hand and on order less backorders and prior bookings before ordering in that period is x. Then the C_j can be calculated inductively from the dynamic-programming recursion

$$C_i(x) = \min_{y \ge x} \{c_i(y-x) + G_i(y) + \mathrm{E}C_{i+1}(y - \sum_{j=i}^{i+L} D_{ij})\}$$

for $i = 1,\ldots,n$ where $C_{n+1} \equiv 0$. Also one optimal choice of the stock $y = y_i(x)$ on hand and on order less backorders and prior bookings at the beginning of period i is a minimizer of the expression in braces on the right-hand side of the above equation subject to $y \ge x$.

Observe that the above dynamic-programming recursion is an instance of (1) of §8.1. Thus the above development reduces the optimal supply problem with advance demand information and a positive lead time to that with no advance demand information and a zero lead time. Hence the results in §8.2 and §8.3, which characterize optimal policies for the latter problem, apply at once to the former, i.e., the present, problem.

9
Myopic Supply Policy with Stochastic Demands

[Ve65c,d,e], [IV69]

1 INTRODUCTION

The results of §8 generalize in various ways to several products. However, the recursion (1) of §8 then involves tabulating functions of N variables where N is the number of products. This is generally not possible to do when $N > 3$ say. Nevertheless, there are a number of important cases where it is possible to compute optimal policies for reasonably large N, say for N in the thousands. One is where a myopic policy is optimal (c.f., Theorem 3 of §8) as the development of this Section shows.

Suppose a supply-chain manager seeks an ordering policy for each of N interacting products that minimizes the expected n-period cost. Assume the demands $D_1, D_2, \ldots$ for the N products in periods $1, 2, \ldots$ are independent random N-vectors with values in $\mathcal{D} \subseteq \Re^N$. At the beginning of each period i, the manager observes the N-vector x_i of *initial stocks* of the N products and orders an N-vector z_i of the N products with immediate delivery. The manager chooses the vector z_i from a given subset Z of $\Re^N_+$. Unsatisfied demands are backordered, so $x_{i+1} = y_i - D_i$ where $y_i \equiv x_i + z_i$ is the N-vector of *starting stocks* in period i. Suppose the cost in each period i is $g_i(y_i, D_i)$ and that the *conditional expected cost* $G_i(y_i) \equiv \mathrm{E}[g_i(y_i, D_i)|y_i]$ given the starting stock y_i in period i is $+\infty$ or real-valued. Assume for now that there is no ordering cost. Section 9.3 discusses the extension to a linear ordering cost.

Assume there is a function $\bar{y}_i : \Re^N \to \Re^N$ such that $\bar{y}_i(x) - x \in Z$ and

$$G_i(\bar{y}_i(x)) = \min_{y-x\in Z} G_i(y)$$

for all $x \in \Re^N$. Call the ordering policy that uses $\bar{y}_1$ in period 1, $\bar{y}_2$ in period 2, etc., *myopic*.

Observe that it is not necessary to tabulate $\bar{y}_i(\,\cdot\,)$ in order to implement the myopic policy. All that is required is to wait and observe the initial inventory x_i in period i and then compute $\bar{y}_i(x_i)$. Thus, for the n-period problem, it suffices to compute only $\bar{y}_1(x_1), \ldots, \bar{y}_n(x_n)$. Hence it is necessary to solve only one N-variable optimization problem in each of the n periods.

2 STOCHASTIC OPTIMALITY AND SUBSTITUTE PRODUCTS [IV69]

It is useful to introduce a very strong form of optimality. Call a policy *stochastically-optimal* if for each x_1 and sequence $D_1, D_2, \ldots$ of demands, the inventory sequence $x_1, y_1, x_2, y_2, \ldots$ that the policy generates has the property that $G_i(y_i) \le G_i(y_i')$ for all i where $x_1, y_1', x_2', y_2', \ldots$ is the inventory sequence that any other ordering policy generates using the same demand sequence. In particular this implies $\sum_1^n \mathrm{E}G_i(y_i) \le \sum_1^n \mathrm{E}G_i(y_i')$, i.e., the first ordering policy minimizes the expected n-period costs for each x_1 and n. The next theorem gives a useful sufficient condition for the myopic policy to be stochastically optimal.

THEOREM 1. Stochastic Optimality of Myopic Policies. *If Z is closed under addition and if*

$$\bar{y}_{i+1}(\bar{y}_i(x) - d) = \bar{y}_{i+1}(x - d) \tag{1}$$

for all $d \in \mathcal{D}$, $x \in \Re^N$ and $i = 1, \ldots, n-1$, then the myopic policy $(\bar{y}_i)$ is stochastically optimal.

Significance of the Hypotheses

The condition that Z is closed under addition holds if it is possible to order each product in any amount or in any multiple of a given (product-dependent) batch size. The condition rules out upper bounds on order sizes.

The condition (1) asserts that for each initial inventory x and demand d in period i, the starting stock in period $i+1$ when the myopic policy is used in both periods i and $i+1$ is the same as the starting stock in period $i+1$ when no orders are placed in period i and the myopic policy is used in period $i+1$. Condition (1) assures that use of the myopic policy in period i not only minimizes the "cost" in period i, but also leaves the initial inventory in a best possible position at the beginning of period $i+1$.

Proof. The proof entails comparison of policies with common random demands $D_1, D_2, \ldots$. Let $x_1, y_1, x_2, y_2, \ldots$ and $x_1' = x_1, y_1', x_2', y_2', \ldots$ be the respective inventory vectors that the myopic policy and an arbitrary policy generate. Also let $x_i'' \equiv x_1 - \sum_{j=1}^{i-1} D_j$ be the initial inventory vector in period i when there are no orders in periods $1, \ldots, i-1$.

The first step is to show that

$$\bar{y}_i(x_i) = \bar{y}_i(x_i''),\ i = 1, 2, \ldots. \tag{2}$$

This is trivially so for $i = 1$. Suppose it is so for i and consider $i+1$. Then from (2) and (1),

$$\bar{y}_{i+1}(x_{i+1}) = \bar{y}_{i+1}(\bar{y}_i(x_i'') - D_i) = \bar{y}_{i+1}(x_i'' - D_i) = \bar{y}_{i+1}(x_{i+1}''),$$

establishing (2).

Now since Z is closed under addition, $x_i' - x_i'' \in Z$ and therefore $\bar{y}_i(x_i') - x_i'' = (\bar{y}_i(x_i') - x_i') + (x_i' - x_i'') \in Z$. Hence by (2),

$$G_i(y_i) = G_i(\bar{y}_i(x_i'')) \le G_i(\bar{y}_i(x_i')) \le G_i(y_i')$$

for all i. ∎

Substitute Products

In order to apply the Stochastic-Optimality-of-Myopic-Policies Theorem, it is necessary to have useful sufficient conditions for (1) to hold. To do this requires the following definition.

Call $\bar{y}_i$ *consistent* if $\bar{y}_i(x') = \bar{y}_i(x)$ whenever $x \le x' \le \bar{y}_i(x)$ and $x' - x \in Z$. This is no real restriction on $\bar{y}_i(x_i')$ if Z is closed under nonnegative differences because $y = \bar{y}_i(x)$ is one minimizer y of $G_i(y)$ subject to $y - x' \in Z$. In particular if $\bar{y}_i(x)$ is the lexicographically least minimizer for each x, then $\bar{y}_i$ is consistent.

THEOREM 2. Substitute Products. *If demands are nonnegative, Z is closed under addition, $\bar{y}_i(x)$ is increasing in i for each x, $\bar{y}_i(x) - x$ is decreasing in x for each i, and $\bar{y}_i$ is consistent for each i, then* (1) *holds. Also the myopic policy* $(\bar{y}_i)$ *is stochastically optimal.*

Proof. The first two hypotheses on $\bar{y}_i$ imply that

$$x - d \le \bar{y}_i(x) - d \le \bar{y}_i(x-d) \le \bar{y}_{i+1}(x-d) \text{ for } d \ge 0 \text{ and all } x.$$

Since $\bar{y}_{i+1}$ is consistent,

$$\bar{y}_{i+1}(\bar{y}_i(x) - d) = \bar{y}_{i+1}(x-d),$$

which establishes (1). The rest follows from Theorem 1. ∎

Substitute-Products Interpretation. It is natural to call Theorem 2 the Substitute-Products Theorem because of the assumption that the myopic order vector $\bar{y}_i(x) - x$ of all products is decreasing in the initial inventory x of those products. Thus initial inventories of all products are substituted for myopic order quantities of all products. This implies that increasing the initial inventory of any product will reduce the myopic starting stocks of all other products, i.e., initial inventories of one product may be substituted for myopic starting inventories of the other products. This is one sense in which the products can be thought of as substitutes.

There is an even more satisfactory sense in which the products can be thought of as substitutes if Z is the nonnegative orthant. Then the myopic starting stock of each commodity j is increasing in the initial stock of that commodity. To see this, let $y = (y^k)$, $x = (x^k)$ and $G_i^j(y^j)$ be the minimum of $G_i(y)$ subject to $y^k \ge x^k$ for all $k \ne j$ and y^j fixed. Then the myopic starting stock for product j is a y^j that minimizes $G_i^j(y^j)$ subject to $y^j \ge x^j$. Since the last (partially-optimized) problem entails minimizing a subadditive (actually an additive) function over a sublattice, an appropriate selection of the myopic starting stock $y^j = \bar{y}_i^j(x)$ is generally increasing in the initial inventory of product j. In this event, the myopic policy calls for substituting starting stocks of product j for starting stocks of the other products.

Here are four examples that satisfy the hypotheses of the Substitute-Products Theorem. In the first three, $Z = \Re_+^N$; and in all four, $\mathcal{D} = \Re_+^N$.

Example 1. Single Product [Ve65d]. Suppose $N = 1$, each G_i is quasiconvex and assumes its minimum on $\Re$ at $\underline{S}_i$, and $\underline{S}_i \le \underline{S}_{i+1}$ for all i (c.f., §8.3). Then $\bar{y}_i(x) - x = (\underline{S}_i - x)^+$ is increasing in i and decreasing in x, and $\bar{y}_i$ is consistent. Of course this example is essentially an instance of the Optimality-of-(s, S)-Policies Theorem with no setup costs.

Example 2. Multi-Product Shared Storage [IV69]. Suppose $y = (y^1, \ldots, y^N)$, $x = (x^1, \ldots,\ x^N)$ and

$$G_i(y) = \sum_{j=1}^N G^j(y^j) + H_i(\sum_{j=1}^N y^j)$$

where the G^j and H_i are convex, and $H_i(z)$ is subadditive in (i, z). Then $\bar{y}_i$ may be chosen to satisfy the hypotheses of the Substitute-Products Theorem. To see this, put $y^{1N} \equiv \sum_{j=1}^N y^j$ and observe that the problem of finding $\bar{y}_i$ reduces to the minimum-convex-cost network-flow problem of minimizing

$$\sum_{j=1}^N G^j(y^j) + H_i(y^{1N})$$

subject to

$$y^{1N} - \sum_{j=1}^N y^j = 0$$

and

$$y^j \ge x^j,\ j = 1, \ldots, N.$$

Figure 1 illustrates the network. Let x^j be the parameter associated with the arc in which y^j flows and i be the parameter associated with the arc in which y^{1N} flows. Hence, since the arcs in which y^{1N} and y^j flow are complements for each j, $\bar{y}_i(x) = (\bar{y}_i^j(x))$ is increasing in i and $\bar{y}_i^j(x) - x^j$ is decreasing in x^j by the Monotone-Optimal-Flow-Selections Theorem. Also, since the arcs in which y^j and y^k flow are substitutes for $j \neq k$, $\bar{y}_i^k(x)$ is decreasing in x^j. Thus $\bar{y}_i(x) - x$ is increasing in $(i, -x)$, so the third and fourth hypotheses of the Substitute-Products Theorem hold. It follows from the Iterated Optimal-Flow Lemma 3 of Appendix 2 that $\bar{y}_i$ is consistent. Thus, the myopic policy is stochastically optimal.

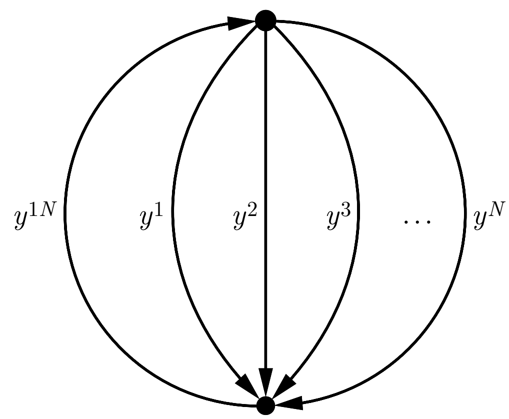


Figure 1. Network for Multi-Product Shared Storage

It should be noted that the function H_i is the only factor preventing the problem from splitting into N separate single-product problems. Indeed splitting is in order when the H_i all vanish.

As a concrete example of H_i, suppose there is an upper bound B_i on the sum of the starting stocks of the N products in period i. Then $H_i(y^{1N}) = \delta_+(B_i - y^{1N})$ is subadditive in (i, y^{1N}) if and only if B_i is increasing in i.

Incidentally, although the theory does not require it, the Monotone Optimal-Flow-Selections Theorem also implies that $\sum_{k=1}^{N} \bar{y}_i^k(x)$ $(= y^{1N})$ and $\bar{y}_i^j(x)$ are increasing in x^j.

Finally, note that this example extends easily to functions $G_i(y)$ that are additive functions of "nested" partial sums of the starting stocks as Figure 2 illustrates with $y^{jk} \equiv \sum_{l=j}^{k} y^l$.

Example 3. Equilibrium [Ve65d]. Suppose first $G_i = G$ for all i and let $\underline{S}$ be a global minimum of G. Then $\bar{y}_i(x) = \underline{S}$ for $x \leq \underline{S}$. Also, the hypotheses of the Substitute-Products Theorem hold for all $x_1 \leq \underline{S}$. Thus, an examination of the proofs of the Stochastic-Optimality-of-Myopic-Policies and Substitute-Products Theorems shows that the myopic policy is stochastically optimal for all $x_1 \leq \underline{S}$.

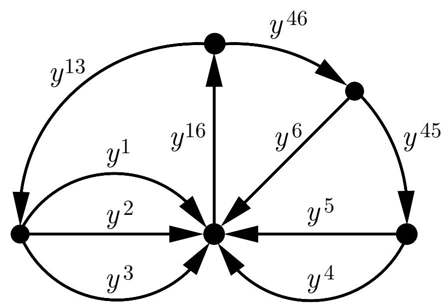


Figure 2. Network for Multi-Product Shared Storage

If in fact G_i is not independent of i, $G_i(y)$ is subadditive in (i, y) and $G_i(\,\cdot\,)$ has a nonempty compact level set, then G_i has a least global minimum $\underline{S}_i$ and, by the Increasing-Optimal-Selections Theorem, $\underline{S}_i$ is increasing in i. Then $\bar{y}_i(x) = \underline{S}_i$ for $x \le \underline{S}_i$ and the hypotheses of the Substitute-Products Theorem hold for all $x_1 \le \underline{S}_1$. Thus the myopic policy is stochastically optimal for $x_1 \le \underline{S}_1$.

Example 4. Optimality of (k, Q) Policies with Batch Orders [Ve65c]. Suppose $N = 1$. So far Z has been the nonnegative real line. However, in many practical situations it is necessary to order in batches of a given size $Q > 0$, e.g., a box, a case, a pallet, a carload, etc., though demands may be for smaller quantities. For this reason it is of interest to consider the case where Z is the set of nonnegative multiples of Q. Then Z is closed under addition. Suppose also that $G_i = G$ is quasiconvex and that there is a k such that $G(k) = G(k + Q)$ with $[k, k + Q]$ containing a global minimum of G as Figure 3 illustrates. Then the myopic policy $\bar{y}_i = \bar{y}$ entails ordering the smallest multiple of Q that brings the starting inventory to at least the *reorder point* k. Call this policy (k, Q). Evidently the hypotheses of the Substitute-Products Theorem hold, so the foregoing myopic (k, Q) policy is stochastically optimal. This result also carries over to the nonstationary case provided that the corresponding reorder points k_i are increasing in time.

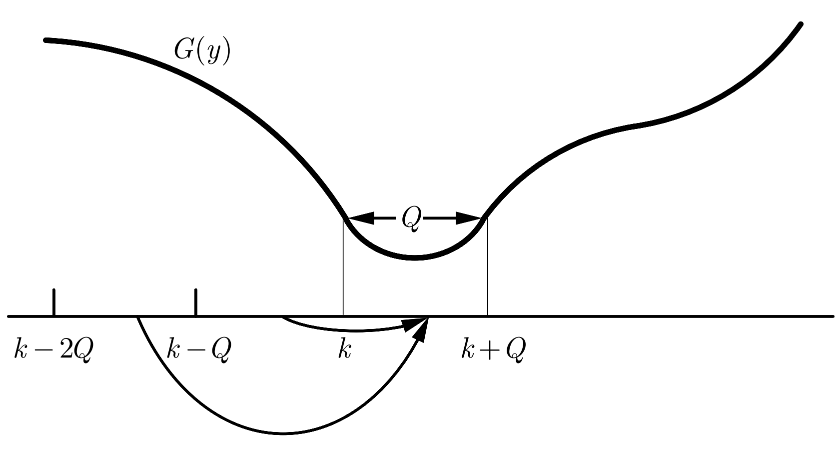


Figure 3. Myopic (k, Q) Policy

3 ELIMINATING LINEAR ORDERING COSTS AND POSITIVE LEAD TIMES

Reducing Linear to Zero Ordering Costs

The assumption made so far in this section is that there is no ordering cost. It is possible to relax this assumption by allowing linear ordering costs. Then it is easy to show how to reduce that problem to one with zero-ordering-costs by appropriately modifying the storage and shortage cost functions in each period.

To see how, suppose that the cost of ordering the N-vector $z \geq 0$ of the N products in period i is the linear functional $c_i z$ for $i = 1, \ldots, n$. Also suppose that the *salvage value* of the N-vector z of (possibly negative) net stocks of each product on hand at the end of period n is $c_{n+1}z$. Finally, assume that there is a storage and shortage cost $h_i(y_i, D_i)$ in period i when the starting stock in that period is y_i and the demand vector therein is D_i.

Thus when the ordering policy generates the inventory sequence $x_1, y_1, x_2, y_2, \ldots,$ the total cost $\mathcal{C}_n$ incurred in periods $1, \ldots, n$ is

$$\mathcal{C}_n = \sum_1^n [c_i(y_i - x_i) + h_i(y_i, D_i)] - c_{n+1}x_{n+1} = \sum_1^n g_i(y_i, D_i) - c_1 x_1$$

where the *modified storage and shortage cost in period* i is

$$g_i(y_i, D_i) \equiv c_i y_i + h_i(y_i, D_i) - c_{i+1}(y_i - D_i)$$

for $i = 1, \ldots, n$. The interpretation is that the modified storage and shortage cost in period i is the cost of ordering enough to raise the initial stock in the period from zero to the starting stock level plus the given cost of storage and shortage in the period minus the salvage value of the net stock on hand at the end of the period.

Hence the problem of finding an ordering policy that minimizes the expected n-period cost $\mathrm{E}\mathcal{C}_n$ is equivalent to that of finding an ordering policy that minimizes the modified expected n-period cost

$$\sum_1^n \mathrm{E}G_i(y_i),$$

where $G_i(y_i) \equiv \mathrm{E}[g_i(y_i, D_i)|y_i]$ for all i.

This is the promised reduction of the linear-to the zero-ordering-cost problem. Thus all results of this section as well as the results on (s, S) policies in the preceding section apply as well to the problem in which the ordering costs also include a linear term.

Reducing Positive to Zero Lead Times

The results on reducing positive to zero lead times in §8.4 carry over immediately to the multiproduct case provided that the lead times for all products coincide. In particular the results of this section extend immediately to that case.

This page is blank.

Appendix

1 REPRESENTATION OF SUBLATTICES OF FINITE PRODUCTS OF CHAINS

Call a subset L of a product S of n chains $S_1, \ldots, S_n$ *i-decreasing* (resp., *i-increasing*) if for each $r \in L$ and $s \in S$ with $r_i = s_i$ and $r \geq s$ (resp., $r \leq s$), one has $s \in L$. The subsets L that are *both* i-decreasing and j-increasing are of special interest because, as we shall soon see, they are

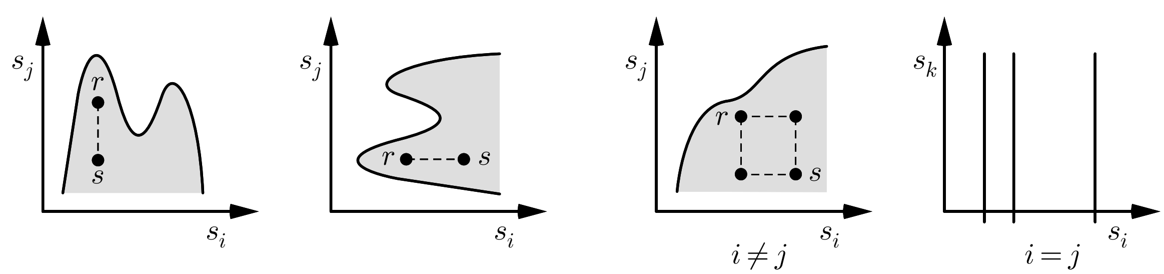


Figure 1. i-Decreasing, j-Increasing and i-Decreasing j-Increasing Sets

sublattices of S. Moreover, each such set L is a cylinder in S with base in $S_i \times S_j$ for $i \neq j$ and in S_i for $i = j$. For if $s \in L$ and $k \neq i, j$, then decreasing (resp., increasing) s_k to a different value in S_k keeps s in L because L is i-decreasing (resp., j-increasing). Thus, the base of the i-decreasing j-increasing set L is one dimensional if $i = j$ and two dimensional in the contrary event.

In practice one often finds a subset of $S_i \times S_j$ represented as the set of solutions to a nonlinear inequality of the form $f(s_i, s_j) \leq 0$ for some real-valued function f on $S_i \times S_j$. The set of such solutions is i-decreasing j-increasing if either $i = j$ or $i \neq j$ and $f(s_i, s_j)$ is decreasing in s_i and increasing in s_j.

The class of i-decreasing j-increasing subsets of S is clearly closed under intersections and contains S. Thus there is a smallest i-decreasing j-increasing set $L_{ij}^{\downarrow\uparrow}$ containing any given subset L of S, viz., the intersection of all i-decreasing j-increasing sets that contain L. The set $L_{ij}^{\downarrow\uparrow}$ is called the *i-decreasing j-increasing hull* of L.

The hull $L_{ij}^{\downarrow\uparrow}$ has the alternate important representation as the projection

$$(1) \qquad L_{ij}^{\downarrow\uparrow} = \pi_S\{(r,s) \in L \times S : r_i \le s_i, r_j \ge s_j\},$$

i.e., $L_{ij}^{\downarrow\uparrow}$ is the set of $s \in S$ for which $r_i \le s_i$ and $r_j \ge s_j$ for some $r \in L$. Figure 2 below illustrates this formula with L being the darkly shaded area and $L_{ij}^{\downarrow\uparrow}$ being the entire (lightly and darkly) shaded area. To establish (1), let K be the right-hand side thereof. Evidently, K is i-decreasing j-increasing and so $K \supseteq L_{ij}^{\downarrow\uparrow}$. Conversely, if $s \in K$, then there is an $r \in L$ satisfying $r_i \le s_i$ and $r_j \ge s_j$. Since $L_{ij}^{\downarrow\uparrow}$ is i-decreasing, the vector r' formed from r by replacing r_j by s_j is in $L_{ij}^{\downarrow\uparrow}$. Then since $L_{ij}^{\downarrow\uparrow}$ is j-increasing, the vector r'' formed from r' by replacing r_i by s_i is in $L_{ij}^{\downarrow\uparrow}$. Finally, since $L_{ij}^{\downarrow\uparrow}$ is a cylinder with base in $S_i \times S_j$ if $i \ne j$ and in S_i if $i = j$, the vector s formed from r'' by replacing its k^{th} component by s_k for each $k \ne i,j$ is in $L_{ij}^{\downarrow\uparrow}$, so $K \subseteq L_{ij}^{\downarrow\uparrow}$. Hence, $K = L_{ij}^{\downarrow\uparrow}$ as claimed.

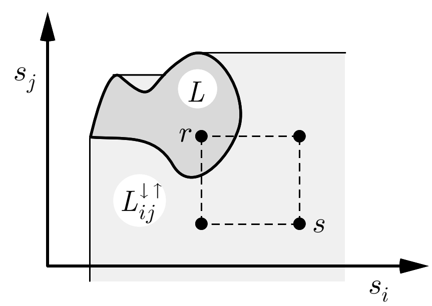


Figure 2. i-Decreasing j-Increasing Hull of L

The next result gives the desired representation of sublattices of a product S of n chains as Figure 3 illustrates.

THEOREM 1. Representation of Sublattices. *The following properties of a subset L of a product S of n chains $S_1, \ldots, S_n$ are equivalent.*

1° *L is a sublattice of S.*

2° *L is the intersection of its i-decreasing j-increasing hulls for all i,j.*

3° *L is the intersection of i-decreasing j-increasing subsets of S for some pairs i,j.*

Proof. 1° $\Rightarrow$ 2°. We have $L \subseteq \bigcap_{i,j} L_{ij}^{\downarrow\uparrow} \equiv K$ because each $L_{ij}^{\downarrow\uparrow}$ is a hull of L. Suppose $s \in K$. Then for each i,j, $s \in L_{ij}^{\downarrow\uparrow}$. Thus from (1) there exist $r^{ij} \in L$ with $r_i^{ij} \le s_i$ and $r_j^{ij} \ge s_j$. Put $r^i \equiv$

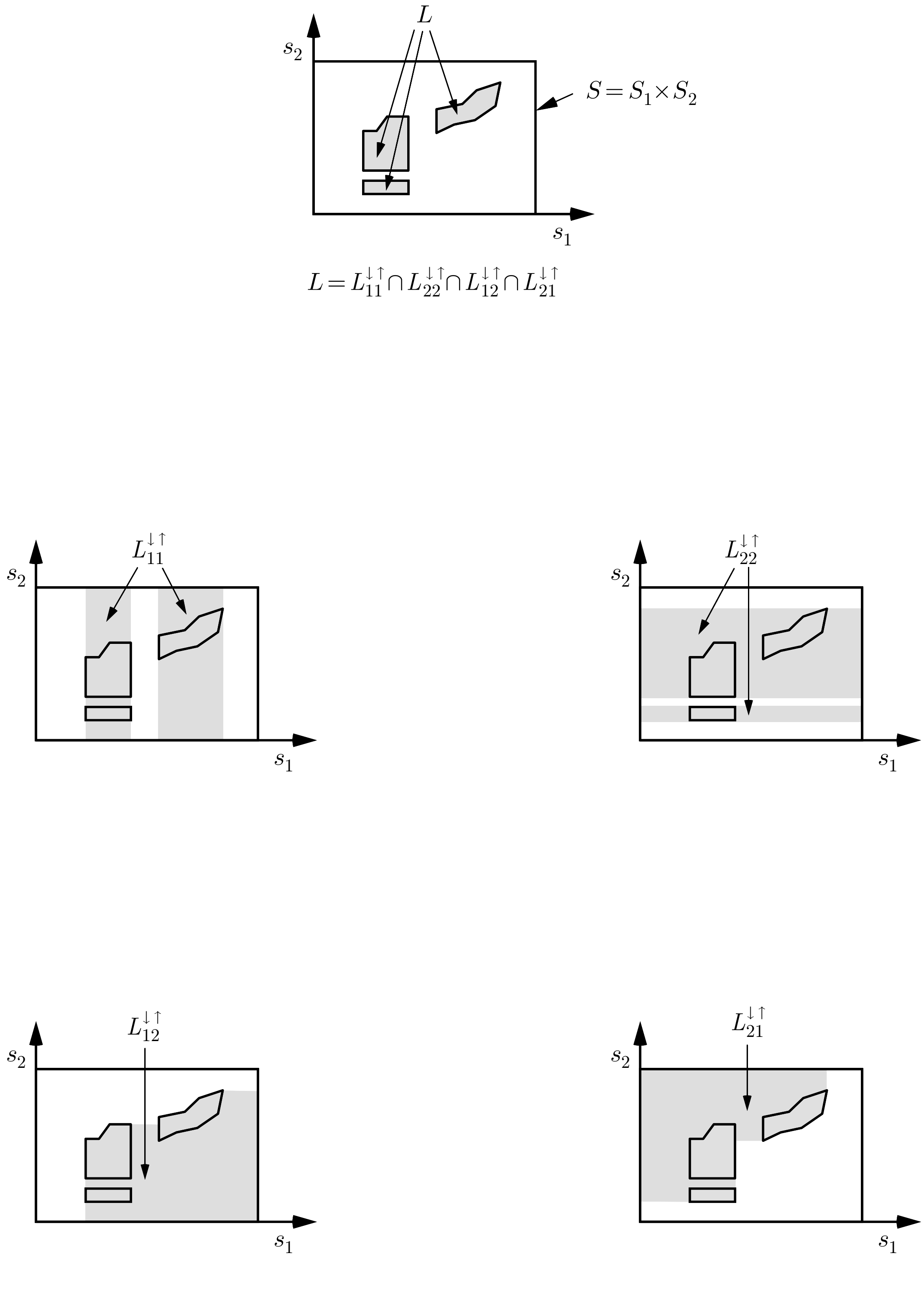


Figure 3. Representation of a Sublattice of a Product of Two Chains

$\vee_j r^{ij}$. Then $r_i^i = \vee_j r_i^{ij} = r_i^{ii} = s_i$, and for $k \neq i$, $r_k^i = \vee_j r_k^{ij} \geq r_k^{ik} \geq s_k$, so $r^i \geq s$ and $r_i^i = s_i$. Since L is a sublattice of S, $s = \wedge_i r^i = \wedge_i \vee_j r^{ij} \in L$, so $L = K$.

$2° \Rightarrow 3°$. Immediate.

$3° \Rightarrow 1°$. Since intersections of sublattices are sublattices, it suffices to show that an i-decreasing j-increasing subset K of S is a sublattice thereof. Suppose $r, s \in K$. Now S_j is a chain, so we can assume $r_j \leq s_j$. Thus $(r \vee s)_j = s_j$, so because $r \vee s \geq s$ and K is j-increasing, $r \vee s \in K$. Similarly, since K is i-decreasing, $r \wedge s \in K$. ▮

Representation of Polyhedral Sublattices

Theorem 1 has many important applications. One is a representation for *polyhedral* sublattices, i.e., sublattices that are intersections of finitely many closed half-spaces. We begin with the simplest case, viz., a single closed half-space.

Example 1. Closed Half-Spaces that are Sublattices. A *closed half-space* $\{s \in \Re^n : as \leq b\}$ is a sublattice of $\Re^n$ if and only if the normal $a \in \Re^n$ has at most one positive and at most one negative element.

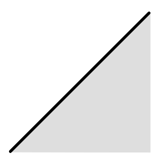


Figure 4a. A Sublattice

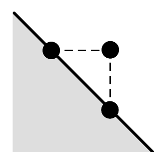


Figure 4b. A Nonsublattice

Since intersections of sublattices are sublattices, intersections of closed half-spaces that are each sublattices are themselves sublattices. The next result asserts that every polyhedral sublattice arises in this way. The proof that 1° implies 2° below consists of applying the Fourier-Motzkin elimination method to carry out the projection in (1) and then applying Theorem 1 to the resulting system of inequalities. The proofs that 2° implies 3° and that 3° implies 1° are immediate from Examples 6 and 5 of §2.2.

THEOREM 2. Representation of Polyhedral Sublattices as Duals of Weighted Distribution Problems. *The following are equivalent.*

1° *L is a polyhedral sublattice of $\Re^n$.*

2° *$L = \{s \in \Re^n : As \leq b\}$ for some matrix $(A \;\; b)$ such that each row of A has at most one positive and at most one negative element.*

3° *L is the intersection of finitely many closed half-spaces, each of which is a sublattice of $\Re^n$.*

Example 2. A Polyhedral Sublattice

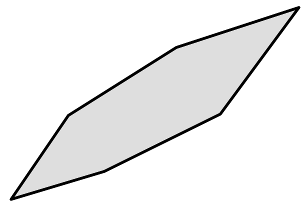


Figure 5. A Polyhedral Sublattice

It is important to note that Theorem 2 does not assert that the closed half-spaces in every representation of a polyhedral sublattice are sublattices, and indeed that is not the case as the following example illustrates.

Example 3. Representations of a Polyhedral Sublattice L

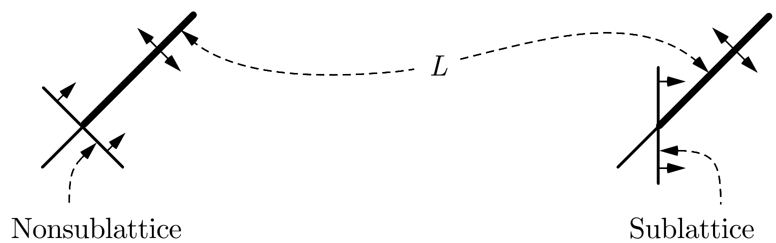


Figure 6. Two Representations of a Polyhedral Sublattice

2 PROOF OF MONOTONE-OPTIMAL-FLOW-SELECTION THEOREM

This section extends the proof of the Monotone-Optimal-Flow-Selection Theorem 4.3 from the case of strictly convex functions to arbitrary convex ones.

Compactness of Level Sets of Convex Functions

LEMMA 1. Compactness of Convex Level Sets. *If f is a $+\infty$ or real-valued lower-semicontinuous convex function on $\Re^n$ and the set of points in $\Re^n$ where f assumes its minimum thereon is nonempty and compact, then every level set of f is compact.*

Proof. Let X^o be the set of points in $\Re^n$ where f assumes its minimum, and let X be a nonempty level set of f. Then $X^o \subseteq X$. Choose $x \in X^o$. If X is unbounded, there is a sequence $\{d_n\}$ of directions in $\Re^n$ with $\|d_n\| = 1$ and nonnegative numbers λ_n with $\lim \lambda_n = \infty$ such that $x + \lambda_n d_n \in X$ for $n = 1, 2, \ldots$. By possibly taking subsequences we can assume $\lim d_n = d$ say, with $\|d\| = 1$. Since X is convex, $x + \lambda d_n \in X$ for $\int$all $0 \le \lambda \le \lambda_n$. And because f is lower semicontinuous, X is closed. Thus since $\lim \lambda_n = \infty$, $x + \lambda d \in X$ for all $0 \le \lambda$. Therefore, since $f(x + \lambda d)$ is increasing and convex in $\lambda \ge 0$, $f(x + \lambda d)$ is constant in $\lambda \ge 0$ and so $x + \lambda d \in X^o$ for all $\lambda \ge 0$. Thus

X^o is unbounded, contradicting the fact X^o is compact. Hence X is bounded. Since X is also closed, it is compact as claimed. ∎

Closedness of Optimal-Response Multi-Function

The *optimal-response* multi-function X^o assigns to each vector y the set $X^o(y)$ of vectors x that minimize $f(x,y)$ subject to $x \in X(y)$. It is important to seek conditions assuring that X^o is "closed". To explore this question requires some definitions.

Let X be a multi-function from a subset Y of $\Re^m$ into a subset X' of $\Re^n$. Call X *closed* if $y, y_n \in Y$, $x_n \in X(y_n)$, and $\lim\,(x_n, y_n) = (x,y)$ imply that $x \in X(y)$. Call X *continuous* if also $y, y_n \in Y$, $\lim y_n = y$, and $x \in X(y)$ imply that there exist $x_n \in X(y_n)$ with $\lim x_n = x$. If X is single-valued, i.e., $X(y)$ is a point-to-point mapping, then by an abuse of language, X is closed (resp., continuous) as a multi-function if and only if X is continuous in the usual sense of point-to-point mappings, viz., $y, y_n \in Y$ and $y_n \to y$ imply that $\lim X(y_n) = X(y)$.

Example 1. A Nonclosed Multi-Function. Let $X' = Y = [0,1]$, $X(y) = [\frac{1}{4}y, y]$ for $y \in [0, \frac{1}{2})$, and $X(y) = [y, 1]$ for $y \in [\frac{1}{2}, 1]$. Then $X(y)$ is not closed from the left at $y = \frac{1}{2}$.

