DISTRIBUTED ASYNCHRONOUS RELAXATION METHODS
FOR CONVEX NETWORK FLOW PROBLEMS[†]

by

Dimitri P. Bertsekas*
Didier El Baz**

Abstract

We consider the solution of the single commodity strictly convex network flow problem in a distributed asynchronous computation environment. The dual of this problem is unconstrained, differentiable, and well suited for solution via Gauss-Seidel relaxation. We show that the structure of the dual allows the successful application of a distributed asynchronous method whereby relaxation iterations are carried out in parallel by several processors in arbitrary order and with arbitrarily large interprocessor communication delays.

*Room No. 35-210, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

**Room No. 35-312B, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139. On leave from LAAS. Toulouse, France, and supported by an INRIA Grant.

1. Introduction

Consider a directed graph with set of nodes N and set of arcs A. Each arc (i,j) has associated with it a cost function $g_{ij}: R \rightarrow (-\infty, +\infty]$. We denote by f_{ij} the flow of the arc (i,j) and consider the problem of minimizing total cost subject to a conservation of flow constraint at each node

$$\text{minimize} \quad \sum_{(i,j) \in A} g_{ij}(f_{ij}) \tag{1}$$

$$\text{subject to} \quad \sum_{(m,i) \in A} f_{mi} - \sum_{(i,j) \in A} f_{ij} = 0, \quad \forall i \in N.$$

We assume that problem (1) has at least one feasible solution. We also make the following standing assumptions on g_{ij}:

a) g_{ij} is strictly convex, and lower semicontinuous.

b)   The conjugate convex function of g_{ij}, defined by

$$g_{ij}^*(t_{ij}) = \sup_{f_{ij}} \{t_{ij} f_{ij} - g_{ij}(f_{ij})\}, \tag{2}$$

is real valued, i.e. $-\infty < g_{ij}^*(t_{ij}) < \infty$ for all real t_{ij}. (Because of the strict convexity assumed in a) above, g_{ij}^* is also continuously differentiable and its gradient denoted $\nabla g_{ij}^*(t_{ij})$ is the unique f_{ij} attaining the supremum in (2)--see [7], pp. 218,253).

It is easily seen from (2) that Assumption b) implies that $\lim_{|f_{ij}| \rightarrow \infty} g_{ij}(f_{ij}) = \infty$. Therefore the objective function of the primal problem (1) has bounded level sets ([7], Section 8). It follows that there exists an optimal solution for problem (1) which must be unique in view of the strict convexity assumed in a).

The problem above is of great practical interest and has been studied for a long time. Except for strict convexity our assumptions are not overly restrictive. For example they are satisfied in the following two cases:

1) The <u>constrained case</u> where g_{ij} is of the form

$$g_{ij}(f_{ij}) = \begin{cases} \infty & \text{if } f_{ij} \notin [\ell_{ij}, c_{ij}] \\ \hat{g}_{ij}(f_{ij}) & \text{otherwise} \end{cases} \tag{3}$$

where ℓ_{ij} and c_{ij} are given lower and upper bounds on the arc flow, and $\hat{g}_{ij}$ is a strictly convex, real valued function defined on the real line R.

2) The <u>unconstrained case</u> where g_{ij} is strictly convex, real valued and its right and left derivatives g_{ij}^+ and g_{ij}^- satisfy

$$\lim_{f_{ij} \to \infty} g_{ij}^+(f_{ij}) = \infty, \quad \lim_{f_{ij} \to -\infty} g_{ij}^-(f_{ij}) = -\infty. \tag{4}$$

A dual problem for (1) is given by

minimize $q(p)$

subject to no constraints on the vector $p = \{p_i | i \in N\}$, $\tag{5}$

where q is the dual functional given by

$$q(p) = \sum_{(i,j) \in A} g_{ij}^*(p_i - p_j). \tag{6}$$

We refer to p as a <u>price vector</u> and its components p_i as <u>prices</u>. The ith price is really a Lagrange multiplier associated with the ith conservation of flow constraint. The duality between problems (1) and (5) is well known and is explored in great detail in the recent book by Rockafellar [1]. The earlier book by Rockafellar [7] gives the necessary and sufficient condition for optimality of a pair (f,p). A feasible flow vector $f = \{f_{ij} | (i,j) \in A\}$ is optimal for (1) and a price vector $p = \{p_i | i \in N\}$ is optimal for (5) if and

only if for all arcs (i,j) ([7], pp. 337-338)

$p_i - p_j$ is a subgradient of g_{ij} at f_{ij}.

An equivalent condition is

$$f_{ij} = \nabla g^*_{ij}(p_i - p_j), \quad \forall \ (i,j) \epsilon A. \tag{7}$$

