Institute of Transportation Studies
University of California at Berkeley

On the Stability of Supply Chains

Carlos F. Daganzo

RESEARCH REPORT
UCB-ITS-RR-2002-2

March 2002
ISSN 0192 4095

On the Stability of Supply Chains

Carlos F. Daganzo
Institute of Transportation Studies and Department of Civil and Environmental Engineering
University of California, Berkeley CA 94720
(March 13, 2002)

ABSTRACT

This paper examines the stability of autonomous supply chains. In an autonomous supply chain, suppliers only communicate with their downstream neighbor; i.e., their immediate customer. The inventory management policies used by the suppliers characterize a chain. An important property of a policy is its "gain", which relates marginal changes in the average steady state inventory to small changes in the steady state demand rate. It is shown that *all* autonomous policies with positive gains are unstable when applied to a homogeneous supply chain if they do not use future order commitments. This generates the so-called "bullwhip" effect. Stability conditions for policies of all types are given.

The paper also discusses a family of stable, autonomous policies that can dynamically maintain any desired inventory level for any demand rate. The family includes just-in-time strategies as a special case. If the target inventories are admissible, i.e. they are large enough to prevent stock-outs in the steady state (with some margin of safety) then the proposed policies avoid stock-outs in the dynamic case. Simulations are used as an illustration.

INTRODUCTION

A supply chain is a network of suppliers that produce goods, both, for one another and for generic customers. Goods travel from origin-suppliers to destination-customers, possibly visiting intermediate suppliers and being altered or recombined in the process. Conservation rules at each supplier, define its outputs as a function of its inputs. The rates at which goods of different types flow over this network depend on the customer demand, the flow of information (orders) across suppliers, and on the algorithms/policies that the suppliers use to place orders and replenish their inventories. This paper examines the stability of serial supply chains with a unique item type, where each supplier j receives all its goods from $j+1$ and ships them to $j-1$; i.e., on what in the inventory control literature would be called a multi-echelon, serial inventory system. It focuses on autonomous chains where j only receives order information from $j-1$. The word "autonomous" is used because the policies of these systems must be local in nature—only using information from a neighbor. The paper has been extracted from a more comprehensive document that analyzes the stability and optimization of more general multi-commodity supply networks (Daganzo, 2001). The extensions to multi-commodity networks developed in this reference are not described here, nor is the optimization theory.

The instability of autonomous chains is a well-known phenomenon; see, for example, the discussion in Simchi-Levi et al. (2000), pp. 82-102. A "beer game" is even used in business schools to illustrate it, and the name "bullwhip effect" has been coined to describe it—perhaps because the suppliers farthest away from the customers feel as if they were at the end of a bullwhip. By instability, it is meant the increased variability in the order sizes and the inventory levels experienced by the suppliers as j increases. The bullwhip effect is so prevalent that the aforementioned beer-game does not have to be structured very carefully, and it invariably leads to instabilities (Erera, 2001).

Despite its importance, our understanding of the bullwhip effect is still incomplete. Although there have been some limited successes in analyzing special cases, e.g., showing that the variance of order size increases exponentially with j for periodic review algorithms with stationary demand (Ryan, 1997), a general theory of supply chains with arbitrary demand inputs has not yet been put forward. This paper attempts to fill this gap. It adds to current knowledge by presenting a systematic analysis of all possible algorithms under general demand conditions—stochastic or deterministic; time-dependent or stationary; ergodic or not. This goal requires that stability be defined as an intrinsic property of an algorithm, independent of the demand. Two definitions are used: "*weak*" and "*strong*" stability. Weak stability means that any infinitesimal perturbation in the customer demand input when the system is in a steady state generates order size and inventory perturbations that are bounded and infinitesimal, as $j \to \infty$. This must be true for all possible steady states and perturbation types. Strong stability means that, in addition, the order sizes and the inventories must be bounded across all j and all times, even as $j \to \infty$, *for any imaginable realization of the* *demand process*. (Clearly, strong stability implies that the variance of the inventory at j for an ergodic demand process is bounded across j.)

This paper develops necessary and sufficient conditions for the weak stability of arbitrary algorithms. It shows that *all* autonomous policies that reduce inventories during periods of low demand are unstable if they do not use future order commitments. The paper also presents a class of strongly stable, commitment-based algorithms that allow suppliers to maintain inventories close to any demand-dependent target of their choice

Mathematical procedures from the field of traffic flow theory are used, with modifications. The techniques are successful because a supply chain is somewhat analogous to a freeway, where the items are cars, and the suppliers, freeway detectors. By looking at the supply chain problem in this way, two separate fields of study are unified. Modifications are needed because the rules of motion (i.e. the management algorithms) in the supply chain management problem are different.

The paper is organized as follows. Section 1 describes the representation of data, algorithm types, and their basic properties. Section 2 examines the weak stability properties of autonomous policies. Section 3 defines a family of strongly stable "targets" for algorithms to track. These results are used in Section 4 to develop a family of strongly stable policies that can track these targets. The family includes "just-in-time" policies as a special case. Section 5 describes the basic properties of this family, and Sec. 6 discusses extensions and possibilities for further work.

1. DEFINITIONS

1.1. The representation of data.

A supply chain is shown in Fig. 1a, where j (j = 0, 1, 2, … J) is the supplier index, and arrows denote the flow of information and physical items. The figure shows that when an order is placed an acknowledgment is sent in return and, some time later, the physical items follow the acknowledgement. As in traffic flow theory, it will be useful to keep track of all these things over time with Newell-curves (N-curves) of cumulative flow. These are curves of cumulative count (integrals of the flow) at different locations, j, with integration constants such that vertical separations between curves represent accumulations between locations. With discrete items, such as cars or most goods, these curves increase in jumps as step functions. With fluids conveyed on pipes they increase continuously.

In our case, two sets of curves will be used. The first set consists of curves, $S_j(t)$, denoted by solid lines on Fig. 1b, that pertain to physical items. (To simplify the figure, the discrete steps corresponding to specific shipments have been smoothed out, as if the item flow was continuous, but nothing in this section requires continuity.) The S_j-curves indicate the item number (#) delivered by time t from j+1 to supplier j. The item inventory of supplier j at time t (including items in transit to j–1), $R_j(t)$, is given by the vertical separation between two of these curves (see Fig. 1b):

$$R_j(t) = S_j(t) - S_{j\text{-}1}(t). \quad (1)$$

The horizontal separation between curves S_j and $S_{j\text{-}1}$ is the time an item spends at j (and in transit to j-1), assuming FIFO; it can never be below a minimum processing time, $P_j \geq 0$. For the moment, P_j is treated as a deterministic constant, which is the sum of (constant) production and handling times at j and a transportation time from j to j-1. Consideration of Fig. 1b, or similar figures including discrete steps, shows that the following constraint is equivalent to assuming that the minimum processing times are not exceeded:

$$S_j(t\text{-}P_j) \geq S_{j\text{-}1}(t) \qquad \text{(feasibility condition.)} \quad (2)$$

Note that (2) implies $R_j(t) \geq 0$.

The second set of curves pertains to orders. We imagine the following. When an order is placed by j (from j+1), an acknowledgement is immediately sent from j+1 to j, as suggested by the dotted arrows on Fig. 1a. The acknowledgement includes the highest item number (#) that will be sent when the order is filled a lead-time later. Note that # is increased by the order size with every request. Curves of this kind, $N_j(t)$, give the acknowledgement number received at t by supplier j. This is also the order number received at t by supplier j + 1. Therefore, the difference between the number of the last order placed by j (from j+1) and the last order received by j (from j–1), i.e., the inventory of order acknowledgements at j, is:

$$K_j(t) = N_j(t) - N_{j\text{-}1}(t). \quad (3)$$

Orders and deliveries between j and j+1are assumed to be related by

$$N_j(t) = S_j(t+L_j), \quad (4)$$

where $L_j \geq 0$ is a deterministic lead-time.[1] Equation (4) rules out order back-logs. It also establishes a 1:1 correspondence between the N- and S-curves. Hence, one of these sets of curves (the S-curves) can be eliminated from the formulation. In terms of the N-curves, the feasibility condition (2) is expressed by combining (2) and (4):

$$N_j(t) \geq N_{j\text{-}1}(t+M_j), \qquad \text{(feasibility condition)} \quad (5)$$

where $M_j = P_j + L_j - L_{j\text{-}1}$. Equation (5) guarantees on-time deliveries; i.e., it prevents stock-outs. The equation says that the lag between the order times for a specific item must exceed M_j. Note that the equation holds for systems with random lead- and processing-times if M_j is defined as the smallest lag that would guarantee on-time delivery in the worst case. Thus, the proposed theory also applies to these systems.

Finally, note that the rates at which N_j and S_j change in an interval are respectively the order rates and replenishment rates at j for the interval. Thus, Fig. 1b embodies all the relevant information of a problem.

[1] If a supplier holds inventories of finished products and ships items as soon as they are ordered, then the lead-time equals the transportation component of the processing time. If a supplier holds inventories of unfinished products, and makes them to order, then the lead-time equals the processing time. Our discussion applies to both forms of operation, and to mixed forms.

1.2. Algorithm types.

An algorithm/policy is a rule that suppliers follow to generate orders. The output of an algorithm is a set of N-curves that meet customer demands, $N_0(t)$. The algorithm is said to be feasible if (5) is satisfied for all possible customer demands. In this paper we will focus on discrete-time algorithms where the $N(t)$ are step functions with jumps on a time lattice, $t_n = nh$, and we will write N_{jn} for $N_j(t_n)$. Here n is an integer and h is the lattice interval. No generality is lost, since by letting h tend to zero the results can be extended to continuous time.

Given the customer demands and an initial set of order inventories $\{K_{j0}\}$ for $t = t_0$, an algorithm must provide the ensuing jumps in the N-curves, i.e., the order sizes for $t_n \geq t_0$, since these jumps determine the N-curves. This, however, can be done in many ways.