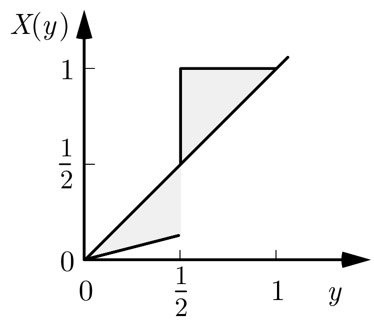


Figure 1. A Nonclosed Multi-Function

Example 2. A Closed Discontinuous Multi-Function. Let $X' = Y = [0,1]$, $X(y) = [y, \frac{1}{4} + y]$ for $y \in [0, \frac{1}{4})$, $X(y) = [0, \frac{3}{4}]$ for $y \in [\frac{1}{4}, \frac{3}{4}]$, and $X(y) = [\frac{5}{4} - y, 1 - y]$ for $y \in (\frac{3}{4}, 1]$.

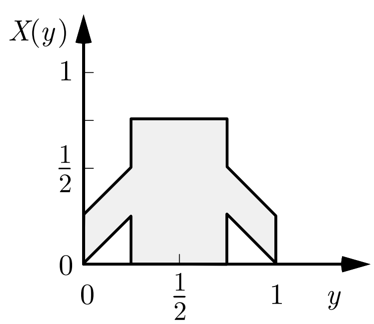


Figure 2. A Closed Discontinuous Multi-Function

Example 3. A Continuous Multi-Function. Let $X' = Y = [0,1]$ and $X(y) = [y, \frac{1}{2} + y]$ for $y \in [0, \frac{1}{2}]$ and $X(y) = [\frac{3}{2} - y, 1 - y]$ for $y \in (\frac{1}{2}, 1]$.

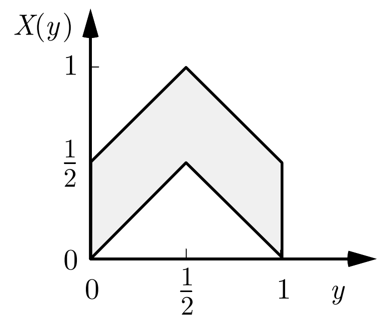


Figure 3. A Continuous Multi-Function

Let $X(\,\cdot\,)$ be a multi-function from a subset Y of $\Re^m$ into a subset X' of $\Re^n$. The *graph* of X is the set $\{(x,y) \in X' \times Y : x \in X(y)\}$. If f is a real-valued function on the graph of X, the *projection* of f is the function g defined by

$$g(y) \equiv \inf_{x \in X(y)} f(x,y) \text{ for } y \in Y,$$

and the *optimal-response multi-function* is the multi-function X^o defined by

$$X^o(y) \equiv \{(x,y) \in X' \times Y : x \in X(y) \text{ and } g(y) = f(x,y)\}.$$

LEMMA 2. Closedness of Optimal-Response Multi-Function. *If X is a continuous nonempty multi-function from a subset Y of $\Re^m$ into a compact subset X' of $\Re^n$ and f is a continuous real-valued function on the graph of X, then the optimal-response multi-function X^o is nonempty and closed, and the projection g of f is continuous.*

Proof. Since X is continuous, it is closed and so $X(y)$ is closed for $y \in Y$. But $X(y) \subseteq X'$ and X' is compact, so $X(y)$ is also compact. Hence since $f(\,\cdot\,, y)$ is continuous on $X(y)$ and $X(y) \neq \phi$, $X^o(y)$ is nonempty and compact for $y \in Y$ so X^o is nonempty.

Now suppose $y, y_n \in Y$, $x_n \in X^o(y_n)$ and $\lim\,(x_n, y_n) = (x,y)$. Since X is closed, $x \in X(y)$. Choose $w \in X^o(y)$. Since X is continuous, there exist $w_n \in X(y_n)$ such that $\lim w_n = w$. Hence since $f(\,\cdot\,,\,\cdot\,)$ is continuous,

$$\begin{aligned}\liminf g(y_n) &= \lim f(x_n, y_n) = f(x,y) \geq g(y) \\ &= f(w,y) = \lim f(w_n, y_n) = \lim g(y_n),\end{aligned}$$

so equality occurs throughout. Thus $g(\,\cdot\,)$ is continuous on Y and X^o is closed. ∎

Iterated Limits

If f is a function from $\Re^n$ into itself, we say $f(x)$ has an *iterated limit* g as $x \to y$ in $\Re^n$, written

$$\operatorname*{Ilim}_{x\to y} f(x) = g$$

if

$$\lim_{x_n\to y_n}\cdots\lim_{x_2\to y_2}\ \lim_{x_1\to y_1} f(x) = g.$$

Observe, that the iterated limit depends on the order of taking limits. If the order of taking limits is not specified, an arbitrary order may be taken.

Now fix $t \in T$, put $\epsilon \equiv (\epsilon_a)$ and let

$$C'(x,\epsilon) \equiv \sum_{a\in\mathcal{A}} c'_a(x_a,\epsilon_a) \tag{1}$$

where

$$c'_a(\xi,\eta) \equiv c_a(\xi,t_a) + \eta e^{-\xi}. \tag{2}$$

Denote by $X'(\epsilon)$ the set of feasible flows that minimize $C'(\,\cdot\,,\epsilon)$. Of course, $X'(0) = X(t)$. Also, write $u \gg 0$ where $u \in \Re^n$ if u is positive.

LEMMA 3. Iterated Optimal Flow. *In a biconnected graph, if $t \in T$, $c_a(\,\cdot\,,t_a)$ is convex and lower semicontinuous for each $a \in \mathcal{A}$, and $X(t)$ is nonempty and bounded, there is a unique $x(\epsilon)$ $\in X'(\epsilon)$ for each $\epsilon \gg 0$ and $\mathrm{Ilim}_{\epsilon\downarrow 0}\ x(\epsilon)$ exists, is in $X(t)$ and has the ripple property.*

Proof. Since the set of optimal flows for the given and perturbed flow costs is unchanged by multiplying the flow cost by a positive scalar, assume without loss of generality that $\epsilon \le 1$. Now show first that $X'(\epsilon)$ is nonempty and the graph of the multi-function X' is compact on the interval $[0,1]$. To that end, there is an $\underline{x} \in X(t)$. Since $C(\underline{x},t)$ is finite, $L \equiv C'(\underline{x},1)$ is finite. Then since $C'(x,\epsilon)$ is increasing in ϵ, the level set $\underline{X}'(\epsilon) \equiv \{\text{flows } x : C'(x,\epsilon) \le L\}$ is nonempty and $\underline{X}'(\epsilon) \subseteq \underline{X}'(0)$ for each $0 \le \epsilon \le 1$. Now $C(\,\cdot\,,t)$ is lower semicontinuous, and $X(t)$ is nonempty and bounded, so $X(t) = X'(0)$ is nonempty and compact. Thus since the level sets of a lower-semicontinuous convex function are all compact if one of them is nonempty and compact by Lemma 1, $\underline{X}'(0)$ is nonempty and compact. Hence $X'(\epsilon)$ is nonempty and compact, whence the same is so of $X'(\epsilon)$. Since for each arc a, $c_a(\,\cdot\,,t_a)$ is convex and continuous on the interval where it is finite, $C'(\,\cdot\,,\,\cdot\,)$ is finite and continuous on the compact set $\underline{X}'(0)\times[0,1]$. Thus it follows from Lemma 2 that the graph of the multi-function $\epsilon \to X'(\epsilon) \subseteq \underline{X}'(0)$ is compact on $[0,1]$.

Now since $C'(\,\cdot\,,\epsilon)$ is strictly convex on $\underline{X}'(0)$ for $0 \ll \epsilon \le 1$, there is a unique $x(\epsilon) \in X'(\epsilon)$ for each such ϵ, and $x(\epsilon)$ has the ripple property. Since c'_a is subadditive for each arc a, it follows from Lemma 4.4 and the Ripple Theorem 4.2 that

$$|x_a(\epsilon') - x_a(\epsilon)| \le x_b(\epsilon') - x_b(\epsilon) \tag{3}$$

for each arc a and $0 \ll \epsilon < \epsilon' \le 1$ differing only in the b^{th} component.

For the remainder of the proof, let $1, 2, \ldots, p$ be any enumeration of the arcs. Now show by induction that $\mathrm{Ilim}_{\epsilon \downarrow 0}\, x(\epsilon)$ exists and is in $X'(0)$. To that end, let ϵ^i be the vector formed from ϵ by replacing its first i components by 0. Now $\lim_{\epsilon_1 \downarrow 0} x(\epsilon)$ exists and is in $X'(\epsilon^1)$. To see this, observe that $x_1(\epsilon)$ is increasing in ϵ_1 by (3) so that $\lim_{\epsilon_1 \downarrow 0} x_1(\epsilon)$ exists and is finite because the graph of X' is compact. Also using (3) and the compactness of the graph of X' again, $\lim_{\epsilon_1 \downarrow 0} x_i(\epsilon)$ exists and is finite for each $i > 1$, so $x(\epsilon^1) \equiv \lim_{\epsilon_1 \downarrow 0} x(\epsilon)$ exists and is in $X'(\epsilon^1)$. Moreover, (3) also holds for $\epsilon_1 = 0$. Now given $x(\epsilon^{i-1})$ and (3) also holding for $\epsilon_1 = \cdots = \epsilon_{i-1} = 0$, define $x(\epsilon^i)$ by the rule $x(\epsilon^i) \equiv \lim_{\epsilon_i \downarrow 0} x(\epsilon^{i-1}) \in X'(\epsilon^i)$. One shows that the limit exists exactly as for the case $i = 1$. Moreover, (3) holds also for $\epsilon_1 = \cdots = \epsilon_i = 0$. Therefore, $x(0) \equiv x(\epsilon^p) \in X'(\epsilon^p) = X'(0) = X(t)$, and $x(0)$ has the ripple property. ∎

The *iterated optimal-flow selection* is the selection that chooses the *iterated optimal flow* $\mathrm{Ilim}_{\epsilon \downarrow 0}\ x(\epsilon)$ defined in Lemma 3 for each $t \in T$.

Proof of Theorem 4.3 Without Strict Convexity

It is now possible to complete the proof of Theorem 4.3. To that end consider the case where the $c_a(\,\cdot\,, t_a)$ are convex, but not necessarily strictly so. Then perturb the flow cost as in (1), (2). By Lemma 3, there is a unique flow $x(t, \epsilon)$ that is optimal for the perturbed flow cost for each $t \in T$ and $\epsilon \gg 0$. From what was shown above, $x(t, \epsilon)$ has the ripple and monotonicity properties in t given in Theorem 4.3. Thus the iterated optimal flow $x(t) \equiv \mathrm{Ilim}_{\epsilon \downarrow 0}\ x(t, \epsilon)$, which exists and is in $X(t)$ for $t \in T$ by Lemma 3, has these properties as well. ∎

3 SUBADDITIVITY/SUPERADDITIVITY OF MINIMUM COST IN PARAMETERS

Section 4.4 examines the variation of the optimal arc flows with the parameter vector t. This section studies instead the behavior of the minimum cost $\mathcal{C}(t)$ as a function of t. In order to motivate the result, recall that in the example of Figure 5 of §4, the flow in arc d was eliminated by expressing it in terms of the flows in arcs a and b that are complements. Since the resulting cost is subadditive in the flows in arcs a and b and their associated parameters, $\mathcal{C}(t)$ is subadditive in those parameters by the Projection Theorem for subadditive functions provided that c_a and c_b are subadditive and $c_d(\,\cdot\,, \tau)$ is convex for each τ. The next Theorem generalizes this conclusion.

For any subset $\mathcal{S}$ of arcs, let $T_{\mathcal{S}} \equiv \times_{a \in \mathcal{S}} T_a$. And if $z = (z_a)$ is a vector whose elements are indexed by the arcs, let $z_{\mathcal{S}}$ be the subvector of z with indices in $\mathcal{S}$.

THEOREM 1. Subadditivity of Minimum-Cost in Parameters of Complements. *In a biconnected graph, if the arcs in $\mathcal{S}$ are complements, $T_{\mathcal{A} \setminus \mathcal{S}} = \{t_{\mathcal{A} \setminus \mathcal{S}}\}$, $c_a(\,\cdot\,, t_a)$ is convex for each $a \in \mathcal{A} \setminus \mathcal{S}$, T is a sublattice, c_a is subadditive for each $a \in \mathcal{S}$, and $\mathcal{C} > -\infty$ on T, then $\mathcal{C}$ is subadditive on T.*

Proof. Suppose that $t, t' \in T$. Put $\underline{t} \equiv t \wedge t'$ and $\overline{t} \equiv t \vee t'$. It is necessary to show that $\mathcal{C}(\underline{t}) + \mathcal{C}(\overline{t}) \le \mathcal{C}(t) + \mathcal{C}(t')$, or equivalently, that for every pair of flows x and x',

$$\mathcal{C}(\underline{t}) + \mathcal{C}(\overline{t}) \le C(x,t) + C(x',t').$$

Observe that if either term on the right-hand side of the above inequality is $+\infty$, then the inequality holds trivially. Thus, assume without loss of generality that both terms are finite.

Let P, Q be the partition of $\mathcal{S}$ for which $x'_P \ll x_P$ and $x'_Q \ge x_Q$. By the Circulation-Decomposition Theorem 4.1, $x' - x$ is a sum of simple conformal circulations. Let y be the sum of those simple circulations whose induced simple cycle contains an arc in P and let z be the sum of the remaining simple circulations. Then y and z are conformal, $y_P = x'_P - x_P$ and $z_P = 0$. Moreover, since the arcs in $\mathcal{S}$ are complements, $z_Q = x'_Q - x_Q$ and $y_Q = 0$. For if $y_a \neq 0$ for some $a \in Q$, the induced cycle of one of the simple circulations forming y, say w, contains a and an arc $b \in P$. Since w is conformal with $x' - x$, $w_a w_b < 0$, which contradicts the fact that a and b are complements. Thus $x_\mathcal{S} \wedge x'_\mathcal{S} = x_\mathcal{S} + y_\mathcal{S}$ and $x_\mathcal{S} \vee x'_\mathcal{S} = x_\mathcal{S} + z_\mathcal{S}$.

Using these facts and our hypotheses on the arc costs,

$$\mathcal{C}(\underline{t}) + \mathcal{C}(\overline{t}) \le C(x+y,\underline{t}) + C(x+z,\overline{t}) \le C(x,t) + C(x',t'). \blacksquare$$

In order to obtain a result comparable to that above for a set of substitute arcs, it is necessary to strengthen the hypotheses of the theorem to require that the T_a be chains and the $c_a(\,\cdot\,,\tau)$ be convex for every $\tau \in T_a$ and $a \in \mathcal{A}$.

THEOREM 2. Superadditivity of Minimum-Cost in Parameters of Substitutes. *In a biconnected graph, if the arcs in $\mathcal{S}$ are substitutes, $T_{\mathcal{A}\backslash\mathcal{S}} = \{t_{\mathcal{A}\backslash\mathcal{S}}\}$, $c_a(\,\cdot\,,\tau)$ is convex for each $\tau \in T_a$ and $a \in \mathcal{A}$, T is a sublattice, c_a is subadditive and T_a is a chain for each $a \in \mathcal{S}$, and $\mathcal{C} > -\infty$ on T, then $\mathcal{C}$ is superadditive on T.*

Proof. Suppose that $t, t' \in T$. Put $\underline{t} \equiv t \wedge t'$ and $\overline{t} \equiv t \vee t'$. It is necessary to show that $\mathcal{C}(t) + \mathcal{C}(t') \le \mathcal{C}(\underline{t}) + \mathcal{C}(\overline{t})$, or equivalently, that for every pair of flows x and x',

$$\mathcal{C}(t) + \mathcal{C}(t') \le C(x,\underline{t}) + C(x',\overline{t}).$$

Observe that if either term on the right-hand side of the above inequality is $+\infty$, then the inequality holds trivially. Thus, assume without loss of generality that both terms are finite.

Since T_a is a chain for each $a \in \mathcal{S}$, there is a unique partition M, P, Q of $\mathcal{S}$ for which $x_M \gg x'_M$, $x_P \le x'_P$ and $t'_P \le t_P$, and $x_Q \le x'_Q$ and $t'_Q \gg t_Q$. By the Circulation-Decomposition Theorem, $x' - x$ is a sum of simple conformal circulations. Let y be the sum of those simple circulations whose induced simple cycle contains an arc in P and let z be the sum of the remaining simple circulations. Then y and z are conformal, $y_P = x'_P - x_P$ and $z_P = 0$. Moreover, since the arcs in $\mathcal{S}$ are substitutes, $z_Q = x'_Q - x_Q$ and $y_Q = 0$. For if $y_a \neq 0$ for some $a \in Q$, the induced cycle of one of the simple circulations forming y, say w, contains a and an arc $b \in P$. Since w is conformal with $x' - x$, $w_a w_b > 0$, which contradicts the fact that a and b are substitutes.

Now observe from these facts that

$$c_a(x_a + y_a, t_a) + c_a(x_a + z_a, t'_a) \leq c_a(x_a, \underline{t}_a) + c_a(x'_a, \bar{t}_a)$$

by convexity and subadditivity for $a \in M$, as an identity for $a \in P \cup Q$, and by convexity for $a \in \mathcal{A} \setminus \mathcal{S}$. Thus adding the above inequalities yields

$$\mathcal{C}(t) + \mathcal{C}(t') \leq C(x + y, t) + C(x + z, t') \leq C(x, \underline{t}) + C(x', \bar{t}). \blacksquare$$

4 DYNAMIC-PROGRAMMING EQUATIONS FOR CONCAVE-COST FLOWS

Proof of Theorem 6.4. The first step is to show that C satisfies (6.1) and (6.2). Since (6.1) holds trivially, it suffices to consider (6.2) with $r_I \neq 0$. Let C_{iI} be an element of C. If $C_{iI} = +\infty$, then (6.2) holds. For if not, either C_{jI} is finite for some $(i, j) \in \mathcal{A}_I$ or B_{iI} is finite. In the former event, there is a feasible flow for the subproblem $i \to I$, viz., the sum of the preflow that sends r_I along arc (i, j) from i to j and a minimum-cost flow for the subproblem $j \to I$. In the latter event, $|I| > 1$, for if not, $B_{iI} = 0$, so $I = \{i\}$, which implies $C_{iI} = 0$, a contradiction. Then C_{iJ} and $C_{i,I \setminus J}$ are finite for some $\emptyset \subset J \subset I$, so there are feasible flows for the subproblems $i \to J$ and $i \to (I \setminus J)$. Hence, the sum of these flows is a feasible flow for the subproblem $i \to I$. In both cases, there is contradiction to the assumption that $C_{iI} = +\infty$. Thus the claim (6.2) holds.

Now suppose C_{iI} is finite. For any preflows x and y with $x + y$ a feasible flow for the subproblem $i \to I$, $C_{iI} \leq c(x + y)$. Hence, by the subadditivity of $c(\,\cdot\,)$,

$$C_{iI} \leq c(x) + c(y). \tag{1}$$

Suppose $(i, j) \in \mathcal{A}_I$ and let x be the preflow that sends r_I along (i, j) from i to j and y be a minimum-cost flow for the subproblem $j \to I$. Then $x + y$ is a feasible flow for the subproblem $i \to I$, so by (1),

$$C_{iI} \leq c_{ij}(r_I) + C_{jI}. \tag{2}$$

If no minimum-cost flow exists for the subproblem $j \to I$, then $C_{jI} = +\infty$, so (2) holds in this case as well.

Let J be a nonempty proper subset of I, and let x and y be minimum-cost flows for the subproblems $i \to J$ and $i \to (I \setminus J)$ respectively. Then $x + y$ is a feasible flow for the subproblem $i \to I$, so (1) becomes

$$(3) \qquad C_{iI} \le C_{iJ} + C_{i,I\setminus J}.$$

If a minimum-cost flow does not exist for one of the subproblems $i \to J$ or $i \to (I \setminus J)$, then (3) holds trivially because the right-hand side thereof is $+\infty$.

Next show that either (2) holds with equality for some j or $C_{iI} = B_{iI}$. The first step is to show that the last equality holds if $i \in I$. If $|I| = 1$, then $I = \{i\}$ and $C_{iI} = 0 = B_{iI}$ because the minimum cost with a zero demand vector is zero by Theorem 3. If $|I| > 1$, then (3) holds with equality by choosing $J = \{i\}$, because then $C_{iJ} = 0$, as was just shown, and trivially $C_{iI} = C_{iI_i}$.

Now suppose $i \in \mathcal{N} \setminus I$. Since C_{iI} is finite, there is a minimum-cost flow z that is extreme for the subproblem $i \to I$ by Theorem 3. The forest that z induces is a union of disjoint trees. Since $i \in \mathcal{N} \setminus I$ and $r_I \ne 0$, one of these trees, say T, contains i and an arc joining i to another node j, say. If i is a leaf of T, then (2) holds with equality for that j. In the contrary event, let J be the nonempty subset of all nodes in I that lie in the maximal subtree of T containing j but not i. Since i is incident to at least two arcs in T, J is a proper subset of I, so (3) holds with equality for that J. Hence C satisfies (6.1) and (6.2) as claimed.

Next show by induction on the cardinality of I that C majorizes every $+\infty$ or real-valued solution C' of (6.1) and $(6.2)'$. To that end observe that when $r_I \ne 0$, C_{iI} is the minimum cost among all chains in $\mathcal{G}'_I$ from i to ν in which the arc costs are as defined above. Thus $C_I \equiv (C_{jI})$ for $j \in \mathcal{N}$ is the greatest $+\infty$ or real-valued solution of $(6.2)'$ given $B_I \equiv (B_{jI})$ for $j \in \mathcal{N}$.

Now if $|I| = 1$, whence $r_I \ne 0$, $C'_{iI} \le C_{iI}$ for all $i \in \mathcal{N}$. Thus assume the claim is so for all $\emptyset \subset I \subset \mathcal{D}$ for which $|I| = k - 1 \ge 1$, and consider I for which $|I| = k$. If $r_I = 0$, then by (6.1) and the induction hypothesis, for each $j \in I$, $C'_{iI} = C'_{jI_j} \le C_{jI_j} = C_{iI}$. Thus suppose $r_I \ne 0$. Then $B'_{iI} \le B_{iI}$ for $i \in \mathcal{N}$ by the induction hypothesis, so $C'_{iI} \le C_{iI}$ for all $i \in \mathcal{N}$ because the greatest $+\infty$ or real-valued solution of $(6.2)'$ is increasing in the arc costs.

It remains to show by induction on the cardinality of I that if every simple circuit in $\mathcal{G}'_I$ has positive cost, then C is the only $+\infty$ or real-valued solution of (6.1) and $(6.2)'$. To that end it suffices to show that for each fixed $\emptyset \subset I \subset \mathcal{D}$ and $r_I \ne 0$, the minimum-cost-chain equations $(6.2)'$ have a unique $+\infty$ or real-valued solution because each simple circuit in $\mathcal{G}'_I$ has positive cost. For then one sees by induction, as in the preceding paragraph, that the inequalities given there hold with equality.

To see that $(6.2)'$ has a unique $+\infty$ or real-valued solution, let C' be any such solution. Now by iterating $(6.2)'$ one sees that C'_i minorizes the cost of each simple chain in $\mathcal{G}'_I$ from i to ν. Thus it suffices to show that either C'_i equals the cost of some such chain or $C'_i = +\infty$, in which case there is no simple chain in $\mathcal{G}'_I$ from i to ν. If the former event does not occur, then one sees by iterating $(6.2)'$ that there is a node j, a simple chain P in $\mathcal{G}'_I$ from i to j, and a simple circuit Q in $\mathcal{G}'_I$ containing j such that $C'_i = c_P + C'_j$ and $C'_j = c_Q + C'_j$ where c_P and c_Q are respectively the costs of traversing P and Q. Since c_Q is positive by hypothesis, it follows that $C'_j = +\infty$, whence $C'_i = +\infty$. ∎

5 SIGN-VARIATION-DIMINISHING PROPERTIES OF TOTALLY POSITIVE MATRICES

[Ka68, Ch. 5]

Let A be a matrix with row and column indices I and J respectively and denote by

$$A\begin{pmatrix} i_1 & \cdots & i_r \\ j_1 & \cdots & j_r \end{pmatrix}$$

the determinant of the submatrix of A formed by deleting therefrom all rows and columns except those labeled $i_1, \ldots, i_r$ in I and $j_1, \ldots, j_r$ in J respectively. The above determinant is called a *minor* of A of order r.

Totally Positive Matrices

If I and J are chains, call A *totally* (resp., *strictly totally*) *positive of order* r, written TP_r (resp., STP_r), if the minors of A of all orders not exceeding r are all nonnegative (resp., positive). Of course the TP_s (resp., STP_s) matrices are TP_r (resp., STP_r) for all $1 \le r < s$.

The TP_1 (resp., STP_1) matrices are precisely the nonnegative (resp., positive) matrices and the TP_r (resp., STP_r) matrices are nonnegative (resp., positive). The TP_2 matrices are the nonnegative matrices (a_{ij}) for which $\log a_{ij}$ is superadditive on $I \times J$. Also the square diagonal matrices of order r whose diagonal elements are nonnegative (resp., positive) are TP_r (resp., STP_r). Moreover, if I' and J' are chains, f and g are respectively increasing functions from I to I' and J to J', and (a_{ij}) is TP_r (resp., STP_r), then so is $(a_{f(i),g(j)})$.

If A, B and C are respectively matrices of orders $m \times n$, $m \times p$ and $p \times n$, and if $A = BC$, then the well known *Cauchy-Binet* formula (Gantmacher (1959), *Matrix Theory,* Vol. I, pp. 9-10) asserts that

$$(1)\qquad A\begin{pmatrix} i_1 & \cdots & i_r \\ k_1 & \cdots & k_r \end{pmatrix} = \sum_{j_1<\cdots<j_r} B\begin{pmatrix} i_1 & \cdots & i_r \\ j_1 & \cdots & j_r \end{pmatrix} C\begin{pmatrix} j_1 & \cdots & j_r \\ k_1 & \cdots & k_r \end{pmatrix}$$

for all $i_1 < \cdots < i_r$ and $k_1 < \cdots < k_r$. It follows from this fact that if B and C are respectively TP_r (resp., STP_r) and TP_s (resp., STP_s), then A is $TP_{r\wedge s}$ (resp., $STP_{r\wedge s}$). In particular, the class of square TP_r (resp., STP_r) matrices is closed under multiplication. Moreover, the class of TP_r (resp., STP_r) matrices is closed under pre- and post-multiplication by diagonal matrices (where defined) whose diagonal elements are nonnegative (resp., positive). The last assertion can be restated as asserting that if x_i and y_j are nonnegative (resp., positive) numbers and if the matrix $A = (a_{ij})$ is TP_r (resp., STP_r), then so is the matrix $(x_i a_{ij} y_j)$. In particular, since the matrix A whose elements are all 1 is TP_r, so is the matrix $(x_i y_j)$. It can be shown that the matrix whose ij^{th} element is $\binom{j}{i}$ is TP_r for every $r \ge 1$. Thus, the binomial distribution $\binom{j}{i}p^i(1-p)^{j-i}$ is TP_r for every fixed $0 \le p \le 1$ and $r \ge 1$. Also, the TP_r (resp., STP_r) matrices are closed under the operations of (i) multiplying one of their rows or columns by a nonnegative (resp., positive) number and (ii) deleting a row or column.

The matrix formed from a TP_r (resp., STP_r) $m \times n$ matrix A by replacing two of its consecutive rows or columns by their sum is TP_r (resp., STP_r). To see this, represent this transformation by pre- or post-multiplying A by a matrix formed by augmenting the identity matrix by inserting one $+1$ just above or below the diagonal. Then apply the Cauchy-Binet formula to the product.

Sign-Variation-Diminishing Properties of Totally Positive Matrices

Denote by S_x the number of sign changes in the vector $x = (x_1, \dots, x_n)$ after deleting the zero elements thereof. Thus for example, if $x = (1 \;\; 0 \;\; \pi \;\; 0 \; -2 \;\; 1 \;\; 3 \; -1 \;\; 0)$, then $S_x = 3$.

THEOREM 1. Sign-Variation-Diminishing Properties of Totally Positive Matrices. *If A is an $m \times n$ TP_r matrix and if $x \ne 0$ is a column n-vector such that $S_x \le r-1$, then $S_{Ax} \le S_x$. If also $S_{Ax} = S_x$, then the first nonzero element of Ax and of x have the same sign.*

Proof. Assume first that A is STP_r. Put $y \equiv Ax$ and $p \equiv S_x$. Now there exist integers $0 = n_0 < n_1 < \cdots < n_p < n_{p+1} = n$ such that $(-1)^{k+1}(x_{n_{k-1}+1}, \dots, x_{n_k}) \ne 0$ has constant sign for all $k = 1, \dots, p+1$. Without loss of generality assume that this constant sign is positive since $S_z = S_{-z}$ for all vectors z. Put $b^k \equiv \sum_{j=n_{k-1}+1}^{n_k} |x_j| a^j$ for $k = 1, \dots, p+1$ where a^j is the j^{th} column of A. Then it is possible rewrite the equation $y = Ax$ as

$$y = \sum_{k=1}^{p+1} (-1)^{k+1} b^k. \qquad (2)$$

Also, $B \equiv (b^1, \dots, b^{p+1})$ is STP_{p+1} since $p \le r-1$ and B is formed from A by deleting columns a^j of A for which $x_j = 0$, multiplying columns of the resulting matrix by positive numbers, and replacing consecutive sequences of columns by their sums.

Let $q \equiv S_y$. Now there exist integers $1 \le i_1 < \cdots < i_{q+1} \le m$ such that $(-1)^{k+1} y_{i_k} \ne 0$ has constant sign for $k = 1, \dots, q+1$.

Now $q \le p$. For if instead $q > p$, then

$$\begin{vmatrix} b_{i_1 1} & \cdots & b_{i_1 p+1} & y_{i_1} \\ \vdots & & \vdots & \vdots \\ b_{i_{p+2} 1} & \cdots & b_{i_{p+2} p+1} & y_{i_{p+2}} \end{vmatrix} = 0$$

since y is a linear combination of the columns of $B = (b_{ij})$ by (2). Expanding the above determinant by elements in the last column and using the fact that B is STP_{p+1} implies

$$0 = \sum_{k=1}^{p+2} (-1)^{p+2-k} y_{i_k} B\begin{pmatrix} i_1 & \cdots & i_{k-1} & i_{k+1} & \cdots & i_{p+2} \\ 1 & \cdots & k-1 & k & \cdots & p+1 \end{pmatrix} \neq 0,$$

a contradiction.

Next suppose $q = p$. Then from (2) again,

$$y_{i_j} = \sum_{k=1}^{p+1} (-1)^{k+1} b_{i_j k},\ j = 1, \ldots, p+1.$$

Solving this equation for the first "variable" $(-1)^2$ by Cramer's rule and expanding the determinant in the numerator by elements in the first column yields

$$1 = (-1)^2 = \frac{\begin{vmatrix} y_{i_1} & b_{i_1 2} & \cdots & b_{i_1 p+1} \\ \vdots & \vdots & & \vdots \\ y_{i_{p+1}} & b_{i_{p+1} 2} & \cdots & b_{i_{p+1} p+1} \end{vmatrix}}{B\begin{pmatrix} i_1 & \cdots & i_{p+1} \\ 1 & \cdots & p+1 \end{pmatrix}}$$

$$= \frac{\sum_{j=1}^{p+1} (-1)^{j+1} y_{i_j} B\begin{pmatrix} i_1 & \cdots & i_{j-1} & i_{j+1} & \cdots & i_{p+1} \\ 2 & \cdots & j & j+1 & \cdots & p+1 \end{pmatrix}}{B\begin{pmatrix} i_1 & \cdots & i_{p+1} \\ 1 & \cdots & p+1 \end{pmatrix}}.$$

Now each of the minors in the last expression is positive since B is STP_{p+1}, whence $(-1)^{j+1} y_{i_j} > 0$ for all j, so $y_{i_1} > 0$.

To extend the proof to the case in which A is TP_r, approximate A by a sequence of STP_r matrices converging to A. See [Ka68, pp. 220, 224] for details. ∎

This page intentionally left blank.

References

BOOKS

Aho, A. V., J. E. Hopcroft, and J. D. Ullman (1974). *The Design and Analysis of Computer Algorithms*. Addison-Wesley, Reading, Massachusetts.

Arrow, K. J., S. Karlin, and H. Scarf (1958). *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, California.

Arrow, K. J., S. Karlin, and H. Scarf (eds.) (1962). *Studies in Applied Probability and Management Science*. Stanford University Press, Stanford, California.

Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, Englewood Cliffs, New Jersey.

Berge, C. (1973). *Graphs and Hypergraphs*. North-Holland, Amsterdam.

Bertsekas, D. P. (1995) *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts.

Bertsimas, D. and J. N. Tsitsiklis (1997). *Introduction to Linear Optimization.* Athena Scientific, Belmont, Massachusetts.

Birkhoff, G. (1948). *Lattice Theory*. American Mathematical Society, Providence, Rhode Island.

Buchan, J. and E. Koenigsberg (1963). *Scientific Inventory Management*. Prentice-Hall, Englewood Cliffs, New Jersey.

Busacker, R. G. and T. L. Saaty (1965). *Finite Graphs and Networks: An Introduction with Applications*. McGraw Hill, New York.

Charnes, A., and W. W. Cooper (1961). *Management Models and Industrial Applications of Linear Programming, Vol*. 1. John Wiley & Sons, New York, 233-245, 274-282.

Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey.

Ford, Jr., L. R., and D. R. Fulkerson (1962). *Flows in Networks*. Princeton University Press, Princeton, New Jersey.

Gantmacher, F. R. (1959). *The Theory of Matrices* (English Translation by K. A. Hirsch of Russian-Language Book *Teoriya Matrits*), Vol. I. Chelsea, New York, 374 pp.

Hadley, G., and T. M. Whitin (1963). *Analysis of Inventory Systems*. Prentice-Hall, Englewood Cliffs, New Jersey.

Hardy, G. H., J. E. Littlewood, and G. Polya (1952). *Inequalities, 2nd Edition*. Cambridge University Press, Cambridge, England.

Holt, C. C., F. Modigliani, J. F. Muth, and H. A. Simon (1960). *Planning Production, Inventories and Work Force*. Prentice-Hall, Englewood Cliffs, New Jersey.

IBM (1972). *Basic Principles of Wholesale IMPACT—Inventory Management and Control Techniques,* 2nd ed., 52pp.

Johnson, L. A., and D. C. Montgomery (1974). *Operations Research in Production Planning, Scheduling, and Inventory Control*. John Wiley & Sons, New York.

Karlin, S. (1966). *A First Course in Stochastic Processes*. Academic Press, New York.

Karlin, S. (1968). *Total Positivity*. Stanford University Press, Stanford, California.

Koehler, G. J., A. B. Whinston, and G. P. Wright (1975). *Optimization over Leontief Substitution Systems*. North-Holland American-Elsevier, Amsterdam/New York.

Lawler, E. L. (1976). *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart and Winston, New York.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. John Wiley & Sons, New York.

Milgrom, P. R. and J. Roberts. (1992). *Economics, Organization, and Management.* Prentice-Hall. Englewood Cliffs, NJ.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Scarf, H. E., D. M. Gilford, and M. W. Shelly (eds.) (1963). *Multistage Inventory Models and Techniques.* Office of Naval Research Monograph on Mathematical Methods in Logistics, Stanford University Press, Stanford, California.

Tayur, S., R. Ganeshan and M. Magazine (eds.) (1999). *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers, Massachusetts.

Veinott, Jr., A. F. (ed.) (1965a). *Mathematical Studies in Management Science*. MacMillan, New York.

Veinott, Jr., A. F. (Forthcoming). *Lattice Programming: Comparison of Optima and Equilibria.* Johns Hopkins University Press, Baltimore, Maryland.

Wagner, H. M. (1969). *Principles of Operations Research*. Prentice Hall, Englewood Cliffs, New Jersey.

Zangwill, W. I. (1969). *Nonlinear Programming: A Unified Approach*. Prentice Hall, Englewood Cliffs, New Jersey, Chapter 2.

ARTICLES and DISSERTATIONS

Aggarwal, A., M. M. Klawe, S. Moran, P. Shor, and R. Wilbur (1987). Geometric Applications of a Matrix Searching Algorithm. *Algorithmica* 2, 195-208.

Aggarwal, A. and J. K. Park (1993). Improved Algorithms for Economic Lot Size Problems. *Operations Res.* 41, 549-571.

Bather, J. A. (1966). A Continuous Time Inventory Model. *J. Appl. Prob.* 3, 538-549.