Any one of these equivalent relations is referred to as the complementary slackness condition, and is shown in Figure 1.
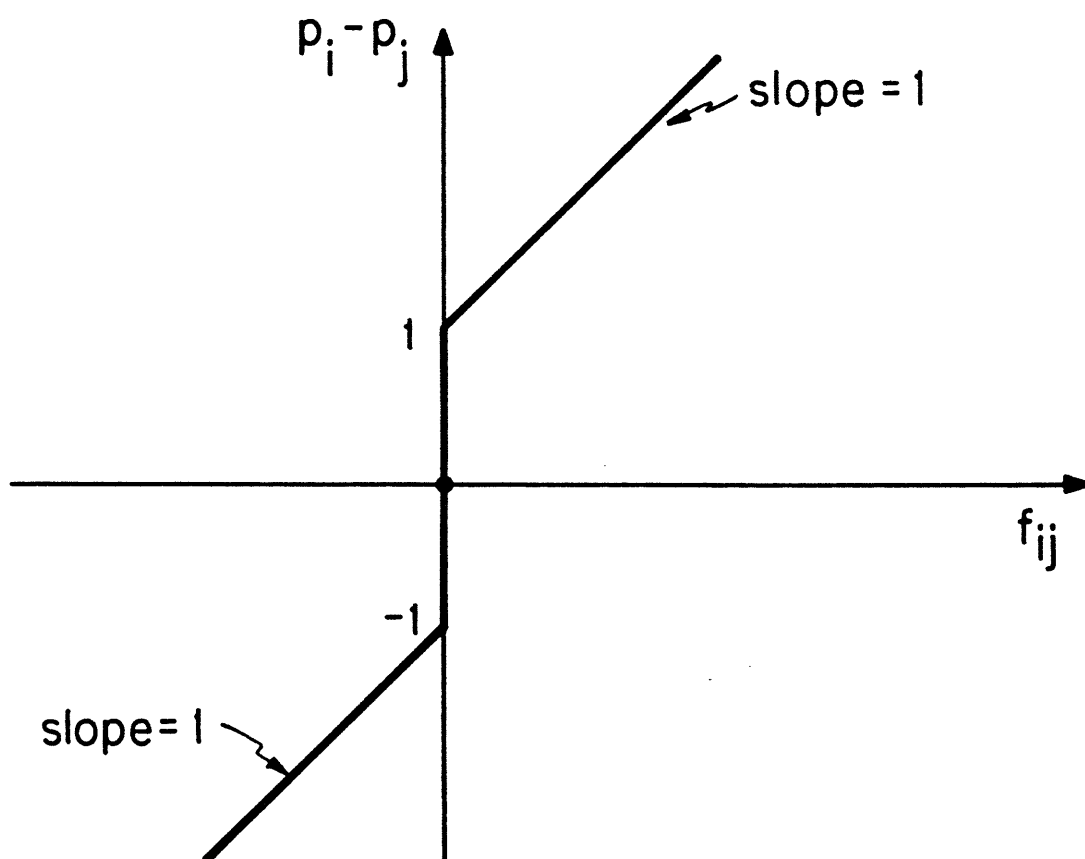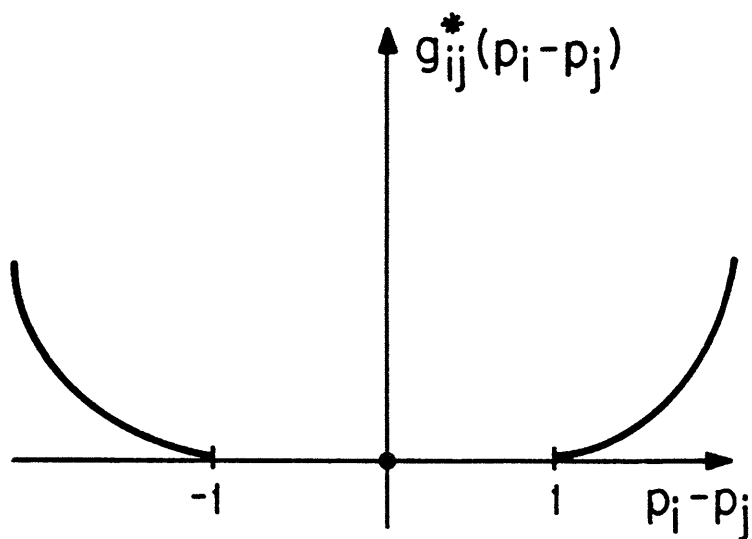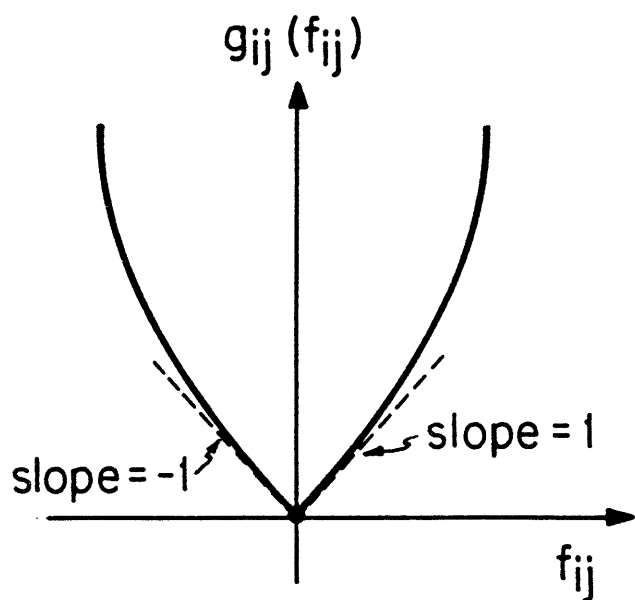
Figure 1: Complementary slackness condition diagram for cost function

$$g_{ij}(f_{ij}) \quad = \quad |f_{ij}| + \frac{1}{2}(f_{ij})^2$$

Since the dual problem is unconstrained and differentiable it is natural to consider algorithmic solution by a descent iterative method. The Gauss-Seidel relaxation method is particularly interesting in this respect since it admits a simple implementation. Given a price vector p, a node i is selected and its price p_i is changed (relaxed) to a value $\hat{p}_i$ such that

$$\sum_{(m,i)\varepsilon A} \nabla g^*_{mi}(p_m - \hat{p}_i) = \sum_{(i,j)\varepsilon A} \nabla g^*_{ij}(\hat{p}_i - p_j) \qquad (8)$$

It is easily seen (compare with the definition (6) of the dual cost q) that this equation is equivalent to $\frac{\partial q}{\partial p_i} = 0$, so the dual cost is minimized at $\hat{p}_i$ with respect to the ith price, all other prices being kept constant. The algorithm proceeds by relaxing the prices of all nodes in cyclic order and repeating the process. The convergence of this algorithm does not follow immediately from standard results on relaxation methods [2], [3], [4] since these results require some assumption that is akin to strict convexity of the dual objective function which does not hold here (for a counterexample see Powell [5]). However Cottle and Pang [6] have shown convergence of a network algorithm based on relaxation. It applies to transportation problems with quadratic cost function, and involves certain restrictions in the way relaxation is carried out. Their results is substantially extended in Bertsekas, Hossein, and Tseng [19].