Example 1: The following expression is an example of an autonomous policy for j where N_j is determined from the $N_{j\text{-}1}$ curve alone:

$$N_{j,n+1} = N_{j-1,n} + \gamma + \beta(N_{j-1,\mathrm{n}} - N_{j-1,\mathrm{n}-B}) \quad (6)$$

Alternatively, (6) can be expressed as:

$$(N_{j,n+1} - N_{j,n}) + (N_{j,n} - N_{j-1,n}) = \gamma + \beta(N_{j-1,\mathrm{n}} - N_{j-1,\mathrm{n}-B}), \quad (7)$$

where the left hand side is the *order inventory position* at j immediately after the order; i.e., the sum of the order size, $Q_{j,n} = N_{j,n+1} - N_{j,n}$, and the order inventory $K_{jn} = N_{j,n} - N_{j-1,n}$. Note from (7) that the order rate can also be expressed as a linear function of the inventory level and the past demand:

$$Q_{j,n} = \gamma - K_{j,n} + \beta(N_{j-1,\mathrm{n}} - N_{j-1,\mathrm{n}-B}). \quad (8)$$

Policy (6-8) adapts the desired order inventory position to past demand with a moving-average rule over B periods—since the adaptation term on the right side of (7) only depends on the number of orders received in the previous B periods. If $\beta = 0$, this policy is simply the "periodic review, order-up-to-level (R, S) system" of inventory control. With $\beta > 0$, the policy reduces inventories during extended periods of low demand.■

Note from (7) that the order inventory position on the left hand side only depends on curve $N_{j\text{-}1}$. Policies with this property will be called "order-based". Policies that use past history to forecast future demand, and then base the orders on the forecast, are in this class. We can also define "inventory-based" policies, by letting the right side of (7) be only a function of the current inventory, $K_{j,n}$.

Example 2: The following policy is inventory-based:

$$Q_{j,n} = \gamma - K_{j,n} + \alpha(N_{j,n} - N_{j-1,n}). \quad (9)$$

$$N_{j,n+1} = \gamma + \alpha N_{j,n} + (1\text{-}\alpha)N_{j-1,n}.\blacksquare \quad (10)$$

Both, (8) and (9) are generalized by the following (mixed) family of policies:

$$Q_{j,n} = F_j(N_{j,n}, N_{j-1,n}, N_{j-1,n-1}, N_{j-1,n-2}\ldots). \quad (11)$$

This is our generic definition of a "non-anticipative," autonomous policy. We stress that even though the goal of adaptation may be to forecast the future, policies based on past history are called "non-anticipative" because their *inputs* are not anticipative. Figure 2a depicts the causality structure of rule (11) by means of a "stencil," where arrows rooted at input points, point to the output. A policy is said to be "smooth" or "linear" if its order function F_j is smooth or linear, as occurred with the prior examples.

Example 3: An example of a non-smooth policy is:

$$Q_{j,n} = \gamma - K_{j,n} + \beta(N_{j-1,n} - N_{j-1,n-B}), \qquad \text{if } K_{j,n} < s \qquad (12a)$$

$$= 0, \qquad \text{otherwise.} \qquad (12b)$$

If h is small and $\beta = 0$, this is the so-called "order point, order-up-to-level (s, S) policy", with $S = \gamma$.■

1.2.1 Anticipative policies

A more general form of (11) would include anticipative data. If supplier j knows the orders in A future periods because supplier j-1 has committed itself to specific order quantities, then it is possible to use an "anticipative" rule,

$$Q_{j,n} = F_j(N_{j,n}, N_{j-1,n+A}, N_{j-1,n+A-1}, N_{j-1,n+A-2}\ldots), \qquad (13a)$$

or

$$N_{j,n+1} = N_{j,n} + F_j(N_{j,n}, N_{j-1,n+A}, N_{j-1,n+A-1}, N_{j-1,n+A-2}\ldots), \qquad (13b)$$

where Ah is the anticipation interval. Figure 2b depicts the stencil for a rule with $A = 2$.

If all the suppliers use rule (13), then to fulfill its commitment to j, supplier j-1 would need commitments from j-2, and j-2 from j-3, etc. The stencils of Fig. 2c show how such a chain would work. Solid dots denote the N-values known immediately after time t_n, circled dots the values that are past history, and squared dots the values that just became committed. When the clock advances by one tick, the stencils are shifted to the right by one lattice spacing, and the customer adds one order at the end of its horizon. With this information and the new stencil, supplier 1 places a new order, which allows supplier 2 to do the same, etc. The key requirement in this scheme is that once a future order has been placed it cannot be changed.

The figure clearly shows that if a chain has J suppliers with the same anticipation, then the "customer" would have to commit itself for $1+(A-1)J$ periods; i.e., the commitment horizon for the chain is:

$$\text{Commitment horizon} = [1+(A-1)J]h. \qquad (14)$$

If the customer is not willing to cooperate, one can ask the first supplier to keep safety stocks and play the role of customer.[2] For simplicity of notation, if a lead supplier is used it will be labeled $j = 0$, and its orders will become the input to our analysis, $N_0(t)$. This is a valid simplification for stability analysis since the stability properties pertain to infinite chains with arbitrary inputs.

Example 4: Examples of anticipative policies are:

Order-based: $$N_{j,n+1} = \gamma + N_{j-1,n+A} + \beta(N_{j-1,n+A} - N_{j-1,n-B}) \qquad (15a)$$

Mixed: $$N_{j,n+1} = \gamma + \alpha N_{j,n} + (1-\alpha)N_{j-1,n+A}.■ \qquad (15b)$$

---

[2] In both cases the committing entity must be compensated for its added expense with cheaper prices. This should be possible with long chains because then total savings are larger than the commitment penalty (Daganzo, 2001).

1.2.2 Flexible commitments

A disadvantage of anticipative algorithms is the rigidity of the commitments required of every supplier. Another disadvantage of smooth policies (such as those in examples 1,2 and 4) is that they require a shipment in every interval. This may not be desirable when h is small. These drawbacks can be eliminated by introducing some flexibility into the commitment process, as explained below. The new policies will be called "flex-time". They are particularly attractive when suppliers produce many different items as part of a network because the added flexibility allows the suppliers to schedule more efficient production runs.

The policies work as follows. An anticipative algorithm like (13) is used to obtain upper bounds, $U_j(t_n)$, for the cumulative orders placed by each supplier, instead of the orders themselves. Suppliers are then allowed to place orders of any size without forewarning (at times not necessarily on the lattice) provided $N_j(t) \leq U_j(t)$ for all t; see Fig. 3. To ensure feasibility, lower bounds, $V_j(t)$, are also defined for the orders. The lower bounds can be chosen in any way, but we require $U_j(t) \geq V_j(t) \geq U_{j\text{-}1}(t+M_j)$. This ensures feasibility, because then $N_j(t) \geq V_j(t) \geq U_{j\text{-}1}(t+M_j) \geq N_{j\text{-}1}(t+M_j)$, in agreement with (5).

Flex-time algorithms in which suppliers trigger an order whenever $N_j(t) = V_j(t)$ are also related to order-point inventory control methods. The only difference is that in a flex-time method the trigger is based on the virtual order curve $U_{j\text{-}1}$ of the downstream supplier, rather than on the orders themselves. This is a logical approach since with flexible commitments the downstream supplier may choose to bring its orders up to the maximum level at any time.

<u>1.3. Steady state properties of discrete-time algorithms</u>

Note first that for an algorithm to be "proper" the output of (13a) should be independent of which order was labeled "zero;" i.e., (13a) must be expressible in terms of the difference between the N-values appearing as arguments and the order number of supplier j-1 at time n, $N_{j\text{-}1,n}$; i.e., order size should be a function of the inventory at time n and the order history of supplier j-1: $Q_{j,n} = H_j(K_{j,n}\ , Q_{j\text{-}1,\text{n}+A\ \text{-}1}, Q_{j\text{-}1,\text{n}+A\ \text{-}2}\ldots)$. Two more requirements of a proper algorithm are: (i) that it should produce bounded inventories for supplier j if the orders from j-1 are steady at Q, and (ii) that the time-average of these inventories should not depend on the initial conditions; i.e., it should be a function of Q. All the examples of Sec. 1.2 are proper.

***<u>K</u>*<u>-functions</u>**: A steady state is said to exist for some order size Q if (13) has a solution with parallel N-curves and steps, Q, as in Fig. 4a. The vertical separation between curves is the steady state inventory, K_j. It is shown in Daganzo (2001) that all smooth, proper algorithms (such as those of examples 1, 2 and 4) have a unique K_j for every Q. The resulting relationship $K_j = \kappa_j(Q)$ will be called the "κ-function" of a (smooth, proper) policy. Because in practical applications it is more physically meaningful to relate the steady state inventory to the demand rate in continuous time, $q = Q/h$, we also define a "K-function", $K_j(q) = \kappa_j(qh)$. The reader can verify that the κ- and K-functions of examples 1, 2 and 4 are:

$$K(Q/h) = \kappa(Q) = \gamma + (\beta B - 1)Q \qquad \text{for example 1} \qquad (16a)$$
$$= (\gamma - Q)/(1-\alpha) \qquad \text{for example 2} \qquad (16b)$$
$$= \gamma + (\beta(B+A) + A - 1)Q \qquad \text{for example 4a} \qquad (16c)$$
$$= AQ + (\gamma - Q)/(1-\alpha) \qquad \text{for example 4b.} \qquad (16d)$$

Note, however, that (non-smooth) example 3 does not have a steady state if $s > \gamma + (\beta B - 1)Q$.

