11:35:04 OCA PAD INITIATION - PROJECT HEADER INFORMATION 07/18/88

Active

Project #: G-37-640 Cost share #: G-37-342 Rev #: 0
Center # : R6535-0A0 Center shr #: F6535-0A0 OCA file #:
Work type : RES
Contract#: DMS-8806946 Mod #: Document : GRANT
Prime #: Contract entity: GTRC

Subprojects ? : N
Main project #:

Project unit: MATH Unit code: 02.010.144
Project director(s):
SPINGARN J E MATH

Sponsor/division names: NATL SCIENCE FOUNDATION / GENERAL
Sponsor/division codes: 107 / 000

Award period: 880615 to 891130 (performance) 900228 (reports)

Sponsor amount	New this change	Total to date
Contract value	27,500.00	27,500.00
Funded	27,500.00	27,500.00
Cost sharing amount		5,510.00

Does subcontracting plan apply ?: N

Title: PROXIMAL AND ELLIPSOID ALGORITHMS IN CONVEX PROGRAMMING

PROJECT ADMINISTRATION DATA

OCA contact: Steven K. Watt 894-4820

Sponsor technical contact	Sponsor issuing office
PETER W. BATES	DAVID P. WESTBROOK
(202)357-3686	(202)357-9843
NATIONAL SCIENCE FOUNDATION	NATIONAL SCIENCE FOUNDATION
MPS/DMS	DGC/MPS
WASHINGTON, D.C. 20550	WASHINGTON, D.C. 20550

Security class (U,C,S,TS) : U
Defense priority rating : N/A
Equipment title vests with: Sponsor

ONR resident rep. is ACO (Y/N): N
NSF supplemental sheet
GIT X

JUL 1988
RECEIVED
SUPPORT SERVICES
OCA

Administrative comments -
INITIATION.

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

NOTICE OF PROJECT CLOSEOUT

Closeout Notice Date 02/21/91

Project No. G-37-640 Center No. R6535-0A0

Project Director SPINGARN J E School/Lab MATH

Sponsor NATL SCIENCE FOUNDATION/GENERAL

Contract/Grant No. DMS-8806946 Contract Entity GTRC

Prime Contract No.

Title PROXIMAL AND ELLIPSOID ALGORITHMS IN CONVEX PROGRAMMING

Effective Completion Date 891130 (Performance) 900228 (Reports)

Closeout Actions Required:	Y/N	Date Submitted
Final Invoice or Copy of Final Invoice	N	
Final Report of Inventions and/or Subcontracts	Y	901127
Government Property Inventory & Related Certificate	N	
Classified Material Certificate	N	
Release and Assignment	N	
Other	N	

Comments98A SATIFIES THE REQUIREMENT OF THE PATENT REPORT. QUESTIONNAIRE WILL NOT BE SENT. FINAL NEGATIVE

Subproject Under Main Project No.

Continues Project No.

Distribution Required:

Project Director	Y
Administrative Network Representative	Y
GTRI Accounting/Grants and Contracts	Y
Procurement/Supply Services	Y
Research Property Managment	Y
Research Security Services	N
Reports Coordinator (OCA)	Y
GTRC	Y
Project File	Y
Other	N
	N

PART IV - SUMMARY DATA ON PROJECT PERSONNEL

NSF Division ____________________

The data requested below will be used to develop a statistical profile on the personnel supported through NSF grants. The information on this part is solicited under the authority of the National Science Foundation Act of 1950, as amended. All information provided will be treated as confidential and will be safeguarded in accordance with the provisions of the Privacy Act of 1974. NSF requires that a single copy of this part be submitted with each Final Project Report (NSF Form 98A); however, submission of the requested information is not mandatory and is not a precondition of future awards. If you do not wish to submit this information, please check this box ☐

Please enter the numbers of individuals supported under this NSF grant.
Do not enter information for individuals working less than 40 hours in any calendar year.