Our main objective in this paper is to explore the convergence properties of distributed versions of the relaxation method just described. Here we assume that each price p_i is under the control of a separate processor who changes p_i to $\hat{p}_i$ on the basis of (8) and communicates the new value to the other processors. One can consider a parallel computation procedure carried out in an orderly manner whereby all processors exchange

their current prices before carrying out their relaxation iteration. Mathematically this would be equivalent to a Jacobi type of relaxation procedure. We would like to consider, however, a much more general procedure whereby the communication between processors is not regular, and the information available at some processors regarding prices of other processors may be arbitrarily out-of-date. In addition we allow some processors to iterate more frequently than others. Models of such asynchronous algorithms have been formulated some time ago and by now there is considerable understanding of their convergence properties (see [8]-[16]; [17] is a survey). It turns out that the dual problem (5) has structure that allows us to show that the asynchronous relaxation method has satisfactory convergence properties. This is particularly true when the dual problem (5) has an essentially unique optimal solution. Otherwise satisfactory convergence depends on the starting point. These results are all new and are shown in Section 3. The next section analyzes the structure of the dual solution set and provides some preliminary analysis.

The results of this paper carry over verbatim to the case where the conservation of flow constraint has the form

$$\sum_{(m,i)\in A} f_{mi} - \sum_{(i,j)\in A} f_{ij} = b_i , \qquad \forall i\in N$$

where b_i are given scalars with $\sum_{i\in N} b_i = 0$. The dual cost of (6) must then include the term $\sum_{i\in N} b_i p_i$, and the relaxation equation (8) must include an additional term b_i in its right side. This extension is important from the practical point of view, but we have restricted attention to the case where $b_i = 0$, $\forall i\in N$ in order to simplify notation.

The results of this paper can also be extended in a simple manner to network problems with positive gains and strictly convex arc costs. This extension was

mentioned to us by P. Tseng who also showed [20] two additional interesting facts. First that Proposition 2 holds even if the strict convexity assumption of a) is removed thereby including the important class of linear minimum cost flow problems. Second that, within the class of monotropic programming problems, the largest class for which the monotonicity property of Proposition 1 holds is the class of network flow problems with positive gains.

Our notational conventions are that a subscript denotes a node or processor index, and a superscript denotes a time or iteration index. All vector inequalities should be interpreted in a coordinatewise sense. In order to simplify notation we have implicitly assumed that there is at most one arc associated with any ordered pair of nodes i and j, so that the arc notation (i,j) has unambiguous meaning. However this assumption is not essential to any of our results.

2. Structure of the Optimal Dual Solution Set

Our standing assumptions a),b) guarantee that the primal problem (1) has a unique optimal solution. Existence of an optimal solution of the dual problem can be guaranteed under an additional (mild) regular feasibility assumption in which case the existence theorem of [1], p. 360 applies. On the other hand the optimal solution of the dual problem is never unique since adding the same constant to all coordinates of a price vector p leaves the dual cost unaffected. We can remove this degree of freedom by constraining the price of one node, say node N, to be zero. (With slight abuse of notation we number nodes as 1,2,...,N). Thus we consider the reduced dual optimal solution set P* defined by

$$P* = \{p* \mid q(p*) = \min_p q(p), \; p_N^* = 0\} \tag{9}$$

where q is the dual objective function

$$q(p) = \sum_{(i,j)\varepsilon A} g_{ij}^*(p_i - p_j) \; . \tag{10}$$

For the most part of the paper, we will operate under the following assumption:

Assumption 1: The reduced dual optimal solution set P* is nonempty and compact.

Assumption 1 is not overly restrictive. For example let $\{f_{ij}^* \mid (i,j)\varepsilon A\}$ be the unique primal optimal solution, and consider the set of arcs

$$\hat{A} = \{(i,j) \mid f_{ij}^* \text{ lies in the interior of the set } \{f_{ij} \mid g_{ij}(f_{ij}) < \infty\}\} \tag{11}$$

Then Assumption 1 is satisfied if the subgraph $(N,\hat{A})$ is connected. To see this note that for all arcs $(i,j)\varepsilon\hat{A}$ we have a bounded set of subgradients

of g_{ij} at f^*_{ij} thereby implying a bounded set of price differences $p_i - p_j$ corresponding to dual optimal solutions [cf. (6)]. Note that in the unconstrained case mentioned in the previous section every arc belongs to $\hat{A}$ so, if the original graph is connected, Assumption 1 is satisfied. The constrained case of the previous section can be converted to the unconstrained case by replacing constraints by nondifferentiable penalty functions (see [18], Section 5.5). For example, assuming a dual optimal solution exists, a constraint $f_{ij} \geq 0$ can be eliminated by adding to the cost g_{ij} a penalty

$c \max\{0, -f_{ij}\} + [\max\{0, -f_{ij}\}]^2$ with c positive and sufficiently large.

Consider now the set

$$P = \{p | p_N = 0\} \tag{12}$$

and for $i=1,\ldots,N-1$, the point-to-set mapping R_i which assigns to a price vector $p \varepsilon P$ the set of all prices $\hat{p}_i$ that minimize the dual cost along the ith price starting from p, i.e. (cf. (8)).

$$R_i(p) = \{\hat{p}_i | \sum_m \nabla g^*_{mi}(p_m - \hat{p}_i) = \sum_j \nabla g^*_{ij}(\hat{p}_i - p_j)\} \quad . \tag{13}$$

It is well known that a real valued convex function having one compact level set, has all its level sets compact ([7], p.70). Therefore under Assumption 1 the sets $R_i(p)$, $p \varepsilon P$ are all nonempty, compact intervals. It follows that under Assumption 1 the (point-to-point) mappings

$$\overline{R}_i(p) = \max_{\hat{p}_i \varepsilon R_i(p)} \hat{p}_i \tag{14}$$

$$\underline{R}_i(p) = \min_{\hat{p}_i \varepsilon R_i(p)} \hat{p}_i \tag{15}$$

are well defined on the set P. We call $\overline{R}_i$ ($\underline{R}_i$) the <u>ith maximal (minimal)</u>
<u>relaxation mapping</u>. It gives the maximal (minimal) minimizing point of the
dual cost along the ith coordinate starting from its argument. The point-to-
set mapping R_i is called the <u>ith relaxation mapping</u>.