Gain and linearity: Figure 4b shows a hypothetical K-function. Its intercept, K_o , is the order inventory during a period of zero demand. Its slope has units of time. It represents the change in inventory for a small change in demand, and will be called the "gain", g. The derivative of the κ-function, $G = g/h$, also measures gain, but does so in terms of the number of intervals; it will be called the "dimensionless gain." Since suppliers prefer to keep lower inventories during low-demand seasons, practical algorithms should have non-negative gains. This happens for (16) if the gains G, satisfy:

$$(\beta B - 1) \geq 0 \qquad \text{for example 1} \qquad (17a)$$
$$-1/(1-\alpha) \geq 0 \qquad \text{for example 2} \qquad (17b)$$
$$\beta(B+A) + A - 1 \geq 0 \qquad \text{for example 4a} \qquad (17c)$$
$$A - 1/(1-\alpha) \geq 0 \qquad \text{for example 4b.} \qquad (17d)$$

For linear algorithms, as occurs with these examples, the K-functions are linear and the gains are constant.

Feasibility: Note from Fig. 4b that the slope of a ray joining a point, "P", on the K-curve with the origin equals K/q. Figure 4a shows that (for supplier j) this ratio is the horizontal separation between curves N_j and N_{j-1} for the given steady state. The steady state is feasible, according to (5), if this separation exceeds or equals M_j; i.e., if as occurs on Fig. 4b, point "P" lies on or above a ray with slope M_j from the origin. A policy with feasible steady states for the range of q's that can occur ($0 \leq q \leq q_{max}$) is said to have a feasible K-function; this happens if the K-curve lies above the ray with slope M_j.

2. WEAK STABILITY

This section derives "weak stability" conditions for discrete-time algorithms of type (13) with homogeneous suppliers and a smooth $F_j \equiv F$, and discusses the role that gain and anticipation play in the causation of instability. Recall that if a system is weakly stable then an infinitesimal deviation in N_0 from a steady state can only generate perturbations to the N_j's that are bounded under the L_∞ norm by an infinitesimal quantity independent of j. This implies that the *maximum* deviation in order size (or inventory) for all members of the chain is always bounded by an infinitesimal constant.

Non-smooth policies such as example 3 are inherently unstable since they can translate infinitesimal input changes into finite output changes. Therefore they are excluded from the analysis. The analysis, however, applies to (smooth) continuous-time policies if one lets $h \to 0$. Since the stability conditions turns out to be independent of h, this extension is trivial. The stability tests can also be applied to flex-time systems if the bounds, $U_j(t)$, obey a smooth version of (13), and the stability of the chain is measured by the decay of small perturbations to the bounds.

Since we are interested in tracking infinitesimal perturbations from a steady state of (13b), this equation can be linearized about the steady state and expressed exclusively in terms of the deviations of the order number from the steady state ε_{jn}. A power series expansion shows that to a first order of approximation

$$\varepsilon_{j,n+1} \cong \alpha\, \varepsilon_{j,n} + \beta_{-A}\varepsilon_{j-1,n+A} + \beta_{1-A}\varepsilon_{j-1,n+A-1} + \ldots \text{ (for integer } n\text{, and } j = 1, 2, \ldots J), \quad (18)$$

where $(\alpha\text{-}1, \beta_{1-A}, \beta_{2-A}, \ldots)$ are the partial derivatives of F evaluated at the steady state. If (13b) is linear; i.e.,

$$N_{j,n+1} = N_{j,n} + \gamma + (\alpha\text{-}1)N_{j,n} + \beta_{-A}N_{j-1,n+A} + \beta_{1-A}N_{j-1,n+A-1} + \ldots, \quad (19)$$

then (18) is exact and applies to large perturbations. We now show that weak stability as determined by an analysis of (18) implies that perturbations of all types (i.e., of orders or inventories) are well behaved.

By linearizing the first forward differences of (13b) with respect to n, or taking first differences in (19) if the system is linear, we find that the order sizes also obey (18); i.e.,

$$Q_{j,n+1} \cong \alpha\, Q_{j,n} + \beta_{-A}Q_{j-1,n+A} + \beta_{1-A}Q_{j-1,n+A-1} + \ldots \text{ (for integer } n\text{, and } j = 1, 2, \ldots J.) \quad (20)$$

Note that for the given steady state $Q \neq 0$ to exist the coefficients on the right side of (20) must add to 1, i.e.:

$$\alpha + \sum_{l=1-A}^{\infty} \beta_l = 1 \qquad \text{(for existence.)} \quad (21)$$

Thus, the identity $[Q = \alpha Q + \beta_{-A}Q + \beta_{1-A}Q + \ldots]$ can be subtracted from both sides of (20). The result is an expression identical to (18) that applies to the order size deviations from the steady state.

By subtracting the j and j-1 instances of (13b) and linearizing the result, or by taking the first differences of (19) with respect to j, the same results are obtained; i.e., (18) also holds for the inventories and for the inventory deviations. Thus, an analysis of (18) is sufficiently general to determine weak stability. In the following we shall use "Q" as the variable name in (18); i.e., we shall work with (20).

Von Neumann's stability test: Since (20) defines a linear and homogeneous set of difference equations it should have solutions of the form $Q_{jn} = \xi(\omega)^j \exp(-n\omega i)$, where ω is real, $\xi(\omega)$ is complex, and i is the imaginary square root of -1. These solutions are called "modes". Linear combinations of modes are also solutions because (20) is linear and homogeneous. Furthermore, a linear combination of modes with $\omega \in [-\pi, \pi)$ that matches the data for $j = 0$ always exists if the total demand is bounded. This is true because for $j = 0$ the modes are of the form, $\exp(-n\omega i)$, and if we let the weighting function be $(2\pi)^{-1/2}L(\omega)$ then the matching condition, $Q_{0,n} = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} L(\omega)e^{-n\omega i}\,d\omega$, can be recognized as the formula for the coefficients of the Fourier series expansion of $L(\omega)$. Hence, $L(\omega) = \frac{1}{\sqrt{2\pi}}\sum_m Q_{0m}e^{m\omega i}$, and the particular solution corresponding to some data is $Q_{j,n} = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} L(\omega)\xi(\omega)^j e^{-n\omega i} d\omega$. This expression is bounded by $\frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} |L(\omega)| d\omega$ for all j (i.e., it is stable) if and only if the modulus of ξ(ω) never exceeds 1. This is Von Neumann's stability test. The expression for $Q_{j,n}$ also shows that if the largest modulus of ξ(ω) exceeds 1, $|\xi^*| > 1$, and this is achieved for a single value, ω*, then the asymptotic solution for large j is determined by $|\xi^*|$ and ω* alone. Thus, independently of the input data, the asymptotic solution should be oscillatory with a period of 2π/ω* intervals, and an amplitude that grows by an amplification factor $|\xi^*|$ with each step up the supply chain. By inserting the modes, $Q_{jn} = \xi(\omega)^j \exp(-n\omega i)$, into (19) and solving for ξ(ω) we find:

$$\xi(\omega) = \sum_{l=-A}^{\infty} \beta_l \exp(l\omega i) / [1-\alpha \exp(\omega i)]. \tag{22}$$

Therefore, the Von Neumann stability condition, $|\xi(\omega)| \leq 1$, is:

$$\left| \sum_{l=-A}^{\infty} \beta_l \exp(l\omega i) \right| \leq \left| 1-\alpha \exp(\omega i) \right|; \ \forall \omega \qquad \text{(for stability.)} \tag{23}$$

For non-linear systems (23) is a weak stability condition, implying that infinitesimal perturbations to order sizes and inventories remain bounded. For linear systems, (23) also implies that the order sizes and the inventories themselves remain bounded for time-dependent customer orders. Therefore, (23) is a strong stability condition for linear systems.

<u>Gain:</u> If the system is linear (obeys 19) an expression for the κ-function can be obtained in terms of the α- and β-coefficients. Since (21) holds, the expression $[N_{j-1,n+1} = \alpha N_{j-1,n+1} + \beta_{-A} N_{j-1,n+1} + \beta_{1-A} N_{j-1,n+1} + ..]$ is an identity that can be subtracted from (19). The result expresses the inventory as a function of the order rates of supplier j-1. In the steady state the result is $[K = \gamma + \alpha(K-Q) + \beta_{1-A}(A-1)Q + \beta_{2-A}(A-2)Q + \ldots]$; i.e.,

$$K = [\gamma - Q(\alpha + \sum_{l=-A}^{\infty} l\beta_l)]/(1-\alpha) \tag{24a}$$

The discrete gain is:

$$G = (\alpha + \sum_{l=-A}^{\infty} l\beta_l)/(\alpha - 1). \tag{24b}$$

Since the gain of a non-linear system describes conditions that are nearly steady, where (13b) can be approximated by (19), we see that (24b) also holds in the non-linear case. The implications of (21), (22) and (24) are now explored for the examples of Sec. 1.

<u>Example 1</u>: This is the special case where all the coefficients are zero, except β_0 and β_B for some $B > 0$. Existence of a steady state requires $\beta_0 + \beta_B = 1$. This was true for the example because $\beta_0 = 1+\beta$ and $\beta_B = -\beta$. Equations (24) reveal, in agreement with (16a) and (17a), that the κ-function and gain are $\gamma + (\beta B - 1)Q$

and (βB–1), respectively. The gain is positive if $\beta \equiv \beta_0 - 1 > 1/B > 0$. This implies that $\beta_0 > 1$ for positive gains, and that the two coefficients must have opposite signs. The complex amplitudes (22) are:

$$\xi(\omega) = \beta_0 + \beta_B \exp(B\omega i) \tag{25a}$$

and we have

$$|\xi(\omega)|^2 = \xi(\omega)\overline{\xi(\omega)} = \beta_0^2 + \beta_B^2 + 2\beta_0\beta_B\cos(B\omega). \tag{25b}$$

Since β_0 and β_B have opposite signs if $G > 0$, this expression is maximized by $\omega^* = \pi/B$, and the amplification factor is: $|\xi^*| = |\beta_0| + |\beta_B| = 1 + 2|\beta_B| > 1$. Thus, the system is unstable if $G > 0$. On the other hand, the periodic review, order-up-to-level (R, S) policy, which has $\beta = 0$ and $G = -1$, is stable.■

The instability result is actually quite general, for it is possible to prove that all non-anticipative, order-based policies with $G > 0$ are unstable; see (Daganzo, 2001). Furthermore, if the analysis is repeated for inventory-based policies (example 2), we obtain the same result: a positive gain without anticipation implies instability. In view of this, one may ask if positive gains without instability can be achieved by anticipative policies. The answer is positive.