*U.S. Citizens/ Permanent Visa	PI's/PD's		Post-doctorals		Graduate Students		Under-graduates		Precollege Teachers		Others	
	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.
American Indian or Alaskan Native												
Asian or Pacific Islander												
Black, Not of Hispanic Origin												
Hispanic												
White, Not of Hispanic Origin	1				1							
Total U.S. Citizens	1				1							
Non U.S. Citizens												
Total U.S. & Non-U.S. ..	1				1							
Number of individuals who have a handicap that limits a major life activity.												

*Use the category that best describes person's ethnic/racial status. (If more than one category applies, use the one category that most closely reflects the person's recognition in the community.)

AMERICAN INDIAN OR ALASKAN NATIVE: A person having origins in any of the original peoples of North America, and who maintains cultural identification through tribal affiliation or community recognition.

ASIAN OR PACIFIC ISLANDER: A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands and Samoa.

BLACK, NOT OF HISPANIC ORIGIN: A person having origins in any of the black racial groups of Africa.

HISPANIC: A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.

WHITE, NOT OF HISPANIC ORIGIN: A person having origins in any of the original peoples of Europe, North Africa or the Middle East.

THIS PART WILL BE PHYSICALLY SEPARATED FROM THE FINAL PROJECT REPORT AND USED AS A COMPUTER SOURCE DOCUMENT. DO NOT DUPLICATE IT ON THE REVERSE OF ANY OTHER PART OF THE FINAL REPORT.

b. Publication Citations

"Network equilibria and the method of successive approximation," preprint.

"Bounds on complexity for nonexpansive fixed points," in preparation.

Lectures and Presentations

"Bounds on complexity for nonexpansive fixed points," Annual meeting of the American Mathematical Society, January 1989, Louisville, Kentucky.

"Approximate fixed points and the ellipsoid algorithm," Annual meeting of the American Mathematical Society, January 1988, Atlanta, Georgia.

"Complexity of fixed points," Mathematics Department Colloquium, Dalhousie University, March 6, 1990.

e. Technical Description of Project and Results

I. Complexity of Fixed Point Computation

Let $q : \Re^d \to \Re^d$ be a nonexpansive mapping, that is

$$| q(x) - q(x) | \leq | x - q(x) | \qquad \forall\, x,y \in \Re^d.$$

Let $B \subset \Re^d$ be a closed ball of radius R. For $\varepsilon > 0$, we wish to investigate how the informational complexity of the problem:

$P(\varepsilon)$ either find $x \in \Re^d$ such that $|x-q(x)| < \varepsilon$, or show that q has no fixed point in B

depends upon ε (while holding d fixed).

We must first describe the class C of algorithms for solving $P(\varepsilon)$ to which our analysis pertains. These algorithms call an oracle which returns, for a given input $x \in \Re^d$, the value $q(x)$. During iteration i (i= 0,1,...), the only decision made by an algorithm $A \in C$ is the choice of the point $x_i \in \Re^d$ that is to be submitted to the oracle. The algorithm is assumed to be *deterministic* in the sense that this choice of x_i must depend only on the information that has already been gathered, namely the values $x_0,\dots,x_{i-1}, q(x_0),\dots,q(x_{i-1})$. This means that if two nonexpansive mappings q and q' have the property that $q(x_0)=q'(x_0),\dots,q(x_{i-1})=q'(x_{i-1})$, then A chooses the same value x_i for both mappings. Since there is no previous information during iteration 0, this implies that A starts always with the same x_0.

A is said to solve $P(\varepsilon)$ in n steps if for *every* nonexpansive mapping $q: \Re^d \to \Re^d$, A produces, after n steps, either an $x_n \in \Re^d$ such that $| x_n - q(x_n) | < \varepsilon$ or concludes that q has no fixed point in B. Define $\kappa(\varepsilon)$, the complexity of the problem $P(\varepsilon)$ to be the minimum n such that there exists a deterministic algorithm A which solves $P(\varepsilon)$ in n steps. We will establish a lower bound for $\kappa(\varepsilon)$.