Some key facts are given in the following proposition.

<u>Proposition 1</u>: Let Assumption 1 hold. The mappings $\overline{R}_i$ and $\underline{R}_i$ are
continuous on P. They are also monotone on P in the sense that for any
p, p'εP, i = 1,...,N-1 we have

$$\overline{R}_i(p) \leq \overline{R}_i(p') \qquad \text{if} \quad p \leq p', \tag{16}$$

$$\underline{R}_i(p) \leq \underline{R}_i(p') \qquad \text{if} \quad p \leq p'. \tag{17}$$

<u>Proof</u>: To show continuity of $\overline{R}_i$ we argue by contradiction. Suppose
there exists a convergent price vector sequence $p^k \to p$ such that the cor-
responding sequence $\{\overline{R}_i(p^k)\}$ does not converge to $\overline{R}_i(p)$. By passing to a
subsequence if necessary suppose that for some $\delta > 0$ we have

$$\overline{R}_i(p) \geq \overline{R}_i(p^k) + \delta, \qquad \forall \, k \tag{18}$$

(the proof is very similar if $\delta < 0$ and the inequality is reversed). By
the definition of $\overline{R}_i$ we have

$$\sum_m \nabla g^*_{mi}(p_m - \overline{R}_i(p)) = \sum_j \nabla g^*_{ij}(\overline{R}_i(p) - p_j), \tag{19}$$

$$\sum_m \nabla g^*_{mi}(p^k_m - \overline{R}_i(p^k)) = \sum_j \nabla g^*_{ij}(\overline{R}_i(p^k) - p^k_j), \qquad \forall \, k. \tag{20}$$

Since $p^k \rightarrow p$ it follows using (18) that for sufficiently large k we have

$$p_m^k - \overline{R}_i(p^k) > p_m - \overline{R}_i(p) \quad , \quad \forall \ (m,i)\epsilon A$$

$$\overline{R}_i(p^k) - p_j^k < \overline{R}_i(p) - p_j \quad , \quad \forall \ (i,j)\epsilon A.$$

Therefore for sufficiently large k we have using the convexity of g_{mi}^*, g_{ij}^*

$$\nabla g_{mi}^*(p_m^k - \overline{R}_i(p^k)) \geq \nabla g_{mi}^*(p_m - \overline{R}_i(p)) \quad , \quad \forall \ (m,i)\epsilon A,$$

$$\nabla g_{ij}^*(\overline{R}_i(p^k) - p_j^k) \leq \nabla g_{ij}^*(\overline{R}_i(p) - p_j), \quad , \quad \forall \ (i,j)\epsilon A.$$

Using these relations together with (19), (20) we obtain for all sufficiently large k

$$f_{mi} \overset{\Delta}{=} \nabla g_{mi}^*(p_m - \overline{R}_i(p)) = \nabla g_{mi}^*(p_m^k - \overline{R}_i(p^k)), \quad \forall \ (m,i)\epsilon A \quad (21)$$

$$f_{ij} \overset{\Delta}{=} \nabla g_{ij}^*(\overline{R}_i(p) - p_j) = \nabla g_{ij}^*(\overline{R}_i(p^k) - p_j^k), \quad \forall \ (i,j)\epsilon A. \quad (22)$$

Consider the intervals I_{mi} and I_{ij} given by

$$I_{mi} = \{t \mid \nabla g_{mi}^*(t) = f_{mi}\} \quad , \quad \forall \ (m,i)\epsilon A$$

$$I_{ij} = \{t \mid \nabla g_{ij}^*(t) = f_{ij}\} \quad , \quad \forall \ (i,j)\epsilon A.$$

For k sufficiently large so that (21), (22) hold we have

$$\overline{R}_i(p) = \max\{\hat{p}_i \mid \hat{p}_i \ \epsilon \ p_m - I_{mi}, \ (m,i) \ \epsilon A, \ \hat{p}_i \epsilon I_{ij} - p_j, \ (i,j)\epsilon A\}$$

$$\overline{R}_i(p^k) = \max\{\hat{p}_i \mid \hat{p}_i \ \epsilon \ p_m^k - I_{mi}, \ (m,i)\epsilon A, \ \hat{p}_i \ \epsilon \ I_{ij} - p_j^k, \ (i,j)\epsilon A\}$$

Since $p^k \to p$, it is evident from these relations that $\overline{R}_i(p^k) \to \overline{R}_i(p)$ thereby contradicting (18).

To show monotonicity of $\overline{R}_i$ we again argue by contradiction. Suppose there exist p and p' such that $p'_j \geq p_j$, $\forall$ j=1,...,N-1 but $\overline{R}_i(p) > \overline{R}_i(p')$. It follows then that

$$p'_m - \overline{R}_i(p') > p_m - \overline{R}_i(p), \qquad \forall (m,i) \epsilon A$$

$$\overline{R}_i(p') - p'_j < \overline{R}_i(p) - p_j, \qquad \forall (i,j) \epsilon A.$$

Therefore

$$\nabla g^*_{mi}(p'_m - \overline{R}_i(p')) \geq \nabla g^*_{mi}(p_m - \overline{R}_i(p)), \qquad \forall (m,i) \epsilon A \qquad (23)$$

$$\nabla g^*_{ij}(\overline{R}_i(p') - p'_j) \leq \nabla g^*_{ij}(\overline{R}_i(p) - p_j), \qquad \forall (i,j) \epsilon A \qquad (24)$$