Example 4: For order-based policy (15a), all the coefficients are zero, except $\beta_{-A} = 1+\beta$ and $\beta_B = -\beta$. Thus, the steady state exists. The dimensionless gain (24b) turns out to be $G = (A-1) + (A+B)\beta$, in agreement with (17c), and the stability condition (23) becomes:

$$|\xi(\omega)|^2 = (1+\beta)^2 + \beta^2 + 2(1+\beta)\beta\cos((B+A)\omega) \leq 1; \quad \forall\omega. \tag{26}$$

This condition is satisfied if and only if β and $1+\beta$ have different signs; i.e., if and only if $0 \leq \beta \leq 1$. Since $\beta = (G+1-A)/(A+B)$, β is in the unit interval if and only if $A \geq G+1$.

If this analysis is now repeated for policy (15b), we again find that the system is weakly stable if and only if $A \geq G+1$. This common result is interesting because it links anticipation and gain: greater anticipation allows greater gains .■

We will see in the following sections that the condition

$$A \geq G+1 \tag{27a}$$

is a necessary condition for the stability of any algorithm. Thus, the least possible anticipation for a chain with gain G is $A = G+1$. Hence, any chain with gains G at all steps needs the following commitment horizon; see (14):

$$\text{Required commitment horizon for stability with gain } G = h[JG+1] \tag{27b}$$

The policies derived in example 4a and 4b are strongly stable because they are linear. They are complete because their parameters γ, A, B and β (or γ, A and α) can be chosen to yield any desired *linear* K-function, as per (16c) and (16d). Hence, example 4 can be used to identify strongly stable policies for linear problems. Note, however, that we have not yet said anything about feasibility. The next three sections show how to derive strongly stable policies for non-linear K-functions, and how to ensure feasibility.

3. THE KINEMATIC WAVE TARGET

Section 2 showed how to test non-linear algorithms for weak stability. But this is not enough. In addition, one should require an algorithm to avoid negative flows, satisfy the feasibility constraint (5), be stable to large perturbations, and stay close to target inventory levels while tracking the customer demand. This will be achieved by introducing a family of continuous-time targets (non-decreasing cumulative curves, $u_j(t)$, with the above-mentioned properties) and then looking for algorithms that can stay close these targets; i.e., such that $N_{j,n} \cong u_j(t_n)$. The present section focuses on the targets and the following two sections on the algorithms.

The targets used in this paper are solutions of the kinematic wave (KW) model of fluids; see Lighthill and Whitham (1955), Richards (1956), Lax (1973) and Newell (1993) for background on this theory. The theory pertains to the flow of generic items such as fluid particles, cars or order acknowledgements over one-dimensional "pipes". Therefore, it is relevant to the supply chain problem if we just imagine that suppliers are located on a "pipe" with spacing p such that their distances from the customer are $x_j = jp$. The solution of a KW problem is expressed by means of a bivariate function, $u(x, t)$, that gives the item number at every point in continuous space-time. The target curves of the supply chain problem can be retrieved from a KW solution by: $u(x_j, t) = u_j(t)$.

There are two reasons for adopting KW theory. First, the KW targets, $u_j(t)$, have all the desired properties mentioned above, and also satisfy an even more stringent condition than is required for strong stability: as shown in Daganzo (1997), the accumulation norms of the KW model, $sup\{u_j(t)–u_{j-1}(t)\}$, are non-increasing in j (not just bounded). This means that a discrete-time policy does not have to track the continuous-time target very accurately to be strongly stable. Second, a discrete-time approximation theory to the KW model already exists, and it will form the basis for the results in Sec. 4.

A KW problem is specified by: (i) boundary data in the form of a non-decreasing customer demand curve, $N_0(t) \equiv u_0(t) \equiv u(0, t)$, and (ii) a K-function. The KW solution must satisfy (i) the boundary condition, and (ii) the following first order partial differential equation everywhere (except perhaps at curves called "shocks" where the partial derivatives of the solution may not exist):

$$\partial u/\partial x \ = \ K(\partial u/\partial t)/p. \qquad (28)$$

The partial derivatives of $u(x, t)$ with respect to x and t are the rates at which the item number increases with space and time. They are respectively called "ρ, density" and "f, flow." In the steady state, ρ and f are constant, and (28) shows that they must satisfy $\rho p = K(f)$. This is consistent with the idea of a K-function since ρp is the order accumulation between consecutive suppliers, i.e., the steady state target inventory.[3] (If the customer demand rate is q_0, then the steady state solution is $f = q_0$, $\rho = K(q_0)/p$, and $u(x, t)$ = constant + $K(q_0)x/p + q_0t$.)

It is well known that the KW solution to a problem of this type is unique; i.e., there is a unique set of target u-curves consistent with the data. As was mentioned earlier, the solution in question has the following properties: it is (i) continuous, (ii) stable, and (iii) non-decreasing with x and t. One can also show (Daganzo, 2001) that the solution is (iv) feasible in the sense of (5) if the K-function is feasible in the sense of Fig. 4b.

One last important feature of the KW model is (v) that the differences in two solutions of a problem with identical input data except for an infinitesimal local perturbation are confined to a line on the (x, t) plane that contains the perturbation. This curve is called a "characteristic" or "wave." It turns out that the flow does not change along the wave, and the velocity of the wave is the following function of its flow:

$$\text{wave velocity} = -p[dK(f)/df]^{-1} = -p/g. \qquad (29)$$

Note that g is the local gain (units of time), and that if the gain is positive, the wave velocity is negative. This equation turns out to be important because, as is explained in the next section, the character of convergent algorithms is strongly influenced by wave behavior.

4. THE ACT POLICY

Section 3 showed that the KW targets were strongly stable, feasible and monotone. Here we look for discrete-time policies (13b) that will approximate the KW target and as a result will also have these desirable properties. We will use inventory-based policies because they turn out to be the simplest. The general non-anticipative form of these policies is derived from the discrete-time version of (28), $K_{j,n} = K(Q_{j,n}/h)$; i.e.:

$$Q_{j,n} = hK^{-1}(K_{j,n}) \Leftrightarrow N_{j,n+1} = N_{j,n} + hK^{-1}(K_{j,n}) \qquad (30)$$

where K is the K-function of the KW target (assumed for the moment to have an inverse).

Equation (30) is related to the so-called "cell-transmission" (CT) algorithm of traffic flow (Daganzo, 1995), which is a finite difference approximation of the KW model in discrete-time. Background on finite difference approximations for conservation laws such as the KW model can be found in LeVeque (1992). For a rule such as (30) to be a valid approximation (i.e., to converge to the continuous-time solution as h and p tend to zero), the rule has to satisfy the so-called Courant stability condition. This is explained below.

4.1 Courant's stability condition

The existence of characteristics/waves with velocity (29) in the KW solution means that the value of f at any given point in space-time (x_a, t_a) cannot be affected by data on points that are not on a wave path

[3] To avoid the proliferation of symbols, we have used "$K(\ )$" for the K-function of the target. This symbol has been used earlier for the K-function of the algorithm. Using a common symbol is reasonable because all algorithms considered here will turn out to have the same steady states as the target. Note, however, that we use f for the flow of the target at any given time instead of q (the actual flow) since the two flows will generally be different.

passing through the point in question. In other words, waves are the set of points, $D(x_a, t_a)$, whose data may influence $f(x_a, t_a)$. This set of points is called the "domain of dependence" of point (x_a, t_a). The domain of dependence for a numerical approximation such as (13b) or (30), D_A, is defined in a similar way; i.e., so that data from points (x, t) not in D_A cannot affect $Q(x_a, t_a)$, regardless of the discretization used. The shadowed wedge in Fig. 5a is D_A for the point at the vertex and a numerical algorithm with the shown stencil. Other stencils would have different domains of dependence; see Figs. 5b and 5c. It should be clear that if data are defined on a region B of the (t, x)-plane and we find that $B \cap D(x_a, t_a) \not\subset B \cap D_A(x_a, t_a)$, then one could vary the initial data in the parts of D not included in D_A so as to change $f(x_a, t_a)$ without having an effect on $Q(x_a, t_a)$. Thus, the finite difference approximation could not approximate the target. The problem can be seen to arise in Figs. 5a and 5b but not in Fig. 5c, for systems where the boundary would include the line $x = 0$. This observation is the basis for Courant's necessary condition for convergence. It states that the domain of dependence of an algorithm must include all possible waves.

Failure to meet the condition means that a numerical algorithm cannot converge to the target solution for smooth but arbitrary initial data in a fixed time-space region, when h and p are decreased toward zero. Equivalently, it means that for fixed h and p the numerical results cannot remain close to the target solution as the number of time and space increments is increased, even if the initial solution is smoothed as the number of steps is increased. In our case, it means that algorithms failing Courant's condition cannot approximate the target as the number of suppliers J is increased.

Figures 5a and 5b clearly show that all non-anticipative algorithms fail Courant's condition if the gain is positive. This result applies to autonomous and non-autonomous algorithms, with stencils as complicated as one wishes. Hence, supply chains without commitments cannot track KW targets with positive gain. This result is consistent with the stability results of Sec. 2.

Figure 5c shows that autonomous, anticipative policies will meet Courant's condition if D_A encloses the wave path when it is shallowest; i.e., when the gain is largest. Thus, the necessary condition is:

$$Ah \geq h + g_{max} \quad \Leftrightarrow \quad A \geq 1 + G_{max} \,, \tag{31}$$

where $g_{max} = sup(dK/df)$. This is consistent with (27a).

It is known that the CT rule (30) converges to the KW target if and only if (31) is satisfied. Since the CT rule is non-anticipative, its convergence condition is $g_{max} \leq -h$, or $G_{max} \leq -1$; i.e., it can only work with negative gains. Hence, the CT policy is not of practical interest for supply chain management. A change of variable is used below to overcome this difficulty.