Our principle tool for accomplishing this is the following

LEMMA: Let $B \subset \Re^d$ be a closed ball of radius $R > 0$. Let $x_0,\dots,x_n \in \Re^d$, $y_0,\dots,y_n \in B$. Define

$$B_n = \{z \in B : | z-y_i | \leq | z-x_i |, i = 0,\dots n\},$$

$$\alpha_n = \min\left\{\frac{|x_i-y_i|}{R} : \text{i=0,\dots,n}\right\},$$

$$v_n = \frac{\mathrm{vol}_d(B_n)}{\omega_d R^d} \qquad (\omega_d = \text{content of unit d-ball}).$$

<u>Suppose that</u>

(Q(n)) i. $|y_i - y_j| \le |x_i - x_j|$ (for all $0 \le i \le n$, $0 \le j \le n$) (in other words, $\{(x_0,y_0),\ldots,(x_n,y_n)\}$ is a *nonexpansive* relation on $\mathfrak{R}^d$)

ii. $v_n > 0$ (that is, the set B_n of possible fixed points for the relation $\{(x_0,y_0),\ldots,(x_n,y_n)\}$ has nonempty interior).

iii. $\alpha_n > 0$.

<u>Then for any</u> $x_{n+1} \in \mathfrak{R}^d$, <u>there exists</u> $y_{n+1} \in B$ <u>such that</u> (Q(n+1)) <u>holds with</u>

(1)
$$\alpha_{n+1} \ge \min\left\{ \beta_n, \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)\right\}, \qquad \left(\beta_n = \frac{v_n\omega_d}{(d+1)\omega_{d-1}}\right)$$

$$v_{n+1} \ge v_n\left(\frac{1}{e} - \frac{1}{d+1}\min\left\{\frac{1}{2}, \frac{\alpha_n}{4}\right\}\right).$$

PROOF OF LEMMA: The <u>width</u> of B_n is the minimum value (taken over all directions in $\mathfrak{R}^d$) of the distance between opposing pairs of parallel supporting hyperplanes to the set B_n. Let $u \in \mathfrak{R}^d$ be a direction for which this minimum is achieved, and let π_u denote orthogonal projection onto a hyperplane orthogonal to u. Since $B_n \subset B$,

$$\begin{aligned} v_n\,\omega_d R^d &= \mathrm{vol_d}(B_n) \\ &\le \mathrm{width}(B_n)\ \mathrm{vol_{d-1}}(\pi_u(B_n)) \\ &\le \mathrm{width}(B_n)\ \omega_{d-1}R^{d-1}. \end{aligned}$$

Define $\beta_n = \dfrac{v_n\omega_d}{(d+1)\omega_{d-1}}$ and let q = centroid(B_n). By [35,p.53], B_n contains the sphere centered at q, with radius $\dfrac{\mathrm{width}(B_n)}{d+1}$ and hence also the smaller sphere centered at q, with radius $\beta_n R$. In particular, $q + \beta_n R\,\dfrac{x_i - y_i}{|x_i - y_i|} \in B_n$, so

$$\left| q + \frac{\beta_n R\,(x_i - y_i)}{|x_i - y_i|} - y_i \right|^2 \le \left| q + \frac{\beta_n R\,(x_i - y_i)}{|x_i - y_i|} - x_i \right|^2 \qquad (0 \le i \le n)$$

which simplifies to

$$|q - x_i|^2 - |q - y_i|^2 \ge 2\beta_n R\,|x_i - y_i| \ge 2\alpha_n\beta_n R^2. \qquad (0 \le i \le n)$$

Since $y_i \in B$ and $q \in B$, we have $|q - y_i| \le 2R$ and so

$$| q\text{-}x_i | - | q\text{-}y_i | \geq \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)R \qquad (0 \leq i \leq n)$$

Define $\rho = \min\left\{ \beta_n, \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)\right\}$. Since

$$\begin{aligned} |z\text{-}q| \leq \rho R \;\Rightarrow \\ |z\text{-}y_i| &\leq |z\text{-}q| + |q\text{-}y_i| \\ &\leq \rho R + |q\text{-}y_i| \leq |q\text{-}x_i| \qquad (i=1,\ldots,n) \end{aligned}$$

the relation $\{(x_0,y_0),\ldots, (x_n,y_n),(q,z)\}$ is nonexpansive for all z such that $|z\text{-}q| \leq \rho R$.