Since by definition we have

$$\sum_m \nabla g^*_{mi}(p'_m - \overline{R}_i(p')) = \sum_i \nabla g^*_{ij}(\overline{R}_i(p') - p'_j) \qquad (25)$$

$$\sum_m \nabla g^*_{mi}(p_m - \overline{R}_i(p)) = \sum_j \nabla g^*_{ij}(\overline{R}_i(p) - p_j) \qquad (26)$$

it follows that equality holds in (23), (24), i.e.

$$f_{mi} \stackrel{\Delta}{=} \nabla g^*_{mi}(p'_m - \overline{R}_i(p')) = \nabla g^*_{mi}(p_m - \overline{R}_i(p))$$

$$f_{ij} \stackrel{\Delta}{=} \nabla g^*_{ij}(\overline{R}_i(p') - p'_j) = \nabla g^*_{ij}(\overline{R}_i(p) - p_j)$$

Consider the intervals

$$I_{mi} = \{t \mid \nabla g^*_{mi}(t) = f_{mi}\}$$

$$I_{ij} = \{t \mid \nabla g^*_{ij}(t) = f_{ij}\},$$

and let

$$\delta = \overline{R}_i(p) - \overline{R}_i(p').$$

We have for all $(m,i)\varepsilon A$

$$p'_m - \overline{R}_i(p')\varepsilon \ I_{mi}, \quad p_m - \overline{R}_i(p)\varepsilon \ I_{mi}$$

and since $p_m \leq p'_m$ we obtain

$$p_m - \overline{R}_i(p) \ \leq \ p'_m - \overline{R}_i(p') - \delta \ \leq \ p'_m - \overline{R}_i(p').$$

Therefore

$$p'_m - \overline{R}_i(p') - \delta \ \varepsilon \ I_{mi}, \qquad \forall \ (m,i)\varepsilon A$$

and similarly

$$\overline{R}_i(p') + \delta - p_j \ \varepsilon \ I_{ij}, \qquad \forall \ (i,j)\varepsilon A.$$

It follows that

$$\overline{R}_i(p') + \delta \ \varepsilon \ R_i(p')$$

thereby contradicting the maximal nature of $\overline{R}_i$ [cf. (14)].

The proof of continuity and monotonicity of $\underline{R}_i$ is analogous with the

one just given for $\overline{R}_i$ and is omitted   Q.E.D.

The monotonicity and continuity of the mappings $\overline{R}_i$ and $\underline{R}_i$ imply a thus far unreported and somewhat surprising property of the optimal dual solution set.

Proposition 2: Let Assumption 1 hold. There exist a maximal and a minimal optimal solution of the dual problem, i.e. there exist $\overline{p} \epsilon P^*$ and $\underline{p} \epsilon P^*$ such that

$$\underline{p} \leq p \leq \overline{p}, \qquad \forall p \epsilon P^* . \qquad (27)$$

Proof: Since P^* is nonempty and compact it contains a noninferior element $\overline{p}$ for which there is no vector $p \epsilon P^*$ such that $p \neq \overline{p}$ and $p_i \geq \overline{p}_i$ for all i. From the definition of $\overline{R}_i$ and the optimality of $\overline{p}$ we have $\overline{p}_i \leq \overline{R}_i(\overline{p})$ for all i. Furthermore for all i the vector $(\overline{p}_1, \ldots, \overline{p}_{i-1}, R_i(\overline{p}), \overline{p}_{i+1}, \ldots, \overline{p}_N)$ belongs to P^* so from noninferiority of $\overline{p}$ it follows that $\overline{R}_i(\overline{p}) \leq \overline{p}_i$. Therefore we have $\overline{p}_i = \overline{R}_i(\overline{p})$ for all i. Let now $\tilde{p}$ be a price vector obtained from $\overline{p}$ according to

$$\tilde{p}_i = \begin{cases} \overline{p}_i + \delta & , \ i=1,\ldots,N-1 \\ \\ 0 & i=N, \end{cases}$$

where $\delta > 0$ is sufficiently large so that

$$\tilde{p} \geq p, \qquad \forall p \epsilon P^* \qquad (28)$$

It is easily seen that we have $\bar{R}_i(\tilde{p}) \leq \tilde{p}_i$, for all i so, using the monotonicity of $\bar{R}_i$ shown in Proposition 1, we obtain

$$\bar{p} \leq \bar{R}^{k+1}(\tilde{p}) \leq \bar{R}^k(\tilde{p}) \quad , \qquad \forall\, k \qquad\qquad (29)$$

where $\bar{R}: R^{N-1} \rightarrow R^{N-1}$ is the mapping

$$\bar{R}(p) = [\bar{R}_1(p),\ldots,\bar{R}_{N-1}(p)] \qquad\qquad (30)$$

and $\bar{R}^k$ is the composition of $\bar{R}$ with itself k times. From (29) we see that the sequence $\bar{R}^k(\tilde{p})$ converges to some $\hat{p}$ and by continuity of $\bar{R}$ we must have $\hat{p} = \bar{R}(\hat{p})$ as well as $\hat{p} \geq \bar{p}$. Since $\hat{p} = \bar{R}(\hat{p})$ implies that $\hat{p} \epsilon P^*$ it follows from the choice of $\bar{p}$ that $\hat{p} \doteq \bar{p}$. Also from (28), (29) and the fact $p \leq \bar{R}(p)$ for all $p \epsilon P^*$ we obtain $\hat{p} = \bar{p} \geq p$ for all $p \epsilon P^*$ which shows that $\bar{p}$ is a maximal element of P^*. The proof for existence of a minimal element $\underline{p}$ is entirely similar.   Q.E.D.