4.2 The change to asynchronous time and the ACT policy

The reader can verify that (30) is invariant under the following transformation to "asynchronous time," where c is a constant with units of "pace" (time/distance) and $f = \partial u/\partial t$ is the item flow:

$$t' = t + cx \,; \quad x' = x \,; \quad u' = u \,; \quad \text{and} \quad K'(f) = K(f) - cpf \,; \tag{32}$$

i.e., if for every distance unit, the clocks are advanced c time units and the K-function is redefined as shown in Figs. 6a and 6b. The transformation of the space-time domain is shown in Figs. 6c and 6d. Invariance means that one can solve the problem in asynchronous time and then undo the transformation to obtain the desired solution. The transformation is useful because with a proper choice of c the transformed problem can be made to satisfy Courant's condition.

To see this note that $g'_{max} = sup(dK'/df) = sup(dK/df) - cp = g_{max} - cp$ can be made as small as needed as long as $g_{max} \neq \infty$ by increasing c. Thus, the CT convergence criterion, $g'_{max}/h \leq -1$, will be satisfied by choosing $cp/h \geq 1 + g_{max}/h$. It follows that the following version of (30) converges:

$$Q_{j,n'} = hK'^{-1}(N_{j,n'} - N_{j-1,n'}) \,. \tag{33}$$

Here n' denotes asynchronous time, $n' = n + Aj$, where $A = cp/h$ is an integer. Note that K' is monotone-decreasing and will have an inverse, even if K does not. In original time, policy (33) becomes:

$$Q_{j,n} = hK'^{-1}(N_{j,n} - N_{j-1,n+A}), \qquad \text{(with } K'(f) = K(f) - Ahf\text{, and integer } A \geq 1 + g_{max}/h.) \tag{34a}$$

This is the proposed asynchronous cell-transmission (ACT) rule. In terms of the N-curves alone, the rule is:

$$N_{j,n+1} = N_{j,n} + hK'^{-1}(N_{j,n} - N_{j-1,n+A}), \ \text{(with } K'(f) = K(f) - Ahf\text{, and integer } A \geq 1 + g_{max}/h.) \tag{34b}$$

Note that orders placed by j are just a function of $N_{j,n} - N_{j-1,n+A}$ at j. We call this quantity, which can be positive or negative, the "asynchronous inventory" of j because it is its inventory in the transformed time coordinates. In original coordinates, the asynchronous inventory is the current inventory minus the future orders to which j-1 has already committed. Thus, the ACT algorithm is neither inventory-based nor order-based, but mixed.

The ACT algorithm is so easy to implement that simulations can be done with spreadsheets. Figure 7 depicts some results for a case with seven suppliers. Part (A) is the result for a linear K-function with positive gain, $g = h$ ($G = 1$). Note the closeness of the curves, and the absence of increased separations with increased j. Parts (B) to (D) depict non-linear results with a concave K-function. Part (B) uses the same customer data as part (A), but it was obtained by setting $A = g_{max}/h$, which violates the stability condition; see (34). Note how the inventories become increasingly chaotic with increasing j. Part (C) repeats the simulation with the same customer data and the same K-function, but using instead $A = 1 + g_{max}/h$, which guarantees stability. The figure clearly shows that the separation between curves is bounded, and that the curves become smoother as j increases despite the appearance of sharp bends or "shocks" -- as predicted in KW theory. The same is observed (even more clearly) in part (D) of the figure, which repeats the simulation with noisier customer data and a higher demand rate. In both cases, (C) and (D), the inventory fluctuations can be seen to decline with increasing j. In fact, the maximum inventory held by a supplier actually decreases with j. In part (D), the maximum inventories from $j = 1$ to $j = 7$ were: 6.54, 5.96, 5.44, 5.31, 5.26, 5.23 and 5.23 units, respectively.

4.3 The linear ACT policy and JIT control

If we use the ACT policy with the generic, linear K-function with gain g, $K = K_o + gq$, and choose A and h to operate with the smallest possible anticipation, i.e., $A = 1 + g/h = 1 + G$, then the transformed K'-function is $K_o + gq - Ahq = K_o + (G-A)hq = K_o - hq$. Therefore, (34b) reduces to $N_{j,n+1} = N_{j,n} - [(N_{j,n} - N_{j-1,n+1+G}) - K_o]$; i.e., to

$$N_{j,n+1} = N_{j-1,n+1+G} + K_o. \tag{35a}$$

The first difference of this expression is:

$$Q_{j,n+1} = Q_{j-1,n+1+G}. \tag{35b}$$

This clearly shows that in the linear case suppliers simply repeat the orders of their downstream neighbors, G intervals earlier. If this procedure is applied with $K_o = 0$ and the least possible anticipation that preserves feasibility (i.e., $G = M/h$), then the result is a lean supply chain of the "just-in-time" type, because the N_j curve is just a horizontally shifted version of the N_{j-1} curve with the smallest feasible shift, $N_{j,n} = N_{j-1,n+M/h}$. (Note that this recipe satisfies (5) strictly.)

5. PROPERTIES OF THE ACT POLICY

This section proves that under certain mild conditions the ACT policy yields monotone, strongly stable and feasible N-curves.

Theorem 1: Monotonicity and boundedness. If the customer data satisfy $0 \le Q_{0n} \le Q_{max}$ for all n and the initial asynchronous inventories satisfy $K'(Q_{max}/h) \le K'_{j0} \le K'(0)$ for all j, then the supplier flows Q_{jn} and their asynchronous inventories, K'_{jn}, satisfy these inequalities too for all j and all n. This means that the N_{jn} are non-decreasing in n.■

Proof: As a preliminary step, note that the asynchronous inventories satisfy the recursion: $K'_{j,n+1} = N_{j,n+1} - N_{j-1,n+1+A} = (N_{j,n} + Q_{j,n}) - (N_{j-1,n+A} + Q_{j-1,n+A}) = K'_{j,n} + Q_{j,n} - Q_{j-1,n+A} = K'_{j,n} + hK'^{-1}(K'_{j,n}) - Q_{j-1,n+A}$. If we now define the function $\psi(x) = x + hK'^{-1}(x)$, the recursion is

$$K'_{j,n+1} = \psi(K'_{j,n}) - Q_{j-1,n+A}. \tag{36a}$$

Note that ψ is monotone--increasing at a rate between 0 and 1. [This is true because ψ is continuous and piecewise differentiable with derivative: $1 + h/[dK'/dq] = 1 + h/[g-cp] = 1 + 1/[g/h - cp/h] = 1 + 1/[g/h - A] \in [0, 1]$. The last inequality follows from the stability condition in Eq. (34), which implies $[g/h - A] \le -1$.] To conclude the preparatory steps, note as well that (34a) may be expressed as:

$$Q_{j,n} = hK'^{-1}(K'_{j,n}), \tag{36b}$$

where hK'^{-1} is monotone-decreasing. The theorem can now be proven.

This is done for supplier $j = 1$ first. Let $n = 0$ in (36) and note from the conditions of the theorem that the right side of (36a) must be in the interval $[\psi(K'(Q_{max}/h)) - Q_{max}, \psi(K'(0))]$, since ψ is monotone and non-decreasing. Furthermore, since $\psi(K'(x)) = K'(x) + xh$, this interval reduces to $[K'(Q_{max}/h), K'(0)]$. Thus, $K'_{1,0}$ satisfies the theorem. If we now iterate (36a) for $n = 1, 2, \ldots$ the same interval is obtained every time. Thus, $K'_{1,n}$ satisfies the theorem. Consider now (36b). Since it is a decreasing relation and its inputs satisfy $K'_{1,n} \in [K'(Q_{max}/h), K'(0)]$, it follows that its outputs satisfy $Q_{1,n} \in [hK'^{-1}(K'(0)), hK'^{-1}(K'(Q_{max}/h))] = [0, Q_{max}]$. Thus, the theorem holds for $j = 1$.

We have just shown that supplier $j = 1$ meets the bounds of the theorem if $j = 0$ meets the bounds. The argument of the previous paragraph can now be repeated to show that $j+1$ meets the bounds if j meets them too, and this concludes the proof.■

Theorem 2: Strong stability. If Theorem 1 holds the ACT algorithm is strongly stable.■

Proof: It was shown in Sec. 2, example 4, that the linearized version of the ACT policy (policy (15b)) was weakly stable. Theorem 1 showed that the ACT flows are bounded. Hence, we just have to show that the inventories are bounded. This is true, however, since $K_{jn} = K'_{jn} + (N_{j-1,n+A} - N_{j-1,n}) \leq K'(0) + AQ_{max}$.■

Conditions guaranteeing the feasibility of the ACT algorithm for targets with moderate gains are given below. It will be shown that the algorithm is feasible if the portion of the K-curve corresponding to possible flows, $q \in [0, q_{max}]$, lies above a ray from the origin with slope $\lceil M/h \rceil h \cong M$ and also above a ray with slope g_{max}. The following lemma is a preliminary step.