Bellman, R. E., I. Glicksberg, and O. Gross (1955). On the Optimal Inventory Equation. *Management Sci.* 2, 83-104.

Berge, C. (1961). Les Problemes de Flot et de Tension. *Cahiers Centre Etudes Recherche Oper.* 3, 69-83.

Bessler, S. A., and A. F. Veinott, Jr. (1966). Optimal Policy for a Dynamic Multi-Echelon Inventory Model. *Naval Res. Logist. Quart.* 13, 355-389.

Brown, G. F., Jr., T. M. Corcoran and R. M. Lloyd (1971). Inventory Models with Forecasting and Dependent Demand. *Management Sci.* 11, 51-63.

Ciurria, I. (1989). *Substitutes and Complements in Multicommodity Flows and Inventories.* Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA, 141 pp.

Ciurria, I., F. Granot and A. F. Veinott, Jr. (1990a). *Substitutes, Complements and Ripples in Multicommodity Flows on Suspension Graphs.* Technical Report, Department of Operations Research, Stanford University, Stanford, CA, 29 pp.

Ciurria, I., F. Granot and A. F. Veinott, Jr. (1990b). *Substitutes, Complements and Ripples in Multicommodity Production and Inventories.* Technical Report, Department of Operations Research, Stanford University, Stanford, CA, 59 pp.

Clark, A. J. and H. E. Scarf (1960). Optimal Polices for a Multi-Echelon Inventory System. *Management Sci.* 6, 4, 475-490.

Dantzig, G. B. (1959). On the Status of Multistage Linear Programming Problems. *Management Sci.* 6, 53-72.

Debreu, G. and H. Scarf (1963). A Limit Theorem on the Core of an Economy. *International Econ. Rev.* 4, 236-246.

DeCroix, G. A. (1992). *Equilibrium Warranties and Prices.* Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA.

DeCroix, G. A. (1999). Optimal Warranties, Reliabilities and Prices for Durable Goods in an Oligopoly. *European J. Operational Res.* 112, 3, 554-569.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Math.* 1, 269-71.

Dreyfus, S. E. (1957). An Analytic Solution of the Warehouse Problem. *Management Sci.* 4, 1, 99-104.

Dreyfus, S. E. and R. A. Wagner (1971). The Steiner Problem in Graphs. *Networks* 1, 195-207.

Duffin, R. J. (1965). Topology of Series-Parallel Networks. *J. Math. Anal. Appl.* 10, 303-318.

Dzielinski, B. P. and R. E. Gomory (1965). Optimal Programming of Lot Sizes, Inventories and Labor Allocations. *Management Sci.* 11, 874-890.

Erickson, R. E., C. L. Monma and A. F. Veinott, Jr. (1987). Send-and-Split Method for Minimum-Concave-Cost Network Flows. *Math. Operations Res.* 12, 634-666.

Federgruen, A. and M. Tzur (1991). A Simple Forward Algorithm to Solve General Dynamic Lot Sizing Models with n Periods in $O(n \log n)$ or $O(n)$ Time. *Management Sci.* 37, 8, 909-925.

Fredman, M. L. and R. E. Tarjan (1984). Fibonacci Heaps and their Uses in Improved Network Optimization Algorithms. *Proc. 25th IEEE Symposium on the Foundations of Computer Science,* 1984, 338-346.

Fulkerson, D. R. (1961). An Out-of-Kilter Method for Minimal Cost Flow Problems. *J. Soc. Indust. Appl. Math.* 9, 18-27.

Gallego, G. and Ö. Özer (2001). *Integrating Replenishment Decisions with Advance Demand Information. Management Sci.* 47, 1344-1360.

Gale, D. and T. Politof (1981). Substitutes and Complements in Network Flow Problems. *Discrete Appl. Math.* 3, 175-186.

Granot, F. and A. F. Veinott, Jr. (1985). Substitutes, Complements and Ripples in Network Flows. *Math. Operations Res.* 10, 471-497.

Harris, F. W. (1913). How Many Parts to Make at Once. *Factory* 10, 135-136 and 152.

Hartung, P. (1973). A Simple Style Goods Inventory Model. *Management Sci.* 19, 1452-1458.

Hirsch, W. and A. J. Hoffman (1961). Extreme Varieties, Concave Functions, and the Fixed Charge Problem. *Comm. Pure Appl. Math.* 14, 355-369.

Hoffman, A. J. (1963). On Simple Linear Programs. In V. Klee (ed.), *Convexity.* Proc. Symp. Pure Math. No. 7, American Mathematical Society, Providence, RI, 317-327.

Hoffman, A. J. and A. F. Veinott, Jr. (1993). *Staircase Transportation Problems with Superadditive Rewards and Capacities. Math. Programming* 62, 199-213.

Hopcroft, J. and R. E. Tarjan (1973). Efficient Algorithms for Graph Manipulation. *Comm. ACM.* 16, 372-378.

Iglehart, D. L. and S. Karlin (1962). Optimal Policy for Dynamic Inventory Process with Nonstationary Stochastic Demand. Chapter 8 in [AKS62], 127-147.

Ignall, E. J. and A. F. Veinott, Jr. (1968). Substitution Properties of Network Flows (Abstract). *Bull. Inst. Management Sci.* 14, 167.

Ignall, E. J. and A. F. Veinott, Jr. (1969). Optimality of Myopic Inventory Policies for Several Substitute Products. *Management Sci.* 15, 284-304.

Johnson, E. L. (1967). Optimality and Computation of (σ,S) Policies in the Multi-Item Infinite Horizon Inventory Problem. *Management Sci.* 13, 475-491.

Johnson, S. M. (1957). Sequential Production Planning Over Time at Minimum Cost. *Management Sci.* 3, 4, 435-437.

Kalymon, B. (1972). A Decomposition Algorithm for Arborescence Inventory Systems. *Operations Res.* 20, 860-874.

Kamae, T., U. Krengel, and G. O'Brien (1977). Stochastic Inequalities on Partially Ordered Spaces. *Ann. Prob.* 6, 899-912.

Karlin, S. (1960). Dynamic Inventory Policy with Varying Stochastic Demands. *Management Sci.* 6, 231-258.

Karush, W. (1959). A Theorem in Convex Programming. *Naval Res. Logist. Quart.* 6, 245-260.

Karush, W. and A. Vazsonyi (1957a). Mathematical Programming and Service Scheduling. *Management Sci.* 3, 2, 140-148.

Karush, W. and A. Vazsonyi (1957b). Mathematical Programming and Employment Scheduling. *Naval Res. Logist. Quart.* 4, 297-320.

Klawe, M. M. (1989). *A Simple Linear Time Algorithm for Concave One-Dimensional Dynamic Programming.* Technical Report 89-16, University of British Columbia, Vancouver, BC., 4 pp.

Konno, H. (1973). Minimum Concave Cost Series Production Systems with Deterministic Demands—A Backlogging Case. *J. Operations Res. Soc. Japan* 16, 246-253.

Krishnarao, P. V. (1993). *Additivity of Minimum Cost in Dual Network Flow Problems.* Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA, 60 pp.

Love, S. (1972). A Facilities in Series Inventory Model with Nested Schedules. *Management Sci.* 18, 327-338.

Manne, A. (1958). Programming of Economic Lot Sizes. *Management Sci.* 4, 115-135.

Maxwell, W. L. and J. A. Muckstadt (1985). Establishing Consistent and Realistic Reorder Intervals in Production-Distribution Systems. *Operations Res.* 33, 1316-1341.

Mello, M. P. (1993). *Generalized Leontief Substitution Systems and Lattice Programming.* Ph.D. Dissertation, Department of Operations Research, Stanford University, Stanford, CA, 78 pp.

Milgrom. P. and J. Roberts (1990). Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities. *Econometrica* 58, 6 1255-1277.

Mitchell, J. S. B. (1987). 98%-Effective Lot-Sizing for One-Warehouse, Multi-Retailer Inventory Systems with Backlogging. *Operations Res.* 3, 399-404.

Modigliani, F. and F. Hohn (1955). Production Planning over Time and the Nature of the Expectation and Planning Horizon. *Econometrica* 23, 46-66.

Nash, J. F., Jr. (1951). Non-Cooperative Games. *Annals of Mathematics*, 54, 289-95.

Notzon, E. (1970). *Minimax Inventory and Queueing Models.* Technical Report No. 7, Department of Operations Research, Stanford University, Stanford, California, 60 pp.

Owen, G. (1975). On the CORE of Linear Production Games. *Math. Programming* 9, 358-370.

Porteus, E. L. (1971). On the Optimality of Generalized (s,S) Policies. *Management Sci.* 17, 411-426.

Porteus, E. L. (1990). *Stochastic Inventory Theory.* Chapter 12 in D. P. Heyman and M. J. Sobel (Eds.), *Handbooks in Operations Research and Management Science* 2, Elsevier Science Publishers B. V. (North-Holland) 2, 605-652.

Puterman, M. L. (1975). A Diffusion Process Model for a Storage System. In M. Geisler (ed.), *Logistics*, North-Holland-TIMS Studies in the Management Sciences 1, 143-159.

Roundy, R. (1985). 98%-Effective Integer-Ratio Lot-Sizing for One-Warehouse Multi-Retailer Systems. *Management Sci.* 31, 1416-1430.

Roundy, R. (1986). A 98%-Effective Lot-Sizing Rule for a Multi-Product, Multi-Stage Production/Inventory System. *Math. Operations Res.* 11, 699-727.

Scarf, H. E. (1959). Bayes Solutions of the Statistical Inventory Problem. *Ann. Math. Statist.* 30, 2, 490-508.

Scarf, H. E. (1960). On the Optimality of (S,s) Policies in the Dynamic Inventory Problem. Chapter 13 in K. J. Arrow, S. Karlin, and P. Suppes (eds.). *Mathematical Methods in the Social Sciences 1959.* Stanford University Press, Stanford, California.

Schäl, M. (1976). On the Optimality of (s,S) Policies in Dynamic Inventory Models with Finite Horizon. *SIAM J. Appl. Math.* 30, 528-537.

Shapley, L. S. (1961). On Network Flow Functions. *Naval Res. Logistics Quart.* 8, 151-158.

Shapley, L. S. (1962). Complements and Substitutes in the Optimal Assignment Problem. *Naval Res. Logist. Quart.* 9, 45-48.

Strassen, V. (1965). The Existence of Probability Measures with Given Marginals. *Ann. Math. Statist.* 36, 423-439.

Tarjan, R. E. (1972). Depth-First Search and Linear Graph Algorithms. *SIAM J. Comput.* 1, 146-160.

Tarski, A. (1955). A Lattice Theoretical Fixpoint Theorem and Applications. *Pac. J. Math* 5, 285-309.

Topkis, D. M. (1970). *Equilibrium Points in Nonzero-Sum n-person Games,* Operations Research Center Tech. Report ORC-70-38, University of California, Berkeley, California.

Topkis, D. M. (1976). The Structure of Sublattices of the Product of n Lattices. *Pac. J. Math.* 65, 525-532.

Topkis, D. M. (1978). Minimizing a Submodular Function on a Lattice. *Operations Res.* 26, 305-321.

Topkis, D. M. and A. F. Veinott, Jr. (1968). *Minimizing a Subadditive Function on a Lattice.* Included as §1 of Chapter III of D. M. Topkis (1968). *Ordered Optimal Solutions.* Ph.D. Dissertation, Stanford University, Stanford, CA, 40-55.

Topkis, D. M. and A. F. Veinott, Jr. (1973a). Monotone Solutions of Extremal Problems on Lattices. *VIII Inter. Symp. Math. Programming,* Abstracts. Stanford University, Stanford, California, 131.

Topkis, D. M. and A. F. Veinott, Jr. (1973b). Meet-representation of Subsemilattices and Sublattices of Product Spaces. *VIII Inter. Symp. Math Programming,* Abstracts. Stanford University, Stanford, California, 136.

Tsay, A. S., S. Nahmias and N. Agrawal (1999). Modeling Supply Chain Contracts: A Review. Chapter 10 in S. Tayur, R. Ganeshan and M. Magazine (Eds.) *Quantitative Models for Supply Chain Management.* Kluwer Academic Publishers, MA., pp. 299-336.

Veinott, Jr., A. F. (1963). Optimal Stockage Policies with Non-Stationary Stochastic Demands. Chapter 4 in [SGS63], 85-115.

Veinott, Jr. A. F. (1964). Production Planning with Convex Costs: A Parametric Study. *Management Sci.* 10, 3, 441-460

Veinott, Jr., A. F. (1965b). Untitled and unpublished manuscript on monotone solutions of extremal problems on sublattices.

Veinott, Jr., A. F. (1965c). Optimal Inventory Policy for Batch Ordering. *Operations Res.* 13, 3, 424-432.

Veinott, Jr., A. F. (1965d). Optimal Policy for a Multi-Product, Dynamic Non-Stationary Inventory Problem. *Management Sci.* 12, 3, 206-222.

Veinott, Jr., A. F. (1965e). Optimal Policy in a Dynamic, Single Product, Non-Stationary Inventory Model with Several Demand Classes. *Operations Res.* 13, 5, 761-778.

Veinott, Jr., A. F. (1966a). On the Optimality of (s,S) Inventory Policies: New Conditions and a New Proof. *SIAM J. Appl. Math.* 14, 1067-1083.

Veinott, Jr., A. F. (1966b). The Status of Mathematical Inventory Theory. *Management Sci.* 12, 745-777.

Veinott, Jr., A. F. (1968a). Extreme Points of Leontief Substitution Systems. *Linear Algebra and Appl.* 1, 181-194.

Veinott, Jr., A. F. (1968b). *Substitutes and Complements in Network Flows.* Presentation to TIMS/ ORSA Joint Meeting, San Francisco, California.

Veinott, Jr., A. F. (1969). Minimum Concave Cost Solution of Leontief Substitution Models of Multi-Facility Inventory Systems. *Operations Res.* 17, 262291.

Veinott, Jr., A. F. (1971). Least d-Majorized Network Flows with Inventory and Statistical Applications. *Management Sci.* 17, 547-567.

Veinott, Jr., A. F. (1974). *Existence and Isotonicity of Equilibrium Points in Noncooperative n-Person Games*, in A. F. Veinott, Jr. (1974). *Lecture Notes in Dynamic and Lattice Programming*, Operations Research 381, Dept. Operations Research, Stanford, California, pp. 69-71,

Veinott, Jr., A. F. (1975). Lattice Programming: Substitutes and Complements in Network Flows. Abstract of presentation to ORSA/TIMS National Meeting, Chicago, IL. *Bull. Operations Res. Soc. Amer. 23, Supplement 1, B*-131.

Veinott, Jr., A. F. (1985). Existence and Characterization of Minima of Concave Cost Functions on Unbounded Convex Sets. *Math. Programming Study* 25, 88-92.

Veinott, Jr., A. F. (1987). *Inventory Policy Under Certainty.* In J. Eatwell, M. Milgate, III, and P. K. Newman (eds.). *The New Palgrave: A Dictionary of Economics* 2, E-J. The Stockton Press, New York, New York, 975-980.

Veinott, Jr., A. F. (1989a). Conjugate Duality for Convex Programs. A Geometric Development. *Linear Algebra Appl.* 114/115, 663-667.

Veinott, Jr., A. F. (1989b). Representation of General and Polyhedral Subsemilattices and Sublattices of Product Spaces. *Linear Algebra Appl.* 114/115, 681-704.

Wagelmans, A., S. Van Hoesel and A. Kolen (1989). *Economic Lot-Sizing: An $O(n \log n)$ Algorithm That Runs in Linear Time in the Wagner-Whitin Case.* CORE Discussion Paper No. 8922, Université Catholique de Louvain, Belgium.

Wagner, H. M. and T. M. Whitin (1958). Dynamic Version of the Economic Lot Size Model. *Management Sci.* 5, 89-96.

Wagner, H. M. (1959). On a Class of Capacitated Transportation Problems. *Management Sci.* 5, 304-318.

Whitney, H. (1932). Non-separable and Planar Graphs. *Trans. Amer. Math. Soc.* 34, 339-362.

Zangwill, W. I. (1968). Minimum Concave-Cost Flows in Certain Networks. *Management Sci.* 14, 429-450.

Zangwill, W. I. (1969). A Backlogging Model and a Multi-Echelon Model of a Dynamic Lot Size Production System—A Network Approach. *Management Sci.* 15, 506-527.

Zheng, Y. S. and A. Federgruen (1991). Finding Optimal (s, S) Policies is About as Simple as Evaluating a Single Policy. *Operations Res.* 39, 4, 654-665.

Homework 1 Due October 7

1. Leontief Substitution Systems. Call an $m \times n$ matrix A *pre-Leontief* if each of its columns has at most one positive element, and *Leontief* if also there is a nonnegative column n-vector $x = (x_i)$ for which Ax is positive. Call two distinct variables x_i and x_j *substitutes* if the i^{th} and j^{th} columns of A have positive elements in the same position.

(a) Extreme Points. If b is a nonnegative column m-vector and A is an $m \times n$ pre-Leontief constraint matrix, then each extreme point x of the *pre-Leontief substitution system* $Ax = b$, $x \geq 0$, has the properties that $x_i x_j = 0$ for each pair (x_i, x_j) of substitute variables and $x_k = 0$ for each x_k for which the k^{th} column of A is nonpositive. Prove this fact where b is positive. [*Hint*: Since extreme points coincide with basic feasible solutions, observe that each extreme point of the system has exactly m positive variables. Moreover, if an extreme point has two positive substitute variables, then some row of the basis has no positive elements.]

(b) Square Leontief Matrices. Show that a square pre-Leontief matrix is Leontief if and only if it is nonsingular and its inverse is nonnegative. [*Hint*: Let A be a square pre-Leontief matrix. Since the class of square Leontief matrices is closed under multiplication by positive scalars and permutation of their columns, it is enough to prove the result for the case that the off-diagonal elements of A are nonpositive and the diagonal elements of A do not exceed one, so $P \equiv I - A \geq 0$. By assumption, there is an $x \geq 0$ for which $b \equiv Ax$ is positive. Thus, $x = b + Px$. Now iterate this equation and conclude that $\sum_{i=0}^{\infty} P^i$ is finite and is the desired inverse of A.]

(c) Feasibility of Leontief Substitution System. Call a pre-Leontief substitution system *Leontief* if its constraint matrix is Leontief. Show that a Leontief substitution system has a Leontief basis, and each such basis is feasible for all nonnegative right-hand sides.

(d) Optimality in Leontief Substitution Systems. Show that if a linear function attains its maximum over the set of feasible solutions of a Leontief substitution system for some positive right-hand side, then there is a Leontief basis that is simultaneously optimal for all nonnegative right-hand sides.

2. Supply Chains with Linear Costs. Consider a collection of *facilities*, labeled $1, \ldots, N$, each producing a single product. Production of each unit at facility j directly consumes $e^{ij} \geq 0$ units of the output of facility i. The time lags in shipments between facilities are negligible, i.e., production at facility j in a period consumes output at other facilities that may be produced at those facilities either before or during the period. There is a given exogenous nonnegative demand s_t^j in period $t = 1, \ldots, T$ for the output of facility j. The demands at each facility in each period are met as they occur. There is a unit cost c_t^j (resp., h_t^j) of producing (resp., storing) each unit at facility j in (resp., at the end of) period t. The unit cost of production at facility j in a period also includes the costs of transporting e^{ij} units of the output of each facility i to facility j in that period. Let x_t^j and y_t^j be the respective amounts produced in and stored at the end of period t at facility j. Let $x_t = (x_t^j)$, $y_t = (y_t^j)$ and $s_t = (s_t^j)$ be respectively the N-element column vectors of production, inventory and demand schedules in period t. Assume that $y_0 \equiv y_T \equiv 0$. The net production at each facility that is available to satisfy exogenous demands in, or to store at the end of, period t is $(I - E)x_t$ where $E \equiv (e^{ij})$. The problem is to find a *schedule* $(x_t, y_t) \geq 0$ that minimizes the *total cost*

$$\sum_{j,t} (c_t^j x_t^j + h_t^j y_t^j) \tag{1}$$

subject to the *stock-conservation constraints*

$$(I - E)x_t + y_{t-1} - y_t = s_t,\ t = 1, \ldots, T. \tag{2}$$

(a) Pre-Leontief. Show that the constraint matrix for the system of equations (2) is pre-Leontief.

(b) Leontief. Show that if $I - E$ is Leontief, then the constraint matrix for the system of equations (2) is Leontief and the set of nonnegative solutions of (2) is bounded.

(c) Upper Triangularity. Show that if E is upper triangular with zero diagonal elements, then $I - E$ is Leontief.

(d) Circuitless Supply Graphs. A *supply graph* is a directed graph whose nodes are the facilities and whose arcs are the ordered pairs $i \to j$ of facilities for which $e^{ij} > 0$. A *simple circuit* in a supply graph is a sequence $i_1, \ldots, i_k$ of distinct facilities for which $i_j \to i_{j+1}$ is an arc for $j = 1, \ldots, k\ (\leq N)$ where $i_{k+1} \equiv i_1$, i.e., each facility in the sequence "consumes" the output of its immediate predecessor in the sequence. A supply graph is *circuitless* if it has no simple circuit. Show that a supply graph is circuitless if and only if it is possible to relabel the facilities so that E is upper triangular with zero diagonal elements. Show by example that circuitless supply graphs encompass assembly and distribution systems.

(e) Extreme Schedules. Show that each *extreme schedule*, i.e., extreme point of the set of nonnegative solutions of (2), satisfies

$$y_{t-1}^j x_t^j = 0 \text{ for } 1 \leq t \leq T \text{ and } 1 \leq j \leq N. \tag{3}$$

Interpret the equations (3).

(f) Dynamic-Programming Equations with $I - E$ Leontief. Suppose that $I - E$ is Leontief and let C_t^j be the minimum cost of satisfying a unit of demand at facility j in period t. Put $c_t = (c_t^j)$, $h_t = (h_t^j)$ and $C_t = (C_t^j)$. Give a dynamic-programming recursion that $C = (C_1, \ldots, C_T)$ satisfies. Show that C is optimal for the dual of the linear program (1), (2). [*Hint*: Show that the dual program is feasible.]

(g) Running Time in a Circuitless Supply Graph. Show that if a supply graph is circuitless, then it is possible to compute C recursively in $O(TN^2)$ time.

(h) Running Time with One-Period Lag. Suppose that there is a one-period lag in delivery of goods from one facility to another. Also assume that production x_1 of the N products in period one represents exogenous procurement in that period with no delay in delivery. Give the appropriate modification of the equations (2). Also show how to compute the corresponding vector C in $O(TN^2)$ time (even if $I - E$ is not Leontief).

Answers to Homework 1 Due October 7

1. Leontief Substitution Systems.

(a) Extreme Points. Suppose $\overline{x}$ is an extreme point of the pre-Leontief substitution system $Ax = b, x \geq 0$ in which A is pre-Leontief and b is positive. Let B be the k, say, columns of A corresponding to positive elements of $\overline{x}$ and $\overline{x}_B$ be the subvector of positive elements of $\overline{x}$. Since extreme points are basic feasible solutions, $k \leq m$. Now B must have at least one positive element in every row. For if not, the i^{th} row B_i, say, of B is nonpositive, whence $0 \geq B_i\overline{x}_B = b_i > 0$, which is impossible. In fact, B has exactly one positive element in each row, whence $\overline{x}_i\overline{x}_j = 0$ for each pair $(\overline{x}_i, \overline{x}_j)$ of substitute variables, and exactly one positive element in each column, whence $\overline{x}_k = 0$ for each $\overline{x}_k$ whose column of coefficients is nonpositive. For if some row of B has at least two positive elements or if some column of B has no positive elements, then because B has at most one positive element in each column, $k > m$ which is impossible.

(b) Square Leontief Matrices. Let A be a square pre-Leontief matrix. Since the class of square Leontief matrices is closed under multiplication by positive scalars and permutation of their columns, it is enough to prove the result for the case that the off-diagonal elements of A are nonpositive and the diagonal elements of A do not exceed one, so $P \equiv I - A \geq 0$.

If A is nonsingular and has a nonnegative inverse, then $x \equiv A^{-1}1 \geq 0$, whence $Ax = 1 \gg 0$, so A is Leontief. Conversely, if A is Leontief, there exists an $x \geq 0$ such that $b \equiv Ax$ is positive, so $x = b + Px$. Iterating the last equation yields

$$x = (\sum_{n=0}^{N} P^n)b + P^{N+1}x \geq (\sum_{n=0}^{N} P^n)b$$

for all $N \geq 0$ since $x \geq 0$ and $P \geq 0$. Thus $(\sum_{n=0}^{N} P^n)b$ is uniformly bounded above by x. Since $(\sum_{n=0}^{N} P^n)b$ is also nondecreasing in N, it must converge, implying that P^n tends to the null matrix as n tends to infinity. Now $(I - P)(\sum_{n=0}^{N} P^n) = I - P^{N+1}$. Also, as N tends to infinity, the right-hand side of this equation tends to the identity matrix. Therefore $A = I - P$ is nonsingular and $A^{-1} = \sum_{n=0}^{\infty} P^n \geq 0$.

(c) Feasibility of Leontief Substitution Systems. Suppose A is Leontief. Then there exists $\overline{x} \geq 0$ such that $\overline{b} \equiv A\overline{x} \gg 0$. Hence there is a basis B and basic variables $\overline{x}_B \gg 0$ such that $B\overline{x}_B = \overline{b} \gg 0$. Also B is Leontief and, from (a), B is square. Hence, from (b), B is nonsingular and has a nonnegative inverse. Also, the basis B is feasible for all nonnegative right-hand sides b because $x_B = B^{-1}b \geq 0$.

(d) Optimality of Leontief Substitution Systems. If the linear function cx attains its maximum over the Leontief substitution system $Ax = \overline{b}$, $x \geq 0$, where $\overline{b} \gg 0$, then, as shown above, there is a square Leontief basis B that is optimal, and associated basic optimal solutions $\overline{x}$ and $\overline{\pi}$ of the primal and dual. The basis B remains optimal for all nonnegative right-hand sides b since the corresponding basic variables $x_B = B^{-1}b$ are nonnegative, $\overline{\pi} = c_B B^{-1}$ is feasible for the dual for $\overline{b}$ and hence also for b, and complementary slackness is maintained.

2. Supply Chains with Linear Costs.

(a) Pre-Leontief. The constraint matrix A takes the form below.

$$
\begin{array}{|cccc|cccc|}
\hline
x_1 & x_2 & \cdots & x_T & y_1 & y_2 & \cdots & y_{T-1} \\
\hline
I-E & & & & -I & & & \\
 & I-E & & & I & -I & & \\
 & & \ddots & & & & \ddots & \\
 & & & & & & & -I \\
 & & & I-E & & & & I \\
\hline
\end{array}
$$

In the column corresponding to x_t^j, all elements are nonpositive, except for $1 - e_{jj}$ which may be of either sign. In the column corresponding to y_t^j, there is exactly one positive element, viz., $+1$, and exactly one negative element, viz., -1. Since each of its columns has at most one positive element, A is pre-Leontief.

(b) Leontief. Suppose $I - E$ is Leontief. Then there is a nonnegative column N-vector ξ for which $\sigma \equiv (I - E)\xi \gg 0$. Consider the column NT-vector $\overline{x}$ defined by $\overline{x}_t = \xi$ for all t, and let $\overline{y}$ be the column $N(T{-}1)$ null vector. Observe that $(\overline{x}, \overline{y})$ is nonnegative and feasible for (2) with σ replacing s_t for each t, so A is Leontief.

To see that the set of nonnegative solutions to (2) is bounded, premultiply (2) by the nonnegative matrix $(I - E)^{-1}$ and sum over all periods yielding $0 \le \sum_{t=1}^{T} x_t = \sum_{t=1}^{T} (I - E)^{-1} s_t \ll \infty$. Since the x_t are nonnegative, they are bounded. Thus, $y_{T-1} = s_T - (I - E)x_T$ is bounded. Similarly, $y_{T-2} = s_{T-1} - (I - E)x_{T-1} + y_{T-1}$ is bounded, etc.

Since the set of feasible solutions is bounded, it follows from 1(d) and the fact that A is Leontief that there is a Leontief basis B that is optimal for all $s \ge 0$. Moreover, since B has a nonnegative inverse from 1(b), it follows that the optimal production and inventory levels corresponding to B are linear and nondecreasing in $s \ge 0$, and that the periods in which it is optimal to produce a product are independent of $s \ge 0$.

(c) Upper Triangularity. Suppose E is upper triangular with zero diagonal elements. Clearly $I - E$ is pre-Leontief. To show that $I - E$ is Leontief, define the nonnegative N-vector ξ recursively by $\xi_j = 1 + \sum_{i=j+1}^{N} e^{ji} \xi_i$ for $j \ge 1$. Since $(I - E)\xi = 1 \gg 0$, $I - E$ is Leontief as claimed.

(d) Circuitless Supply Graphs. Suppose that the supply graph is circuitless. Initially, all nodes are unlabeled. Choose an unlabeled facility j_1, then an unlabeled facility j_2 for which $j_2 \to j_1$ is an arc, then an unlabeled facility j_3 for which $j_3 \to j_2$ is an arc, etc. Proceed in this way until an unlabeled facility j_k is found for which there is no unlabeled facility i for which $i \to j_k$ is an arc. Give the facility j_k a label equal to the cardinality of the set of labeled nodes plus one. Repeat this process until all nodes are labeled. Under this labeling, the resulting graph has the property that $e^{ij} > 0$ only if $i < j$. Thus, E is upper triangular with zero diagonal elements as claimed. Conversely, if E is upper triangular and one labels its rows $1, \ldots, N$, the supply graph is circuitless.

Consider a manufacturer that has s suppliers and several r retail outlets. Label the suppliers $1,\ldots,s$, the manufacturer $s+1$, and the retailers $s+2,\ldots,s+1+r$. This supply graph is one with simple assembly and distribution, and it is circuitless.

(e) Extreme Schedules. Consider the column of the constraint matrix corresponding to x_t^j. If this column is nonpositive, i.e., $1-e_{jj}\le 0$, then $x_t^j=0$ by 1(a). In the contrary event, the column has exactly one positive element, i.e., $1-e_{jj}>0$, so x_t^j and y_{t-1}^j are substitute variables. In either event, $x_t^j y_{t-1}^j=0$ by 1(a). Hence, in any extreme schedule—including any optimal extreme schedule—one produces a product only in a period in which there is no entering inventory of the product.

(f) Dynamic Programming Equations with $\boldsymbol{I-E}$ Leontief. The dual program is that of choosing C that maximizes

$$\sum_{t=1}^{T} C_t s_t$$

subject to

$$C_t \le c_t + C_t E \qquad \text{(supply from current production)}$$

and

$$C_t \le h_{t-1} + C_{t-1} \qquad \text{(supply from storage)}$$

for $t=1,\ldots,T$ where $h_0\equiv 0$ and $C_0\equiv 0$. Now from (b), the primal is feasible its feasible set is bounded. Thus, from linear programming theory and 1(d), there is a Leontief optimal basis, and that basis is optimal for all $s\ge 0$. Consequently, the corresponding basic optimal solution C of the dual is independent of $s\ge 0$, so C_t^j is the minimum unit cost of satisfying demands for product j in period t. Moreover, since production and storage of a product in a period are substitute variables and the basis associated with an extreme point of a Leontief substitution system does not include columns corresponding to substitute variables, it follows from complementary slackness that C satisfies the dynamic-programming equations

$$C_t = \min\left(c_t + C_t E, h_{t-1} + C_{t-1}\right)$$

for $t=1,\ldots,T$. Thus, one may compute $C_1,\ldots,C_T$ recursively in that order. Since the work to find C_t given C_{t-1} depends on N, but not on t, it is possible to compute C in $O(T)$ time with N fixed.[1] Incidentally, notice that for the case of a single product that does not consume itself, this equation specializes to the recursion (4) on page 5 of *Lectures in Supply-Chain Optimization.*

[1]This result can be sharpened to show in fact that the overall running time is at most $O(TN^3)$. To see this, observe that the above equation for fixed t is the dynamic-programming equation for a stopping problem. Given C_{t-1}, Initiate the simplex method on the primal program for period t starting with the basis that produces no product in period t, i.e., stores every product from period $t-1$. Show that the simplex method requires at most N iterations to find C_t because once production of a product in period t is introduced by the simplex method, that activity is never removed from the basis and so is optimal. This finds C_t in $O(N^3)$ time.

The interpretation of the above equation is that the minimum unit cost C_t^j of supplying a unit of product j in period t is the cheaper of two options, viz., (i) producing product j in period t or (ii) supplying product j in period $t-1$ at minimum cost and storing it until period t. The unit cost of the first option is the sum of the direct unit production cost c_t^j and the minimum indirect unit production cost $\sum_{i=1}^N e^{ij}C_t^i$ of producing the goods consumed in making a unit of product j in period t. The cost of the second option is the sum of the minimum unit cost C_{t-1}^j of supplying product j in period $t-1$ and the unit cost h_{t-1}^j of storing product j to period t.

(g) Running Time with E Upper Triangular. Since $I-E$ is upper triangular with zero diagonal elements, the above dynamic-programming equation reduces to

$$C_t^j = \min\,(c_t^j + \textstyle\sum_{i=1}^{j-1} e^{ij}C_t^i, h_{t-1}^j + C_{t-1}^j)$$

for $j = 1, \ldots, N$ and $t = 1, \ldots, T$. Given C_{t-1}, one may compute $C_t^1, \ldots, C_t^N$ in that order from this recursion. For a fixed pair (j,t), computing C_t^j requires $O(N)$ additions and multiplications, and one comparison. Since there are TN such pairs (j,t), the C_t can all be computed in $O(TN^2)$ time.

(h) Running Time with One-Period Lag. When there is a one-period lag in delivery, equation (2) must be modified so

$$x_t - Ex_{t+1} + y_{t-1} - y_t = s_t$$

for $t = 1, \ldots, T$ where $x_{T+1} = y_0 = y_T = 0$. The resulting constraint matrix is given below.

x_1	x_2	$\cdots$	x_T	y_1	y_2	$\cdots$	y_{T-1}
I	$-E$			$-I$			
	I			I	$-I$		
		$\ddots$				$\ddots$	
			$-E$				$-I$
			I				I

An argument like that in (a) and (b) shows that this system is Leontief. The dynamic-programming equations for the C_t then reduce to the recursion

$$C_t = \min\,(c_t + C_{t-1}E,\, h_{t-1} + C_{t-1})$$

for $t = 1, \ldots, T$. Given t and C_{t-1}, computing C_t requires $O(N^2)$ multiplications and additions, and $O(N)$ comparisons. Since there are T values of t, the total time to compute the C_t is $O(TN^2)$.

Homework 2 Due October 14

1. Assembly Supply Chains With Convex Costs. Consider the problem of choosing production schedules at each of N facilities, labeled $1, \ldots, N$ and structured as an assembly system, to minimize total production and storage costs incurred over n periods. Facility $1 \le k < N$ manufactures a single (intermediate) product that is consumed in making the product manufactured at it's unique *follower*, i.e., facility f_k with $k < f_k \le N$. Facility N assembles the final product that is used to satisfy the given cumulative (in time) nonnegative sales schedule $S = (S_1, \ldots, S_n)$ for assembled product with S_i being increasing in $1 \le i \le n$. There are no time lags in manufacture of a product or its delivery to the follower. Stock produced at a facility is retained there until it is consumed at its follower. Each product is measured in units of final assembled product, so production of one unit of product at facility k in a period consumes one unit of product in the period from each of its *predecessors*, i.e., facilities in the set $f_k^{-1} \equiv \{j : f_j = k\}$ whose follower is k.

Backorders, lost sales and negative production are not permitted. There is no initial or final inventory. There is an upper bound U_i^k on the cumulative (in time) production at each facility k in each period i. Put $U^k = (U_i^k)$. There is a feasible schedule, i.e., $S \le U^k$ for $1 \le k \le N$.

The cost of producing (resp., storing) z units at facility k in (resp., at the end of) period i is $c_i^k(z)$ (resp., $h_i^k(z)$) with $c_i^k(\,\cdot\,)$ (resp., $h_i^k(\,\cdot\,)$) being continuous and convex in $z \ge 0$. The problem is to choose cumulative (in time) production schedules $X^k = (X_i^k)$ at each facility $1 \le k \le N$ that minimize the total cost.

(a) Monotonicity. Show that there is an optimal production matrix $X = (X^1, \ldots, X^N)$ that is increasing in the vector S. Show that delaying an increase in final sales of assemblies has the effect of reducing the amount by which optimal cumulative production increases in each period at each facility. Explain why these two results remain valid for the case where production at each facility k in each period must be a multiple of a fixed batch size $Q_k > 0$.