3. Convergence Analysis of Asynchronous Relaxation

The model of distributed asynchronous computation we adopt is described in [11], [12]. With each node $i = 1,...,N-1$ we associate a processor that computes from time to time some element of $R_i(p)$ (here p is the latest price vector available to processor i), and sets the price p_i to this element. This price is then communicated at some later time to all other processors. Computation and communication at the various processors need not be synchronized. The precise model is as follows:

At each time instant, node i can be in one of three possible states compute, transmit, or idle. In the compute state node i computes a new price p_i. In the transmit state node i communicates the price p_i obtained from its own latest computation to one or more nodes $m(m \neq i)$. In the idle state node i does nothing related to the solution of the problem.

We assume that computation and transmission for each node takes place in time intervals $[t_1,t_2]$ with $t_1 < t_2$, but do not exclude the possibility that a node may be simultaneously transmitting to more than one node nor do we assume that the transmission intervals to these nodes have the same origin and/or termination. We also make no assumptions on the length, timing and sequencing of computation and transmission intervals other than the following:

Assumption 2: For every node i and time $t \geq 0$ there exists a time $t' > t$ such that $[t,t']$ contains at least one computation interval for i and at least one transmission interval from i to each node m such that $(m,i)\epsilon A$ or $(i,m)\epsilon A$.

Assumption 2 is very natural. It states in essence that no node "drops out of the algorithm" permanently--perhaps due to a hardware failure. Without this assumption there is hardly anything we can hope to prove.

Each node i has a buffer B_{im} for each $m \neq i$ where it stores the latest transmission from m, as well as a buffer B_{ii} where it stores its own price estimate p_i. The contents for each buffer B_{im} at time t are denoted $p_m^t(i)$. Thus $p_m^t(i)$ is, for every t, i and m an estimate of the price p_m available at node i at time t. It is important to realize in what follows that the buffer contents $p_m^t(i)$, and $p_m^t(i')$ at two different nodes i and i' need not coincide at all times. If $i \neq m$ and $i' \neq m$ the buffer contents $p_m^t(i)$, and $p_m^t(i')$ need not coincide at any time t. The vector of all buffer contents of node i is denoted $p^t(i)$, i.e.,

$$p^t(i) = \{p_m^t(i) \mid m = 1,\ldots,N-1\}.$$

The rules according to which the buffer contents $p_m^t(i)$ are updated are as follows:

(1) If $[t_1,t_2]$ is a transmission interval from node m to node i, the contents of the buffer B_{mm} at time t_1 are transmitted and entered in the buffer B_{im} at time t_2, i.e.

$$p_m^{t_2}(i) = p_m^{t_1}(m). \tag{31}$$

(2) If $[t_1,t_2]$ is a computation interval for node i, the content of the buffer B_{ii} is replaced at time t_2 with an element of $R_i(p^{t_1}(i))$, i.e.

$$p_i^{t_2}(i) \in R_i(p^{t_1}(i)) \tag{32}$$

(3) The contents of a buffer B_{ii} can change only at the end of a computation interval for node i. The contents of a buffer B_{im}, $i \neq m$ can change only at the end of a transmission interval from m to i.

The algorithm based on (32) will be called <u>Asynchronous Relaxation</u> <u>Method (ARM)</u>.

Our objective is to derive conditions under which limit points of the sequences $\{p^t(i)\}$ are optimal solutions of the dual problem (5). The following proposition is our main result. The proof is based on a general convergence theorem given in [12] (see also [17]) and applicable to asynchronous iterative algorithms such as the one just described. The key property that makes asynchronous convergence possible is the monotonicity of the mappings $\overline{R}_i$ and $\underline{R}_i$ shown in Proposition 1. This property is also present in dynamic programming models and has been similarly exploited to show the validity of asynchronous versions of the successive approximation method [11].

<u>Proposition 3</u>: Let Assumptions 1 and 2 hold. For any initial buffer contents $p^o(i) \in P$, $i = 1,\ldots,N-1$, each limit point of the sequences $\{p^t(i)\}$ generated by the ARM belongs to the set

$$\overline{P} = \{p \mid \underline{p} \leq p \leq \overline{p}\} \tag{33}$$

where $\overline{p}$ and $\underline{p}$ are the maximal and minimal dual optimal solutions. In particular, if the reduced dual optimal solution set P* consists of a unique vector p* we have

$$\lim_{t \to \infty} p^t(i) = p^* \quad , \quad i = 1,\ldots,N-1. \tag{34}$$

<u>Proof</u>: Let $\underline{p}$, $\tilde{p} \in P$ be price vectors such that

$$\underline{p} \leq p^o(i) \leq \tilde{p} \quad , \quad \forall \, i = 1,\ldots,N-1$$

and such that

$$\underset{\sim}{p} \leq \underline{R}(\underset{\sim}{p}) \leq \underline{p} \leq \overline{p} \leq \overline{R}(\tilde{p}) \leq \tilde{p},$$

$$\lim_{k \to \infty} \underline{R}^k(\underset{\sim}{p}) = \underline{p} \quad , \quad \lim_{k \to \infty} \overline{R}^k(\tilde{p}) = \overline{p}.$$

(The existence of such vectors was established in the proof of Proposition 2). Consider the sets

$$\overline{P}^k = \{p \mid \underline{R}^k(\underset{\sim}{p}) \leq p \leq \overline{R}^k(\tilde{p})\} \quad , \quad k = 1,2,\ldots \tag{35}$$

Note that the sequence $\{\overline{P}^k\}$ is nested and that the common intersection of the sequence is the set $\overline{P}$ of (33).