Feasibility lemma: The ACT algorithm with the least possible anticipation ($A = cp/h = 1 + g_{max}/h$) is feasible if $g_{max}/h \geq \lceil M/h \rceil$, and $K(q) \geq g_{max}q$.■

Proof: It suffices to show that $N_{j,n+1} \geq N_{j-1,n+1+\lceil M/h \rceil}$. To this end, let $\psi(x) = x + hK'^{-1}(x)$, as before, and rewrite (34b) as follows:

$$N_{j,n+1} = N_{j-1,n+1+\lceil M/h \rceil} + (N_{j-1,n+A} - N_{j-1,n+1+\lceil M/h \rceil}) + \psi(N_{j,n} - N_{j-1,n+A}).$$

Thus, it suffices to show that the last two terms on the right side of this equality are non-negative. Since the N-curves are non-decreasing (Theorem 1), the second term on the right side will be non-negative if $A \geq 1+\lceil M/h \rceil$. This is true because $A = 1 + g_{max}/h$ and $g_{max}/h \geq \lceil M/h \rceil$. The third term will be non-negative if $\psi(x) = x + hK'^{-1}(x)$ is non-negative; i.e., if $hK'^{-1}(x) \geq -x$ for all x. Since $K'(x)$ is monotonic decreasing, we apply this transformation to both sides of the inequality after dividing by h and obtain the equivalent condition: $x \leq K'(-x/h) \Leftrightarrow -qh \leq K'(q) \Leftrightarrow K(q) \geq (cp-h)q = g_{max}q$. Since $K(q) \geq g_{max}q$ by assumption, the third term is also non-negative, and the lemma is proven.■

Theorem 3: Feasibility. The ACT algorithm is feasible if the portion of the K-curve corresponding to possible flows, $q \in [0, q_{max}]$, lies above a ray from the origin with slope $\lceil M/h \rceil h \cong M$ and also above a ray with slope g_{max}.■

Proof: Assume that the portion of the K-curve corresponding to possible flows lies above the two rays, so that $K(q) \geq \lceil M/h \rceil hq$ and $g_{max}h$,and that the lemma cannot be invoked; i.e., that the K-curve violates the condition $g_{max}/h \geq \lceil M/h \rceil$. A K-curve with such a low maximum slope can always be approximated arbitrarily well from above by a new curve with a larger maximum slope; e.g., obtained from a small perturbation in an infinitesimal neighborhood around some q. Let the new maximum slope satisfy $[g_{max}]^{new} = \lceil M/h \rceil h$ so that the new curve satisfies the first condition of the theorem. Since $K_{new}(q) \geq K(q) \geq \lceil M/h \rceil hq$, it follows that $K_{new}(q) \geq K(q) \geq [g_{max}]^{new} q$. Thus, the new curve satisfies the second condition of the lemma and yields feasible results if the anticipation is $A^{new} = 1 + [g_{max}]^{new}/h = 1 + \lceil M/h \rceil$. Since the results of the new curve should be indistinguishable from the old, we can conclude that the original results will be feasible if one chooses $A = 1 + \lceil M/h \rceil$.■

Corollary. For linear problems the ACT algorithm is feasible if the K-function is feasible.■

Proof: If a linear K- function is feasible, it satisfies the conditions of Theorem 3.■

The ACT algorithm is also feasible with large-gain feasible targets (i.e., where the g-ray crosses the K-curve) if a sufficiently small h can be used. This is true because the ACT algorithm tracks the exact KW solution to within a small time error, η. Therefore, it will be feasible if the exact solution, u, satisfies (5) with M replaced by $M + \eta$. We mentioned in Sec. 3 (property (iv)) that this occurs if the K–function lies above the ray from the origin with slope $M + \eta$. This is a reasonable thing to expect in practice since a feasible K-function cannot dip below the ray with slope M from the origin; see Fig. 4b. In pathological cases where the K-function dips below the $M + \eta$ ray but not below the M-ray, one should replace the K-function by the upper envelope of the K-curve and the ray of slope $M + \eta$ to achieve the desired effect. Since the upper bounds of a flex-time system can be determined with arbitrarily small h (so that $\eta \to 0$), no changes are necessary for this type of application.

6. DISCUSSION

This paper has demonstrated that the instability of supply chains is intimately related to the concept of "gain". It has shown that non-anticipative chains cannot be stable if the gain is positive. Formulae were also given for the amplitude and period of the oscillations of unstable chains. For order-based systems the oscillation period was shown to be a multiple of the number of intervals used in the adaptation term (see example 1 of section 2). For systems where it is expensive to change the production rates, suppliers have an incentive to adapt sluggishly to the demand by using many intervals in their forecasting procedures. In industries where orders are placed infrequently (e.g., on the order of months) the total adaptation interval may be comparable with a year. Thus, the supply chain effects discussed in this paper could induce oscillations with a period of multiple years. They may be an important cause of business cycles.

Using concepts from traffic flow theory, the paper also demonstrated how to construct strongly stable policies for homogeneous chains. The results apply to both discrete-time and "flex-time" systems. As in the case of traffic flow theory, the results can be easily extended to production networks of one and multiple commodities, and to cases where suppliers do not share the same time lattice. Space restrictions prevent a development of these ideas, but they can be found in Daganzo (2001). This reference also discusses the optimization of supply chains, subject to stability constraints, and some duality results pertaining to the stability of queuing networks.

This paper has assumed perfect reliability in communications and execution, and this cannot be expected in practice. Therefore, an understanding of the occasional system failures and recoveries arising in mildly unreliable systems should also be developed, e.g., when the lags M_j are chosen to be large, but not equal to the one-hundredth percentile of the random process times, or when some items are defective. Our initial research on this subject shows that mildly unreliable systems designed to fail infrequently and to be strongly stable when operating normally, continue to be strongly stable despite the failures. A detailed analysis of mildly unreliable systems including cost estimation is beyond the scope of this paper, however.

ACKNOWLEDGEMENTS

I wish to thank Professors Michael Cassidy, Phil Kaminsky and Samer Madanat of the University of California, Berkeley, and PhD candidate Alejandro Lago, for offering very valuable feedback after patiently sitting through a series of lectures in which these ideas were roughly described. A group of advanced graduate students attending the seminars also offered valuable suggestions.

REFERENCES

Daganzo, C.F. (1995) "A finite difference approximation of the kinematic wave model of traffic flow" *Trans. Res.* 29B(4), 261-276.

Daganzo, C.F. (1997) "A simple traffic analysis procedure" ITS Working Paper, UCB-ITS-WP-97-4, Univ. of California, Berkeley, CA. Abridged in *J. Nets and Spatial Theory* (in press).

Daganzo, C.F. (2001) *A Theory of Supply Chains*, Institute of Transportation Studies Research Report UCB-ITS-RR-2001-7, Univ. of California, Berkeley, CA. (In press, Springer, Heidelberg, Germany.)

Erera, A. (2001) "Private communication."

Lax, P.D. (1973), *Hyperbolic systems of conservation laws and the mathematical theory of shockwaves*, SIAM Regional Conference Series in Applied Mathematics. J.W. Arrowsmith Ltd., Bristol, U.K.

LeVeque, R.J. (1992), *Numerical methods for conservation laws*, (2nd edition), Birkhauser-Verlag, Boston, MA.

Lighthill, M.J. and G.B. Whitham (1955), "On kinematic waves. I flow movement in long rives. II A theory of traffic flow on long crowded roads," Proc. Roy. Soc., A. 229, 281-345.

Newell, G.F. (1993), "A simplified theory of kinematic waves in highway traffic, I general theory, II queuing at freeway bottlenecks, III multi-destination flows", Trans. Res., 27B, 281-313.

Richards, P.I. (1956), "Shockwaves on the highway," Opns. Res., 4, 42-51.

Ryan, J. K. (1997) "Analysis of Inventory Models with Limited Demand Information," Ph.D. thesis, IEMS Dept., Northwestern University, Evanston, IL.

Simchi-Levi, D., Kaminsky, P. and Simchi-Levi, E. (2000) "Designing and Managing the Supply Chain"; McGraw-Hill, New York, N.Y.

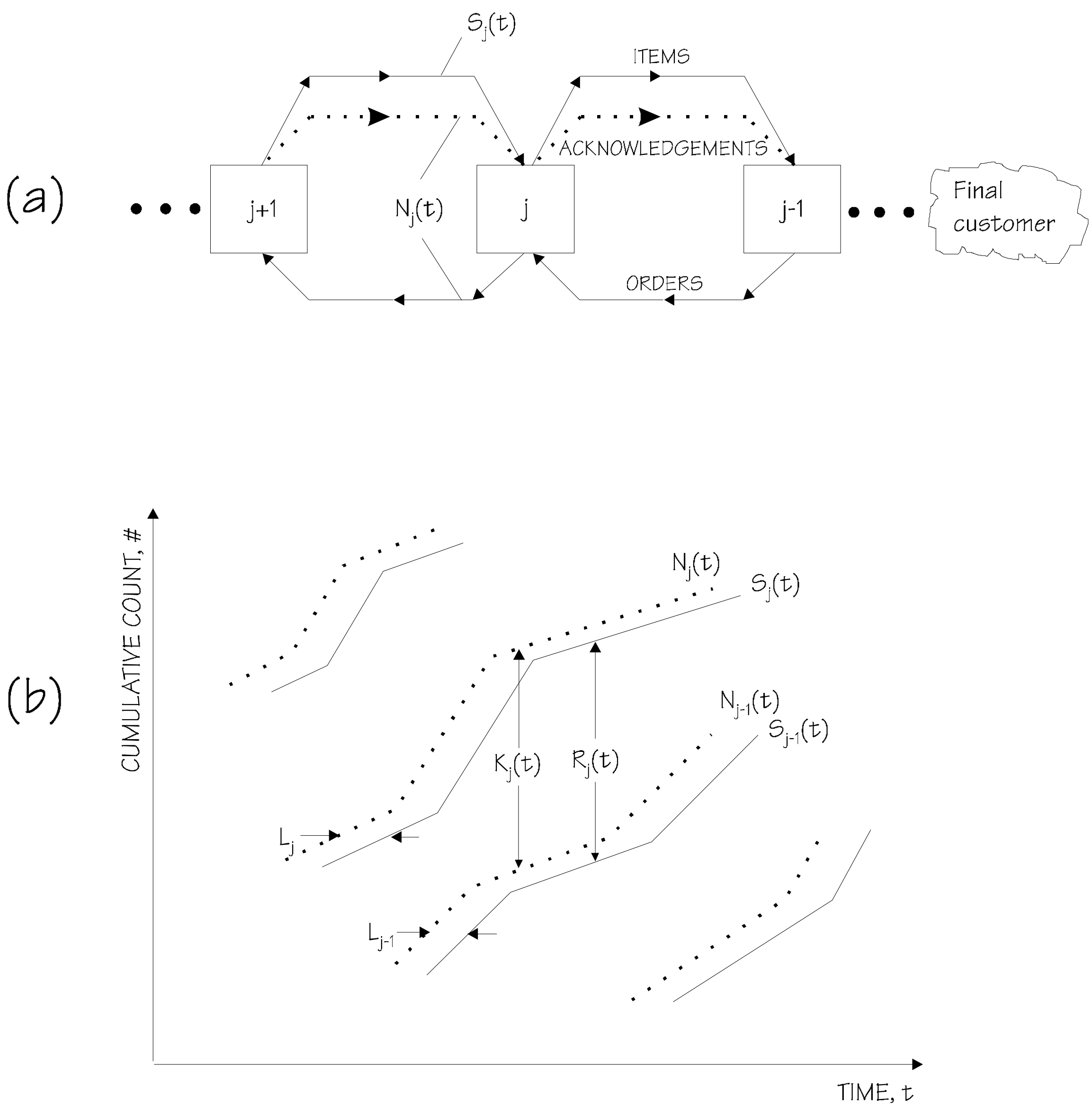
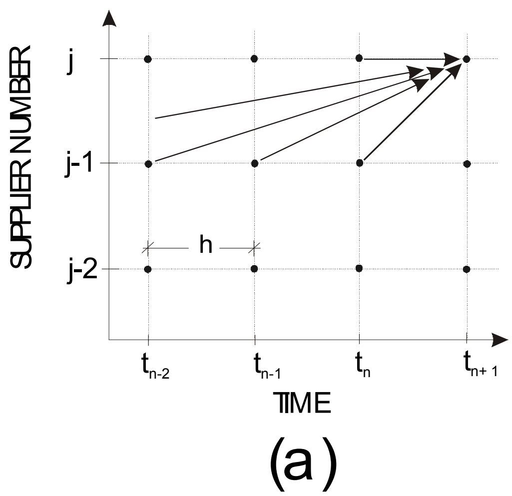
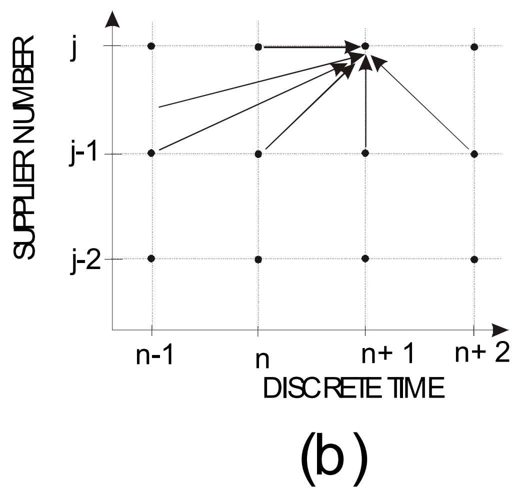