Now fix $x_{n+1} \in \mathfrak{R}^d$. Choose any $q' \in \mathfrak{R}^d$ such that $|q'\text{-}q| = \rho R$ and such that q lies on the line segment $[x_{n+1}, q']$ (this choice is unique when $x_{n+1} \neq q$). Since $|q'\text{-}q| \leq \rho R$, the relation $\{(x_0,y_0),\ldots, (x_n,y_n),(q,q')\}$ is nonexpansive. Also, since $\rho \leq \beta_n$, $q' \in B_n \subset B$. By Kirszbraun's Theorem, it is possible to choose y_{n+1} such that $\{(x_0,y_0),\ldots, (x_n,y_n),(q,q'),(x_{n+1},y_{n+1})\}$ is also nonexpansive. Furthermore, an examination of the proof of Kirszbraun's theorem reveals that y_{n+1} may be chosen so that $y_{n+1} \in \mathrm{conv}\{y_0,\ldots,y_n,q'\} \subset B$. For such y_{n+1}, we claim that (Q(n+1)) holds.

That (Q(n+1))i holds is obvious, since $\{(x_0,y_0),\ldots,(x_{n+1},y_{n+1})\}$ is a subset of the nonexpansive relation $\{(x_0,y_0),\ldots, (x_n,y_n),(q,q'),(x_{n+1},y_{n+1})\}$. Now,

$$\begin{aligned} |x_{n+1}\text{-}y_{n+1}| &\geq |x_{n+1}\text{-}q'| - |y_{n+1}\text{-}q'| \\ &\geq (\rho R + |x_{n+1}\text{-}q|) - |x_{n+1}\text{-}q| \\ &= \rho R \end{aligned}$$

so

$$\begin{aligned} \alpha_{n+1} &= \min\left\{\alpha_n, \frac{|x_{n+1}\text{-}y_{n+1}|}{R}\right\} \\ &\geq \min\{\alpha_n, \rho\} \\ &= \min\left\{\alpha_n, \beta_n, \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)\right\} \\ &= \min\left\{\beta_n, \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)\right\}, \end{aligned}$$

where the last equality follows from

$$\left(\sqrt{4+2\alpha_n\beta_n} - 2\right) = \frac{2\alpha_n\beta_n}{\sqrt{4+2\alpha_n\beta_n}+2} \leq \alpha_n$$

(since $\beta_n \le \dfrac{\omega_d}{(d+1)\omega_{d-1}} \le 1$ for all d). This completes the proof of the first part of (1).

We must now estimate the volume of B_{n+1}. Now, $B_{n+1} = B_n \cap H$, where H is the closed halfspace

$$H = \{\, z : |z-y_{n+1}| \le |z-x_{n+1}| \,\}$$

Note that

$$\begin{aligned} \left|\frac{q+q'}{2}-y_{n+1}\right| &\le \left|\frac{q+q'}{2}-q'\right| + |q'-y_{n+1}| \\ &\le \left|\frac{q+q'}{2}-q\right| + |q-x_{n+1}| \\ &= \left|\frac{q+q'}{2}-x_{n+1}\right|. \end{aligned}$$

so $\frac{q+q'}{2} \in H.$ If $q = \text{centroid}(B_n)$ is contained in H then the proposition below implies

$$\text{vol}(B_{n+1}) \ge \frac{\text{vol}_d(B_n)}{e} = \frac{v_n\omega_d R^d}{e}.$$