We will apply now a convergence theorem given in Section 3 of [12] (or Prop. 3.1 of [17]). According to this theorem the desired result will be proved if the following three conditions are satisfied. (Rather than consulting the references just cited, the reader may wish to think through the proof of this since it is rather simple).

a) If $p \varepsilon \overline{P}^k$ then for every i the vector p' with coordinates

$$p'_j = \begin{cases} p_j & \text{if } j \neq i \\ R_i(p) & \text{if } j = i \end{cases}$$

(cf. the equation (32) associated with computation at node i), also belongs to $\overline{P}^k$.

b) If $p \varepsilon \overline{P}^k$ and $\hat{p} \varepsilon \overline{P}^k$ then, for every i and m, the vector p' with coordinates

$$p'_j = \begin{cases} p_j & \text{if } j \neq m \\ \hat{p}_m & \text{if } j = m \end{cases}$$

(cf. equation (31) associated with transmission from node m to node i), also belongs to $\vec{P}^k$.

c) If $p(1),\ldots,p(N-1)$ belong to $\vec{P}^k$ then the vector p' with coordinates

$$p_j' = R_j(p(j)) \qquad , \qquad j = 1,\ldots,N-1$$

$$p_N' = 0$$

(cf. a computation (32) at each node followed by a transmission to every other node), belongs to $\vec{P}^{k+1}$.

It is easily seen that all the conditions stated above are satisfied in our case so the desired conclusion follows. $\qquad$ Q.E.D.

Proposition 3 shows that the ARM has satisfactory convergence when P^* has a unique element. One way to guarantee this is to consider the optimal solution f^* of the primal problem (1) and the set of arcs

$$\tilde{A} = \{(i,j)\varepsilon A \mid g_{ij} \text{ is differentiable at } f_{ij}^*\}.$$

Then, if the graph $(N,\tilde{A})$ is connected, P^* consists of a unique point in view of the complementary slackness condition (7). In order to improve the convergence properties when P^* has more than one point it is necessary to modify the ARM so that a computation at node i replaces p_i with $\overline{R}_i(p)$ (not just any element of $R_i(p)$). We call this the __maximal ARM__. If in place of $\overline{R}_i(p)$ we use $\underline{R}_i(p)$ the resulting method is called the __minimal ARM__.

__Proposition 4:__ Let Assumptions 1 and 2 hold. Assume that the starting buffer contents satisfy

$$p^o(i) \geq \bar{p} \ , \qquad \forall \ i = 1, \ldots, N-1 \qquad (36)$$

Then if $\{p^t(i)\}$ is generated by the maximal ARM we have

$$\lim_{t \to \infty} p^t(i) = \bar{p} \ , \qquad i = 1, \ldots, N-1. \qquad (37)$$

Proof: The proof is identical to the one of Proposition 3 except that the set $\bar{P}^k$ of (35) should be replaced by

$$\bar{P}^k = \{p | \quad \bar{p} \leq p \leq \bar{R}^k(\tilde{p})\}. \qquad \text{Q.E.D.}$$

There is a similar result for the minimal ARM whereby $\bar{p}$ is replaced by $\underline{p}$ and condition (36) is replaced by $p^o(i) \leq \underline{p}$ for all i. The following example demonstrates that the results of Proposition 3 and 4 cannot be improved.

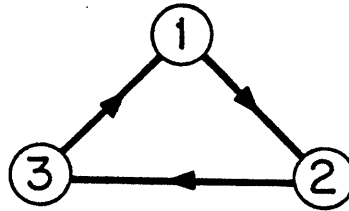Example: Consider the 3-node network shown in Figure 2.

Figure 2

The arc costs are

$$g_{12}(f_{12}) = (f_{12})^2, \ g_{23}(f_{23}) = |f_{23}| + (f_{23})^2, \ g_{31}(f_{31}) = |f_{31}| + (f_{31})^2$$

and the optimal primal solution is

$$f_{12}^* = f_{23}^* = f_{31}^* = 0.$$

The reduced dual optimal solution set is derived from condition (7) and is given by

$$P^* = \{p \mid p_3 = 0, \; p_1 = p_2, \; -1 \leq p_1 \leq 1, \; -1 \leq p_2 \leq 1\}.$$