(b) Decomposition. There is no loss in generality in assuming that $U^k \ge U^{f_k}$ for $1 \le k < N$. Assume also that $c_i^k(\,\cdot\,) = a_k c_i(\,\cdot\,)$ for some positive number a_k for $1 \le i \le n$ and $1 \le k \le N$. In addition, assume that $h_i^k(z) = h^k z$ in each period i with $h^k > 0$ at each facility k. Further, assume that the *incremental unit storage cost* $h^k - \sum_{j \in f_k^{-1}} h^j$ at facility k equals $a_k h_k$ for some $h_k > 0$. The constant a_k can be thought of as an index of the value added to the product at facility k. Show that there is a least production schedule $X^k = \underline{X}^k$ for facility $1 \le k \le N$ that minimizes ($X_0^k \equiv 0$)

$$\sum_{i=1}^{n} \left[c_i(X_i^k - X_{i-1}^k) + h_k X_i^k\right]$$

subject to

$$X_i^k \geq X_{i-1}^k \text{ for } i = 1, \ldots, n, \; U^k \geq X^k \geq S \text{ and } X_n^k = S_n.$$

Show that if also $h_k \leq h_{f_k}$ for all $1 \leq k < N$, then the production matrix $\underline{X} = (\underline{X}^1, \ldots, \underline{X}^N)$ is optimal for the N-facility problem. [*Hint*: Show that $\underline{X}^k$ is increasing in (h_k^{-1}, U^k), and so $\underline{X}^k \geq \underline{X}^{f_k}$ for $1 \leq k \leq N$, where $f_N \equiv N+1$ and $\underline{X}^{N+1} \equiv S$.]

2. Multiple Assignment Problem with Subadditive Costs. Let p be a real $n \times m$ matrix whose ij^{th} element is p_i^j and whose i^{th} row is p_i. Let S_j be the set of elements in the j^{th} column of p and $S \equiv \times_{j=1}^m S_j$. Let c be a real-valued function on S. The *multiple assignment problem* is to choose an $n \times m$ matrix π that minimizes $C(\pi) \equiv \sum_{i=1}^n c(\pi_i)$ (the *cost* of π) among those for which each column of π is a (possibly different) permutation of the elements of the corresponding column of p. The special case in which $m = 2$ is a form of the *assignment problem.*

(a) Subadditive Costs. Show that if c is subadditive on S, then one optimal choice of π is *increasing*, i.e., the i^{th} row π_i of π is increasing in i. [*Hint*: If π is not increasing, then there exist $i < j$ for which $\pi_i \not\leq \pi_j$. Show that replacing the i^{th} and j^{th} rows of π respectively by $\pi_i \wedge \pi_j$ and $\pi_i \vee \pi_j$ produces a matrix π' with lower cost.]

(b) Superadditive Costs. Establish an analog of (a) where $m = 2$ and c is superadditive on S.

(c) Order of Issuing. Consider a stockpile of n items of initial ages $a_1 \leq \cdots \leq a_n$. It is possible to issue the items in any order to satisfy unit demands for the product occurring at times $t_1 \leq \cdots \leq t_n$. The income from issuing an item of age a to satisfy a unit demand at a given time is $I(a)$ where I is a real-valued function on $\Re$. Determine the order of issuing items that maximizes the total income from the stockpile where I is convex. Also do this where I is concave.

(d) Order of Selling Securities. An investor owns n shares of a security bought previously at prices $b_1 \leq \cdots \leq b_n$ respectively. He plans to sell one share in each of the next n years $1, \ldots, n$ and expects the tax rates in those years to be $t_1, \ldots, t_n$. In what order should he sell the shares so as to minimize his total taxes over the n years? (Assume that if he sells the j^{th} share in year i, he pays $t_i(I_i - b_j)$ in taxes that year where I_i is the given taxable income in the year.)

(e) Order of Introducing Energy Technologies. Suppose that n new technologies for producing energy, labeled $1, \ldots, n$, are available and plans call for introducing exactly one new technology in each of the next n decades. The cost of introducing technology i in decade j is c_{ij}. Assume that the cost $c_{i,j+1} - c_{ij}$ of deferring the introduction of technology i in decade j for one more decade diminishes with i for each j. What order of introducing technologies minimizes total cost?

3. Representation of Additive Functions. Justify the representation of real-valued additive functions on a finite product of chains given in Theorem 7 of §2.4. [*Hint*: Prove the "only if" part by induction on n. Let f be real-valued and additive on the product of the chains $S_1, \ldots, S_n$. For $n = 2$, fix $\sigma = (\sigma_1, \sigma_2) \in S_1 \times S_2$ and observe that since f is additive,

$$f(s_1, s_2) = f(s_1, \sigma_2) + f(\sigma_1, s_2) - f(\sigma_1, \sigma_2) \text{ for all } (s_1, s_2) \in S_1 \times S_2.]$$

4. Projections of Additive Convex Functions are Additive Convex.

(a) Additivity of Minimum Cost. Suppose f is a real-valued additive convex function on $\Re^n$ and let $F(a, b)$ be the minimum value of

$$f(x)$$

subject to

$$a \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq b$$

for each pair of real numbers $a \leq b$ where $x = (x_i)$. Show that there exist real-valued convex functions g and h on the real line with g being increasing and h being decreasing thereon for which $F(a, b) = g(a) + h(b)$ for each $a \leq b$. Thus conclude that F is additive and convex on the sublattice $L \equiv \{(a, b) \in \Re^2 : a \leq b\}$. [*Hint*: Prove the result by induction on n.]

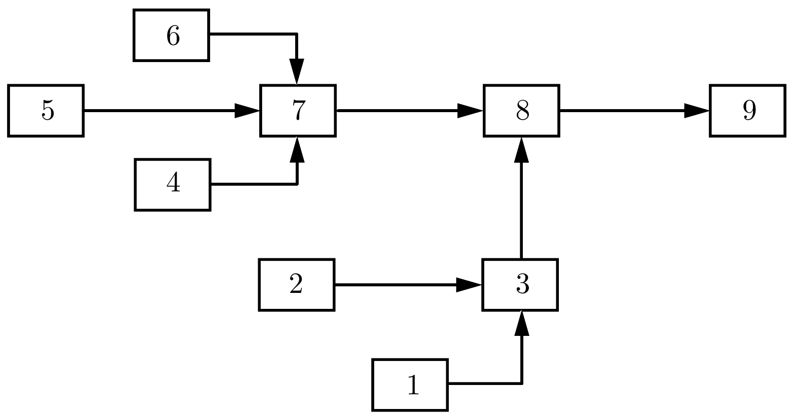
Remark. This important result will be used later in the course to decompose the problem of optimizing serial-supply-chains with uncertain demands into a seqeunce of single-facility problems.

(b) Rent-a-Car Fleet Expansion. A rent-a-car fleet manager wants to gradually expand the firm's fleet of cars over the next n days to minimize the total cost during that interval. Let $x_i \geq 0$ ($x_i \geq x_{i-1}$) be the number of cars in the fleet on day i. Let $a_i > 0$ be the unit cost of adding cars to the fleet, $m_i > 0$ be the unit cost of maintaining cars in the fleet and $r_i(x_i)$ be the revenue the firm earns by pricing the cars to rent the entire fleet x_i of cars on day i with the r_i being concave. Cars rented on a day are returned in time to be rented the following day. Let $\mathcal{C}(x_1, x_k, x_n)$ be the minimum cost over n days given that there are x_1 cars the first day, the fleet size must be x_k cars on a fixed day k ($1 \leq k \leq n$), and the terminal fleet size must be x_n cars on day n. Discuss whether or not $\mathcal{C}$ is additive on $L \equiv \{(x_1, x_k, x_n) \in \Re^3 : 0 \leq x_1 \leq x_k \leq x_n\}$. Describe how the incremental minimum cost of increasing the target car fleet size on day k by $p\%$ (≥ 0) depends on the initial and terminal car fleet sizes on days one and n assuming that $(1 + \frac{p}{100})x_k \leq x_n$.

Answers to Homework 2 Due October 14

1. Assembly Supply Chains with Convex Costs.

(a) Monotonicity. Let X_i^k be the cumulative production in periods $1,\ldots,i$ at facility k for $1 \le k \le N$ and $1 \le i \le n$. Interpret $X_i^{N+1} \equiv S_i$ as the cumulative sales in periods $1,\ldots,i$ at a dummy facility $N+1$ for $1 \le i \le n$. The system is illustrated in the figure below for the case $N = 8$. Let $c_i^k(z)$ and $h_i^k(z)$ be the (continuous convex) costs respectively of producing and stor-



ing $z \ge 0$ units at facility $1 \le k \le N$ in period $1 \le i \le n$. Put $X^k = (X_i^k)$ for $1 \le i \le n$ and $1 \le k \le N{+}1$, $X \equiv (X^k)$ for $1 \le k \le N$, and $S \equiv (S_i)$ for $1 \le i \le n$. The problem is to choose X that minimizes ($f_N \equiv N+1$ and $X_0^k \equiv 0$ for $0 \le k \le N+1$)

$$C(X,S) \equiv \sum_{k=1}^{N}\sum_{i=1}^{n}[c_i^k(X_i^k - X_{i-1}^k) + h_i^k(X_i^k - X_i^{f_k})] \tag{1}$$

subject to

$$X_i^k \ge X_{i-1}^k,\ 1 \le i \le n \text{ and } 1 \le k \le N, \tag{2}$$

$$U^k \ge X^k \ge X^{f_k} \text{ for } 1 \le k \le N, \text{ and} \tag{3}$$

$$X_n^k = S_n \text{ for } 1 \le k \le N. \tag{4}$$

Now the set L of pairs (X,S) satisfying (2)-(4) is a closed polyhedral sublattice of the set of all $n(N{+}1)$-tuples of real numbers by Example 6 of §2. Also, the section L_S of L at S is nonempty (e.g., one can put $X^k = S$ for $1 \le k \le N$) and bounded for each S in the set $\mathcal{S}$ of S that satisfy $0 \le S_1 \le \cdots \le S_n$ and $S \le U^k$ for $1 \le k \le N$ by (3) and (4). Thus since a convex function of the difference of two variables is subadditive and sums of subadditive functions are subadditive, it follows from (1) that $C(X,S)$ is finite, subadditive and continuous in (X,S) on L.

Hence by the Increasing-Optimal-Selections Theorem, there is a least $X = X(S)$ minimizing $C(\,\cdot\,, S)$ over $X \in L_S$ for $S \in \mathcal{S}$, and $X(S)$ is increasing in S on $\mathcal{S}$ as was to be shown.

In general the actual production in a period (other than the first or last) is not increasing in final sales in other periods when there are at least two facilities and at least three periods.

Consider now the effect of increasing the final sales in period j by $\mu > 0$ units, say. Let μ_j be the n vector whose i^{th} component is 0 or μ according as $1 \le i \le j$ or $j < i < n$. Then $S + \mu_j$ is decreasing in j so $X(S + \mu_j)$ is decreasing in j by what was shown above. Thus the effect of delaying an increase in final sales is to reduce the amount by which cumulative production increases at all facilities in all periods.

When production at each facility k in each period is a multiple of Q_k, we must add the constraint $X_i^k \in \{0, Q_k, 2Q_k, \dots\}$ for all i, k, which is a sublattice. The above results therefore carry over without change since intersections of sublattices are sublattices.

(b) Decomposition. Consider the problem of finding a production schedule $X^k \equiv \underline{X}^k$ that minimizes

$$\sum_{i=1}^{n} [c_i(X_i^k - X_{i-1}^k) + r_k^{-1} X_i^k] \tag{5}$$

subject to

$$X_i^k \ge X_{i-1}^k \text{ for } i = 1, \dots, n,\ U^k \ge X^k \ge S \text{ and } X_n^k = S_n \tag{6}$$

where $r_k \equiv h_k^{-1}$. The problems (5)-(6) for $1 \le k \le N$ form a relaxation of the problem (1)-(4) in which the constraints $X^k \ge X^{f_k}$ for $1 \le k \le N$ are relaxed to $X^k \ge S$. It turns out that the former constraint is automatically satisfied by the least optimal solutions to (5) and (6). To see this, observe that since a convex function of the difference of two variables is subadditive, the product of a decreasing and an increasing function of one variable is subadditive, and sums of subadditive functions are subadditive, (5) is subadditive in (X^k, r_k, U^k). Also, the set of such triples satisfying (6) is a nonempty sublattice. Hence, there is a least $X^k \equiv \underline{X}^k$ that minimizes (5) subject to (6), and $\underline{X}^k$ is increasing in (r_k, U_k), or equivalently, increasing in U^k and decreasing in h_k, by the Increasing-Optimal-Selections Theorem. Hence, since $U^k \ge U^{f_k}$ and $h_k \le h^{f_k}$, it follows that $\underline{X}^k \ge \underline{X}^{f_k}$ for $1 \le k \le N$, as claimed.

2. Multiple Assignment Problem with Subadditive Costs.

(a) Subadditive Costs. Let Π be the set of $n \times m$ matrices π such that each column of π is a (possibly different) permutation of the elements of the corresponding column of p. Let $S(\pi)$ be the $n \times m$ matrix whose j^{th} row is $\sum_{i=1}^{j} \pi_i$.

If $\pi \in \Pi$ is not monotone, then there exist $i < j$ for which $\pi_i \not\leq \pi_j$. Let π' be the matrix formed from π by replacing the i^{th} and j^{th} rows of π respectively by $\pi_i \wedge \pi_j$ and $\pi_i \vee \pi_j$. Then $\pi' \in \Pi$ and, because c is subadditive,

$$c(\pi_i \wedge \pi_j) + c(\pi_i \vee \pi_j) \leq c(\pi_i) + c(\pi_j),$$

whence $C(\pi') \leq C(\pi)$. Also, since $\pi_i \wedge \pi_j < \pi_i$ and $(\pi_i \wedge \pi_j) + (\pi_i \vee \pi_j) = \pi_i + \pi_j$, it follows that $S(\pi') < S(\pi)$. Since $S(\pi)$ decreases strictly at each step, no π can recur. Also, since Π has only finitely many elements, the process must terminate in finitely many steps with a monotone matrix $\sigma \in \Pi$. Moreover, since $C(\pi)$ decreases at each step, $C(\sigma) \leq C(\pi)$. Since all monotone matrices in Π have the same cost, σ has minimum cost.

(b) Superadditive Costs. We claim that if $m = 2$ and c is superadditive in (a), then one $\pi = (\pi_i^j)$ that minimizes $C(\,\cdot\,)$ over Π has the property that π_i^1 is increasing in i and π_i^2 is decreasing in i. To see this, let p^* be the matrix formed from p by replacing the elements of the second column of p by their negatives. Let Π^* be the set of $n \times 2$ matrices π^* such that each column of π^* is a (possibly different) permutation of the elements of the corresponding column of p^*. Let $S^* \equiv S_1 \times (-S_2)$ and define c^* on S^* be the rule $c^*(\gamma, \delta) \equiv c(\gamma, -\delta)$. Now since c is superadditive on S, c^* is subadditive on S^*. Consequently, from (a), C^* attains its minimum over Π^* at a π^* that is monotone. Thus, replacing the elements of the second column of π^* by their negatives yields a matrix π that minimizes $C(\,\cdot\,)$ over Π. Since π^* is monotone, π_i^1 is increasing in i and π_i^2 is decreasing in i as claimed.

(c) Order of Issuing. Let p be the matrix whose i^{th} row is (a_i, t_i). The cost (negative income) of issuing an item of age a at time t is $c(a, t) \equiv -I(t + a)$. The problem is to choose a π (i.e., an *issuing policy* or order of issuing items) having minimum cost. Now c is subadditive or superadditive according as I is convex or concave. Thus, by (a), if I is convex, then $\pi = p$, i.e., a LIFO (last-in-first-out) issuing policy, is optimal. Similarly, by (b), if I is concave, then π is the matrix whose i^{th} row is (a_i, t_{n+1-i}), i.e., a FIFO (first-in-first-out) issuing policy, is optimal.

(d) Order of Selling Securities. Let p be the matrix whose i^{th} row is (t_i, b_i) with $b_i \equiv p_i$ for all i. Let I_i be the gross income in year i. Then the tax paid in year i is $t_i(I_i - b) = t_iI_i - t_ib$ when a share bought at the price b is sold in year i. Since the term t_iI_i is independent of the choice of shares sold in each year, it suffices to consider the incremental tax $c(t, b) = -tb$ in a year when the tax rate is t and a share bought at the price b is sold that year. The problem is to choose a matrix π that minimizes total incremental taxes. Since c is subadditive, it follows from (a) that a monotone π is optimal, i.e., sell the share with i^{th} lowest price in the year of the i^{th} lowest tax rate.

(e) Order of Introducing Energy Technologies. Let p be the matrix whose i^{th} row is (i,i). The two columns of p are respectively the indices of the n technologies and the n decades in which they are to be introduced. Let $c(i,j)=c_{ij}$ be the cost of introducing technology i in decade j. By hypothesis, c is subadditive whence $\pi = p$ is optimal. Thus it is optimal to introduce the energy technology i in decade i for $i=1,\ldots,n$.

3. Representation of Additive Functions. The "if" part is immediate since a real-valued function on a chain is additive and sums of additive functions are additive. The proof of the "only if" part is by induction on n. The result is certainly true for $n=1$. Suppose the result is so for $n-1$ and consider n. By the induction hypothesis, there exist real-valued additive functions $f_1,\ldots,f_{n-1}$ on $S_1\times S_n,\ldots,S_{n-1}\times S_n$ respectively for which

$$f(s)=\sum_{i=1}^{n-1} f_i(s_i,s_n) \text{ for } s\in S.$$

Now apply the hint to each f_i.

4. Projections of Additive Convex Functions are Additive Convex.

(a) Additivity of Minimum Cost. Let $F(a,b)\equiv \min_{a\le x_1\le\cdots\le x_n\le b} f(x)$. The proof is by induction on n. Suppose $n=1$. If f is not monotone, then since f is convex, f attains its minimum over $\Re$ at x^0 say. Thus, $F(a,b)=f((a\vee x^0)\wedge b)=g(a)+h(b)$ where $g(a)\equiv f(a\vee x^0)-\frac{1}{2}f(x^0)$ and $h(b)\equiv f(b\wedge x^0)-\frac{1}{2}f(x^0)$. If f is monotone, then $F(a,b)=g(a)+h(b)$ where $g\equiv 0$ and $h\equiv f$ when f is decreasing and $g\equiv f$ and $h\equiv 0$ when f is increasing. From the convexity of f, it follows that g is increasing, h is decreasing and both are convex on $\Re$, establishing the result for $n=1$.

Next suppose the result is true for up to $n-1$ variables and consider n. It follows from Problem 3 that $f(x)=\sum_{i=1}^n f_i(x_i)$ for some real-valued functions $f_1,\ldots,f_n$ on the real line. Since f is convex, so are $f_1,\ldots,f_n$. On fixing x_1, minimizing with respect to $x_2,\ldots,x_n$ and using the induction hypothesis, $F(a,b)=\min_{a\le x_1\le b}[f_1(x_1)+g_1(x_1)]+h_1(b)$ where g_1 is increasing, h_1 is decreasing and both are convex. Since f_1 is convex, so is f_1+g_1, and the result follows from the induction hypothesis for the case of a single variable.

(b) Rent-a-Car Fleet Expansion. The minimum cost $\mathcal{C}(x_1,x_k,x_n)$ is additive on L as one sees by applying the result of (a) separately with the respective constraints $x_1\le x_2\le\cdots\le x_k$ and $x_k\le x_{k+1}\le\cdots\le x_n$ and the variables x_1,x_k and x_k,x_n held fixed as parameters respectively. Since $\mathcal{C}$ is additive, the incremental minimum cost of increasing the fleet size on day k is independent of the initial and terminal fleet sizes x_1 and x_n respectively.

Homework 3 Due October 21

1. Monotone Minimum-Cost Chains. Consider a (directed) graph (S, A) in which the set S of nodes is a finite sublattice of $\Re^n$ containing a distinguished node τ, the set A of arcs is a sublattice of S^2 and contains (τ, τ), and the (finite) cost c_{st} of arc (s, t) is subadditive on A with $c_{\tau\tau} = 0$. Assume that there are no negative-cost circuits and that there is a chain from each node to node τ. (A *chain* is a sequence of nodes $s_1, \ldots, s_m$ and corresponding arcs $(s_1, s_2), \ldots, (s_{m-1}, s_m)$. If $s_1 = s_m$, call the chain a *circuit*. The *cost* of a chain is the sum of its arc costs.) Let C_s be the minimum cost among all chains from node s to τ. Show that C_s is subadditive in s on S. Also show that the first node t_s visited after leaving node s on one minimum-cost chain from s to τ is increasing in s on S. [*Hint*: To prove these results, let C_s^k be the minimum cost among all k-arc chains from s to τ. Then

$$C_s^k = \min_{(s,t)\in A} [c_{st} + C_t^{k-1}] \text{ for } s \in S \text{ and } k = 1, \ldots, \sigma{-}1$$

where $C_\tau^0 \equiv 0$, $C_s^0 \equiv \infty$ for $s \in S \setminus \{\tau\}$ and σ is the cardinality of S. Since there are no negative-cost circuits, $C_s = C_s^{\sigma-1}$. Now establish the claim for each C_s^k, and hence for C_s.]

2. Finding Least Optimal Selection. Suppose S, T are chains of real numbers each containing n elements and f is real-valued on $S \times T$. Consider the problem of finding an $s = s_t$ in S that minimizes $f(s, t)$ over S for each $t \in T$. For an arbitrary function f, this entails evaluating f at all n^2 points in its domain. Where f is subadditive on $S \times T$, as we assume in the sequel without further mention, it is possible to reduce the number of necessary function evaluations significantly. Suppose that the least optimal selection s_t is desired.

(a) Finding Least Optimal Selection with $O(n \log_2 n)$ Function Evaluations by Interval-Bisection. Give an algorithm for finding s_t for each $t \in T$ that requires at most $n\left\lceil \log_2(n+1)\right\rceil$ $+ 2n$ function evaluations and comparisons where $\lceil x \rceil$ is the *ceiling* of x, i.e., the least integer majorant of x. [*Hint*: Let $t_1 \in T$ be the midpoint of T, e.g., the $\left\lceil \frac{n}{2} \right\rceil^{th}$ largest element of T. Find s_{t_1} by evaluating the n values of $f(\,\cdot\,, t_1)$ on S. Now let t_2 and t_3 be respectively the midpoints of the sets $\{t \in T : t < t_1\}$ and $\{t \in T : t > t_1\}$. Repeat the above procedure to find *both* s_{t_2} *and* s_{t_3} using what you know about the order of s_{t_1}, s_{t_2} and s_{t_3} with a total of at most $n+1$ function evaluations. Repeat the procedure inductively. Actually, there is an even better algorithm that requires at most about $7n$ function evaluations.]

(b) Finding Least Minimizer of f with $O(n \log_2 n)$ Function Evaluations. Show how to use the results of (a) above to find the least minimizer (s, t) of f over $S \times T$ with at most $n\left\lceil \log_2(n+1)\right\rceil + 2n$ function evaluations and $n\left\lceil \log_2(n+1)\right\rceil + 3n$ comparisons.

3. Research and Development Game. Consider the R&D game in which a finite set I of firms are each racing to make a discovery. If firm $i \in I$ makes the discovery first, the firm earns R_i and patents the discovery, thereby precluding the others firms from benefiting. If firm i devotes effort $e_i \geq 0$ to this research, the time to make the discovery is exponentially distributed with mean $1/e_i$ and the cost per unit time of so doing is $c_i(e_i)$. Assume that whatever the effort each firm allocates, the times the firms take to make the discovery are independent. Also assume that for each i, $c_i(0) = 0$ and $c_i(\cdot)$ is lower semicontinuous. Each firm's goal is to maximize its expected profit from the research. Show that there is no loss in generality in assuming that $R_i e_i - c_i(e_i) > 0$ for some $e_i > 0$. The firm wishes to restrict the choice of effort levels for firm i to a compact subset E_i of positive numbers e_i for which $R_i e_i - c_i(e_i) > 0$. Discuss whether or not there are Nash equilibria for this game. If there exist Nash equilibria, discuss the variation of the equilibria with an increase in R_i and a percent increase in c_i. (By a percent increase in c_i is meant multiplying c_i by a number $\theta_i > 1$.) Do this under the assumption that $a_i(e_i) \equiv c_i(e_i)/e_i$ is increasing (resp., decreasing) on E_i for each i.

Answers to Homework 3 Due October 21

1. Monotone Minimum-Cost Chains. Let C_s^k be the minimum cost among all k-arc chains from S to τ. Then

$$(*) \qquad C_s^k \equiv \min_{(s,t)\in A} [c_{st} + C_t^{k-1}] \text{ for } s \in S \text{ and } k = 1, \ldots, \sigma.$$

where $C_\tau^0 \equiv 0$ and C_s^0 for $s \in S \setminus \{\tau\}$. (Incidentally, the conditions that $(\tau, \tau) \in A$ and $c_{\tau\tau} = 0$ are used to assure that $(*)$ is valid for $s = \tau$.) Now C_s^0 is the indicator function of the sublattice $\{\tau\}$ of S and so is subadditive in s on S. Thus suppose that C_t^{k-1} is subadditive in t on S. Then since c_{st} is subadditive in (s,t) on A and sums of subadditive functions are subadditive, the bracketed term on the right-hand-side of $(*)$ is subadditive on the sublattice A. Hence, by the Projection Theorem for subadditive functions, C_s^k is subadditive in s on S. Also, by the Increasing-Optimal-Selections Theorem, there is a least t_s^k such that (s, t_s^k) achieves the minimum in $(*)$, and t_s^k is increasing in s on S. Finally, since there are no negative-cost circuits, $C_t^\sigma = C_t$, and so $t_s = t_s^\sigma$ is increasing in s on S.

2. Finding Least Optimal Selection. For each $\Sigma \subseteq S$ and $t \in T$, let $f(\Sigma, t) = \{f(s) : s \in \Sigma\}$ be the image of $f(\,\cdot\,, t)$ under Σ, $\min f(\Sigma, t)$ be the least element of $f(\Sigma, t)$ and $\{s \in \Sigma : f(s) = \min f(\Sigma, t)\}$ be the set of minimizers of $f(\,\cdot\,, t)$ over Σ. Also let $|\Sigma|$ be the cardinality of Σ.

(a) Finding Least Optimal Selection with $O(n \log_2 n)$ Function Evaluations by Interval-Bisection. Since f is real-valued and subadditive on $S \times T$, the least element s_t of the set of minimizers of $f(S,t)$ is increasing in t on T by the Increasing-Optimal-Selections Theorem. In the sequel choose distinct points $t_1, t_2, \ldots$ in T and find $s_i \equiv s_{t_i}$ for $i = 1, 2, \ldots$. In particular, let t_1 be the midpoint of T. Find s_1 by evaluating the $n = |S|$ values of $f(\,\cdot\,, t_1)$ on S. Now let t_2 and t_3 be respectively the midpoints of the sets $T_2 = \{t \in T : t < t_1\}$ and $T_3 = \{t \in T : t_1 < t\}$ if the sets are nonempty. Let $S_2 = \{s \in S : s \le s_1\}$ and $S_3 = \{s \in S : s \ge s_1\}$. Since $s_2 \le s_1 \le s_3$, it follows that $s_i \in S_i$ for $i = 2, 3$. Thus to find s_i, it suffices to evaluate the $|S_i|$ values of $f(\,\cdot\,, t_i)$ on S_i for $i = 2, 3$. Hence the total function evaluations and comparisons to find *both* s_2 *and* s_3 is $|S_2| + |S_3| = n + 1$ since $S_2 \cup S_3 = S$, $S_2 \cap S_3 = \{s_{t_1}\}$ and $|S| = n$. Let t_4, t_5, t_6, t_7 be respectively the midpoints of the sets $T_4 \equiv \{t \in T_2 : t < t_2\}$, $T_5 \equiv \{t \in T_2 : t_2 < t\}$, $T_6 \equiv \{t \in T_3 : t < t_3\}$ and $T_7 \equiv \{t \in T_3 : t_3 < t\}$ if the sets are nonempty. Let $S_4 = \{s \in S_2 : s \le s_2\}$, $S_5 = \{s \in S_2 : s \ge s_2\}$, $S_6 = \{s \in S_3 : s \le s_3\}$ and $S_7 = \{s \in S_3 : s \ge s_3\}$. Now since $s_4 \le s_2 \le s_5 \le s_1 \le s_6 \le s_3 \le s_7$, it follows that $s_i \in S_i$ for $i = 4, 5, 6, 7$. Thus since $\bigcup_{i=4}^7 S_i = S$ and only s_2, s_1, s_3 belong to at least —indeed at most—two of the sets S_i for $i = 4, 5, 6, 7$, it is possible to find the four points s_4, s_5, s_6, s_7 with at most $n + 3$ function evaluations and comparisons. More generally, the k^{th} step computes s_i for up to 2^k different values of the index i and requires at most $n + 2^{k-1} - 1$ function evaluations and comparisons. The total number of steps is at most the smallest index m such that $n \le 2^m - 1$, i.e., $m = \left\lceil \log_2 (n+1) \right\rceil$. Thus the total number of function evaluations and comparisons is at most

$$\sum_{i=1}^{m}(n+2^{i-1}-1)=m(n-1)+2^m-1\le n\left\lceil\log_2(n+1)\right\rceil-\left\lceil\log_2(n+1)\right\rceil+2^{\left\lceil\log_2(n+1)\right\rceil}-1$$
$$\le n\left\lceil\log_2(n+1)\right\rceil-\left\lceil\log_2(n+1)\right\rceil+2^{\log_2 2(n+1)}-1\le n\left\lceil\log_2(n+1)\right\rceil+2n.$$

(b) Finding Least Minimizer of f with $O(n\log_2 n)$ Function Evaluations. The desired minimum is found by simply comparing the n function values $f(s_t,t)$ for each $t\in T$. This requires at most $n-1\le n$ additional comparisons.

3. Research and Development Game. The time to the first discovery is the minimum of independent exponentially distributed random variables with expected values $1/e_i,\ i\in I$, and so is exponentially distributed with the expected time to the first discovery being $1/\sum_j e_j$. Also the probability that firm i makes the discovery first is $e_i/\sum_j e_j$. Thus, the expected profit to firm i is

$$(R_ie_i-c_i(e_i))/\textstyle\sum_j e_j.$$

If $R_ie_i-c_i(e_i)\le 0$ for all $e_i\ge 0$, then $e_i=0$ is optimal for firm i and the firm can be eliminated from the problem. Thus, there is no loss in generality in assuming for each remaining firm i, that $R_ie_i-c_i(e_i)>0$ for some $e_i>0$. Now it suffices to maximizing the natural logarithm $u_i(e)$ of the expected profit to firm i, viz.,

$$u_i(e)\equiv\ln(R_ie_i-c_i(e_i))-\ln(\sum_j e_j)$$

where $e=(e_i)$. The first term is a function of only one variable and so is superadditive. The second term is superadditive as well since $\ln(\,\cdot\,)$ is concave. Thus since $c_i(\,\cdot\,)$ is lower semicontinuous on E_i, $u_i(\,\cdot\,)$ is upper semicontinuous and superadditive on the compact sublattice $E\equiv\times_i E_i$ of $\Re^{|I|}$. Thus, by Theorem 1 of §3.2, there exist least and greatest Nash equilibria in E.

Let $a_i(e_i)\equiv c_i(e_i)/e_i$ and multiply the cost $c_i(e_i)$ by a positive number θ_i. Then on expressing the dependence of u_i on e, R_i and θ_i, it follows that

$$u_i(e,R_i,\theta_i)\equiv\ln(R_i-a_i(e_i)\theta_i)+\ln(e_i)-\ln(\sum_j e_j).$$

Now if $a_i(\,\cdot\,)$ is increasing, it follows that $u_i(e,R_i,\theta_i)$ is superadditive in (e_i,R_i) for each i, so the least and greatest Nash equilibria rise with $R=(R_i)$ by Theorem 1 of §3.2 again, i.e., the higher the stakes, the higher the Nash equilibria. Similarly under the same hypothesis, $u_i(e,R_i,\theta_i)$ is subadditive in (e_i,θ_i) for each i, so the least and greatest Nash equilibria fall with $\theta=(\theta_i)$ by Theorem 1 of §3.2 again, i.e., increasing the discovery costs by a given percent at all effort levels causes the Nash equilibria to fall. If instead $a_i(\,\cdot\,)$ is decreasing, dual results hold.

Homework 4 Due October 28

1. Computing Nash Equilibria of Noncooperative Games with Superadditive Profits.

(a) Computing Fixed Points of Increasing Mappings. Suppose $\emptyset \neq S \subseteq \Re^n$ is a compact lattice and that $\sigma(\,\cdot\,)$ is an increasing mapping of S into itself. Define s_N inductively by the rule $s_0 = \vee S$ and $s_{N+1} = \sigma(s_N)$ for $N = 0, 1, \ldots$. Show that if

$$(*) \qquad \sigma(\wedge C) \geq \wedge\sigma(C) \text{ for each chain } \emptyset \neq C \subseteq S \text{ of excessive points of } \sigma,$$

then the sequence $\{s_N\}$ converges downwards to the greatest fixed point of σ. For the case where S is a finite set, discuss whether or not $(*)$ holds. Give an analog of the above results starting instead with $\wedge S$. Give an example where $(*)$ fails to hold and $s^* = \lim_{N\to\infty} s_N$ is not a fixed point.

(b) Computing Nash Equilibria with Superadditive Profits. Give conditions that allow you to apply the above results to compute the least and greatest Nash equilibria of noncooperative games with superadditive profits (Theorem 1 in §3.2) as a limit of a sequence. Give an economic interpretation of this algorithm as an iterative process in which all firms announce strategies, find optimal replies and announce them, find optimal replies and announce them again, etc. [*Hint*: Use Lemma 2 of Appendix 2 of *Lectures in Supply-Chain Optimization*.]

2. Cooperative Linear Programming Game. Consider the cooperative linear programming (LP) game studied in §3.3.

(a) Core of Cooperative Linear Programming r-Subsidiary Game. Show that each profit allocation in the core of the cooperative LP r-subsidiary game allocates the same profit to all subsidiaries of a firm. Also, show that the resulting profit allocation to each firm is in the core of the cooperative LP game.

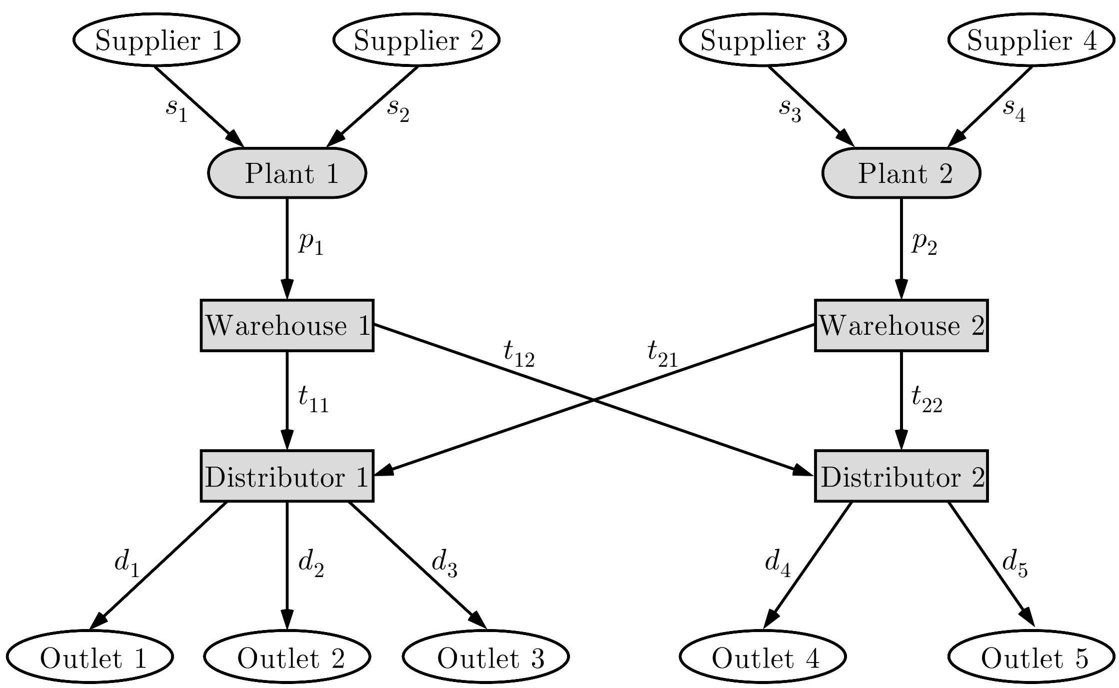
(b) Is Cooperation Beneficial? Determine in each of the following cases whether or not there is a potential benefit to the firms from cooperation?