Figure 1. Flows on a supply chain: (a) Physical diagram; (b) *N*-curve representation.





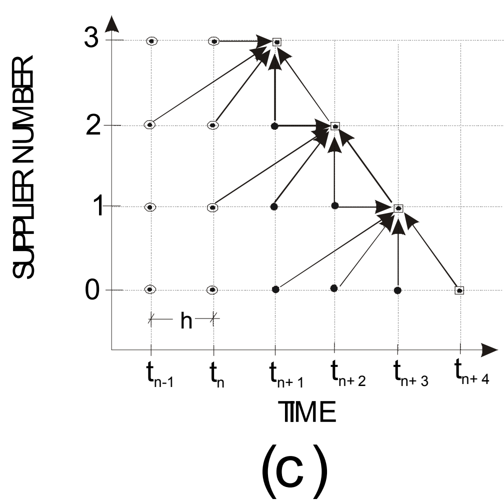


Figure 2. Stencils for different discrete-time, autonomous algorithms: (a) Non-anticipative, pull stencil; (b) Anticipative, pull stencil with A=2; (c) Influence diagram for anticipative, pull algorithm with J=3 and A=2 (solid dots = cumulative counts determined by time n; circled dots = cumulative counts realized by time n).

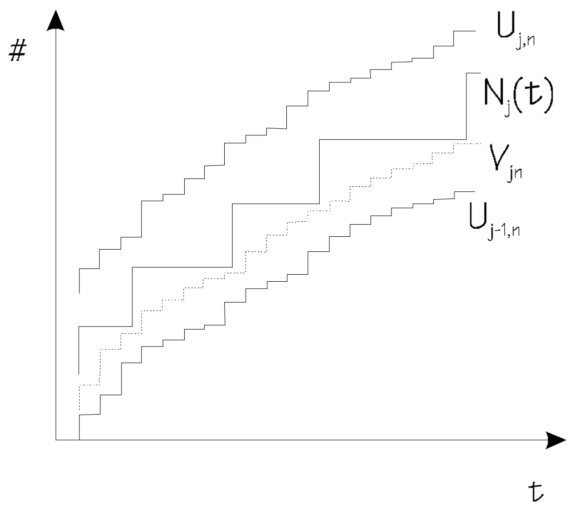


Figure 3. Bounds and *N*-curves for a flex-time system.

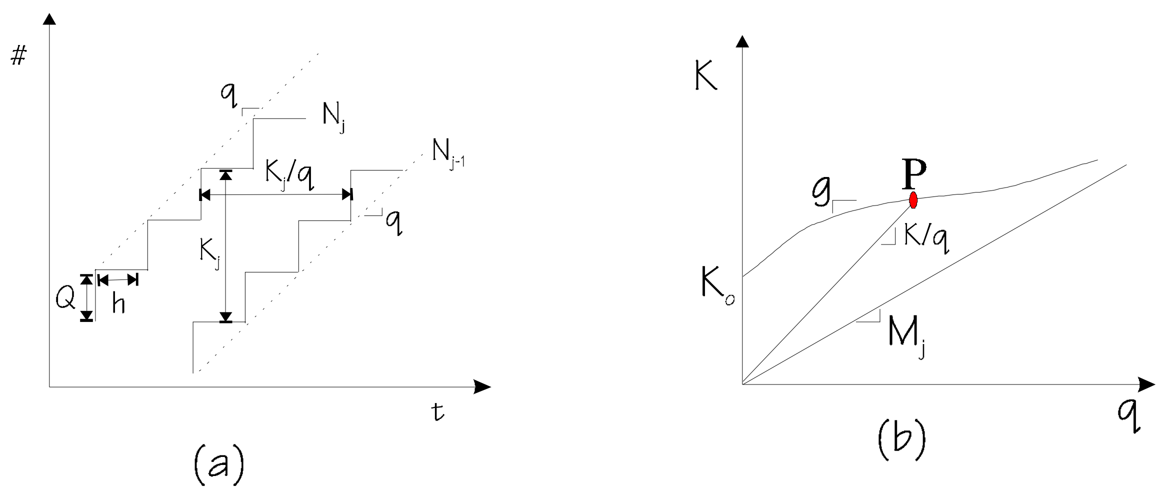


Figure 4. Steady state diagrams: (a) *N*-curves; (b) A feasible *K*-function

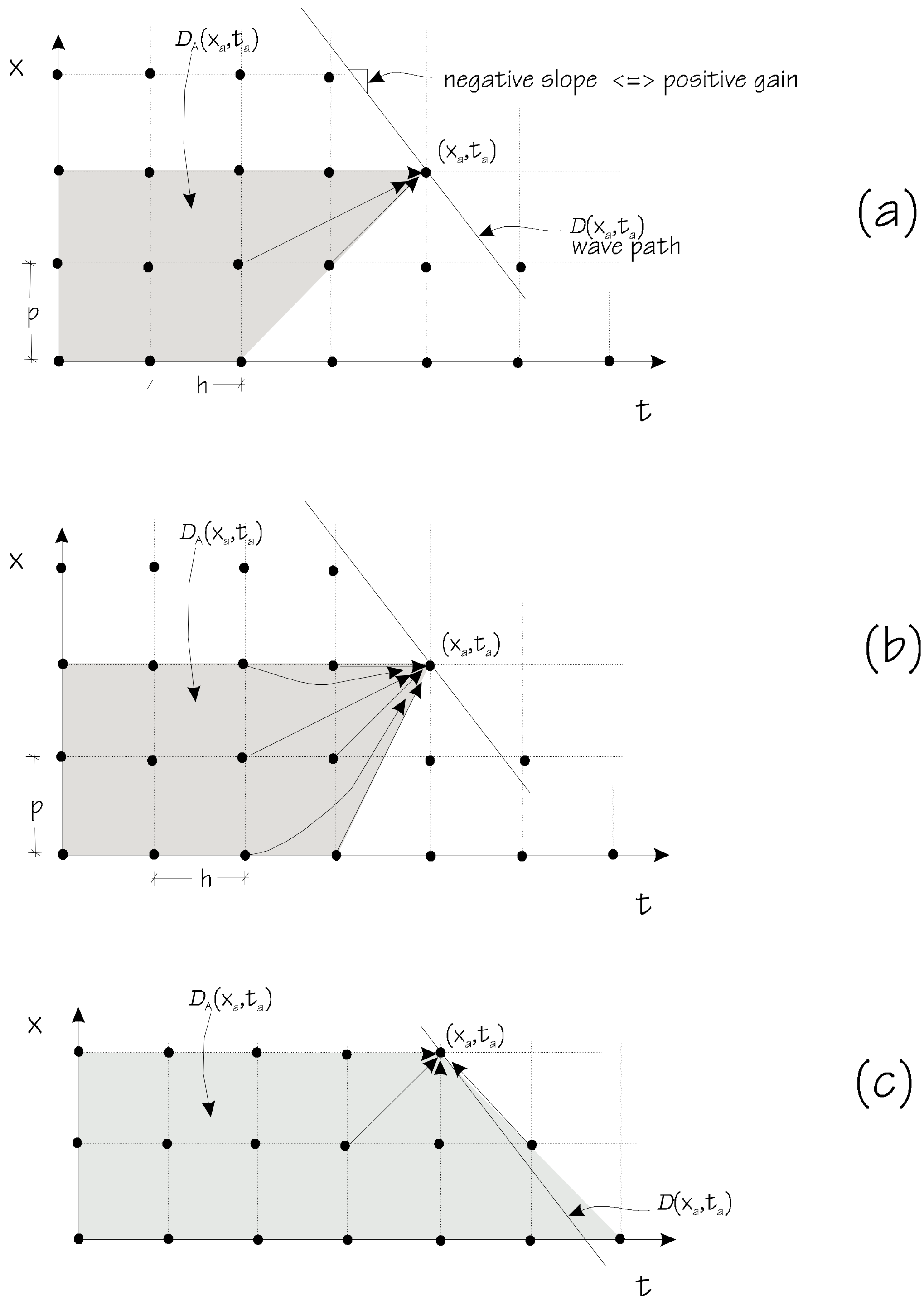


Figure 5. Domains of dependence for different algorithms (positive gain problems): (a) Autonomous, non-anticipative; (b) Non-autonomous, non-anticipative (c) Autonomous, anticipative.