The results of Proposition 3 and 4 are illustrated in Figure 3. To see that the ARM as well as the maximal and minimal ARM may not converge to a dual optimal solution, let the buffer contents of processors 1 and 2 be both equal to (-1,1) and let both processors update the respective price coordinates and then exchange the results of the computation. Then the buffer contents will be (1,-1), and by repeating this process one more time the buffer contents will become again (-1,1) thereby completing a cycle. Therefore in general we cannot expect convergence of the ARM to the optimal solution set if the latter contains more than one element. Similarly the maximal and minimal ARM need not converge to $\bar{p}$ and $\underline{p}$ respectively if the initial buffer contents do not belong to the appropriate regions [cf. (36)]. Note that this counterexample applies also to a synchronous Jacobi method.
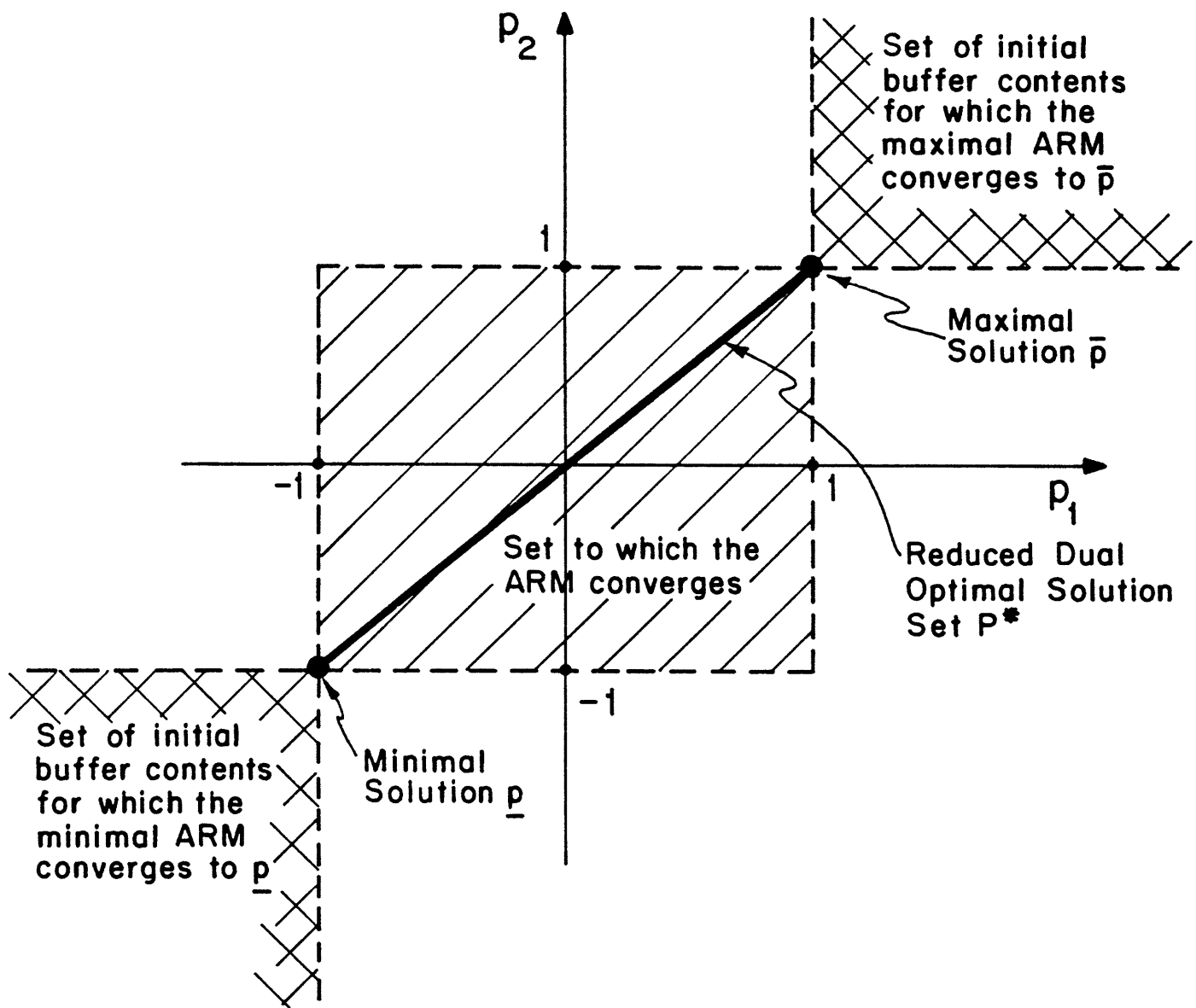
Figure 3: Structure of the optimal solution set, and convergence regions of the ARM, the maximal ARM, and the minimal ARM.

References

[1] Rockafellar, R. T., Network Flows and Monotropic Optimization, Wiley, N.Y., 1984.

[2] Zangwill, W. J., Nonlinear Programming, Prentice Hall, Englewood Cliffs, N.J., 1969.

[3] Sargent, R.W.H, and Sebastian, D. J., "On the Convergence of Sequential Minimization Algorithms", J. Opt. Theory and Applications, Vol. 12, 1973, pp. 567-575.

[4] Polak, E., Computational Methods in Optimization: A Unified Approach, Academic Press, N.Y., 1971.

[5] Powell, M.J.D., "On Search Directions for Minimization Algorithms", Math. Programming, Vol. 4, 1973, pp. 193-201.

[6] Cottle, R. W. and J. S. Pang, "On the Convergence of a Block Successive Overrelaxation Method for a Class of Linear Complementarity Problems", Math. Programming Study 17, 1982, pp. 126-138.

[7] Rockafellar, R. T., Convex Analysis, Princeton Univ. Press, Princeton, N.J., 1970.

[8] Chazan, D. and W. Miranker, "Chaotic Relaxation", Linear Algebra and Its Applications, Vol. 2, 1969, pp. 199-222.

[9] Miellou, J. C., "Iterations Chaotiques a Retards, Etude de la Convergence dans le Cas d'Espaces Partiellement Ordonnes", Comptes Rendus de l'Academie des Sciences, Paris, Serie A, 280, 1975, pp. 233-236.

[10] Baudet, G. M., "Asynchronous Iterative Methods for Multiprocessors", Journal of the ACM, Vol. 2, 1978, pp. 226-244.

[11] Bertsekas, D. P., "Distributed Dynamic Programming", IEEE Trans. on Automatic Control, Vol. AC-27, 1982, pp. 610-616.

[12] Bertsekas, D. P., "Distributed Asynchronous Computation of Fixed Points", Mathematical Programming, Vol. 27, 1983, pp. 107-120.

[13] Tsitsiklis, J. N., D. P. Bertsekas, and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms", IEEE Trans. on Auto. Control, (submitted).

[14] Tsitsiklis, J. N., "Problems in Decentralized Decision Making and Computation", Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., 1984.

[15] El Baž, D., "Etude d'algorithmes iteratifs de calcul parallele ap-
plication a la resolution distribuee du probleme du routage optimal
dans un resau maille a commutation de paquets", These de Docteur
Ingenieur, Toulouse, 1984.

[16] Authie, G., Bernussou, J., and ElBaz, D., "Distributed Asynchronous
Iterative Control Algorithms, Optimal Routing Application", IFAC
Symposium: Components and Instruments for Distributed Control Systems,
December 9-11, 1982, Paris.

[17] Bertsekas, D. P., J. N. Tsitsiklis, and M. Athans, "Convergence Theories
of Distributed Iterative Processes: A Survey", Laboratory for Information
and Decision Systems Report LIDS-P-1342, M.I.T., September 1984.

[18] Bertsekas, D. P., Constrained Optimization and Lagrange Multiplier
Methods, Academic Press, N. Y., 1982.

[19] Bertsekas, D.P., Hossein, P., and Tseng, P., "Relaxation Methods for Network
Flow Problems with Convex Arc Costs," LIDS Report, MIT, Dec. 1985.

[20] Tseng, P., "Relaxation Methods for Monotropic Programming Problems,"
Ph.D. Thesis, Operation Research Center, MIT, 1986.