- The resource vectors b^i are positive multiples of one another.
- The resource vectors b^i are nonnegative and A is Leontief (Ningxiong Xu).
- The matrix A is a node-arc incidence matrix.

Show that the core of the cooperative LP game is a singleton set for the first two of the above examples.

(c) Supply-Chain Games. Supply chains in which the facilities have different owners who wish to form an alliance to maximize overall profits can be often be represented as a Leontief substitution system as Problem 2 of Homework 1 illustrates. However, in these cooperative games, some activities are available only if the owners of different facilities agree. For example, a supplier and a buyer must agree on a transfer of goods from the former to the latter. For this reason, though the price allocations to the firms in a supply chain generated by the optimal dual prices for the grand alliance of firms in the supply chain is in the core of the game, there are other elements of the core of the LP game and of the LP r-subsidiary game. Illustrate this fact with the following example of a single supplier (facility 1) and a single buyer (facility 2) in a single period. The buyer faces demand $s_2 > 0$, can sell the product at the price p_2, and can purchase the product at the price $p_1 < p_2$ on the open market. Alternately, the buyer can purchase the product from the supplier whose unit production cost is $0 < c_1 < p_1$ provided that both agree. Find the optimal dual prices of the supplier-buyer alliance and show that the price allocation to the firms (resp., their r subsidiaries) based on these prices does not constitute the entire core of the game. Moreover, the entire benefit of the alliance accrues entirely to the buyer with that allocation.

3. Multiplant Procurement/Production/Distribution. A firm manufactures a single product at two plants and has five retail outlets at which to sell the product as depicted below.



Each plant orders a single raw material which is available from several suppliers. The cost $c_i(s_i)$ of purchasing $s_i \geq 0$ units from supplier i is continuous and convex. Each plant uses one unit of the raw material to produce one unit of finished product. The labor cost $l_i(p_i)$ of producing $p_i \geq 0$ units of the product at plant i is continuous and convex. Production at plant i enters warehouse i. Warehouse i ships $t_{ij} \geq 0$ units to distributor j at a unit cost c_{ij}. Each distributor serves several retail outlets. The demand $d_i \geq 0$ at retail outlet i is determined by the price $\pi_i \geq 0$ established there through the demand curve $d_i = \alpha_i - \beta_i \pi_i$, where α_i, $\beta_i > 0$ and $\pi_i \leq$

α_i/β_i. The revenue received at outlet i when the price is π_i there is $\pi_i d_i$. Assume that there are upper bounds on the amounts of raw materials available from each supplier, the amounts produced at each plant, and the amounts shipped from each warehouse to each distributor. Also assume all cost functions are increasing and vanish at the origin. There are no inventories anywhere in the system. The aim is to choose levels of procurement, production, transportation and prices that minimize total costs less revenues.

(a) Reduction to Wheel on Five Nodes. Show that this is a minimum-cost-flow problem and that the associated graph can be reduced to the wheel on five nodes by a sequence of series-parallel contractions. [*Hint*: Append a hub node H with arcs from each outlet to H and arcs from H to each supplier, and consider the minimum-cost-flow problem on the resulting graph with no demand at H.]

Discuss the effect of changes in the parameters below on each of the following optimal decision variables: $s_1, s_2, s_4, p_1, t_{21}, \pi_1, \pi_4$. [*Hint*: The easiest way to do this is to study the associated wheel on five nodes in which each arc corresponds to a series-parallel subgraph of the original graph.]

(b) Adding Supplier. Adding another supplier to Plant 1.

(c) Price Increase. A 5% increase in prices charged by Supplier 2.

(d) Strike. A strike at Plant 1 that prevents production there.

(e) Transportation Cost Increase. An increase in the unit transportation cost c_{21}.

(f) Upward Shift in Demand Curve. An increase in α_4 (which translates the demand curve at outlet 4 upwards).

4. Projections of Convex Functions are Convex. Suppose that f is a $+\infty$ or real-valued convex function on $\Re^{n+m}$ and the projection g of f defined by

$$g(t) = \inf_{s \in \Re^n} f(s,t),\ t \in \Re^m,$$

does not equal $-\infty$ anywhere. Show that g is convex on $\Re^m$. [*Hint*: Adapt the proof of the Projection Theorem for subadditive functions.]

Answers to Homework 4 Due October 28

1. Computing Nash Equilibria of Noncooperative Games with Superadditive Profits.

(a) Computing Fixed Points of Increasing Mappings. Since S is a nonempty compact lattice in $\Re^n$, it has a greatest element s_0 and a least element. Thus since $\sigma(\,\cdot\,)$ maps S into itself, $s_{N+1} \equiv \sigma(s_N) \in S$ for $N = 0, 1, \ldots$, so $s_1 = \sigma(s_0) \le s_0$. Now suppose that $s_N \le s_{N-1}$. Then since $\sigma(\,\cdot\,)$ is increasing, $s_{N+1} = \sigma(s_N) \le \sigma(s_{N-1}) = s_N$ so s_N is decreasing in N. Since s_N is bounded below by the least element of S and S is compact, $s_N \downarrow s^*$, say, in S. Thus, $s^* \le s_N$ and so $\sigma(s^*) \le \sigma(s_N)$ $= s_{N+1}$ for each $N = 0, 1, \ldots$. Thus $\sigma(s^*) \le s^*$. Now by hypothesis,

$$(*) \qquad \sigma(\wedge C) \ge \wedge\sigma(C) \text{ for each chain } \emptyset \ne C \subseteq S \text{ of excessive points of } \sigma.$$

Let C be the chain $\{s_N\}$. The elements of C are excessive. Since $\sigma(C) \cup \{s_0\} = C$, it follows from $(*)$ that $s^* = \wedge C = \wedge\sigma(C) \le \sigma(\wedge C) = \sigma(s^*)$, whence s^* is a fixed point of σ. It remains to show that s^* is the greatest fixed point of σ. To that end, suppose s is a fixed point of σ. Then $s \le s_0$ and so $s = \sigma(s) \le \sigma(s_0) = s_1$. Now by induction, $s \le s_N$ for $N = 0, 1, \ldots$, so $s \le s^*$, i.e., s^* is the greatest fixed point of σ. If S is a finite set, $(*)$ holds with equality because then $\wedge C \in C$.

There is an analog of the above results starting from the least element $s_0 \equiv \wedge S$ of S where σ instead satisfies the *dual* of $(*)$, i.e., reverse the inequality, replace meets by joins, and replace excessive by deficient in $(*)$. Then s_N converges upwards to the least fixed point of σ. If S is a finite set, the dual of $(*)$ holds with equality because then C is finite and contains $\vee C$.

Here is an example to show that if $(*)$ fails to hold, then $s^* = \lim_{N\to\infty} s_N$ need not be a fixed point of $\sigma(\,\cdot\,)$. Let $S = [0, 1]$, so S is a nonempty compact lattice in $\Re$ whose greatest element is 1. Let

$$\sigma(s) = \begin{cases} \frac{1}{3}s, \text{ if } 0 \le s \le \frac{1}{2} \\ \frac{1}{3}(s{+}1), \text{ if } \frac{1}{2} < s \le 1 \end{cases}.$$

Then $\sigma(\,\cdot\,)$ is an increasing mapping of S into itself, $s_N \downarrow s^* = \frac{1}{2}$, $C = \{s_N\}$ and $\wedge C = \frac{1}{2} = \wedge\sigma(C)$. However, $\sigma(s^*) = \sigma(\frac{1}{2}) = \frac{1}{6} < \frac{1}{2} = s^*$, so $(*)$ does not hold and s^* is not a fixed point of $\sigma(\,\cdot\,)$.

(b) Computing Nash Equilibria with Superadditive Profits. For this part, let $\mathcal{T}$ be a singleton set $\{t\}$ and suppress t. The required conditions include the hypotheses of Theorem 1 of §3.2 and the additional hypotheses that $\mathcal{S}_{si}$ and $u_i(s)$ are continuous in s on S. Let $\sigma(s)$ be the greatest (resp., least) optimal reply to $s \in S$. Then from the Increasing-Optimal-Selections Theorem 8 of §2.5, $\sigma(\,\cdot\,)$ is an increasing mapping of S into itself. Now by Lemma 2 of Appendix A.2, it follows for each nonempty chain C of excessive (resp., deficient) points of σ that $\wedge\sigma(C)$ (resp., $\vee\sigma(C)$) is an optimal reply to $\wedge C$ (resp., $\vee C$). Thus the greatest (resp., least) optimal reply $\sigma(\wedge C)$ (resp., $\sigma(\vee C)$) to $\wedge C$ (resp., $\vee C$) majorizes (resp., minorizes) $\wedge\sigma(C)$ (resp., $\vee\sigma(C)$) , i.e., $(*)$ (resp., the dual of $(*)$) holds. Now let s_0 be the greatest (resp., least) element of S. Then as in

(a) above, set $s_{N+1} \equiv \sigma(s_N)$ for $N = 0, 1, \ldots$. Then $s^* \equiv \lim_{N\to\infty} s_N$ is the greatest (resp., least) fixed point of σ. The interpretation of the s_N is as described in the problem.

2. Cooperative Linear Programming Game.

(a) Core of Cooperative Linear Programming r-Subsidiary Game. Consider a partition of the $r \cdot |I|$ subsidiaries of the r-subsidiary game into r alliances each of which each has exactly one subsidiary of each firm. Now the resource vector available to each of these alliances is the fraction r^{-1} of the total resources available to the grand alliance and each alliance has identical unit profit and constraint matrices. Thus, the maximum profit each such alliance can assure itself is the fraction r^{-1} of the maximum profit that the grand alliance can earn. Since each such alliance must earn at least this amount from each allocation in the core and cannot earn more, each such alliance earns exactly this amount from each allocation in the core. Now the same is so if two subsidiaries of the same firm swap the alliances to which they belong. Thus, the amount that each of these two subsidiaries earns from any allocation in the core must be identical. Thus, each subsidiary of a firm must receive the same amount from each allocation in the core.

(b) Is Cooperation Beneficial?

• *The resource vectors b^i are positive multiples of a one another.* Cooperation is not beneficial. To see this, observe that each b^i is a positive fraction $\lambda^i > 0$ of the vector b available to the grand alliance, and $\sum_I \lambda^i$. Thus, each firm i can earn the same fraction λ^i of the maximum profit M that the grand alliance can earn and no more. Consequently allocating each firm i the profit $\lambda^i M$ is the unique element of the core.

• *The resource vectors b^i are nonnegative and A is Leontief (Ningxiong Xu).* Cooperation is not beneficial. To see this, observe that there is a Leontief basis and corresponding optimal dual prices π that are simultaneously optimal for all nonnegative right-hand sides. Let $b = \sum_I b^i$ the resource vector available to the grand alliance and its maximum profit is πb. Now the maximum profit each firm i can earn is πb^i. Consequently allocating each firm i the profit πb^i is the unique element of the core.

• *The matrix A is a node-arc incidence matrix.* Cooperation is generally beneficial. For example, suppose one firm has unit supply in San Francisco and unit demand in Boston, and a second firm has the reverse. Also suppose the cost to ship within either city is a small fraction of the cost to ship from one city to the other. Then an alliance of the two firms benefits each firm since each can supply the demand of the other more cheaply than the firm can satisfy its own demand.

(c) Supply-Chain Games. Let w be the amount the seller produces, x be the amount the seller sells to the buyer, y be the amount the buyer purchases on the open market and z be the amount the buyer sells to meet demand. The problem that the supplier-buyer grand alliance faces entails choosing $w, x, y, z \geq 0$ to maximize

$$
\begin{array}{rrrrrl}
 & -c_1 w & & -\, p_1 y & +\, p_2 z & \\
\text{subject to} & & & & & \\
 & w & -\, x & & & = 0 \\
 & & x & +\, y & -\, z & = 0 \\
 & & & & z & = s_2
\end{array}
$$

The optimal dual prices $\alpha_1, \alpha_2, \beta_2$ are $\alpha_1 = \alpha_2 = -c_1$ and $\beta_2 = p_2 - c_1$.

The grand alliance can earn the maximum profit $(p_2 - c_1)s_2$, but the supplier can guarantee only the maximum profit 0 while the buyer can guarantee only the maximum profit $(p_2 - p_1)s_2$. Thus the core is the set of profit allocations $\pi = (\pi_1 \;\; \pi_2) \geq 0$ to the two firms for which $\pi_1 + \pi_2 = (p_2 - c_1)s_2$, $\pi_1 \geq 0$ and $\pi_2 \geq (p_2 - p_1)s_2$. However, allocating profits to the supplier and buyer with the optimal dual prices produces $\pi_1 = 0$ and $\pi_2 = (p_2 - c_1)s_2$, so the buyer gets the entire profit of the grand alliance. This allocation is in the core, but there are others that give the seller up to $(p_1 - c_1)s_2$ and the buyer as little as $(p_2 - p_1)s_2$.

The profit allocation to each subsidiary of a firm in the corresponding r-subsidiary game is the fraction r^{-1} of the profit allocation the firm gets. Thus, the profits allocation to the firms is unchanged by forming subsidiaries.

3. Multiplant Procurement/Production/Distribution. The problem is one of finding a minimum-cost circulation on the planar graph in Figure 1 formed by coalescing suppliers and retail outlets for both plants at a "hub" node H. In this event P_i is the plant i production node, W_i is the warehouse i storage node, and D_i is the distributor i distribution node for $i = 1, 2$.

(a) Reduction to Wheel on Five Nodes. Making series-parallel contractions in the above graph produces the series-parallel-free graph W_5 in Figure 2, viz., the wheel on five nodes. Notice that the rim arcs of the wheel coincide with those of the above graph. The spoke arcs of the wheel are all series-parallel contractions of subgraphs of the above graph. Observe that two arcs in W_5 are conformal if and only one of the following holds: the two arcs (i) are both rim arcs, (ii) are both spoke arcs, or (iii) are incident.

The cost of transportation is $T_{ij}(t_{ij}) = c_{ij}t_{ij}$, $t_{ij} \geq 0$. The cost of filling demands is $D_i(d_i) = -d_i\pi_i = \frac{1}{\beta_i}d_i^2 - \frac{\alpha_i}{\beta_i}d_i$, $d_i \in [0,\alpha_i]$, α_i, $\beta_i > 0$. Note that $c_i(\,\cdot\,)$, $l_i(\,\cdot\,)$, $T_{ij}(\,\cdot\,)$, $D_i(\,\cdot\,)$ are continuous, convex, increasing and vanish at the origin. Upper bounds on the amounts supplied, produced, and shipped are treated by adding a δ function, each preserving convexity, subadditivity and lower semicontinuity. The upper bounds on all the flows imply the constraint set $X(t)$ is nonempty and compact. Hence, by the lower semicontinuity of the cost functions, there exists an optimal flow. Thus, the Monotone-Optimal-Flow-Selection Theorem applies. The two graphs are biconnected and Table 1 describes relevant pairs of substitutes and complements for the original graph.

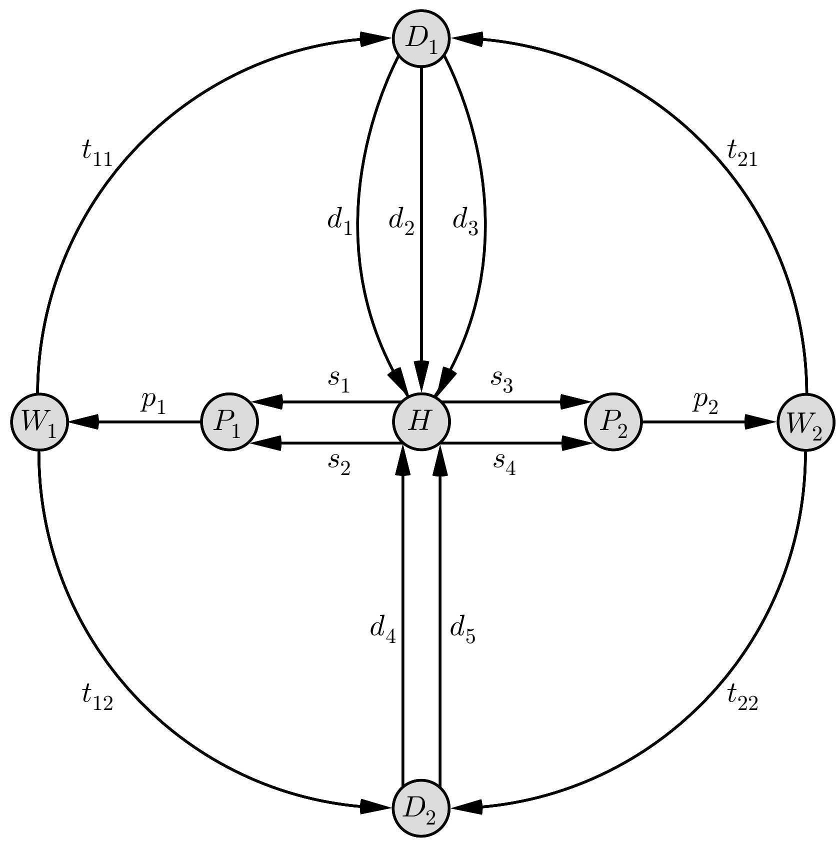


Figure 1

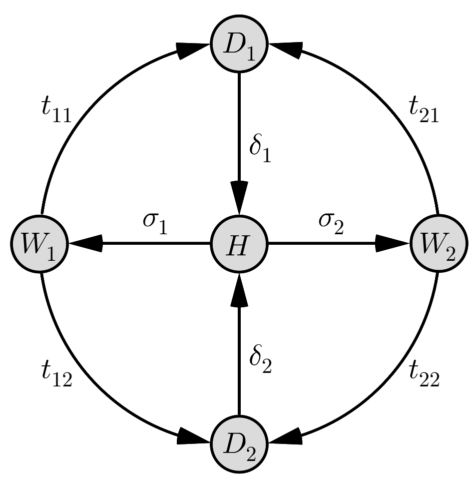


Figure 2

Table 1

	s_0	s_2	p_1	t_{21}	d_4
s_1	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{C}$	—	$\mathcal{C}$
s_2	$\mathcal{S}$	$\mathcal{C}$	$\mathcal{C}$	—	$\mathcal{C}$
s_4	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{C}$	$\mathcal{C}$
p_1	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	—	$\mathcal{C}$
t_{21}	—	—	—	$\mathcal{C}$	—
d_1	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{S}$
d_4	$\mathcal{C}$	$\mathcal{C}$	$\mathcal{C}$	—	$\mathcal{C}$

(b) Adding Supplier. Adding a supplier to plant P_1 corresponds to making a copy of the arc associated with s_1, introducing the flow s_0 in that arc, and increasing the upper bound τ_0 on that flow from zero to a positive level. The associated arc-cost function, $c_0(s_0,\tau_0) = c_0(s_0) + \delta_+(\tau_0 - s_0)$ is subadditive in s_0, τ_0, and convex and lower semicontinuous in s_0. The increase in τ_0 implies that the optimal s_0, p_1, d_1 and d_4 rise; s_1, s_2 and s_4 fall; and t_{21} is undetermined. Also, since d_1 (resp., d_4) rises, π_1 (resp., π_4) falls.

(c) Price Increase. Set $c_2(s_2,\tau_2) = \tau_2 c_2(s_2)$, so because $c_2(\,\cdot\,)$ is increasing, $c_2(\,\cdot\,,\cdot\,)$ is superadditive. Now $\tau_2 \geq 0$ and $c_2(\,\cdot\,)$ convex imply $c_2(\,\cdot\,,\tau_2)$ is convex. Thus, a 5% increase in prices at S_2 implies τ_2 increases from 1 to 1.05. Hence, the optimal s_1 and s_4 rise; s_2, p_1, d_1, and d_4 fall; and t_{21} is undetermined. Also, since d_1 (resp., d_4) falls, π_1 (resp., π_4) rises.

(d) Strike. Set $l_1(p_1,\tau_1) = l_1(p_1) + \delta_+(\tau_1 - p_1)$ where τ_1 is the upper bound on production at P_1. A strike decreases τ_1 to 0. Thus, the optimal s_4 rises; s_1, s_2, p_1, d_1 and d_4 fall, the first three to 0; and t_{21} is undetermined. Also, since d_1 (resp., d_4) falls, π_1 (resp., π_4) rises.

(e) Transportation Cost Increase. Evidently, the function $T_{21}(\,\cdot\,,\cdot\,)$ is superadditive, and $T_{21}(\,\cdot\,,c_{21})$ is convex. Thus, increasing c_{21} implies the optimal t_{21}, s_4 and d_1 fall; and s_1, s_2, p_1 and d_4 are undetermined. Also, since d_1 falls (resp., d_4 is undetermined), π_1 rises (resp., π_4 is undetermined).

(f) Upward Shift in Demand Curve. Set $D_4(d_4,\alpha_4) = \dfrac{1}{\beta_4} d_4^2 - \dfrac{\alpha_4}{\beta_4} d_4$, where α_4, $\beta_4 > 0$, $d_4 \geq 0$. Then $D_4(\,\cdot\,,\cdot\,)$ is subadditive, and $D_4(\,\cdot\,,\alpha_4)$ is convex (since $\beta_4 > 0$). Thus, increasing α_4 implies the optimal d_4, s_1, s_2, s_4, and p_1 rise; d_1 falls, so π_1 rises; and t_{21} is undetermined. Note that $\pi_4 = \dfrac{\alpha_4 - d_4}{\beta_4}$. If α_4 rises, then d_4 rises, but d_4 does not increase as much as does α_4. To see this, check that if $c(x,t) = \dfrac{1}{\beta}(x^2 - tx)$, then $c^{\#}(y,t) = \dfrac{1}{\beta}(y^2 - ty)$ is subadditive, so Lemma 5 of §4.5 of *Lectures in Supply-Chain Optimization* applies. Thus, π_4 rises.

4. Projections of Convex Functions are Convex. Suppose $t, \tau \in \Re^m$, $a, \alpha \geq 0$, and $a + \alpha = 1$. Since f is convex, then for each $s, \sigma \in \Re^n$

$$g(at + \alpha\tau) \leq f(as + \alpha\sigma, at + \alpha\tau) \leq af(s,t) + \alpha f(\sigma,\tau).$$

Now take infima over $s, \sigma \in \Re^n$.

Homework 5 Due November 4

1. Guaranteed Annual Wage. Consider the n-period production-planning problem of choosing a production vector $x = (x_i)$, inventory vector $y = (y_i)$ and sales vector $s = (s_i)$ that minimize the total cost

$$\sum_{i=1}^{n} [c_i(x_i) + h_i(y_i) + r_i(s_i)]$$

subject to the inventory-balance equations

$$x_i - s_i + y_{i-1} - y_i = 0,\ i = 1, \ldots, n$$

where $y_0 = y_n = 0$, $c_i(z)$ and $h_i(z)$ are respectively the costs of producing and storing z units in period i, and $r_i(z)$ is the negative of the revenue from selling z units in period i. Assume $c_i(\,\cdot\,)$, $h_i(\,\cdot\,)$ and $r_i(\,\cdot\,)$ are convex and lower semicontinuous on the real line with each equaling $+\infty$ for negative arguments, and $c_i(\,\cdot\,)$ is increasing.

(a) Guaranteed Production Levels. Let t be a common lower bound for x_i in each period. Let $x(t)$, $y(t)$ and $s(t)$ denote suitable minimum-cost production, inventory and sales vectors given t. One interpretation of a *guaranteed annual wage* (GAW) is often to increase the lower bound on production from 0 to $t > 0$. With this interpretation, show that the effect of a GAW is that for each $1 \le i \le n$, $t \le x_i(t) \le t \vee x_i(0)$. Thus, in this sense, employment is stabilized.

(b) Guaranteed Production Costs. A more flexible interpretation of a GAW is simply to put a floor $c_i(t)$ on production costs, but not on production levels, so that the cost of producing x_i in period i becomes instead $c_i(x_i \vee t)$. Show that $x_i(t) \le t \vee x_i(0)$ still holds for all i, but that $x_i(t) < t$ is possible. For given t, which leads to lower costs, guaranteed production levels or costs?

(c) Effect of Increasing Guarantee Level. Show that optimal sales in each period in (a) and (b) is increasing in t, but average optimal sales over n periods does not increase faster than t does. Also explain why end-of-period inventories in each period cannot be expected to vary monotonically with t.

(d) GAW vs. Higher Wage Rates. Establishing a GAW and raising wage rates, c.f., Example 9 in §4.8, both increase labor costs. Compare their impact on optimal sales.

2. Quadratic Costs and Linear Decision Rules. Consider the n-period quadratic-cost production planning problem of choosing a production vector $x = (x_i)$ and an inventory vector $y = (y_i)$ that minimize the total cost

$$\sum_{i=1}^{n} [c_i(x_i) + h_i(y_i)]$$

subject to the inventory-balance constraints

$$x_i - s_i + y_{i-1} - y_i = 0,\ i = 1, \ldots, n$$

where s_i is the given sales in period i, y_0 is the given initial inventory, $c_i(z) = \frac{1}{2}c_i z^2 + d_i z$ is the cost of producing z units in period i, $h_i(z) = \frac{1}{2}h_i z^2 + k_i z$ is the cost of storing z units at the end of period i, and c_i and h_i are positive for $i = 1, \ldots, n$. Note that all variables, including y_n, are unconstrained.

Use calculus—Lagrange multipliers is one way—to establish by backward induction on i that the optimal (x, y) is a *linear decision rule*, i.e., for each $i = 1, \ldots, n$,

$$x_i = -\beta_i^i y_{i-1} + \sum_{j=i}^{n} \beta_i^j s_j + \gamma_i \tag{1}$$

where the constants β_i^j and γ_i depend only on the coefficients of the cost functions in periods $i, \ldots, n$. Also show that

$$0 < \beta_i^n < \beta_i^{n-1} < \cdots < \beta_i^i < 1. \tag{2}$$

Interpret the results. Show also that if it is assumed that (1) holds, then the theory of substitutes, complements and ripples implies that β_i^j is decreasing in $j \geq i$.

3. Balancing Overtime and Storage Costs. Consider the n-period production planning problem in which in each period $i = 1, \ldots, n$ the sales of a product is a given integer $s_i \geq 0$, there is a unit cost $c \geq 0$ of normal production, a unit cost of overtime production that exceeds that of normal production by $d > 0$, and a unit storage cost $h > 0$. There is an integer maximum capacity $t > 0$ for normal production and unlimited production capacity for overtime available in each period. The problem is to choose integer nonnegative production and inventory vectors $x = (x_i)$ and $y = (y_i)$ respectively that minimize the n-period total cost

$$\sum_{i=1}^{n} [cx_i + d(x_i - t)^+ + hy_i]$$

subject to the inventory-balance constraints

$$x_i + y_{i-1} - y_i = s_i,\ i = 1, \ldots, n,$$

where $y_0 = y_n = 0$.

(a) Irrelevance of Production Costs. Show that the set of optimal production schedules is independent of $c \geq 0$.

(b) Optimal Production Scheduling Rule. Show that one optimal production schedule may be found inductively with the aid of the following rule. Assume that a production schedule has been found that optimally meets sales in periods $1, \ldots, i-1$ and (possibly) an integer portion of the sales in period i.

> Satisfy the next unit of sales in period i by producing the unit in the latest period k, $i - d/h < k \leq i$, in which there is unused normal production capacity when such a k exists, and by producing the unit on overtime production in period i otherwise.

[*Hint*: Apply the Unit-Parameter-Changes Theorem of the theory of substitutes, complements and ripples by parametrically increasing sales in unit increments.]

Answers to Homework 5 Due November 4, 2005

1. Guaranteed Annual Wage. It suffices to establish the result for the case where $c_i(\,\cdot\,)$ is strictly convex for each i. Let $\bar{t} \equiv (t, \ldots, t) \in \Re^n$ where $t \in \Re$. Let $x(t_1, \ldots, t_n)$ be optimal with the vector (t_i) of lower bounds on x. Then $x(t) = x(\bar{t})$.

(a) Guaranteed Production Levels. To study the effect on $x(t_1, \ldots, t_n)$ of raising (t_i) from the null vector 0 to $\bar{t}$, use monotonically step-connected changes, i.e., first raise t_i to t, then raise the t_j to t, $j \neq i$, one at a time. Put $c_i(x_i, t_i) \equiv c_i(x_i) + \delta_+(x_i - t_i)$. Observe that since $c_i(\,\cdot\,)$ and $\delta_+(\,\cdot\,)$ are convex, $c_i(x_i, t_i)$ is doubly subadditive, and $c_i(\,\cdot\,, t_i)$ is convex. Let 1_i be the i^{th} unit vector. Since x_i is a self-complement, $x_i(\,t1_i)$ equals $x_i(0)$ for $0 \leq t \leq x_i(0)$, and is increasing in t for $t > x_i(0)$ with the increase in x_i not exceeding that of t by the smoothing theorem, so $x_i(\,t1_i)$ $= t \vee x_i(0)$. Now since x_i and x_j are substitutes for $i \neq j$, it follows from the Monotone-Optimal-Flow-Selections Theorem that increasing t_j reduces the optimal value of x_i. Combining these facts yields $x_i(t) \leq t \vee x_i(0)$. Also, by definition of t, $t \leq x_i(\,t)$.

(b) Guaranteed Production Costs. Put $c_i(x_i, t_i) = c_i(x_i \vee t_i)$. Since $c_i(\,\cdot\,)$ is convex and increasing, $c_i(x_i, t_i)$ is doubly subadditive and $c_i(\,\cdot\,, t_i)$ is convex. Also, $c_i(x_i, t_i)$ is constant in t_i for $0 \leq t_i \leq x_i(0)$. Now the argument given in part (a) shows $x_i(t) \leq t \vee x_i(0)$. In this case it is possible that optimal production is less that the guarantee level is some period. Indeed this must be so if t exceeds the average sales $n^{-1}\sum_{i=1}^{n} s_i$ during the n periods because if not, total production during the n periods would exceed total sales during those periods. This can also happen if demand is low in a period in which storage is expensive. The minimum cost with guaranteed production costs minorizes that with guaranteed production levels. This is because an optimal schedule with guaranteed production levels is feasible and has the same costs with guaranteed production costs, but the latter allows additional options with possibly lower costs.

(c) Effect of Increasing the Guarantee Level. Since x_j and s_i are complements, s_i is increasing in t_j for each j. Thus, s_i is increasing in t. Now consider a change Δt in t and corresponding changes Δx_i and Δs_i in optimal x_i and s_i. Since the c_i are doubly subadditive, $\Delta x_i \leq \Delta t$ for all i. Thus since total production equals total sales, $n^{-1}\sum_{i=1}^{n} \Delta s_i = n^{-1}\sum_{i=1}^{n} \Delta x_i \leq \Delta t$, i.e., average optimal sales does not increase faster than t does. Finally, the end-of-period inventories y_i can't be expected to vary monotonically with t because y_i is a complement or substitute of x_j according as $j \leq i$ or $j > i$, and changes are made in the parameters for each x_j.

(d) GAW vs. Higher Wage Rates. With either interpretation of a GAW, each production cost $c_i(x_i, t_i)$ is subadditive and production in a period is a complement of sales in every period. By contrast, as Example 9 of §4.8 discusses, each production cost $c_i(x_i, t_i)$ is superadditive. Thus, optimal sales rise in each period under a GAW and fall in each period as wage rates rise.

2. Quadratic Costs and Linear Decision Rules.

Rewrite the inventory-balance constraints as

$$y_i = y_{i-1} + x_i - s_i,\ i = 1, \ldots, n \tag{1}$$

where y_0 is given. Denote the Lagrangian by

$$L(x, y, \lambda) = \sum_{i=1}^{n} \{\frac{1}{2} c_i x_i^2 + d_i x_i + \frac{1}{2} h_i y_i^2 + k_i y_i + \lambda_i [y_i - y_{i-1} - x_i + s_i]\}.$$

The optimal solution satisfies (1) and $\dfrac{\partial L}{\partial x_i} = \dfrac{\partial L}{\partial y_i} = 0$ for each i, so

$$\frac{\partial}{\partial x_i} L(x, y, \lambda) = c_i x_i + d_i - \lambda_i = 0,\ i = 1, \ldots, n \tag{2}$$

$$\frac{\partial}{\partial y_i} L(x, y, \lambda) = h_i y_i + k_i + \lambda_i - \lambda_{i+1} = 0,\ i = 1, \ldots, n \tag{3}$$

where $\lambda_{n+1} \equiv 0$. Setting $c_{n+1} \equiv d_{n+1} \equiv 0$ and using (2) to eliminate λ_i from (3) yields

$$c_i x_i + d_i + h_i y_i + k_i = c_{i+1} x_{i+1} + d_{i+1},\ i = 1, \ldots, n, \tag{4}$$

i.e., *the marginal cost of producing an item in period i and storing it to period $i + 1$ equals the marginal cost of producing the item in period $i + 1$.*

The next step is to show by backward induction on i that the optimal (x, y) satisfies,

$$x_i = -\beta_i^i y_{i-1} + \sum_{j=i}^{n} \beta_i^j s_j + \gamma_i,\ i = 1, \ldots, n, \tag{5}$$

where the constants $\beta_i^j > 0$ and γ_i depend only on the coefficients of the cost function in periods $i, \ldots, n$. On defining $\beta_{n+1}^{n+1} \equiv 1$, $\gamma_{n+1} \equiv 0$ and $x_{n+1} \equiv -y_n$, it follows that (5) holds for $i = n + 1$. Suppose (5) holds for $1 < i + 1 \le n + 1$ and consider i. Using (5) to replace x_{i+1} in (4) and then (1) to replace y_i in the resulting equation gives

$$\delta_i x_i + \epsilon_i y_{i-1} - \epsilon_i s_i - c_{i+1} \sum_{j=i+1}^{n} \beta_{i+1}^j s_j - \eta_i = 0$$

where $\epsilon_i \equiv h_i + c_{i+1} \beta_{i+1}^{i+1} > 0$, $\delta_i \equiv c_i + \epsilon_i > 0$ and $\eta_i \equiv -d_i - k_i + d_{i+1} + c_{i+1} \gamma_{i+1}$ for $i = 1, \ldots, n$. Thus (5) holds with $\beta_i^i \equiv \epsilon_i \delta_i^{-1} > 0$, $\beta_i^j \equiv c_{i+1} \beta_{i+1}^j \delta_i^{-1} > 0$ for $j = i{+}1, \ldots, n$, and $\gamma_i \equiv \eta_i \delta_i^{-1}$ for $i = 1, \ldots, n$. The coefficients β_i^j and γ_i depend only on the coefficients of the cost functions in periods i and $i + 1$, and on β_{i+1}^j and γ_{i+1}. And by the induction hypothesis, β_{i+1}^j and γ_{i+1} depend only on the coefficients of the cost functions in periods $i{+}1, \ldots, n$, so (5) holds for i.

It remains to show that

$$\beta_i^n < \beta_i^{n-1} < \cdots < \beta_i^i < 1,\ i = 1, \ldots, n. \tag{6}$$

Observe first that $\delta_i > \epsilon_i$ since $c_i > 0$, so $\beta_i^i = \epsilon_i \delta_i^{-1} < 1$. Next prove (6) by backward induction on i. The condition (6) was just shown for $i = n$. Now assume (6) holds for $1 < i+1 \le n$ and consider i. Since $\beta_i^j = c_{i+1}\beta_{i+1}^j \delta_i^{-1}$ for $j = i+1, \ldots, n$ and $c_{i+1}, \delta_i > 0$, it follows from the induction hypothesis that $\beta_{i+1}^n < \cdots < \beta_{i+1}^{i+1}$. Thus, it remains only to show that $\beta_i^{i+1} < \beta_i^i$. Since $h_i > 0$, whence $\epsilon_i > c_{i+1}\beta_{i+1}^{i+1}$, it follows that $\beta_i^{i+1} = c_{i+1}\beta_{i+1}^{i+1}\delta_i^{-1} < \epsilon_i \delta_i^{-1} = \beta_i^i$ as claimed.

The above results can be interpreted as follows. Optimal production in period i depends linearly on

- end of period inventories in period $i-1$ and on
- sales in current and future periods $i, \ldots, n$.

Optimal production in period i is less dependent upon sales in the later periods, i.e., optimal production in period $i \le j$ increases more if s_j increases by a unit than if s_{j+1} increases by a unit. Finally, if sales increases in a current or future period, then optimal production in the current period increases, but not as much as sales increases.