If $q \notin H,$ let L be the halfspace containing H whose bounding hyperplane contains q and is parallel to the bounding hyperplane of H. The distance between these two hyperplanes is no more than $\text{distance}(q,H) \le \left|\frac{q+q'}{2}-q\right| = \frac{\rho R}{2}$. Now,

$$\begin{aligned} \text{vol}_d(B_n \cap (L-H)) &\le \frac{\rho R}{2}\,\omega_{d-1} R^{d-1} \qquad\qquad (\text{since } B_n \subset B\,) \\ &= \tfrac{1}{2}\min\left\{ \beta_n, \left(\sqrt{4+2\alpha_n\beta_n} - 2\right)\right\} \omega_{d-1} R^d \\ &\le \tfrac{1}{2}\min\left\{ \beta_n, \frac{2\alpha_n\beta_n}{\sqrt{4+2\alpha_n\beta_n} + 2}\right\} \omega_{d-1} R^d \\ &\le \min\left\{\tfrac{1}{2}, \tfrac{\alpha_n}{4}\right\} \beta_n\omega_{d-1}R^d \end{aligned}$$

Hence,

$$\begin{aligned} \text{vol}_d(B_{n+1}) &= \text{vol}_d(B_n \cap L) - \text{vol}_d(B_n \cap (L-H)) \\ &\ge \frac{\text{vol}_d(B_n)}{e} - \text{vol}_d(B_n \cap (L-H)) \\ &\ge \frac{v_n\omega_d R^d}{e} - \min\left\{\tfrac{1}{2}, \tfrac{\alpha_n}{4}\right\} \beta_n\omega_{d-1}R^d \end{aligned}$$

$$= \frac{v_n \omega_d R^d}{e} - \min\left\{\frac{1}{2}, \frac{\alpha_n}{4}\right\} \omega_{d-1} R^d \frac{v_n \omega_d}{(d+1)\omega_{d-1}}$$

$$= v_n \omega_d R^d \left(\frac{1}{e} - \frac{1}{d+1} \min\left\{\frac{1}{2}, \frac{\alpha_n}{4}\right\}\right)$$

which is the second half of (1). We have used the following:

PROPOSITION [37]: Let $C \subset \Re^d$ be a convex body, $q = \text{centroid}(C)$. Let H be a hyperplane passing through q, and let C^+ be one of the two parts into which C is divided by H. Then $\text{vol}(C^+) \geq \left(\frac{d}{d+1}\right)^d \text{vol}(C) \geq \frac{\text{vol}(C)}{e}$

The induction may be initialized as follows:

Proposition: <u>For every</u> $x_0 \in \Re^d$ $(d \geq 2)$ <u>there exists</u> $y_0 \in$ B <u>such that</u> (Q(0)) <u>holds with</u> $v_0 \geq \frac{1}{4}$ <u>and</u> $\alpha_0 \geq \frac{\pi\sqrt{2}}{4\sqrt{d}}$

Proof: Assume B is centered at 0. Let $\delta = \frac{\pi\sqrt{2}}{4\sqrt{d}}$. If $x_0 = 0$, set $y_0 = (\delta R, 0 \ldots, 0)$. Then $B_0 = \{z = (z_1, \ldots, z_d) \in B : z_1 \geq \frac{\delta R}{2}\}$ and

$$\begin{aligned} \text{vol}_d(B_0) &= \frac{1}{2}\text{vol}_d(B) - \text{vol}_d(\{z = (z_1, \ldots, z_d) \in B : 0 \leq z_1 \leq \frac{\delta R}{2}\}) \\ &\geq \frac{1}{2}\omega_d R^d - \frac{1}{2}\delta\omega_{d-1}R^d = \frac{1}{2}\omega_d R^d (1 - \frac{\delta\omega_{d-1}}{\omega_d}) \\ &\geq \frac{1}{4}\omega_d R^d. \end{aligned}$$

Thus $v_0 \geq \frac{1}{4}$ and $\alpha_0 = \frac{|x_0 - y_0|}{R} = d = \frac{\pi\sqrt{2}}{4\sqrt{d}}$.