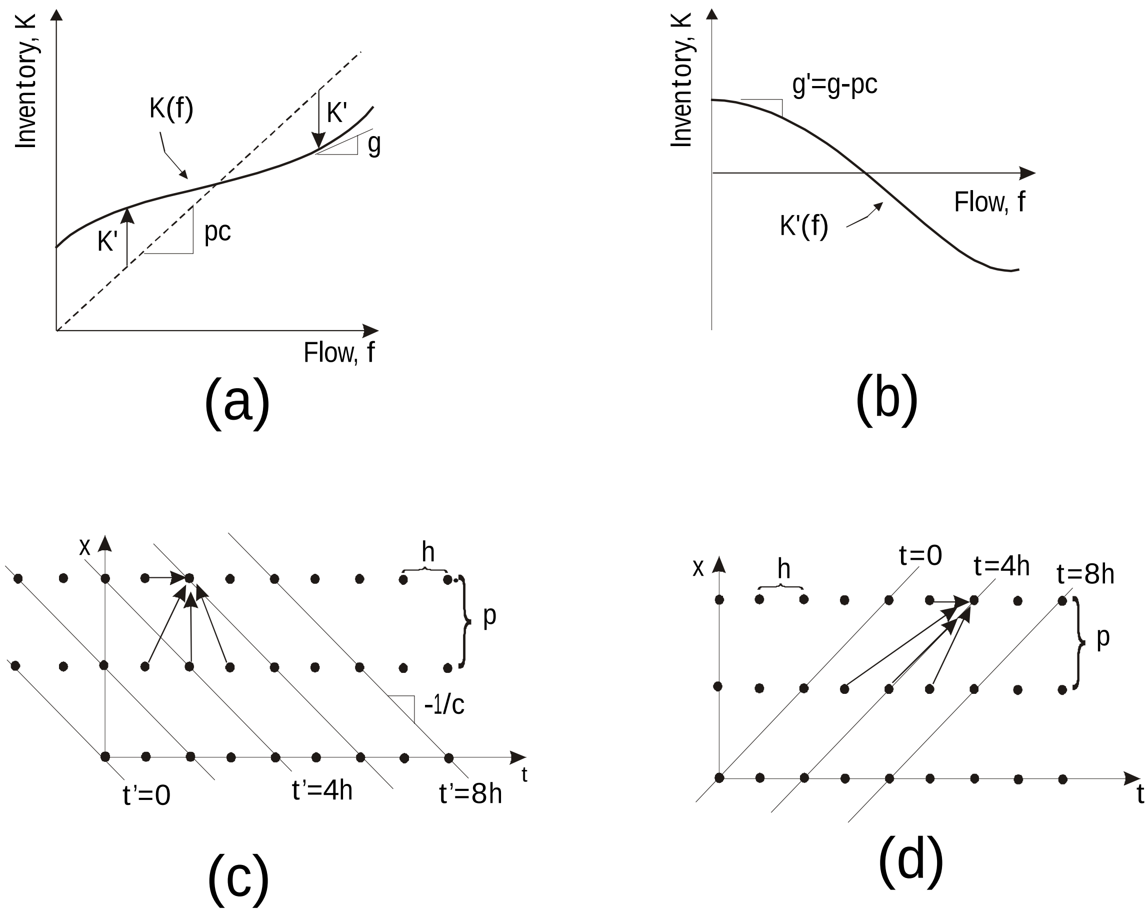
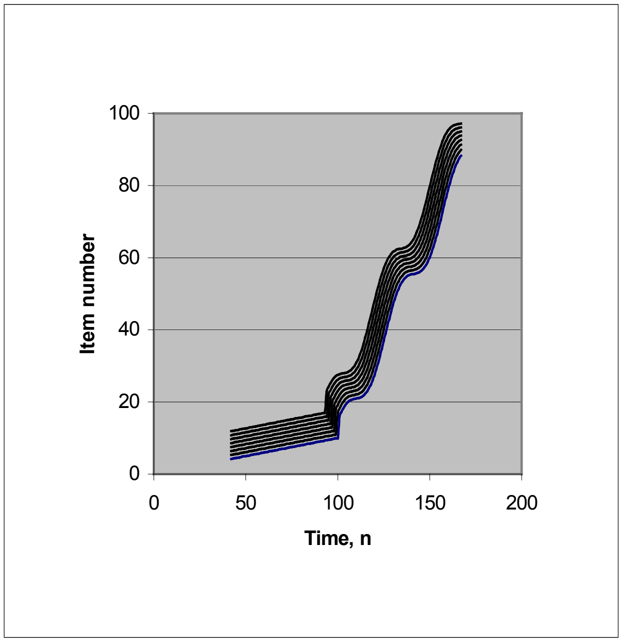
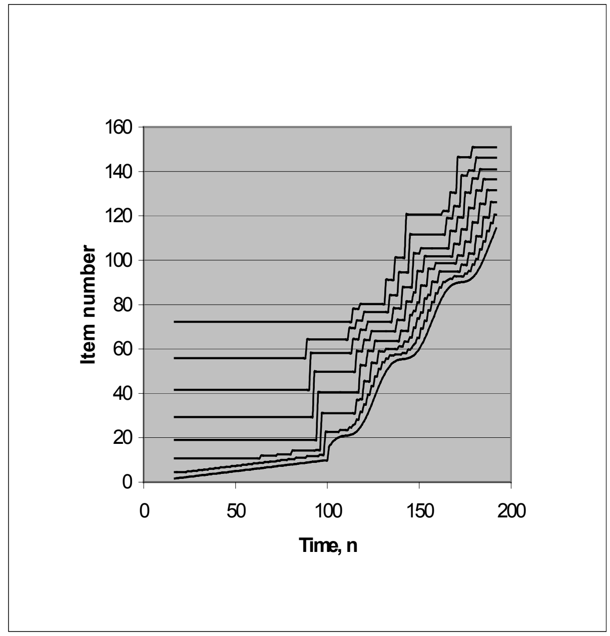


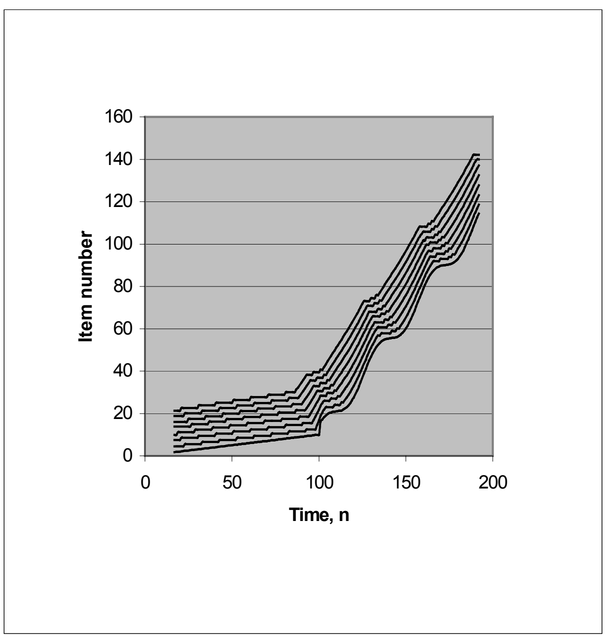
Figure 6. The transformation to asynchronous time: (a) Non-linear *K*-function; (b) Transformed *K*'-function; (c) Space-time diagram for original problem (*A*=2); (d) Transformed space-time diagram (*A*'=0).



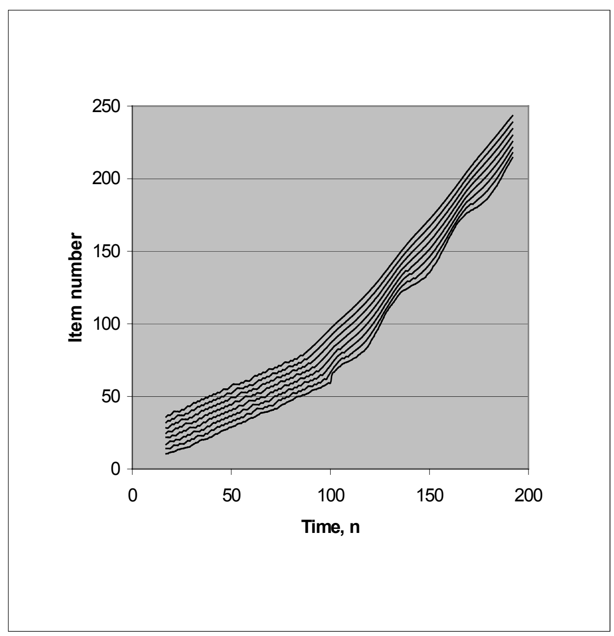
(A)



(B)



(C)



(D)

Figure 7. Cumulative *N*-curves for seven suppliers: (A) linear model with exact algorithm; (B) non-linear model with ACT algorithm, cp-g_{max} = 0 (unstable); (C) non-linear model with ACT algorithm, cp-$g_{max} = h$ (stable); (D) non-linear model with noisy data and ACT algorithm, cp- $g_{max} = h$ (stable).