Now use the theory of substitutes and complements to show that

$$0 \le \beta_i^n \le \beta_i^{n-1} \le \cdots \le \beta_i^i \le 1. \tag{7}$$

The costs $c_i(\,\cdot\,)$ and $h_i(\,\cdot\,)$ are clearly convex for all i. Therefore $c_i(\,\cdot\,)$ and $h_i(\,\cdot\,)$ are doubly subadditive. Also $c_i(\,\cdot\,)$ and $h_i(\,\cdot\,)$ are continuous. Fix the sales s_j equal to $\underline{s}_j$ by appending the revenue cost function $\delta_0(s_j - \underline{s}_j)$ which is convex, doubly subadditive, and lower-semi-continuous. To allow end-of-period n inventory to be variable, append period $n+1$ with zero ordering and storage costs, and assume without loss of generality that $y_{n+1} = 0$. The graph is illustrated Figure 4.18 for $n = 3$, so $n+1 = 4$, except that sales $\underline{s}_1$ in period 1 is reduced by the initial inventory y_0. Since the graph is biconnected, and the arcs x_i and s_j are complements, x_i is increasing in s_j for $j \ge i$. Also, since x_i is less biconnected to s_j than s_j is, the Ripple Theorem implies that x_i increases slower than s_j, $j \ge i$. This shows that $0 \le \beta_i^j \le 1$ for all j.

It remains to establish the middle inequalities in (7). To that end, let $X_i = \sum_{j=1}^i x_j$ be the cumulative production and $S_i = \sum_{j=1}^i s_j$ the cumulative sales in period i and consider the network illustrated in Figure 4.20. Then x_i and X_j are complements for $j \ge i$, so x_i is increasing in S_j for $j \ge i$. Thus, as discussed in §2.6, x_i increases faster in s_j than in s_{j+1}, which implies the desired inequality.

3. Balancing Overtime and Storage Costs

(a) Irrelevance of Production Costs. The inventory-balance constraints imply that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} s_i$, so the normal production cost $c\sum_{i=1}^{n} x_i = c\sum_{i=1}^{n} s_i$ is independent of the schedule.

(b) Optimal Production Scheduling Rule. In view of (a), set $c = 0$ without loss in generality. The problem is an instance of the minimum-cost-flow problem discussed in §3.9 where v_i is the fixed sales in period $i = 1, \ldots, n$,

- the flow cost in arc $(0, i)$ is $d(x_i - t)^+$, $i = 1, \ldots, n$;
- the flow cost in arc $(i, i{+}1)$ is $h_i(y_i) = hy_i$, $i = 1, \ldots, n{-}1$; and
- the flow cost in arc $(i, 0)$ is $r_i(s_i, v_i) = \delta_0(s_i - v_i)$ $i = 1, \ldots, n$.

The costs in all arcs are affine between integers, convex and doubly subadditive by Example 8.

Now assume that an integer production schedule (x, y) has been found that optimally meets sales in periods $1, \ldots, i{-}1$ and an integer portion v_i of the sales in period i. Now increase v_i to $v_i' = v_i + 1$. Since the graph is biconnected, Theorem 4.7 implies that one new optimal flow (x', y') has the property that either $(x', y') = (x, y)$ or $(x', y') - (x, y)$ is a unit simple circulation whose induced cycle contains the arc $(i, 0)$. The former is impossible since $v' > v$. Thus the latter is so, and the structure of the graph and the nonnegativity assumptions on the (x, y) implies that the simple circulation consists of the arcs $(i, 0)$, $(0, k)$ and, for $k < i$, the arcs $(k, k{+}1), \ldots, (i{-}1, i)$ for some $k \le i$. The sum of the unit costs in the latter arcs is $h(i - k)$. The cost in the arc $(i, 0)$ is zero, and the cost in the arc $(0, k)$ is either 0 or d, depending on whether there is some unused normal production capacity in k.

The cheapest such simple circulation is sought. Since the cost of each circulation is nonnegative, if normal production is available in period i, i.e., $v_i < t$, then $k = i$ yields the cheapest circulation since it is feasible and incurs zero cost. If $v_i \ge t$, then the circulation that entails unit overtime production in period i incurs the cost d. This circulation is certainly cheaper than any other that incurs overtime production costs in period $k < i$. The only other possibility is that in some period $k < i$, there is normal production capacity available, i.e., $v_k < t$. The cost of this circulation is then the holding cost $h(i{-}k)$. For this to be cheaper than overtime production in period i, it is necessary and sufficient that $h(i{-}k) < d$, i.e., $i - d/h < k$. And of course the largest such $k < i$ is cheapest. This establishes the optimality of the indicated production rule.

Homework 6 Due November 11

1. Production Planning: Taut String. An inventory manager forecasts that the sales for a single product in periods $1, \ldots, 5$ will be respectively 1, 1, 2,−4, 2. He wishes to maintain his end-of-period inventories in the five periods between the respective lower bounds 2, 2, 3, 0, 3 and upper bounds 3, 4, 4, 2, 3.

(a) Optimal Production Schedule. Suppose the costs of producing z units in periods $1, \ldots, 5$ are respectively $\frac{1}{3}z^2$, z^2, $\frac{1}{2}z^2$, z^2, $\frac{1}{3}z^2$. Find the least-cost amounts to produce in periods $1, \ldots, 5$.

(b) Planning Horizon. In part (a) above, what is the earliest period k for which the optimal production level in period one is independent of the sales and of the upper and lower bounds on inventories in periods $k, \ldots, 5$?

(c) Solution of the Dual Problem. Give a compact form of the dual problem. Also, find the optimal dual variables for the data in part (a).

2. Plant vs. Field Assembly and Serial Supply Chains: Taut String. A firm has designed the supply chain for one of its products to admit assembly in n successive stages labeled $1, \ldots, n$. Each stage of assembly can be carried out at the firm's plant or in the field. The demand D for the product during a quarter must be satisfied and is a nonnegative random variable whose known strictly increasing continuous distribution is Φ. The demand can be satisfied by assembly at the plant before the demand is observed, by assembly in the field after the demand is observed or by some combination of both. The inventory manager has been asked to determine the amounts of the product to assemble at the plant to each stage that minimizes the expected total cost of plant and field assemblies. The unit cost of assembly at stage i is $p_i > 0$ at the plant and $f_i > p_i$ in the field, $i = 1, \ldots, n$. Thus the unit cost to assemble the product to stage i at the plant and after stage i in the field is $\sum_{j=1}^{i} p_j + \sum_{j=i+1}^{n} f_j$. Plant assemblies that are not needed have no salvage value. For each $i = 1, \ldots, n$, let a_i be the number of plant assemblies to at least stage i, i.e., the sum of the plant assemblies to stages $i, \ldots, n$. Thus a_n is the number of plant assemblies of the finished product and $a_i - a_{i+1}$ is the number of plant assemblies to stage i, $1 \le i < n$.

(a) Formulation. Show that the minimum expected-cost assembly problem is equivalent to choosing the plant assembly schedule $a = (a_i)$ to *maximize*

$$\sum_{i=1}^{n} [g_i a_i - f_i \mathrm{E}(a_i - D)^+]$$

subject to

$$a_1 \ge \cdots \ge a_n \ge 0$$

for suitably chosen numbers $g_1, \ldots, g_n$. *Be sure to define these numbers.*

(b) Monotonicity in Parameters. Discuss the monotonicity of one optimal plant assembly schedule $\bar{a} = (\bar{a}_i)$ in the parameters f_i, p_i.

(c) Optimal Solution. Show that there are *fill probabilities* $1 > \pi_1 \ge \pi_2 \ge \cdots \ge \pi_n > 0$ depending on (g_i) and (f_i), but not Φ, for which one optimal plant assembly schedule $\bar{a} = (\bar{a}_i)$ has the property that $\bar{a}_i$ is the $100\pi_i^{th}$ percentile of the demand distribution Φ, i.e., $\bar{a}_i = \Phi^{-1}(\pi_i)$ for each i. This schedule satisfies all demand from plant assemblies to at least stage i with probability π_i.

(d) Variation of Optimal Assemblies with Location and Variability of Demand Distribution. Let Ψ and Ω be distribution functions. The distribution Ψ is *stochastically smaller* than Ω if for each with $0 < u < 1$, $\Psi^{-1}(u) \le \Omega^{-1}(u)$. The distribution Ψ is *spread less than* Ω if for each u, v with $0 < u < v < 1$, $\Psi^{-1}(v) - \Psi^{-1}(u) \le \Omega^{-1}(v) - \Omega^{-1}(u)$. Determine the plant assembly stages i at which the optimal amount to make to stage i (resp., stage i or more) at the plant increases as Φ increases stochastically. Also do this where Φ increases its spread.

(e) Example. Suppose that $n = 5$ and that the g_i and f_i are as given in Table 1. The constants $G_i \equiv \sum_{j=1}^{i} g_j$ and $F_i \equiv \sum_{j=1}^{i} f_j$ are also tabulated there. Determine the fill probabilities $\pi_1, \ldots, \pi_5$ using the taut-string solution.

Table 1

i	1	2	3	4	5
f_i	40	40	20	40	20
g_i	10	32	18	10	10
F_i	40	80	100	140	160
G_i	10	42	60	70	80
π_i					

(f) Serial Supply Chain. Give an alternate interpretation of this problem in terms of a single-quarter serial-supply chain.

3. Just-In-Time Supply-Chain Production: Taut String. Consider the problem of finding minimum-cost n-period production schedules for each of N facilities labeled $1, \ldots, N$. Each facility j produces a single product labeled j. Facility j *directly consumes* $c_{ij} \ge 0$ units of product i in making one unit of product j. Facility j (directly or indirectly) *consumes* product i if there is a sequence of products that begins with i and ends with j such that each product in the sequence directly consumes its immediate predecessor in the sequence. The inventory manager projects that the vector of *cumulative external sales of product* j in the n periods will be $S^j = (S_i^j) = \sigma_j S^*$ for some number $\sigma_j \ge 0$ and *standard sales* vector S^* not depending on j, i.e., the cumulative external sales vectors of the products differ only by scale factors. Put $C = (c_{ij})$ and assume $\lim_{i\to\infty} C^i \to 0$, whence $(I - C)^{-1}$ exists and equals $\sum_{i=0}^{\infty} C^i$. Let π_j be the total amount of product j that would have to be produced to sell the vector $\sigma = (\sigma_1, \ldots, \sigma_N)^{\mathrm{T}}$ of the N products. Then $\pi = (\pi_1, \ldots, \pi_N)^{\mathrm{T}}$ is the unique solution of $\pi = \sigma + C\pi$, i.e., the total production of each product is the sum of the amounts of the product sold and consumed internally in producing other products.

There is a given *upper-bound vector* $U^j = (U_i^j) = \pi_j U^*$ on cumulative production at each facility j in each period for some *standard upper-bound vector* U^* on cumulative production. Assume that $S^* \le U^*$, $S_0^* = U_0^*$ and $S_n^* = U_n^*$. There are no time lags in production or delivery. Stock of a product at a facility is retained there until it is sold or directly consumed at some internal facility. There is no initial or final inventory of any product.

The *cost of producing* $z \ge 0$ units at facility j in period i is $d_i c^j(\frac{z}{d_i})$ where $d_1, \ldots, d_n$ are given positive numbers and c^j is a real-valued convex function on $\Re_+$ for each j. There are no direct storage costs at any facility. Negative production is not permitted. The goal is to find *cumulative production schedules* X^j at each facility j that minimize the total production cost at all facilities subject to the constraint that each facility meets projected external sales and internal consumption requirements for its product without backlogging or shortages. Assume that there exist feasible schedules for all facilities. Let X (resp., S, U) be the $N \times (n+1)$ matrix whose j^{th} row is X^j (resp., S^j, U^j).

(a) Feasibility. Show that X is feasible if and only if $S + CX \le X \le U$ and X_i^j is nondecreasing in i for each j.

(b) Positive Homogeneity of Optimal Single-Product Schedule. Let $X^* = X^*(S^*, U^*) = (X_i^*) \in \Re^n$ minimize $\sum_i d_i^{-1}(X_i^* - X_{i-1}^*)^2$ subject to $S^* \le X^* \le U^*$ and X_i^* nondecreasing in i. Show that $X^*(S^*, U^*)$ is *positively homogeneous of degree one* in (S^*, U^*), i.e., $X^*(\alpha S^*, \alpha U^*) = \alpha X^*(S^*, U^*)$ for each $\alpha \in \Re_+$.

(c) Reduction of N-Facility to Single-Facility Problem. Show that $X = \pi X^*$ is a minimum-cost schedule for the N-facility problem. [*Hint*: Show that $X = \pi X^*$ is feasible for the N-facility problem and that $X^j = \pi_j X^*$ minimizes $\sum_i d_i c^j(d_i^{-1}(X_i^j - X_{i-1}^j))$ subject to $\pi_j S^* \le X^j \le \pi_j U^*$ and X_i^j nondecreasing in i.]

(d) Just-in-Time Supply-Chain Production. Show that maintaining zero inventories is optimal at each facility j having no projected external sales, i.e., $S^j = 0$.

4. Stationary Economic Order Interval. Suppose that there is a constant demand rate $r > 0$ per unit time for a single product. There is a fixed setup cost $K > 0$ for placing an order, a holding cost $h > 0$ per unit stored per unit time and a penalty cost $p > 0$ per unit backordered demand per unit time. Each time the net inventory level (i.e., inventories less backorders) reaches s (< 0), an order for D $(> -s)$ units is placed that brings the net inventory on hand immediately up to $S \equiv s + D$.

(a) Optimal Ordering Policy. Show that the values of s, S and D that minimize the long-run average cost per unit time are given by

$$D = \sqrt{\frac{2Kr}{h} \cdot \frac{h+p}{p}},$$

$$S = \frac{p}{h+p} D \text{ and}$$

$$s = -\frac{h}{h+p} D.$$

(b) Limiting Optimal Ordering Policy as Penalty Cost Increases to Infinity. Find the limiting values of s, S and D as $p \to \infty$. Show that the limiting value of D agrees with the Harris square-root formula given in §1.2 of *Lectures on Supply-Chain Optimization.*

(c) Independence of Optimal Fraction Backordered and the Demand Rate and Setup Cost. What fraction f of the time will there exist no backorders under the policy in part (a)? Show that f does not depend on the demand rate or setup cost. Explain this fact by showing that for any fixed order interval, the same fraction f is optimal.

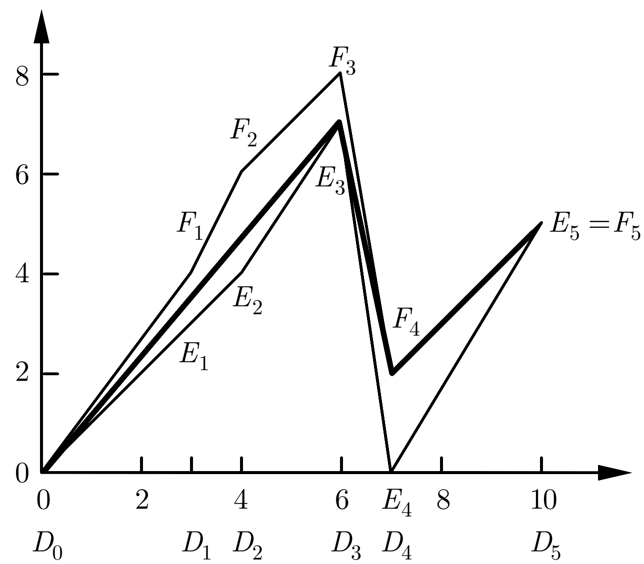
Answers to Homework 6 Due November 11

1. Production Planning: Taut String. When formulated in terms of cumulative production as described in §4.6 of the *Lecture in ...*, the problem is an instance of $\mathbb{P}$ with $A = (2\ 2\ 3\ 0\ 3)$, $B = (3\ 4\ 4\ 2\ 3)$ and $S = (1\ 2\ 4\ 0\ 2)$, so $E = S + A = (3\ 4\ 7\ 0\ 5)$ and $F = S + B = (4\ 6\ 8\ 2\ 5)$ with $E_0 = F_0 = 0$, $E_5 = F_5 = 5$, $\widehat{f}(z) = z^2$ and d-additive convex objective function

$$\sum_{i=1}^{5} d_i \widehat{f}(\frac{x_i}{d_i}) = \frac{1}{3}x_1^2 + x_2^2 + \frac{1}{2}x_3^2 + x_4^2 + \frac{1}{3}x_5^2$$

where $d = (3\ \ 1\ \ 2\ \ 1\ \ 3)$, so $D = (3\ \ 4\ \ 6\ \ 7\ \ 10)$.

(a) Optimal Production Schedule. The bold polygonal line in the figure below gives the taut-string solution.



The optimal solution is $X = (\frac{7}{2}\ \ \frac{14}{3}\ \ 7\ \ 2\ \ 5)$, so $x = (\frac{7}{2}\ \ \frac{7}{6}\ \ \frac{7}{3}\ \ -5\ \ 3)$.

(b) Planning Horizon. Observe from the figure that the data in periods $1, 2, 3, 4$ determine the optimal x_1, so the planning horizon is 4. And the smallest k such that the optimal choice of x_1 is independent of the data in periods $k, \ldots, 5$ is 5.

(c) Solution of the Dual Problem. The dual problem $\mathbb{P}^*$ is to maximize

$$\sum_{i=1}^{5} [E_i(t_{i+1} - t_i)^- - F_i(t_{i+1} - t_i)^+ - d_i \cdot \frac{t_i^2}{4}]$$

where $t_6 \equiv 0$, so $\widehat{f}^*(y) = \frac{1}{4}y^2$. Observe that $\partial\widehat{f}(y) = \{\frac{1}{2}y\}$. Now choose the optimal t_i to satisfy $\frac{x_i}{d_i} \in \partial\widehat{f}^*(t_i)$, whence it follows that $t_i = \frac{2x_i}{d_i}$ for $i = 1, \ldots, 5$. Thus $t = (\frac{7}{3}\ \ \frac{7}{3}\ \ \frac{7}{3}\ -10\ \ 2)$.

2. Plant vs. Field Assembly and Serial Supply Chains: Taut String.

(a) Formulation. Let a_i be the number of plant assemblies to stages $i, \ldots, n$ and D be the nonnegative random demand for the product. The manager chooses the number of plant assemblies to each stage before observing the demand D. Thus, a_i and $(D-a_i)^+$ are respectively the numbers of assemblies that perform stage i respectively at the plant and in the field. The problem is to choose $a=(a_1,\ldots,a_n)$ that minimize

$$(1)\qquad \sum_{i=1}^{n}\left[p_i a_i + f_i \mathrm{E}(D-a_i)^+\right]$$

subject to

$$(2)\qquad a_1 \ge \cdots \ge a_n \ge 0.$$

Now on substituting $x^+ = x + x^- = x - (-x)^+$, (1) reduces to

$$\sum_{i=1}^{n}\left[(p_i - f_i)a_i + f_i \mathrm{E}(D-a_i)^-\right] + \text{constant} = -\sum_{i=1}^{n}\left[g_i a_i - f_i \mathrm{E}(a_i - D)^+\right] + \text{constant}$$

where $g_i \equiv f_i - p_i$ for each i. Thus, the problem is to find an assembly schedule $a=(a_i)$ to maximize

$$(1)'\qquad \sum_{i=1}^{n}\left[g_i a_i - f_i \mathrm{E}(a_i - D)^+\right]$$

subject to (2).

(b) Monotonicity in Parameters. Now $f_i \mathrm{E}(D-a_i)^+$ is the product of an increasing function of f_i and decreasing function of a_i, and so is subadditive in those variables. Thus, (1) is subadditive in (a,f) where $f=(f_1,\ldots,f_n)$. Also, the set of vectors a that satisfy the constraints (2) is a sublattice of $\Re^n$ and is bounded below. Finally, since $p_i, f_i > 0$, it follows that (1) approaches ∞ as $a_i \to \infty$. Thus, from the Increasing-Optimal-Selections Theorem, there is a least minimizer $a(f)$ of (1) and $a(f)$ is increasing in f. In short, the higher the cost of assembly at each stage in the field, the greater is the optimal number of assemblies at the plant to at least each stage.

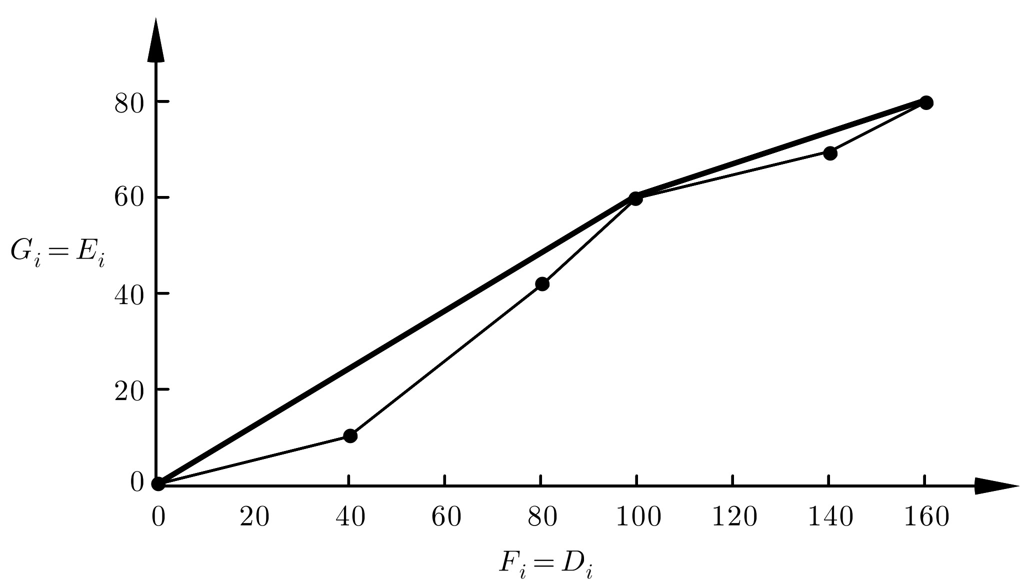
(c) Optimal Solution. The above problem whose objective function is reformulated in $(1)'$ is an instance of the dual program $\mathbb{P}^*$ that §5.6 discusses where $d_i = f_i$ and $E_i = \sum_{j=1}^{i} g_j$ for $1 \le i \le n$, $F_i = +\infty$ for $0 < i < n$, and $\widehat{f}^*(\alpha) \equiv \mathrm{E}(\alpha - D)^+$. Thus, $D_i = \sum_{j=1}^{i} d_j = \sum_{j=1}^{i} f_j$. Now draw a graph of the E_i versus the D_i for $i=0,\ldots,n$, and let $X=(X_i)$ be the taut-string solution of $\mathbb{P}$. Since the F_i in §5.6 have $F_i = +\infty$ for all $0<i<n$, the taut string is the least concave majorant of the (D_i, E_i) and so the slope $\pi_i \equiv x_i/d_i$ of the taut string to the left of D_i is decreasing in i where $x_i = X_i - X_{i-1}$. Then as §5.6 discusses, one optimal solution $(\bar{a}_i)$ of $\mathbb{P}^*$ is to choose the least $a_i = \bar{a}_i$ so $\pi_i \in \partial \widehat{f}^*(a_i)$ for each i, i.e., the least a_i so that $\pi_i \le \Phi(a_i)$. This is equivalent to choosing $\bar{a}_i = \Phi^{-1}(\pi_i)$ for each i. Thus since π_i is decreasing in i, Φ^{-1} is increasing, and the demand is nonnegative, $\bar{a}_1 \ge \cdots \ge \bar{a}_n \ge 0$. Also, the π_i depend only on the f_i and p_i.

(d) Variation of Optimal Assemblies with Location and Variability of Demand Distribution. The optimal amount to assemble to stage i (but not further) at the plant is $\Phi^{-1}(\pi_i) - \Phi^{-1}(\pi_{i+1})$ for $1 \le i < n$ and is $\Phi^{-1}(\pi_n)$ for $i = n$. The former increases as the spread of Φ increase, while the latter increases as Φ increases stochastically. Informally, assemblies in the plant to the final stage increases as the level of demand rises while assemblies to each stage before the final one increases as the variability of demand rises. The optimal amount to assemble to stage i or more is $\Phi^{-1}(\pi_i)$ and so increases as Φ increases stochastically.

(e) Example. Suppose that $n = 5$ and that the g_i and f_i are as given in Table 1. That table also tabulates the constants $G_i \equiv \sum_{j=1}^{i} g_j$ and $F_i \equiv \sum_{j=1}^{i} f_j$ (these F_i are not the upper bounds in §5.6). The taut-string is the heavy solid line in the figure below. The slopes of this line to the left of $D_1, \ldots, D_5$ are $\pi_1 = \pi_2 = \pi_3 = \frac{G_3}{F_3} = \frac{3}{5}$ and $\pi_4 = \pi_5 = \frac{G_5 - G_3}{F_5 - F_3} = \frac{1}{3}$. These slopes are the fill probabilities, and are recorded in Table 1. The interpretation is that it is optimal to completely assemble (to stage n) at the plant enough to satisfy the entire demand with probability $\frac{1}{3}$. Also it is optimal to assemble to stage 3 at the plant (and finish assembling in the field) an amount that, when added to the amount that is completely assembled at the plant, will satisfy all demand with probability $\frac{3}{5}$. It is optimal to satisfy the remaining demand by assembly entirely in the field.

Table 1

i	1	2	3	4	5
f_i	40	40	20	40	20
g_i	10	32	18	10	10
F_i	40	80	100	140	160
G_i	10	42	60	70	80
π_i	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{1}{3}$	$\frac{1}{3}$



(f) Serial Supply Chain. Consider a serial-supply chain with facilities labeled $1, \ldots, n$. Facility i supplies facility $i+1$ for each $i < n$ and facility n supplies a nonnegative random demand D with finite expectation in a single-quarter. There is a zero-period lead time for delivery to a facility from its supplier (external if facility one and internal otherwise). Demand that cannot be satisfied at a facility is passed to its predecessor until it can be satisfied, or if not, reaches facility 1 and is back-ordered if necessary. Let h_i be the amount by which the unit storage cost at facility i exceeds that at facility $i-1$ ($h_0 = 0$) and s_i be the amount by which the unit shortage cost at facility i exceeds that at $i+1$ ($s_{n+1} = 0$). Let a_i be the sum of the starting stocks on hand at facilities $i, \ldots, n$ after ordering at the beginning of the quarter. The problem is to choose $a = (a_i)$ to minimize the expected storage and shortage cost

$$\sum_{i=1}^{n} [h_i \mathrm{E}(a_i - D)^+ + s_i \mathrm{E}(D - a_i)^+].$$

subject to (2). This problem is equivalent to choosing $a = (a_i)$ to maximize $(1)'$ subject to (2) where $f_i = h_i + s_i$ and $g_i = s_i$ for all i.

3. Just-In-Time Multi-Facility Production: Taut String. The problem is to minimize

(1) $$\sum_{j=1}^{N}\sum_{i=1}^{n} d_i c^j (d_i^{-1}(X_i^j - X_{i-1}^j))$$

subject to

(2) $$S + CX \le X \le U$$

and

(3) $$X_i^j \text{ is nondecreasing in } i \text{ for each } j.$$

(a) Feasibility. Observe that X is feasible if and only if (2) and (3) hold. For the left-hand inequality in (2) assures that each facility meets the projected external sales and internal consumption requirements for its product without backorders or shortages. The right-hand inequality in (2) assures that cumulative production at each facility in each period does not exceed the given upper bound thereon. And condition (3) rules out negative production.

(b) Positive Homogeneity of Optimal Single-Product Production Schedule. The claim is immediate from the facts that the objective function $f(X^*)$ is homogeneous of degree two, i.e., $f(\alpha X^*) = \alpha^2 f(X^*)$ for all α, and the set of triples $(S^*\ X^*\ U^*)$ satisfying $S^* \le X^* \le U^*$ and X_i^* nondecreasing in i is a convex cone.

(c) Reduction of N-Facility to Single-Facility Problem. Since $C \ge 0$, $(I - C\,)^{-1} = \sum_{i=0}^{\infty} C^i \ge 0$. Thus since $\pi = \sigma + C\pi$ and $\sigma \ge 0$, $\pi = (I - C\,)^{-1}\sigma \ge 0$.

We show first that $X = \pi X^*$ satisfies (2) and (3). Since $\sigma \ge 0$ and $S^* \le X^*$,

$$S + CX = \sigma S^* + C\pi X^* \le (\sigma + C\pi)X^* = \pi X^* = X.$$

Also, since $\pi \geq 0$ and $X^* \leq U^*$, $X = \pi X^* \leq \pi U^* = U$, so (2) holds. Finally $X_i^j = \pi_j X_i^*$ satisfies (3) since $\pi \geq 0$ and X_i^* is nondecreasing in i.

Now it is enough to show that $X = \pi X^*$ is optimal for the *relaxation* of the N-facility problem in which one seeks X that minimizes (1) subject to

$$(2)' \qquad \pi S^* \leq X \leq \pi U^*.$$

This is a relaxation of the problem of minimizing (1) subject to (2) and (3) since the left-hand inequality in $(2)'$ is a relaxation of that in (2). To see this, subtract CX from both sides of the left-hand inequality in (2), premultiply by $(I-C)^{-1} \geq 0$ and substitute $S = \sigma S^*$ and $\pi = (I-C)^{-1}\sigma$. Moreover, $X = (X^j) = \pi X^*$ is optimal for the relaxation if and only if for each $j = 1, \ldots, N$, $X^j = \pi_j X^*$ minimizes

$$(1)' \qquad \sum_{i=1}^{n} d_i c^j (d_i^{-1}(X_i^j - X_{i-1}^j))$$

subject to

$$(2)'' \qquad \pi_j S^* \leq X^j \leq \pi_j U^*.$$

But since the function $(1)'$ is d-additive convex, it follows by applying the Invariance Theorem to this instance of $\mathbb{P}$ that it is enough to show $X^j = \pi_j X^*$ minimizes

$$(1)'' \qquad \sum_{i=1}^{n} d_i^{-1}(X_i^j - X_{i-1}^j)^2$$

subject to $(2)''$ for each $j = 1, \ldots, N$. But that is immediate from part (b).

(d) Just-in-Time Production. Let $y \equiv (y^j)$ be the optimal vector of inventories at the N facilities and $y^* \equiv X^* - S^*$ be the optimal vector of inventories for the standard facility. Then

$$y = X - (S + CX) = (I - C)\pi X^* - \sigma S^* = \sigma(X^* - S^*) = \sigma y^*.$$

This implies that optimal inventories at each facility j in each fixed period are proportional to the scale factor σ_j for external demand at the facility. In particular, if $S^j = 0$, in which case we can assume $\sigma_j = 0$ since $S^j = \sigma_j S^*$, $y^j = 0$. Thus *just-in-time* is optimal at each facility at which there is no external sales. Also, the pattern of temporal variation of inventories is the same at each facility!

Homework 7 Due November 18

1. Extreme Flows in Single-Source Networks. Consider the minimum-additive-concave-cost network-flow problem on the graph $(\mathcal{N}, \mathcal{A})$ with demand vector $r = (r_i)$ in which the flows are required to be nonnegative and there is a single source $\sigma \in \mathcal{N}$, i.e., $r_\sigma < 0$ and $r_i \geq 0$ for $i \in \mathcal{N} \setminus \{\sigma\}$. Use a graph-theoretic argument to establish the equivalence of 1°-3° below about a nonnegative flow $x = (x_{ij})$. Also show that 1° implies the second assertion of 3° using Leontief substitution theory from problem 1(a) of Homework 1.

1° x is an extreme flow.

2° The subgraph induced by x is a tree with all arcs directed away from σ.

3° The subgraph induced by x is connected and contains no arc whose head is σ, and $x_{ij}x_{kj} = 0$ for all arcs $(i, j), (k, j) \in \mathcal{A}$ for which $i \neq k$.

2. Cyclic Economic Order Intervals. Consider the problem of scheduling orders, inventories and backorders of a single product over periods $1, 2, \ldots$ so as to minimize the (long-run) average-cost per period of satisfying given demands $r_1, r_2, \ldots$ for the product in periods $1, 2, \ldots$. There is a real-valued concave cost $c_i(x_i)$ of ordering $x_i \geq 0$, $h_i(y_i)$ of storing $y_i \geq 0$ and $b_i(z_i)$ of backordering $z_i \geq 0$ in each period $i \geq 1$. Assume that $c_i(0) = h_i(0) = b_i(0) = 0$ for each $i \geq 1$ and that the data are *n-periodic*, i.e.,

$$(c_i, h_i, b_i, r_i) = (c_{i+n}, h_{i+n}, b_{i+n}, r_{i+n}) \text{ for } i = 1, 2, \ldots.$$

Clearly, there is a feasible schedule if and only if $\sum_1^n r_i \geq 0$. Assume this is so in the sequel.

(a) Reduction to Network-Flow Problem on a Wheel. Show that the problem of finding an ordering, inventory and backorder schedule that minimizes the average-cost in the class of n-periodic schedules is a minimum-additive-concave-cost uncapacitated nearly-1-planar network-flow problem. Show also that apart from duplicate arcs between pairs of nodes, the graph is a wheel with the hub node labeled 0 and the nodes associated with demands in periods $1, \ldots, n$ labeled by those periods cyclically around the hub. Thus, the node that immediately follows n in the cyclic order is, of course, node 1. Denote by $(i, k]$ the interval of periods in the cyclic order that begins with the node following i and ends with k for $1 \leq i, k \leq n$.

(b) Existence of Optimal Periodic Schedules. Give explicit necessary and sufficient conditions for the existence of a minimum-average-cost n-periodic schedule in terms of the derivatives at infinity of the cost functions (c_i, h_i, b_i) for $1 \leq i \leq n$.

(c) Extreme Periodic Schedules. Show that a feasible n-periodic schedule $(x_1, y_1, z_1, \ldots, x_n, y_n, z_n)$ is an extreme point of the set of such schedules if and only if

- inventories and backorders do not occur in the same period, i.e., $y_i z_i = 0$ for $i = 1, \ldots, n$,
- there is a period $1 \le l \le n$ with no inventories or backorders, i.e., $y_l = z_l = 0$, and
- between any two distinct periods $1 \le i, k \le n$ of positive production, there is a period j in the interval $[i, k)$ with no inventories or backorders, i.e., $y_j = z_j = 0$.

Show that if also the demands are all nonnegative, then $y_{i-1}x_i = z_i x_i = y_{i-1}z_i = 0$, for $1 \le i \le n$ where $y_0 \equiv y_n$.

(d) An $O(n^3)$ Running-Time Algorithm. Use the second condition of (c) to show that if there is a minimum-average-cost n-periodic schedule, then such a schedule can be found with at most $\frac{3}{2}n^3 + O(n^2)$ additions and $n^3 + O(n^2)$ comparisons by solving n n-period economic-order-interval problems. This implementation improves upon the $O(n^4)$ running time of straightforward application of the send-and-split method. [*Hint*: First show how to compute the minimum costs c_{ik} incurred in the interval $(i, k]$ with zero inventories and backorders in periods i and k, and with at most one order placed in the interval $(i, k]$ for all $1 \le i, k \le n$ in $O(n^3)$ time.]

(e) Constant-Factor Reduction in Running Time with Nonnegative Demands and No Backorders. Show that if also the demands are nonnegative and no backorders are allowed, then the running time in (d) can be reduced to $\frac{1}{2}n^3 + O(n^2)$ additions and a like number of comparisons.

3. 94%-Effective Order Intervals for One Warehouse and Three Retailers. Find a power-of-two ordering policy that has at least 94% effectiveness for the one-warehouse three-retailer economic-order-interval problem in which the setup cost for placing an order at each facility is $25 and the weekly demand rate at each retailer and the weekly unit storage cost rates at the warehouse and the retailers are as given in the table below. Also, give an improved *ex post* estimate of the worst-case effectiveness of the power-of-two policy that you find.