If $x_0 \neq 0$, let $y_0 = -\delta R \frac{x_0}{|x_0|}$. The resulting values for v_0 and α_0 are larger than in the case where $x_0 = 0$.

We may use the inductive lemma to define sequences (α_k), (β_k), and (v_n) inductively. For each k, we then can make the following assertion:

Theorem: <u>No deterministic algorithm</u> A <u>can solve</u> $P(\alpha_k)$ <u>in k steps</u>. Furthermore, k grows with α_k on the order of $k \approx \sqrt{\log(1/\alpha_k)}$.

Proof: To start, suppose A chooses x_0. By the Proposition, there exists $y_0 \in B$ such that (Q) holds for n=0 with $\alpha=\alpha_0=1$ and $\beta=\beta_0=.25$. There is a nonexpansive mapping $q_0 : \Re^d \to \Re^d$ such that $q_0(x_0) = y_0$. Applied to this q_0, it follows by Q that A does not solve $P(\alpha_0)$ in zero steps.

Based on knowledge of x_0 and y_0, suppose that A chooses to query the oracle with x_1. By the Inductive Lemma, there exists $y_1 \in B$ such that (Q) holds for n=1 with $\alpha=\alpha_1$ and $\beta=\beta_1$. There is a nonexpansive mapping $q_1 : \Re^d \to \Re^d$ such that $q_1(x_0) = y_0$ and $q_1(x_1) = y_1$. Applied to this q_1, it is not possible by Q that A solve $P(\alpha_1)$ in α_1 steps. The proof continues inductively in this manner, producing a sequence $q_0, q_1, \ldots$ of nonexpansive mappings such that A does not solve $P(\alpha_k)$ in α_k steps. Since for each k, $q_k(x_k) = q_{k+1}(x_k) = q_{k+2}(x_k)\ldots$,the fact that A is a deterministic algorithm guarantees that A makes the same choices when applied to q_{k+i} during the first k steps. To complete the proof, it can then be shown that k increases with α_k on the order of $\sqrt{\log(1/\alpha_k)}$.

II. Network Equilibria and the Method of Successive Approximations

A. Introduction.

In a separate project, we have described a solution and modeling approach for network equilibria. The problems to which it has been applied have traditionally been solved by a common family of closely related methods, including conversion to an equivalent variational inequality, optimization problem, complementarity, or fixed point problem. Those that we have worked with include spatial economic equilibria, location problems, and traffic equilibria.

Our method exposed here is based on the method of successive approximations (with averaging) [17], which computes a fixed point of a nonexpansive mapping $q : R^d \to R^d$ as the limit of a sequence

$$x_{n+1} = \frac{x_n + q(x_n)}{2}. \tag{0.1}$$

An equivalent method (i.e., one which produces the same sequences) is the proximal point algorithm [26] which computes a zero of a maximal monotone multifunction $S : R^d \rightrightarrows R^d$ as the limit of a sequence

$$x_{n+1} = \mathrm{prox}_S(x_n), \tag{0.2}$$

where $\mathrm{prox}_S = (I+S)^{-1}$ is the proximal mapping [23] for S. The procedures we have described are instances of the method of partial inverses [29]-[31], which can be viewed as yet another expression of (0.1) and (0.2).

Although fixed point methods have been applied to most of these network problems, the basic iteration (0.1) has been overlooked. This is especially evident in the computation of economic equilibria, where extensive use has been made in recent years of fixed point methods. However, these applications have primarily exploited the properties of *continuous*, rather than *nonexpansive* mappings. The reason for this oversight is that lacking the equivalent partial inverse formulation, it has not been obvious what the nonexpansive mapping should be.

The major advantages of the methods discussed here are that they are globally convergent, require no differentiabity assumptions or use of derivatives, do not rely on an equivalent reformulation as an equivalent optimization problem and therefore require no integrability assumptions, and they produce decompositions that readily lend these problems to solution by parallel processor. The major disadvantages are that they often require a very large number of iterations, so that the models must be carefully constructed to avoid use of proximal mappings that are hard to compute.