	Facility			
		Retailers		
	Warehouse	1	2	3
Weekly Demand Rate	0	75	50	10
Weekly Unit Storage Cost Rate	.4¢	2.0¢	.7¢	.5¢

4. Nearly-Optimal Multiproduct Dynamic Order Intervals with Shared Production. Suppose that N products are produced at a plant over n periods labeled $1, \ldots, n$. The projected sales of each product j in each period i is $s_i^j \ge 0$ with all products being measured in a common unit. Let $x_i^j \ge 0$ be the amount of each product j produced by the plant in each period i. The cost of producing (resp., storing) $z \ge 0$ units of each product j in (resp., at the end of) each period i, denoted $c_i^j(z)$ (resp., $h_i^j(z)$), is real-valued, nonnegative and concave on $\Re_+$. There is no initial or ending inventory of any product. There is a given vector $u = (u_1, \ldots, u_n)$ of upper bounds on total production of all products at the factory in each period. No backlogging is permitted. Let X^j be the set of production schedules $x^j = (x_1^j, \ldots, x_n^j)$ that are feasible for each product j ignoring the upper bounds on capacity in each period. Assume that X^j is nonempty for all j. Let $c^j(x^j)$ be the cost of the schedule x^j for product j. The problem is to choose a production schedule $x = (x^j)$ for the N products that minimizes the total cost

$$(1) \qquad \mathcal{C}^N(x) \equiv \sum_{j=1}^{N} c^j(x^j)$$

subject to the plant capacity constraint

$$(2) \qquad \sum_{j=1}^{N} x^j \le u$$

and each product's feasibility constraints

$$(3) \qquad x^j \in X^j \text{ for } 1 \le j \le N.$$

Instead of seeking a minimum-cost schedule, it is easier to find a schedule with high guaranteed effectiveness. To that end, observe that each $x^j \in X^j$ is a convex combination of the extreme points of X^j, i.e.,

$$(4) \qquad x^j = \sum_k \alpha_{jk} x^{jk} \text{ for } 1 \le j \le N,$$

where the x^{jk} are extreme points of X^j for all k,

$$(5) \qquad \sum_k \alpha_{jk} = 1 \text{ and all } \alpha_{jk} \ge 0.$$

One way of approximating the problem (1)-(3) is to replace (1) by the linear function

$$(1)' \qquad C^N(\alpha) \equiv \sum_{j,k} \alpha_{jk} c^j(x^{jk}),$$

and consider the approximate problem $\mathbb{P}$ of finding $\alpha = (\alpha_{jk})$ that minimizes $(1)'$ subject to (5) and

$$(2)' \qquad \sum_{j,k} \alpha_{jk} x^{jk} \le u.$$

Observe that $x = (x^j)$ is feasible for the original problem if and only if there is a feasible solution $\alpha = (\alpha_{jk})$ of the approximate problem satisfying (4).

(a) Objective Function of Approximate Problem Minorizes that of Original Problem. Show that for each such pair x and α of feasible solutions, $\mathcal{C}^N(x) \geq C^N(\alpha)$ with equality occurring if the α_{jk} are all integer. Also show that if the optimal α for the approximate problem is a zero-one matrix, then x is optimal for the original problem.

(b) At Most $2n$ Noninteger Variables in each Extreme Optimal Solution of Approximate Problem. Show that there is an optimal solution of the linear program $\mathbb{P}$ with at most $2n$ of the α_{jk} not being zero or one.

(c) Effectiveness of Solution Induced by Optimal Solution of Approximate Problem. Show that the effectiveness of the solution of the original problem determined from (4) by the optimal solution to the approximate problem is at least 100% times $(1+\frac{Mn}{mN})^{-1}$ (and so converges to 100% as $N \to \infty$) provided that there exist numbers $0 < m,M$ such that for all $x^j \in X^j$ and $j = 1,\ldots,N$ and for all $N > 0$, the following conditions hold.

- No Dominant Product. $c^j(x^j) \leq M$.
- Total Costs Grow At Least Linearly in N. $\mathcal{C}^N(x) \geq mN$.
- Feasibility. The capacity vector u grows with N fast enough so that the original problem is feasible for all N.

(d) Pricing all Variables in $O(Nn^2)$ Time. Show that if the revised simplex method is used to solve the approximate problem and if $\pi = (\pi_1,\ldots,\pi_n)$ is a vector of prices of capacities in each period corresponding to the basic solution of the problem at some iteration, then one can price out all variables and use the usual criterion to select the new variable to become basic in $O(Nn^2)$ time. [*Hint*: Price out each product j separately by finding a schedule $x^j \in X^j$ that minimizes the adjusted cost $c^j(x^j) - \pi x^j$.]

Answers to Homework 7 Due November 18

1. Extreme Flows in Single-Source Networks.

$1^\circ \Rightarrow 2^\circ$. By Theorem 6.2, a flow is extreme if and only if its induced subgraph $\mathcal{T}$ is a forest. Then $\mathcal{T}$ is a tree. For if not, there is a nonempty subtree T of $\mathcal{T}$ not containing σ. Thus, by flow conservation, the sum of the demands in T is zero. Hence, because the demands are all nonnegative at nodes in T, all demands in T are zero. Thus the flow in all arcs of T is zero, contradicting the fact that T is a nonempty subtree of the induced subgraph $\mathcal{T}$. It remains to show that all arcs in the tree $\mathcal{T}$ are directed away from σ. If not, there is an arc (i,j) in $\mathcal{T}$ that is not directed away from σ. Let flow $x_{ij} > 0$ be the flow in (i,j). Delete arc (i,j) from $\mathcal{T}$ thereby forming two trees T_i and T_j, with T_i containing i but not σ, and T_j containing j and σ. Add (resp., subtract) x_{ij} to (resp., from) the demand at node i (resp., j). This preserves feasibility of the flows in the two trees. But now the sum of the demands in T_i is positive, which contradicts conservation of flow in T_i.

$2^\circ \Rightarrow 3^\circ$. Since the subgraph $\mathcal{T}$ induced by x is a tree with all arcs directed away from σ, $\mathcal{T}$ is connected and contains no arc whose head is σ. If there is a node j such that $x_{ij}x_{kj} \neq 0$ for some $i \neq k$, $\mathcal{T}$ contains arcs (i,j) and (k,j), so $j \neq \sigma$. Now there exist two internally node disjoint simple chains from σ to j, one from σ to i to j and the other from σ to k to j. But this contradicts the fact there is a unique simple chain from σ to j.

$3^\circ \Rightarrow 1^\circ$. By Theorem 6.2, it is enough to show that the subgraph $\mathcal{F}$ induced by x is a forest. If not, there is a simple cycle $\mathcal{C}$ in $\mathcal{F}$. We claim that there is a node in $\mathcal{F}$ that is the head of two arcs in $\mathcal{F}$. This is certainly so if $\mathcal{C}$ is not a circuit. Suppose instead that $\mathcal{C}$ is a circuit. Then $\mathcal{C}$ does not contain σ because $\mathcal{F}$ contains no arc whose head is σ. And since $\mathcal{F}$ is connected, there is a simple path $\mathcal{P}$ from σ to some node l in $\mathcal{C}$. If $\mathcal{P}$ is not a chain, $\mathcal{P}$ contains a node that is the head of two arcs in $\mathcal{P}$. If $\mathcal{P}$ is a chain, then l is the head of two arcs in $\mathcal{F}$, viz., an arc in $\mathcal{P}$ and one in $\mathcal{C}$. Thus in all cases there is a node j in $\mathcal{F}$ that is the head of two arcs in $\mathcal{F}$, say (i,j) and (k,j). Thus, $x_{ij}x_{kj} \neq 0$, which is a contradiction.

It remains to use Leontief substitution theory to show that 1° implies that $x_{ij}x_{kj} = 0$ for all arcs (i,j) and (k,j) in $\mathcal{A}$ with $i \neq k$. The set of feasible flows is the set of solutions $x = (x_{ij})$ of the flow-conservation equations

$$\sum_{i \neq j} x_{ij} - \sum_{i \neq j} x_{ji} = r_j,\ j \in \mathcal{N} \setminus \{\sigma\} \tag{4}$$

associated with nodes other than σ for which

$$x \geq 0. \tag{5}$$

Of course, the equation for node σ is redundant and so can and is omitted. Now (4)-(5) form a pre-Leontief substitution system since the right-hand side of (4) is nonnegative and each x_{ij} enters at most two equations in (4), the j^{th} with a coefficient of $+1$ and the i^{th} with a coefficient of

-1. Thus, x_{ij} and x_{kj} are substitutes for $i \neq k$. Consequently, by the result of Problem 1(a) in Homework 1, $x_{ij}x_{kj} = 0$ for $i \neq k$.

2. Cyclic Economic-Order-Interval

(a) Reduction to Network-Flow Problem on a Wheel. The problem is to choose levels $x_1, \ldots, x_n$ of production, $y_1, \ldots, y_n$ of inventories, and $z_1, \ldots, z_n$ of backorders in an interval of n periods that minimizes the average-cost per period. This is equivalent to minimizing the average-cost per n-period interval

$$\sum_{i=1}^{n} [c_i(x_i) + h_i(y_i) + b_i(z_i)]$$

subject to

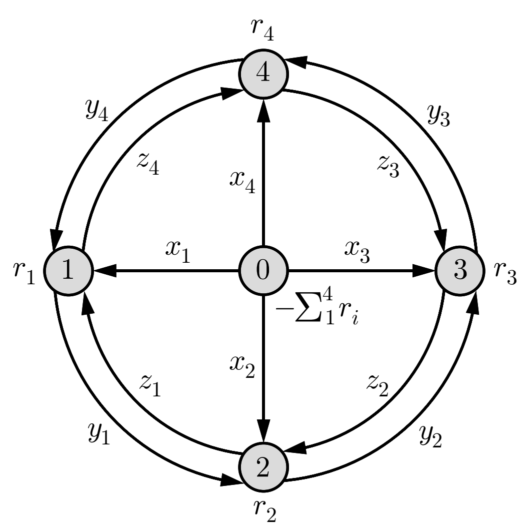
$$y_{i-1} - z_{i-1} + x_i - y_i + z_i = r_i, i = 1, \ldots, n$$

where $y_0 = y_n,\ z_0 = z_n$ and $(x_i, y_i, z_i) \geq 0$. This is a minimum-additive-concave-cost uncapacitated network-flow problem on the graph with $n+1$ nodes as the figure below illustrates for $n = 4$. Observe that the network in nearly 1-planar since the graph is planar and all nodes but the hub lie in the outer face of the graph. Also, apart from duplicate arcs y_i, z_i joining nodes i and $i+1$ for each i, the graph is a wheel.

(b) Existence of Optimal Periodic Schedules. There are three types of simple circuits, viz.,

- the arcs y_i, z_i (there are n of these "bicircuits", one for each i),
- the rim arcs $y_1, \ldots, y_n$, and
- the rim arcs $z_1, \ldots, z_n$.

By Theorem 6.3, a minimum-cost schedule exists if and only if the sum of the derivatives of the flows costs at infinity in arcs around each of these simple circuits is nonnegative. In particular



$$\dot{h}_i(\infty) + \dot{b}_i(\infty) \geq 0 \text{ for all } i,$$

$$\sum_{i=1}^{n} \dot{h}_i(\infty) \geq 0$$

and

$$\sum_{i=1}^{n} \dot{b}_i(\infty) \geq 0.$$

(c) Extreme Periodic Schedules. There are three types of simple cycles in the given graph, viz.,

- the arcs y_i, z_i (there are n of these "bicycles", one for each i),
- the rim arcs $w_1, \ldots, w_n$ where each w_i is either y_i or z_i, and
- the arcs $x_i, w_i, \ldots, w_{k-1}, x_k$ where each w_j with is either y_j or z_j for $j \in [i, k)$.

A flow is extreme if and only if it induces a graph that is a forest. Equivalently, a flow is extreme if and only if the graph it induces contains no simple cycles. This is so if and only if on each simple cycle of the graph, at least one of the arc flows is zero. Thus from what was shown above, a feasible flow is extreme if and only if

- $y_i z_i = 0$ for each i,
- $y_l = z_l = 0$ for some l,
- for each $i \neq k$ with $x_i x_k > 0$, $y_j = z_j = 0$ for some $j \in [i, k)$.

If also the demands are nonnegative in each period, then the hub node of the wheel has nonpositive demand and the rim nodes each have nonnegative demand. Then each extreme flow induces an arborescence whose root is the hub node and whose arcs are directed away from the hub node. In that event, at each rim node i, at most one of the arc flows y_{i-1}, x_i, z_i entering i can be positive, i.e., $y_{i-1}x_i = x_i z_i = y_{i-1} z_i = 0$.

(d) An $O(n^3)$ Running Time Algorithm. Without loss of generality, assume that $c_i(0) = h_i(0) = b_i(0)$ for all i. Now as we have shown above, there is a period $1 \leq l \leq n$ with $y_l = z_l = 0$. For each such l, one can delete the arcs y_l, z_l and solve the dynamic (noncyclic) n-period inventory problem that begins in period l and ends in period $l-1$. Since there are n possible choices of l, it suffices to solve n such problems, one for each l.

Let c_{ik}^j be the sum of the cost of ordering in period $j \in (i, k]$ enough to satisfy the demands in the interval $(i, k]$ and the costs of backorders and storage in the interval $(i, k]$ for each distinct $1 \leq i, k \leq n$. Put $h_i(z) \equiv b_i(-z)$ for $z \leq 0$ and all i. Extend the definitions given in §6.7 of *Lectures on Supply-Chain Optimization* from the n- to the $2n$-period problem. This allows tabulation of the R_j for $1 \leq j \leq 2n$ with $2n$ additions. This also permits computation of the $R_{ij} = R_j - R_i$ for $1 \leq i < j \leq 2n$ and $j - i \leq n$ in $O(n^2)$ time. Finally, this allows recursive calculation of the B_{ij} and H_{ij} for $1 \leq i < j \leq 2n$ and $j - i \leq n$ by the recursions given in §6.7. Now define R_{ij}, B_{ij} and H_{ij} for $1 \leq j < i \leq n$ by the rules $R_{ij} = R_{i,j+n}, B_{ij} = B_{i,j+n}$ and $H_{ij} = H_{i,j+n}$. These definitions are appropriate because the demands and costs are n-periodic. Thus it is possible to compute the R_{ij}, B_{ij} and H_{ij} for all distinct $1 \le i, j \le n$ in $O(n^2)$ time.

Next observe that just as for the n-period problem,

$$c_{ik}^j = B_{ij} + c_j(R_{ik}) + H_{jk}$$

for all $j \in (i,k]$ and distinct $1 \le i,k \le n$. Now there are $\frac{1}{2}n^3 + O(n^2)$ such triples (i,j,k) and each requires two additions. Thus these c_{ik}^j can all be computed with $n^3 + O(n^2)$ additions.

Next it is necessary to compute the $c_{ik} = \min_{j\in(i,k]} c_{ik}^j$. This requires $\frac{1}{2}n^3 + O(n^2)$ comparisons, one for each triple (i,j,k) with $j \in (i,k]$.

Finally it is necessary to solve the n-period problem for each $1 \le l \le n$. Now §6.7 of *Lectures on Supply-Chain Optimization* shows how to do this for one n-period problem with $\frac{1}{2}n^2 + O(n)$ additions and a like number of comparisons, and so for all n such problems with $\frac{1}{2}n^3 + O(n^2)$ additions and a like number of comparisons.

Thus it is possible to carry out the entire computation with $\frac{3}{2}n^3 + O(n^2)$ additions and $n^3 + O(n^2)$ comparisons.

(e) Constant-Factor Improvement in Running Times with Nonnegative Demands and No Backorders. If the demands are nonnegative and there are no backorders, then the z_i are all zero. Delete those arcs from the graph. Thus, if the inventories at the ends of distinct periods i and k vanish and one orders only once in the interval, $[i,k)$, the order must be in period $i+1$. Thus it suffices to compute the c_{ik}^j only for $j = i+1$. As a consequence, computation of all the c_{ik}^{i+1} and c_{ik} requires $O(n^2)$ time. Thus, the algorithm in (d) requires at most $\frac{1}{2}n^3 + O(n^2)$ additions and a like number of comparisons.

3. 94% Effective Lot-Sizing. Choose the units of each product so the sales rate per unit time at each retailer is two. Measuring costs in dollars yields

$$\begin{array}{lll} h_1' = .75 & h^1 = .15 & h_1 = .6 \\ h_2' = .175 & h^2 = .10 & h_2 = .075 \\ h_3' = .025 & h^3 = .02 & h_3 = .005. \end{array}$$

Step 1. Calculate and Sort the Breakpoints.

Compute the breakpoints $\tau_n' = \sqrt{K_n/h_n'}$, $\tau_n = \sqrt{K_n/h_n}$. Each $K_n = 25$. Thus

$$\begin{array}{ll} \tau_1' = \ 5.77 & \tau_1 = \ 6.45 \\ \tau_2' = 11.95 & \tau_2 = 18.26 \\ \tau_3' = 31.62 & \tau_3 = 70.71 \end{array}$$

Step 2. Initialize E, G, L, K and H.

Let $E = \emptyset$, $G = \emptyset$, $L = \{1,2,3\}$, $K = 25$ and $H = \sum_{i=1}^3 h^i = .27$.

Step 3. Cross the Largest Uncrossed Breakpoint.

• Then $\tau = \tau_3 = 70.71 > \sqrt{K/H} = \sqrt{92.59} = 9.62$, so $E = \{3\}$, $L = \{1,2\}$, $H \leftarrow H + h_3 = .275$, $K \leftarrow K + K_3 = 50$.

• Also, $\tau = \tau_3' = 31.62 > \sqrt{K/H} = \sqrt{181.82} = 13.48$, so $E = \emptyset$, $G = \{3\}$, $H \leftarrow H - h_3 = .25$, $K \leftarrow K - K_3 = 25$.

• Moreover, $\tau = \tau_2 = 18.26 > \sqrt{K/H} = \sqrt{100} = 10$, so $E = \{2\}$, $L = \{1\}$, $H \leftarrow H + h_2 = .325$, $K \leftarrow K + K_2 = 50$.

• Finally, $\tau = \tau_2' = 11.95 \le \sqrt{K/H} = \sqrt{153.85} = 12.40$, so $T^* \in [11.95,\ 18.26]$.

Step 4. Calculate T^* and $\mathbb{B}$.

Evidently, $E = \{2\}$, $L = \{1\}$, $G = \{3\}$, $T^* = 12.40$, $T_1^* = \tau_1 = 6.45$, $T_2^* = 12.40$, $T_3^* = \tau_3' = 31.62$. Since $c_n(T,T_n) = \frac{K_n}{T_n} + h_nT_n + h^n(T \vee T_n)$,

$$\begin{aligned}
c_1(T^*,T_1^*) &= \frac{25}{6.45} + 0.6(6.45) + 0.15(12.40) = 9.61,\\
c_2(T^*,T_2^*) &= \frac{25}{12.40} + 0.075(12.40) + 0.1(12.40) = 4.19, \text{ and}\\
c_3(T^*,T_3^*) &= \frac{25}{31.62} + 0.005(31.62) + 0.02(31.62) = 1.58, \text{ so}
\end{aligned}$$

$$\mathbb{B} = \frac{K_0}{T^*} + \sum\nolimits_{i=1}^{3} c_i(T^*, T_i^*) = 17.39$$

Now compute the desired integer-ratio policy $\mathcal{T}$ as in §6.8 of *Lectures on Supply-Chain Optimization.*

- Since $2 \in E$, $r_2 = 1$.
- Then $r_1^* = T_1^*/T^* = .5202 \in (2^{-1},2^0] \equiv (r_-,r]$. Since $r_1^* \le \sqrt{rr_-} = 0.7071$, $r_1 = 0.5$.
- Also $r_3^* = T_3^*/T^* = 2.55 \in (2^1,2^2] \equiv (r_-,r]$. Since $r_3^* \le \sqrt{rr_-} = 2.83$, $r_3 = 2$.

Thus, $T = 12.4$, $T_1 = \frac{1}{2}T^* = 6.2$, $T_2 = T^* = 12.4$, $T_3 = 2T^* = 24.8$.

Finally, compute an improved posterior estimate of the worst case effectiveness of the above power-of-two policy as follows. Evidently $K \equiv K_0 + K_2 = 50$, $H \equiv h_2' + h^1 = .325$, $H_1 = .6$, $H_3 = .025$, $M = 8.06$, $M_1 = 7.75$, $M_3 = 1.58$, $\mathbb{B} = 17.39$, $q_1 = .9612$, $q_3 = .7843$ $(q_n = r_n/r_n^*)$, $e(q_1) = .9992$ and $e(q_3) = .97$. Hence

$$\frac{c(T^*)}{c(T)} = \left[\frac{M}{\mathbb{B}} + \frac{M_1}{\mathbb{B}e(q_1)} + \frac{M_3}{\mathbb{B}e(q_3)}\right]^{-1} = [.46 + .45 + .09]^{-1} = \frac{1}{1.0037} \simeq 0.997 > e(q_1) \wedge e(q_3).$$

4. Dynamic Lot-Sizing with Shared Production

(a) Objective Function of Approximate Problem Minorizes that of $\mathcal{P}$. Let $\mathcal{P}$ be the original problem. Suppose $x = (x^j)$ is feasible for $\mathcal{P}$ and $\alpha = (\alpha_{jk})$ satisfies (4) and (5) with the given x. Then

$$C^N(\alpha) = \sum_{j,k} \alpha_{jk} c^j(x^{jk}) \le \sum_j c^j(\sum_k \alpha_{jk} x^{jk}) = \sum_j c^j(x^j) = \mathcal{C}^N(x)$$

with equality holding if and only if the α_{jk} are all integers. This is because for each j there is exactly one k for which $\alpha_{jk} > 0$, and that $\alpha_{jk} = 1$. This implies that the cost of optimal solution of $\mathbb{P}$ minorizes that of $\mathcal{P}$, with equality obtaining if α is integer. Thus if α is optimal for $\mathbb{P}$ and is integer, the corresponding x defined by (4) is optimal for $\mathcal{P}$.

(b) At Most $2n$ Noninteger Variables in each Extreme Optimal Solution of $\mathbb{P}$. Add slack variables to the inequalities $(2)'$. Then $\mathbb{P}$ has N equations in (5) and n in $(2)'$. Thus each basic feasible solution of $\mathbb{P}$ has at most $n + N$ basic variables. In any such solution, all N of the equations (5) contain at least one basic variable, so at most n of them can contain two or more basic variables. Thus at least $N - n$ of those equations contain exactly one basic variable whose value is necessarily one. Thus in each basic solution of $\mathbb{P}$, at most $n + N - (N - n) = 2n$ of the α_{jk} are not 0-1.

(c) Effectiveness of Solution Induced by Optimal Solution of $\mathbb{P}$. Let x^* be optimal for $\mathcal{P}$ and α^* be optimal for $\mathbb{P}$. Let $x(\alpha^*)$ be the corresponding feasible solution of $\mathcal{P}$ given by (4). Now by the hypotheses

$$mN \le \mathcal{C}^N(x^*) \le \mathcal{C}^N(x(\alpha^*)) \le C^N(\alpha^*) + Mn \le \mathcal{C}^N(x^*) + Mn.$$

Thus,

$$\frac{\mathcal{C}^N(x^*)}{\mathcal{C}^N(x(\alpha^*))} \ge \frac{\mathcal{C}^N(x^*)}{\mathcal{C}^N(x^*) + Mn} \ge \frac{mN}{mN + Mn} = \Big(1 + \frac{Mn}{mN}\Big)^{-1}.$$

Hence, the effectiveness of $x(\alpha^*)$, i.e., the left-hand side of the above inequality, approaches one as $N \to \infty$ because that is so of the right-hand side of that inequality.

(d) Pricing all Variables in $O(Nn^2)$ Time. Let $\sigma = (\sigma_1, \ldots, \sigma_N)$ and $\pi = (\pi_1, \ldots, \pi_n)$ be respectively basic prices associated with the constraints (5) and $(2)'$. To price out the variables associated with product j, it suffices to check that

$$\min_k \,[c^j(x^{jk}) - \pi x^{jk}] \ge \sigma_j.$$

But the minimum on the left is the minimum cost for the single-product concave-cost no-backorders problem in which the ordering cost for product j in each period i is replaced by the (concave) function $c_i^j(x_i) - \pi_i x_i$. Consequently, the minimum cost for product j can be found in $O(n^2)$ time using the recursion for the no-backorders problem given in §6.7 of *Lectures on Supply-Chain Optimization*. Thus it is possible to price out all N products in $O(Nn^2)$ time.

Homework 8 Due December 2

1. Total Positivity. A real-valued nonnegative function f on a product $X \times Y$ of two sets of real numbers is called *totally-positive of order* n (TP_n) if

$$\begin{vmatrix} f(x_1, y_1) & \cdots & f(x_1, y_m) \\ \vdots & & \vdots \\ f(x_m, y_1) & \cdots & f(x_m, y_m) \end{vmatrix} \geq 0$$

for each $x_1 < x_2 < \cdots < x_m$ in X, $y_1 < y_2 < \cdots < y_m$ in Y, and $m = 1, \ldots, n$. Observe that the TP_n functions are TP_m for $m = 1, \ldots, n-1$. Also, the TP_1 functions are the nonnegative functions, so the TP_n functions are nonnegative for $n = 1, 2, \ldots$. Also, since the determinants of a square matrix and its transpose coincide, the function g defined by $g(y, x) \equiv f(x, y)$ is TP_n on $Y \times X$ if and only if f is TP_n on $X \times Y$.

(a) Characterization of TP_2 Functions. Show that the following statements about a nonnegative function f on $X \times Y$ are equivalent:

1° f is TP_2.

2° f has *monotone likelihood ratio*, i.e., $\dfrac{f(x, y_2)}{f(x, y_1)}$ is increasing in x on X for all $y_1 < y_2$ in Y for which the ratio is well defined.

3° $\ln f$ is superadditive.

(b) Stochastic Monotonicity of TP_2 Probabilities. Suppose $X = \Re$ and $f(\,\cdot\,, y)$ is a density function for each $y \in Y$. Let $F(w, y) = \int_{-\infty}^{w} f(x, y)\, dx$ for $w \in \Re$. Show that if f is TP_2, then $F(\,\cdot\,, y)$ is stochastically increasing in y on Y. [*Hint*: Integrate the inequality $f(x_1, y_1) f(x_2, y_2) \geq f(x_1, y_2) f(x_2, y_1)$ with respect to x_1 on $(-\infty, w]$ and with respect to x_2 on $[w, \infty)$ for each fixed w and $y_1 < y_2$.]

(c) Gamma Density. Let $f(\,\cdot\,, r, \lambda)$ be the gamma density function with shape parameter $r \geq 1$ and scale parameter $\lambda > 0$. Then

$$f(x, r, \lambda) = \begin{cases} 0 & , x \leq 0 \\ \dfrac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} & , x > 0 \end{cases}$$

Determine whether the gamma distribution is stochastically increasing or decreasing or neither in $r \geq 1$ and $\lambda > 0$.

Sign-Variation-Diminishing Property. Totally positive functions have important sign-variation-diminishing properties. To describe them, suppose X and Y are sets of real numbers, f is TP_n, σ is real-valued and increasing on X, g is real-valued on X, and $g(x)$ changes sign at most $n - 1$ times as x traverses X (excluding zeroes of g). (For example, $\sin(x)$ changes sign twice on the interval $[0, 3\pi]$, viz., from + to − to +.) It is known (Karlin (1968), Chapter 5) that if the function G defined on Y by $G(y) \equiv \int_X g(x) f(x,y)\, d\sigma(x)$ exists and is finite on Y, then the number of sign changes of G on Y does not exceed that of g on X. Moreover, if g and G have the same number of sign changes and if g is first positive (resp., negative), then so is G. Section 5 of the Appendix to *Lectures in Supply-Chain Optimization* proves these facts for the case where X and Y are finite sets.

(d) Convexity Preservation by TP_3 Distributions. Show that a function g on a set X of real numbers is convex if and only if for every affine function a on X, the function $g - a$ changes sign at most twice on X; and when there are two sign changes, they are from + to − to +. Now suppose f is real-valued on a product $X \times Y$ of two sets of real numbers, σ is real-valued and increasing on X, and the function G defined on Y by $G(y) \equiv \int_X g(x) f(x,y)\, d\sigma(x)$ exists and is finite on Y. The question arises under what conditions on f is it true that convexity of g implies convexity of G. For this to be so it is necessary (but not sufficient) that if g is affine, then g and $-g$ are convex. Thus G and $-G$ are convex, so G is affine. A sufficient condition for G to be affine is that

$$(1) \qquad \int_X f(x,y)\, d\sigma(x) = \alpha \text{ and } \int_X x f(x,y)\, d\sigma(x) = \beta y$$

exist and are finite for all $y \in Y$ and some constants $\alpha, \beta \neq 0$. Show that if (1) holds and f is TP_3, then convexity of g implies convexity of G. An important example of such a function f is the binomial distribution $f(x,y) = \binom{y}{x} p^x (1-p)^{y-x}$ where $0 \le p \le 1$, X and Y are the nonnegative integers and $\sigma(x) = \lfloor x \rfloor$.

2. Production Smoothing. An inventory manager seeks a policy for producing a product that minimizes the expected n-period costs of production, production changes, storage and shortage. The demands $D_1, \ldots, D_n$ for the product in periods $1, \ldots, n$ are independent random variables with known distributions. At the beginning of each period i, the manager first observes the initial inventory x on hand in period i and the nonnegative production z in period $i-1$. The manager then chooses the nonnegative amount z' to produce in period i. There are costs $c_i(z')$ of production, $d_i(z' - z)$ of changing the rate of production, and expected cost $G_i(x + z')$ of storage and shortage in period i. The manager satisfies demands in period i from the starting stock $y = x + z$ as far as possible, and backorders any remaining excess demand. Assume the c_i, d_i, and G_i are convex; $c_i(|z|)$, $d_i(z)$, and $G_i(z)$ converge to ∞ as $|z| \to \infty$; and all relevant expectations exist and are finite. Let $C_i(x, z)$ be the minimum expected cost in periods $i, \ldots, n$ when x is the initial inventory in period i and z is the production in period $i-1$, $i = 1, \ldots, n$ ($C_{n+1} \equiv 0$).

(a) Dynamic-Programming Recursion. Give a dynamic-programming recursion expressing C_i in terms of C_{i+1} and from which one can find the least optimal production level $z_i(x,z)$ in period i when x is the initial inventory in period i and z is the production in period $i-1$.

(b) Subadditivity of Dual of Superadditive Function. Show that if a real-valued superadditive function f on a rectangle in $\Re^2$ is convex in its first coordinate, then the dual $f^{\#}$ is subadditive.

(c) Superadditivity and Convexity of Minimum Expected Cost. Show by induction that C_i is convex and superadditive, and its dual is subadditive. [*Hint*: Make the change of variables $x' = -x$ and show that $C_i'(x', z) \equiv C_i(-x', z)$ is subadditive in (x', z).]

(d) Monotonicity of Optimal Starting Stock and Reduction in Production. Show that the quantities $x + z_i(x,z)$ and $z - z_i(x,z)$ are increasing in x and $z \geq 0$. Give an intuitive rationale for these results.

3. Purchasing with Limited Supplies. An inventory manager seeks an ordering policy for a product that minimizes the expected n-period costs of ordering, storage and shortages with limited future supplies. The demands $D_1, \ldots, D_n$ for the product in periods $1, \ldots, n$ are independent random variables with known distributions. At the beginning of each period i, the manager observes the initial inventory x of the product in period i and purchases a nonnegative amount z not exceeding the given supply s_i in the period. The manager then satisfies demands in period i from the starting stock $y = x + z$ as far as possible and backorders any remaining excess demand. There is a convex cost $c(z)$ of ordering $z \geq 0$ units in the period and unit costs $h > 0$ and $p > 0$ of storage and shortage respectively at the end of the period. Assume that all relevant expectations exist and are finite, and that the appropriate functions are continuous.

Let $z(x, S_i)$ be the least optimal amount to purchase in period i given that x is the initial inventory in the period and $S_i = (s_i, \ldots, s_n) \geq 0$ is the vector of supplies in periods $i, \ldots, n$, and let $C(x, S_i)$ be the associated minimum expected cost in periods $i, \ldots, n$. Assume that there are no costs after period n, so $C(\,\cdot\,, S_{n+1}) \equiv 0$.

(a) Dynamic-Programming Recursion. Give a dynamic programming recursion expressing $C(\,\cdot\,, S_i)$ in terms of $C(\,\cdot\,, S_{i+1})$ from which it is possible to finding $z(x, S_i)$.

(b) Convexity and Superadditivity of Minimum Expected Cost. Show that $C(x, S_i)$ is convex in (x, S_i) and superadditive in (x, s_j) for $1 \leq i \leq j \leq n$. [*Hint*: The cases $j = i$ and $j > i$ require different arguments. Also, it is *not* so that $C(x, S_i)$ is superadditive in (x, S_i).]

(c) Monotonicity of Optimal Policy. Show that $z(x, S_i)$ is increasing in s_i and decreasing in (x, s_j) for $i < j \leq n$. Also show that $s_i - z(x, S_i)$ is increasing in s_i for $1 \leq i \leq n$. Establish the same results for the case of deterministic demands using the theory of substitutes and complements in network flows. Explain these results intuitively.

Answers to Homework 8 Due December 2

1. Total Positivity

(a) Characterization of TP_2 Functions. Suppose f is nonnegative on $X \times Y$. Then f is TP_2 if for each $x_1 < x_2$ in X and $y_1 < y_2$ in Y,

$$\begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{vmatrix} \geq 0,$$

or equivalently

$$f(x_1, y_1) f(x_2, y_2) \geq f(x_1, y_2) f(x_2, y_1). \tag{1}$$

Also, f has monotone likelihood ratio if

$$\frac{f(x_2, y_2)}{f(x_2, y_1)} \geq \frac{f(x_1, y_2)}{f(x_1, y_1)} \tag{2}$$

for each $x_1 < x_2$ in X and $y_1 < y_2$ in Y for which both of the above ratios are well defined.

We first show that f is TP_2 if and only if f has monotone likelihood ratio. To see this, observe that either (a) $f(x_1, y_1), f(x_2, y_1) > 0$, (b) $f(x_2, y_1) = 0$, or (c) $f(x_2, y_1) > 0$ and $f(x_1, y_1) = 0$. If (a) holds, (1) and (2) are equivalent. If (b) holds, then (1) holds and either the left-hand side of (2) is $+\infty$ or is undefined, so either (2) holds or is undefined. If (c) holds, then (1) holds if and only if $f(x_1, y_2) = 0$, and that is so if and only if (the right-hand side of) (2) is undefined. This establishes the claim.

Finally, f is TP_2 if and only if $\ln f$ is superadditive on $X \times Y$ because (1) is equivalent to

$$\ln f(x_1, y_1) + \ln f(x_2, y_2) \geq \ln f(x_1, y_2) + \ln f(x_2, y_1).$$

(b) Stochastic Monotonicity. If f is TP_2, then for all $x_1 < x_2$ in X and $y_1 < y_2$ in Y,

$$f(x_1, y_1) f(x_2, y_2) \geq f(x_1, y_2) f(x_2, y_1).$$

Suppose $w \in \Re$. Integrating this inequality with respect to $x_1 \in (-\infty, w]$ and $x_2 \in [w, +\infty)$ yields

$$F(w, y_1)[1 - F(w, y_2)] \geq F(w, y_2)[1 - F(w, y_1)],$$

or equivalently, $F(w, y_1) \geq F(w, y_2)$. Thus $F(\,\cdot\,, y)$ is stochastically increasing in y on Y.

(c) Gamma Density. Evidently

$$\ln f(x, r, \lambda) = \begin{cases} -\infty & \text{if } x \leq 0 \\ \ln \lambda - \ln \Gamma(r) + (r-1)(\ln \lambda + \ln x) - \lambda x & \text{if } x > 0 \end{cases}$$

First fix $\lambda > 0$ and show that $\ln f(\,\cdot\,,\,\cdot\,,\lambda)$ is superadditive on the sublattice $\{(x,r) : r \geq 1, x \in \Re\}$. It suffices to establish that fact on the sublattice $L \equiv \{(x,r) : r \geq 1,\ x > 0\}$ where $\ln f(\,\cdot\,,\,\cdot\,,\lambda)$ is finite. On L, $\ln f(x,r,\lambda) = r\ln x + g(r) + h(x)$ for some functions g and h. The first term is superadditive since it is a product of two increasing functions, each of a single distinct variable. The other terms are additive. Since superadditive functions are closed under addition, $\ln f(\,\cdot\,,\,\cdot\,,\lambda)$ is superadditive on L. Thus by (a), $f(\,\cdot\,,\,\cdot\,,\lambda)$ is TP_2 on L, so by (b), $f(\,\cdot\,,r,\lambda)$ is stochastically increasing in $r \geq 1$.

Now fix $r \geq 1$ and let $X_\lambda \geq 0$ be a gamma random variable whose distribution depends on the scale parameter $\lambda^{-1} > 0$. Then X_λ has density $f(\,\cdot\,,r,\lambda)$. Since nonnegative random variables increase stochastically with their scale parameters, $f(\,\cdot\,,r,\lambda)$ is stochastically increasing in $\lambda^{-1} > 0$, and so stochastically decreasing in $\lambda > 0$.