B. Background and Notation.

A network G is a triple (N,A,e). The finite sets N and A are the nodes and (directed) arcs of G. If we number the nodes and arcs in some arbitrary order, we can form the $|N|\times|A|$ node-arc incidence matrix $E=(e_{ia})$ by setting $e_{ia}=-1$ if i is the initial node of arc a, $e_{ia}=+1$ if i is the terminal node of arc a, and $e_{ia}=0$ otherwise. This allows multiple arcs but not loops. We will write a~(i,j) to indicate that i is the initial node and j the terminal node of arc a, and write a=(i,j) to indicate that a is the *only* arc from i to j. A flow in G is a function $x : A \to R^d$. The flow in arc a will be written as x_a=x(a). If m>1, x is a multicommodity flow. The divergence y = div x of the flow x is the function $y : N \to R^d$ defined for each node i by $y_i = \Sigma_a e_{ia}x_a$ (or, in matrix notation, y = Ex). If div x = 0, x is a circulation in G. A potential in G is a function $u : N \to R^d$. A potential determines, in a natural way, a function $\Delta u : A \to R^d$ called the differential of u by $(\Delta u)_a = u_j - u_i$ (a~(i,j)) or, in matrix form, $\Delta u = E^t u$. The flow v is called a differential if $v = \Delta u$ for some potential u. The spaces of all circulations, differentials, and potentials in G will be denoted C, T, and U, respectively.

A multifunction $S : R^d \rightrightarrows R^d$ is monotone if $\langle x-x', y-y'\rangle \geq 0$ whenever $y \in S(x)$ and $y' \in S(x')$. S is maximal monotone if, in addition, graph(S) = $\{(x,y) : y \in S(x)\}$ is not properly contained in the graph of another monotone multifunction. Whenever S is maximal monotone, there is a function prox_S, the proximal mapping for S, which assigns to each $x \in R^d$ the unique x' such that $x-x' \in S(x')$ [22].

The methods discussed in this paper specialize to a network setting the method of partial inverses [31]. Given a maximal monotone multifunction $S : R^d \rightrightarrows R^d$ and a subspace X of R^d, the method is a procedure for solving the following problem:

(1.1) to find $x \in X$ and $y \in X^\perp$ such that $y \in S(x)$.

To solve (1.1), the method starts with arbitrary $x_0 \in X$ and $y_0 \in X^\perp$ and constructs sequences $x_n \in X$ and $y_n \in X^\perp$ in such a way that (writing $x_{n+1}=(x_n)^+$, $y_{n+1}=(y_n)^+$):

(1.2) $x^+ = \mathrm{proj}_X(x')$ and $y^+ = \mathrm{proj}_{X^\perp}(y')$, where x' and y' are chosen so that $x'+y'=x+y$ and $y' \in S(x')$. (Or, in other words, $x'=\mathrm{prox}_S(x+y)$ and $y'=x+y-x'$.)

It is possible to regard this procedure, depending on one's point of view, as a special case of the proximal point algorithm [26], or as a special case of Krasnoselski's averaged iterate method for finding a fixed point of a nonexpansive mapping [17]. The relationship between the partial inverse method and these others is clarified in [18]. The main result regarding convergence of (1.2) is the following

THEOREM [2]. Let $x_n \in X$ and $y_n \in X^\perp$ be sequences of iterates produced by the method of partial inverses (1.2). It will always happen either that

i) $x_n \to x$ and $y_n \to y$ for some solution x,y to the problem (1.1), and the distance from $x_n + y_n$ to the set {x+y: x,y solves (1.1)} is nonincreasing, or that

ii) $|x_n + y_n| \to \infty$ and (1.1) has no solutions.

The vectors x' and y' found in each iteration form sequences x_n' and y_n', and it can be shown that these converge to the same limiting values as the sequences x_n and y_n, respectively.