Alternately, it is possible to establish the last result by showing that $\ln f(\,\cdot\,,r,\,\cdot\,)$ is subadditive on the positive plane P, a sublattice. Evidently $\ln f(x,r,\lambda) = -\lambda x + g(x) + h(\lambda)$ for some functions g and h. The first term is subadditive since it is the negative product of two variables. The other terms are additive. Since subadditive functions are closed under addition, $\ln f(\,\cdot\,,r,\,\cdot\,)$ is subadditive on P. Thus by (a), $f(x,r,\lambda)$ is TP_2 in (x,λ^{-1}) for $\lambda > 0$, so by (b), $f(\,\cdot\,,r,\lambda)$ is stochastically decreasing in $\lambda > 0$.

(d) Convexity Preservation by TP_3 Densities. Show first that g is convex on X if and only if for each affine function a on X, the function $g - a$ changes sign at most twice on X and, when there are two sign changes, they are from $+$ to $-$ to $+$. To establish the necessity, suppose to the contrary that g is convex on X and for some affine function a on X and some $x < y < z$ in X, the function $h \equiv g - a$ satisfies

$$h(x) < 0,\ h(y) > 0 \text{ and } h(z) < 0. \tag{3}$$

Then on choosing $0 < \alpha < 1$ so $\alpha x + (1-\alpha)z = y$, it follows that

$$h(\alpha x + (1-\alpha)z) > \alpha h(x) + (1-\alpha)h(z) \tag{4}$$

since the left- and right-hand sides of (4) are respectively positive and negative. Thus h, and hence $g = h + a$, is not convex on X, which is impossible. Conversely, suppose g is not convex on X. Then there exist numbers $x < z$ in X and $0 < \alpha < 1$ such that $y \equiv \alpha x + (1-\alpha)y \in X$ and (4) holds with $h \equiv g$. Choose $\epsilon > 0$ small enough so (4) continues to hold when ϵ is added to both $h(x)$ and $h(z)$. Let a be the affine function on X whose graph passes through the points $(x, h(x){+}\epsilon)$ and $(z, h(z){+}\epsilon)$. By definition of ϵ, a lies below h at y. Thus, $h' \equiv g - a$ satisfies (2) (with h' replacing h), contradicting the fact that two sign changes must be from $+$ to $-$ to $+$.

For the second part, suppose g is convex. It suffices to show that for every affine function A on Y, the function $G - A$ changes sign at most twice on Y; and when there are two sign changes, they are from $+$ to $-$ to $+$. To see this, observe first that there exist numbers α', β' such that $A(y) = \alpha' + \beta' y$ for $y \in Y$. Now since $\alpha \neq 0$ and $\beta \neq 0$ by hypothesis in (1) of the problem statement, the function $a(x) \equiv \dfrac{\alpha'}{\alpha} + \dfrac{\beta'}{\beta} x$ is well defined on X. Then from (1), it follows that

$$A(y) = \int_X a(x) f(x,y)\, d\sigma(x)$$

for all $y \in Y$. Thus,

$$G(y) - A(y) = \int_X [g(x) - a(x)] f(x,y)\, d\sigma(x)$$

for $y \in Y$, so because $g - a$ changes sign at most twice on X and when there are two sign changes they are from $+$ to $-$ to $+$, the same is so of $G - A$ on Y since f is TP_3 and TP_3 functions preserve this property. Thus, G is convex from the result of the first part.

2. Production Smoothing.

(a) Dynamic-Programming Recursion. The dynamic programming recursion is

$$(1) \quad C_i(x,z) = \min_{z'} \{\delta_+(z') + c_i(z') + d_i(z' - z) + G_i(x + z') + \mathrm{E}C_{i+1}(x + z' - D_i, z')\},$$

for all x, $z \geq 0$ and $i = 1, \ldots, n$, where $C_{n+1} \equiv 0$.

(b) Subadditivity of Dual of Superadditive Function. Suppose $x < x'$ and $y < y'$. Then by the convexity of $f(\,\cdot\,, z)$ for each z and the superadditivity of f,

$$\begin{aligned} f^{\#}(x',y) + f^{\#}(x,y') &= f(y - x', y) + f(y' - x, y') \\ &= [f(y - x', y) + f(y' - x, y)] + [f(y' - x, y') - f(y' - x, y)] \\ &\geq [f(y - x, y) + f(y' - x', y)] + [f(y' - x', y') - f(y' - x', y)] \\ &= f(y - x, y) + f(y' - x', y') = f^{\#}(x,y) + f^{\#}(x',y'). \end{aligned}$$

Thus the *left dual* $f^L \equiv f^{\#}$ of f is subadditive as claimed. By symmetry, if f is superadditive and $f(z,\,\cdot\,)$ is convex for each z, the *right dual* $f^R(x,y) \equiv f(x, x - y)$ of f is also subadditive.

(c) Superadditivity and Convexity of Minimum Cost. The proof that C_i is convex and superadditive is by induction. This is certainly so for $C_{n+1} = 0$. Suppose C_{i+1} is convex and superadditive with $i \leq n$ and consider i. Set $C_i'(x', z) \equiv C_i(-x', z)$ where $x' = -x$, and rewrite (1) as

$$(2) \qquad C_i'(x',z) = \min_{z'} \{\delta_+(z') + c_i(z') + d_i(z' - z) + G_i(z' - x') + \mathrm{E}C_{i+1}^L(x' + D_i, z')\}.$$

Since C_{i+1} is convex and superadditive, its left dual C_{i+1}^L is convex and subadditive by (b), so $\mathrm{E}C_{i+1}^L(x' + D_i, z')$ is convex and subadditive in (x', z'). Also the remaining terms of the expression $B(z', z, x')$ in braces on the right-hand side of (2) are convex functions of a single variable or are convex functions of differences of two variables, so they are convex and subadditive in (z', z, x').

Hence B is convex and superadditive. Thus from the Projections-of-Convex-and-Subadditive-Functions Theorems, C_i' is convex and subadditive, whence C_i is convex and superadditive.

(d) Monotonicity of Optimal Starting Stock and Reduction in Production. Now $B(\,\cdot\,,\,\cdot\,,x')$ is doubly subadditive, so $z_i(x,z)$ and $z-z_i(x,z)$ are increasing in z. It remains to show that $y_i(x,z) \equiv x+z_i(x,z)$ and $x-y_i(x,z)=-z_i(x,z)$ are increasing in x. The latter follows from (2) because $B(\,\cdot\,,z,\,\cdot\,)$ is subadditive. It remains to show the former. To that end, rewrite (1) as

$$(3) \qquad C_i(x,z)=\min_y\{\delta_+(y-x)+c_i(y-x)+d_i(y-z-x)+G_i(y)+\mathrm{E}C_{i+1}^R(y-D_i,x-D_i)\}.$$

Since C_{i+1} is superadditive by (c), its right dual C_{i+1}^R is subadditive by (b), so $\mathrm{E}C_{i+1}^R(y-D_i,x-D_i)$ is subadditive in (x,y). Also, the remaining terms in braces on the right-hand side of (3) are all subadditive in (x,y) since they are either functions of y or convex functions of $y-x$, so the expression is braces is subadditive in (x,y). Hence by the Increasing-Optimal-Selections Theorem, $y_i(x,z)$ is increasing in x.

The intuition behind these results is as follows. As initial inventory rises, the marginal cost of starting inventory (resp., production) at a given level falls (resp., rises), so it is beneficial to raise starting inventory (resp., lower production). Also, as prior production rises, the marginal cost of production at a given level falls, so it is beneficial to raise current production. On the other hand, if production rises by more than prior production, the marginal cost of production rises, so it is beneficial to increase production by less than prior production rises.

3. Purchasing with Limited Supplies. A supply manager seeks an ordering policy for a product that minimizes the expected n-period costs of ordering, storage and shortage with limited future supplies $s_1,\ldots,s_n$ in those periods. The demands $D_1,\ldots,D_n$ for the product in periods $1,\ldots,n$ are independent random variables with known distributions. There is a convex cost $c(z)$ of ordering $z\geq 0$ units in the period and unit costs $h>0$ and $p>0$ of storage and shortage respectively at the end of the period. Let $g(z)=hz^+ + pz^-$ and $G_i(y)=\mathrm{E}g(y-D_i)$. Let $z(x,S_i)$ be the least optimal amount to purchase in period i given that x is the initial inventory in the period and $S_i=(s_i,\ldots,s_n)\geq 0$ is the vector of supplies in periods $i,\ldots,n$. Let $C(x,S_i)$ be the associated minimum expected cost in periods $i,\ldots,n$. Assume that there are no costs after period n.

(a) Dynamic-Programming Recursion. The dynamic-programming recursion for finding an optimal supply policy is

$$(1) \qquad C(x,S_i)=\min_{0\leq z\leq s_i}[c(z)+G_i(z+x)+\mathrm{E}C(z+x-D_i,S_{i+1})]$$

for $i=1,\ldots,n$ where $C(\,\cdot\,,S_{n+1})=0$.

(b) Convexity and Superadditivity of Minimum Expected Cost. The function $C(x,S_i)$ is convex in (x,S_i) and superadditive in (x,s_j) for $1\leq i\leq j\leq n$.

Convexity. Clearly $C(x,S_{n+1})=0$ is convex. Suppose $C(x,S_{i+1})$ is convex and $1<i\leq n$. Then the bracketed term on the right-hand side of (1) is convex because c and G_i are convex, and

convex functions of affine functions and expectations of convex functions are convex. Also, the constraint set $0 \le z \le s_i$ is convex. Hence $C(x, S_i)$ is convex by the Projection-of-Convex-Functions Theorem.

Superadditivity. Clearly $C(x, S_{n+1}) = 0$ is superadditive in (x, s_j) for $n+1 \le j \le n$ (vacuously). Suppose $C(x, S_{i+1})$ is superadditive in (x, s_j) for $i+1 \le j \le n$.

Case 1: $j = i$. The bracketed term on the right-hand side of (1) is subadditive in $(z, -x, s_i)$ because that term is independent of s_i, $c(z)$ is additive, and $G_i(z+x) + \mathrm{E}C(z+x-D_i, S_{i+1})$ is subadditive in $(z, -x)$. The last sum is subadditive because it is convex in $z + x = z - (-x)$ since $G_i(\,\cdot\,)$ and $C(\,\cdot\,, S_{i+1})$ are convex and a convex function of the difference of two variables is subadditive therein. Also, $0 \le z \le s_i$ is a sublattice in $(z, -x, s_i)$. Thus by the Projection-of-Subadditive-Functions Theorem, $C(x, S_i)$ is subadditive in $(-x, s_i)$, and so superadditive in (x, s_i).

Case 2: $j > i$. Put $y = z + x$, so (1) becomes

$$C(x, S_i) = \min_{0 \le y - x \le s_i} [c(y-x) + G_i(y) + \mathrm{E}C(y - D_i, S_{i+1})]. \tag{2}$$

Now the bracketed term on the right-hand side of (2) is subadditive in $(y, x, -s_j)$ since $c(y-x)$ is subadditive in (y, x) because c is convex and a convex function of the difference of two variables subadditive, $G_i(y)$ is additive, and, because $C(w, S_{i+1})$ is superadditive in (w, s_j), $\mathrm{E}C(y - D_i, S_{i+1})$ is superadditive in (y, s_j) and so subadditive in $(y, -s_j)$. Also, $0 \le y - x \le s_i$ is a sublattice in (y, x) and so, because that set is independent of s_j, in $(y, x, -s_j)$ as well. Thus, by the Projection-of-Subadditive-Functions Theorem, $C(x, S_i)$ is subadditive in $(x, -s_j)$, and so superadditive in (x, s_j).

(c) Monotonicity of Optimal Policy. The optimal order quantity $z(x, S_i)$ is increasing in s_i and decreasing in (x, s_j) for $i < j \le n$, and $s_i - z(x, S_i)$ is increasing in s_i for $1 \le i \le n$. The intuitive rationale for these results is as follows. The optimal order quantity $z(x, S_i)$ in period i rises as supply s_i in the period rises because there is more available, but falls as initial inventory in the period or future supplies rise because both are substitutes for ordering. However, $s_i - z(x, S_i)$ also rises with s_i reflecting the fact that orders and unused supply in a period are substitutes.

As we saw in (b), the right-hand side of (1) entails minimizing a subadditive function of $(z, -x, s_i)$ over a sublattice therein, so $z(x, S_i)$ is increasing in s_i and decreasing in x. Also, the right-hand side of (1) is doubly subadditive in (z, s_i), so $s_i - z(x, S_i)$ is increasing in s_i. Finally, the term in brackets on the right-hand side of (1) is superadditive in (z, s_j), $j \ge i+1$, and so is subadditive in $(z, -s_j)$, $j \ge i+1$. Thus, $z(x, S_i)$ is decreasing in s_j, $j \ge i+1$.

These results for the case of deterministic demands follows from the theory of substitutes and complements of network flows. To see this consider the extension of the production planning network in Figure 1 of §1.2 to n periods in which one coalesces the nodes $1, \ldots, i-1$ with node 0 and deletes loops, i.e., arcs with common head and tail. Then replace each s_j by D_j, each x_j by z_j and impose the upper bound s_j on z_j, and replace each y_j by x_j. Let x be a fixed value of x_i. Label the arcs by the labels of the flows in them. Then the arc z_i is a complement of itself and a substitute of the arcs x and s_j for $j > i$. Also, add the cost $\delta_+(s_j - z_j)$ to the cost $c(z_j)$ of arc z_j. Let $\delta_0(x - x_i)$ be the cost of arc x_i and let $g(x_j)$ be the cost of arc x_j for $j > i$. Since the arc costs are convex in their flows, the arc costs for arcs x_i and z_i are doubly subadditive, it follows from the Smoothing Theorem 6 of §4.5 that $z(x, S_i)$ is increasing in s_i and decreasing in (x, s_j) for $i < j \leq n$, and $s_i - z(x, S_i)$ is increasing in s_i for $1 \leq i \leq n$. In addition $x + z(x, S_i)$ is increasing in x.

Homework 9 Due December 9

If you have taken MS&E 251, do problems 2 and 3; otherwise do problems 1 and 3.

1. Optimality of (s, S) Policies. Call a real-valued function g on the real line K-*convex* if for some number $K \ge 0$, $K + g(y) - g(x) - (y - x)[g(x) - g(x - b)]/b \ge 0$ for all $y > x$ and $b > 0$. The following properties of K-convex functions are easy to verify; assume them in the sequel:

1° *Convexity.* Ordinary convexity is equivalent to 0-convexity.

2° *Translations.* If $g(x)$ is K-convex in x, so is $g(x + h)$ for any fixed h.

3° *Increase* K. If g is K-convex and $K \le L$, then g is L-convex.

4° *Positive Combinations.* If $g(y, u)$ is $K(u)$-convex in y for each fixed u and if Ψ is increasing, then $\int g(y, u)d\Psi(u)$ is $\int K(u)d\Psi(u)$-convex in y.

Now consider the inventory equation ($C_{n+1} \equiv 0$)

$$C_i(x) = \min_{y \ge x} \{c_i(y - x) + G_i(y) + \mathrm{E}C_{i+1}(y - D_i)\},\ i = 1, \ldots, n,$$

and each $x \in \Re$. Assume that $D_1, \ldots, D_n$ are independent random variables, $c_i(0) = 0$ and $c_i(z) = K_i$ for $z > 0$ and $1 \le i \le n$ with $K_1 \ge \cdots \ge K_n \ge K_{n+1} = 0$, and G_i is finite and convex with $G_i(y) \to \infty$ as $|y| \to \infty$ for each $1 \le i \le n$. It can be shown that $J_i(y) \equiv G_i(y) + \mathrm{E}C_{i+1}(y - D_i)$ is continuous in y for $1 \le i \le n$. Finally, suppose that $1 \le i \le n$.

(a) K_i-Convexity of J_i. Show that if C_{i+1} is K_{i+1}-convex, then J_i is K_{i+1}-convex, and hence also K_i-convex. (Of course, $C_{n+1} = 0$ is K_{n+1}-convex.)

(b) (s_i, S_i)-Optimal Policy. Show that if J_i is K_i-convex, then there is an (s_i, S_i) optimal policy in period i.

(c) K_i-Convexity of C_i. Show that if J_i is K_i-convex, then C_i is K_i-convex.

2. Supplying a Paper Mill. The Georgia-Pacific Corporation has a paper mill with known positive requirements $d_1, \ldots, d_n$ for cords of wood in weeks $1, \ldots, n$. These requirements must be met as they arise. The mill has two sources of supply, viz., its forest and the open market. Because of unpredictable weather conditions, the nonnegative maximum numbers $m_1, \ldots, m_n$ of cords of wood that could be cut from the forest in weeks $1, \ldots, n$ are independent nonnegative random variables with known distributions. Because the open market is usually expensive, it is used mainly to assure that weekly requirements at the mill are met. The mill maintains an inventory of wood to buffer random fluctuations in supply from the forest and to reduce the need to buy wood on the open market. In each week there is a convex increasing cost $p(w)$ (resp., $s(w)$) of cutting (resp., buying) w cords of wood from the forest (resp., open market). There is a convex increasing cost $h(w)$ of storing w cords of wood at the mill at the end of a week. The problem is to find a supply policy with minimum expected n-week cost that meets each week's requirements.

At the beginning of week i $(= 1, \ldots, n)$, the mill observes the number $x \geq 0$ of cords of wood on hand at the mill and maximum number $m_i = m$ of cords of wood that could be cut from the forest that week. The mill then chooses the numbers of cords of wood to cut from the forest and to buy on open market that week. Let $C_i(x, M)$ be the minimum expected cost in weeks $i, \ldots, n$ when the mill has $x \geq 0$ cords of wood on hand and $M \equiv x + m \geq x$ cords on hand or available from its forest in week i. There are no costs after period n, whence $C_{n+1}(\,\cdot\,,\,\cdot\,) \equiv 0$. Let $y_i(x, M)$ (resp., $z_i(x, M)$) be the least optimal number of cords of wood on hand in week i given (x, M) after ordering from its forest (resp., both the forest and open market) with immediate delivery, but before demand, in that week.

(a) Dynamic Programming Recursion. Give a dynamic-programming recursion for calculating $C_i(x, M)$ for $0 \leq x \leq M$ and $1 \leq i \leq n$.

(b) Subadditivity of $C_i(x, M)$. Show that $C_i(x, M)$ is subadditive in (x, M) for $0 \leq x \leq M$ and $1 \leq i \leq n$.

(c) Monotonicity of Optimal Policy. Show that $y_i(x, M)$, $z_i(x, M)$ and $y_i(x, M) - z_i(x, M)$ are increasing in (x, M) for $0 \leq x \leq M$ and $1 \leq i \leq n$. Also show that $x - y_i(x, M)$ is increasing in x for $0 \leq x \leq M$ and $M - y_i(x, M)$ is increasing in $M \geq 0$ for $1 \leq i \leq n$.

3. Optimal Supply Policy with Fluctuating Demands. A supply manager seeks a minimum-expected-cost ordering policy for a single product over the next n months. Demands $D_1, \ldots, D_n$ for the product in months $1, \ldots, n$ arise from aggregating requests of members of a homogeneous population of size N. Each member of the population requests a single unit of the product in month i with probability p_i independently of the requests of other individuals in the month and of all individuals in other months. At the beginning of each month, the manager observes the initial inventory x of the product in the month before any information about the demand for the month becomes available. The manager than orders $z \geq 0$. The cost $c(z) \geq 0$ for so doing is convex. Delivery is immediate. Demands in a month are met in so far as possible from stock on hand after delivery of orders in the month. Unsatisfied demands in a month are backordered. There is a unit cost $h > 0$ of storing product left over at the end of a month, and a unit cost $s > 0$ of each unit on backorder at the end of the month. Let $C(x, P_i)$ be the minimum expected cost in months $i, \ldots, n$ when x is the initial inventory of the product in month i and $P_i = (p_i, \ldots, p_n)$. Denote by $z(x, P_i)$ the corresponding least optimal order quantity and by $y(x, P_i) \equiv x + z(x, P_i)$ the corresponding least optimal starting stock in period i. Assume that all relevant expectations exist and are finite.

(a) Dynamic-Programming Recursion. Give a dynamic-programming recursion for finding an optimal ordering policy where $C(x, P_{n+1}) = 0$.

(b) Monotonicity of Optimal Ordering Policy. Discuss the monotonicity properties of $y(x, P_i)$ and $z(x, P_i)$ in x and p_j for each $i \le j \le n$. Justify your answers.

(c) Optimality of Myopic Ordering Policies. Suppose $c(z) = 0$ for all $z \ge 0$. Determine the optimal ordering policy as explicitly as possible for the case where individual customer demand is rising, i.e., $p_1 \le \cdots \le p_n$.

(d) Nonhomogeneous Population. Suppose the population is nonhomogeneous, so there are N types of customers. Each customer of type $j = 1, \ldots, N$ in a period independently requests a single unit of the product with probability p_{ij}. Briefly outline the generalization of the above results to this situation and indicate the key differences in the proofs.

Answers to Homework 9 Due December 9

1. Optimality of (s, S) Policies.

(a) K_i-Convexity of J_i. Let $H_{i+1}(y) \equiv \mathrm{E}C_{i+1}(y - D_i)$ and Φ_{i+1} be the distribution function of D_i. If C_{i+1} is K_{i+1}-convex, then $C_{i+1}(y-t)$ is K_{i+1}-convex in y for each fixed t by 2°. Thus, H_i is K_{i+1}-convex by 4° because $\int K_{i+1} d\Phi_{i+1}(t) = K_{i+1}$. Now since G_i is convex, it is 0-convex. Hence, by 4° again, $J_i = G_i + H_i$ is K_{i+1}-convex. And since $K_i \ge K_{i+1}$, J_i is K_i-convex by 3°.

(b) (s_i, S_i)-Optimal Policy. Since $G_i(y) \to \infty$ as $|y| \to \infty$ for each i, it is routine to check by backward induction on i that C_i is bounded below and $J_i(y) \to \infty$ as $|y| \to \infty$ for each i. Thus, since J_i is continuous, there is an S_i that minimizes J_i on the real line and a greatest $s_i \le S_i$ such that $J_i(s_i) = K_i + J(S_i)$. We claim that the (s_i, S_i) policy is optimal in period i.

Case 1. $x < s_i$. First, show that $J_i(x) \ge K_i + J_i(S_i)$, so it is optimal to order to S_i. If not, there exists an $\overline{x} < s_i$ such that $J_i(\overline{x}) < K_i + J_i(S_i)$. If $s_i = S_i$, then $J_i(\overline{x}) < J_i(S_i)$, which contradicts the fact S_i minimizes J_i. If instead $s_i < S_i$, then $J_i(s_i) - J_i(\overline{x}) = K_i + J_i(S_i) - J_i(\overline{x}) > 0$. Hence,

$$K_i + J_i(S_i) - J_i(s_i) - (S_i - s_i)\left[\frac{J_i(s_i) - J(\overline{x})}{s_i - \overline{x}}\right] = -(S_i - s_i)\left[\frac{J_i(s_i) - J_i(\overline{x})}{s_i - \overline{x}}\right] < 0,$$

which contradicts the K_i-convexity of J_i.

Case 2. $s_i \le x \le S_i$. By definition of s_i and the continuity of J_i, it follows that $K_i + J_i(S_i) \ge J_i(x)$. Thus, it is optimal not to order.

Case 3. $S_i < x$. Now show that $K_i + J_i(y) \ge J_i(x)$ for all $y > x$, so it is optimal not to order. If not, there is a $y > x$ such that $K_i + J_i(y) < J_i(x)$. Also, $J_i(x) \ge J_i(S_i)$ because S_i minimizes J_i. Thus,

$$[K_i + J_i(y) - J_i(x)] - (y - x)\left[\frac{J_i(x) - J_i(S_i)}{x - S_i}\right] < 0,$$

which contradicts the K_i-convexity of J_i.

(c) K_i-Convexity of C_i. From part (b), $C_i(x) = J_i(x \vee s_i)$. Suppose $y > x$ and $b > 0$.

Case 1. $s_i \le x - b$. Since $C_i = J_i$ on $[s_i, \infty)$, the K_i-convexity inequality for C_i follows from that for J_i.

Case 2. $x - b < s_i \le S_i < x$. Since $C_i = J_i$ on $[s_i, \infty)$, $C_i(x-b) = J_i(s_i) \ge J_i(S_i)$, $J_i(x) \ge J_i(S_i)$ and $b > x - S_i > 0$, it follows from the K_i-convexity of J_i that

$$C_i(y) - C_i(x) - (y-x)\left[\frac{C_i(x) - C_i(x-b)}{b}\right] \ge J_i(y) - J_i(x) - (y-x)\left[\frac{J_i(x) - J_i(S_i)}{x - S_i}\right] \ge -K_i.$$

Case 3. $x - b < s_i \le x \le S_i$. Since $C_i(y) - C_i(x) = J_i(y) - J_i(x)$, $C_i(x-b) = J_i(s_i) \ge J_i(x)$ $= C_i(x)$ and $J_i(x) \le J_i(y) + K_i$, it follows that

$$C_i(y) - C_i(x) - (y-x)\left[\frac{C_i(x) - C_i(x-b)}{b}\right] \ge J_i(y) - J_i(x) \ge -K_i.$$

Case 4. $x < s_i \le y$. Clearly $C_i(x-b) = C_i(x) = J_i(s_i)$ and $C_i(y) = J_i(y)$, so

$$K_i + C_i(y) - C_i(x) - (y-x)\left[\frac{C_i(x) - C_i(x-b)}{b}\right] = K_i + J_i(y) - J_i(s_i) \ge 0.$$

Case 5. $y < s_i$. Clearly $C_i(x-b) = C_i(x) = C_i(y) = J_i(s_i)$, so

$$K_i + C_i(y) - C_i(x) - (y-x)\left[\frac{C_i(x) - C_i(x-b)}{b}\right] = K_i \ge 0.$$

2. Supplying a Paper Mill. At the beginning of each week $i = 1, \ldots, n$, the mill observes the number $x \ge 0$ of cords of wood on hand and maximum number m_i of cords of wood that could be cut from the primary source that week. The mill then chooses the numbers of cords of wood to cut from the primary source and to buy from the secondary source that week. Let $C_i(x, M)$ be the minimum expected cost in weeks $i, \ldots, n$ when the mill has $x \ge 0$ cords of wood on hand and $M \equiv x + m \ge x$ on hand or available from the primary source in week i. No costs are incurred after period n, whence $C_{n+1}(\,\cdot\,,\,\cdot\,) \equiv 0$. Given (x, M) in week i, let $y_i(x, M)$ (resp., $z_i(x, M)$) be the least optimal number of cords of wood on hand after ordering from the primary source (resp., both the primary and secondary sources) with immediate delivery, but before demand, in that week.

(a) Dynamic Programming Recursion. Let y be the sum of the number of cords of wood initially on hand in week i and the amount that will be cut from the primary source that week. Let z be the sum of y and the amount that will be bought from the secondary source that week. Then

$$(1) \quad C_i(x, M) = \min_{\substack{y \le M \\ x \le y \le z \\ d_i \le z}} \left[s(z-y) + p(y-x) + h(z-d_i) + \mathrm{E}C_{i+1}(z-d_i, z-d_i+m_{i+1})\right]$$

for $i = 1, \ldots, n$ where $C_{n+1}(\,\cdot\,,\,\cdot\,) \equiv 0$.

(b) Subadditivity of $C_i(x, M)$. The proof that C_i is subadditive is by induction on i. Since this is trivially so for $i = n+1$, suppose it is so for $i+1$ and consider i. Since s and p are convex, and a convex function of the difference of two variables is subadditive, the first two terms in brackets on the right side of (1) are subadditive. The other two terms in brackets depend only on the single variable z and so are additive, and thus subadditive. Since subadditive functions are closed under addition, the sum of the terms in brackets on the right side of (1) is subadditive in (x, y, z, M). Also, the constraints on the right side of (1) form a sublattice in those variables. Consequently, by the projection theorem for subadditive functions, C_i is subadditive.

(c) Monotonicity of Optimal Policy. The first step is to show by induction on i that C_i is convex. Since this is trivially so for $i = n+1$, suppose it is so for $i+1$ and consider i. Since s, p, h and C_{i+1} are convex, convex functions of affine functions are convex and expectations of random translates of convex functions are convex, the bracketed term on the right side of (1) is convex in (x, y, z, M). The constraints on the right-hand side of (1) form a polyhedral convex set in those same variables. Consequently, by the Projection Theorem for Convex Functions, C_i is convex.

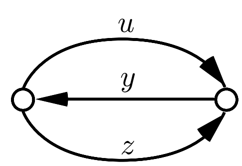
Substitutes and Complements in Network Flows. The easiest way to establish the monotonicity results is to apply the theory of substitutes and complements in network flows. To that end, put $H(z) \equiv h(z - d_i) + \mathrm{E}C_{i+1}(z - d_i, z - d_i + m_{i+1})$ and the introduce the variable $u \equiv y - z$ because one goal is to show that one optimal u is increasing in (x, M). Then the problem of minimizing the right-hand side of (1) amounts to choosing u, y, z to minimize

(2) $$[s(-u) + \delta_+(-u)] + [p(y - x) + \delta_+(y - x) + \delta_+(M - y)] + [H(z) + \delta_+(z - d_i)]$$

subject to

$$\begin{aligned} u + z - y &= 0 \\ -u - z + y &= 0. \end{aligned}$$

The first of these equations is a restatement of the definition of u and the second equation is the negative of the first. The resulting system is a network-flow problem with two nodes and three arcs as the following figure illustrates.



Note from (2) that the (bracketed) flow cost on each arc is convex in its flow. Also, the flow cost in arc y is doubly subadditive in both (y, x) and (y, M). Moreover, u, y and z are complements of y. Thus from the Smoothing Theorem 6 of §4.5, there is an optimal flow selection $u_i(x, M)$, $y_i(x, M)$ and $z_i(x, M)$ with the properties that $u_i(x, M) = y_i(x, M) - z_i(x, M)$, $y_i(x, M)$, $z_i(x, M)$, are increasing in (x, M), $x - y_i(x, M)$ is increasing in x and $M - y_i(x, M)$ is increasing in M.

Direct Lattice Programming. It is also possible to prove these results directly from lattice programming by using the Increasing Optimal Selections Theorem 8 of §2.5 and changing variables several times. To simplify the discussion, it is convenient to introduce the (convex) functions $e \equiv s + \delta_+$, $f \equiv p + \delta_+$ and $g(z) \equiv H(z) + \delta_+(z - d_i)$ for all z.

Replace u in (2) by $y - z$. Then (2) becomes

(2)′ $$e(z - y) + f(y - x) + \delta_+(M - y) + g(z).$$

This function is subadditive in (x, y, z, M), so $y_i(x, M)$ and $z_i(x, M)$ are increasing in (x, M).

Let $w = x - y$ and $v = w + z$, and use them to eliminate y and z from (2)′. Then (2)′ becomes

$$e(v - x) + f(-w) + \delta_+(M - x + w) + g(v - w).$$

This function is subadditive in (v, w, x), so $w_i(x, M) = x - y_i(x, M)$ in increasing in x.

Let $p = M - y$ and $q = p + z$, and use them to eliminate y and z from (2)′. Then (2)′ becomes

$$e(q - M) + f(M - p - x) + \delta_+(p) + g(q - p).$$

This function is subadditive in (p, q, M), so $p_i(x, M) = M - y_i(x, M)$ in increasing in M.

Replace z in (2) by $y-u$. Then (2) becomes

$$e(-u)+f(y-x)+\delta_+(M-y)+g(y-u).$$

This function is subadditive in (u,x,y,M), whence $u_i(x,M)=y_i(x,M)-z_i(x,M)$ is increasing in (x,M).

3. Optimal Supply Policy with Fluctuating Demands. Let D_i be the total demand in period i. Let $g(z)=hz^+ + sz^-$ and $G(y,p_i)\equiv \mathrm{E}g(y-D_i)$ be the conditional expected storage and shortage cost in period i given that starting stock on hand after receipt of orders in period i is y. Let $c(z)$ be the convex cost of ordering $z\ge 0$ in a period. Let $C(x,P_i)$ be the minimum expected cost in months $i,\dots,n$ when x is the initial inventory of the product in month i and $P_i=(p_i,\dots,p_n)$ is the vector of probabilities that an individual requests the product in periods $i,\dots,n$. Denote by $z(x,P_i)$ the corresponding least optimal order quantity and by $y(x,P_i)\equiv x+z(x,P_i)$ the corresponding least optimal starting stock in period i.

(a) Dynamic-Programming Recursion. The dynamic-programming recursion for finding an optimal ordering policy is

$$C(x,P_i)=\min_{y\ge x}\left[c(y-x)+G(y,p_i)+\mathrm{E}C(y-D_i,P_{i+1})\right] \tag{1}$$

for $i=1,\dots,n$ where $C(x,P_{n+1})=0$.

(b) Monotonicity of Optimal Ordering Policy. The optimal starting stock $y(x,P_i)$ in period i is increasing in x and P_i while the optimal order quantity $z(x,P_i)$ in period i is decreasing in x and increasing in P_i. These monotonicity results in x were shown in §8.2 of *Lectures on Supply-Chain Optimization*. In addition is was shown there that $C(\,\cdot\,,P_i)$ is convex in x.

It remains only to show that $y(x,P_i)$ is increasing in P_i since then $z(x,P_i)=y(x,P_i)-x$ is increasing in P_i. The proof entails showing that $C(x,P_i)$ is subadditive in (x,p_j) for each $i\le j\le n$. This is vacuously true for $i=n+1$. Thus suppose the claim is so for $i+1$ and consider i. It is necessary to show that the bracketed term on the right-hand side of (1) is subadditive in (x,y,p_j) for $i\le j\le n$. Now $c(y-x)$ is subadditive in (x,y) because c is convex. Also, D_i is binomially distributed with parameters p_i and N, and so is stochastically increasing in p_i by the Equivalence-of-Stochastic-and-Pointwise-Monotonicity Theorem 2 of §7.2. Thus, by the Subadditivity-Preservation-by-Stochastic-Monotonicity Corollary 5 of §7.3, $G(y,p_i)$ is subadditive in (y,p_i) since g is convex and D_i is stochastically increasing in p_i. Similarly, $\mathrm{E}C(y-D_i,P_{i+1})$ is subadditive in (y,p_i) since $C(\,\cdot\,,P_{i+1})$ is convex and D_i is stochastically increasing in p_i. Moreover, by the induction hypothesis, $C(w,P_{i+1})$ is subadditive in (w,p_j) for $i+1\le j\le n$, so $\mathrm{E}C(y-D_i,P_{i+1})$ is subadditive in (y,p_j) as well. Since sums of subadditive functions are subadditive, the bracketed term on the right-hand side of (1) is subadditive in (x,y,p_j) for $i\le j\le n$. Also, since the minimum is over the sublattice $x\le y$, it follows from the Increasing-Optimal-Selections-Theorem 8 of §2.5 that $y(x,P_i)$ is increasing in P_i and from the Projections-of-Subadditive-Functions Theorem 9 that $C(x, P_i)$ is subadditive in (x, p_j) for each $i \le j \le n$.

(c) Optimality of Myopic Ordering Policies. Since individual customer demand is rising, i.e., $p_1 \le \cdots \le p_n$, it follows that $G(y, p_i)$ is subadditive in (y, i) and the least minimizer $y = y(p_i)$ of $G(y, p_i)$ is increasing in i. Also, $y(p_i) - D_i \le y(p_{i+1})$ for each i since D_i is nonnegative. Thus, since $c(z) = 0$ for all $z \ge 0$, the optimal starting stock $y(x, P_i)$ in each period i is myopic and base-stock with $y(x, P_i) = x \vee y(p_i)$ as §8.3 and §9.2 discuss.

(d) Nonhomogeneous Population. Let $p_i = (p_{i1}, \dots, p_{iN})$ be the vector probabilities that each of the N individuals wishes to purchase the product in period i. Then all the above notation and results carry over at once to this new situation. The only changes required are the following. Observe first that although D_i is not binomially distributed, D_i is stochastically increasing in p_i by the Equivalence-of-Stochastic-and-Pointwise-Monotonicity Theorem 2 of §7.2. Also in part (b), one must show that $C(x, P_i)$ is subadditive in (x, p_{jk}) for each $i \le j \le n$ and $1 \le k \le N$ and in both parts (b) and (c), one must show that $G(y, p_i)$ is subadditive in (y, p_{ik}) for each $1 \le k \le N$. With these changes, the proofs are the same as before.