For methods based upon the proximal point algorithm to be successful, it is important that the relevant proximal mappings be simple to compute. This is a major limitation of such methods, and can sometimes make them impractical. Typically, proximal methods require a large number of iterations, so it is essential that a single iteration involve only minimal computational effort. This observation has been bourne out by recent numerical experiments with partial inverse algorithms

[2], [14], [19]. The chief advantages are global convergence assuming only convexity and monotonicity, simplicity and flexibility in modelling, and stability.

In this article, R^d denotes Euclidean space equipped with the standard inner product. If $K \subset R^d$ is closed and convex, then $\mathrm{proj}_K(x)$ denotes the nearest point to x in K. (Although, if K is a subspace, we will prefer the notation x_K to denote this projection.) The normal cone to the convex set K at x is the set $N_K(x) := \{y: \langle y,x-z\rangle \ge 0, \forall z \in K\}$ if $x \in K$, and the empty set if $x \notin K$. The multifunction $N_K : R^d \rightrightarrows R^d$ is maximal monotone and its proximal mapping is proj_K. For $x,y \in R^d$, $\max\{x,y\}$ shall denote the vector whose i^{th} component is $\max\{x_i,y_i\}$. The nonnegative orthant in R^d is denoted as $R^d_+ = \{x \in R^d : x \ge 0\}$. Its normal cone mapping is

$$N_{R^d_+}(x) = \begin{cases} \{y: y \le 0 \text{ and } \langle x,y\rangle = 0\} & \text{if } x \ge 0 \\ \varnothing & \text{otherwise} \end{cases}$$

whose proximal mapping is $\mathrm{prox}(x) = \max\{0,x\}$. For any set K, the characteristic function of K is the function $\Psi_K(x)$ whose value is 0 if $u \in K$ and $+\infty$ if $u \notin K$. If $S : R^d \rightarrow R^d$ is linear, then S is monotone if and only if $\langle x,Sx\rangle \ge 0$ for all $x \in R^d$; the proximal mapping for S is $(I+S)^{-1}$. If f is a closed proper convex function, the subdifferential of f is maximal monotone and its proximal mapping is $\mathrm{prox}_f(x) = \mathrm{argmin}_y \frac{1}{2} |y-x|^2 + f(y)$ [23], [25,31.5.2]. It is usually a straightforward matter to compute the proximal mapping for a one-dimensional $S : R \rightrightarrows R$; the most common cases which arise occur when S is piecewise linear or a step function.

C. Spatial Price Equilibrium.

In the spatial price equilibrium problem we are given, at each of n markets, a relationship between supply and price (which in the single-commodity case would be called a "supply curve") and between demand and price ("demand curve") prevalent at that market. We also know the costs incurred for shipping goods through links that join one market to another. The problem is to determine the competitive equilibrium -- the amounts supplied or demanded at each market and the amounts shipped through the transportation links.

It is a problem with a long history, a two-market problem having been formulated as early as 1838 by Cournot [4]. In 1951, Enke [6] argued that the model could be simulated by an appropriate electrical network. When the current stabilized, prices and trade flows would correspond to voltage drops and currents and could therefore be "computed" by reading values from meters. Due to the disappearance of analogue computers, that approach nowadays strikes us as technologically quaint. But the underlying idea has had an important influence. Samuelson [27], observed that Enke's problem could be converted into a maximum problem. He reasoned that since equilibrium in a passive electric network can be described in terms of an extremum principle (the minimization of total power loss), the same must hold for the economic problem. Building on this analogy, he defined a "net social pay-off" function, and showed how it could be maximized to solve a linear spatial equilibrium problem by the then new simplex method.

Samuelson's idea to formulate the problem as an extremal problem whose optimality conditions coincide with the equilibrium conditions has served as the basis for a great deal of research in mathematical economics. It is the approach expounded in the book by Takayama and Judge [32] and by many others. This optimization approach has some clear strengths: it aids greatly in formulation and exposition, makes available a wide choice of solution algorithms, and provides tools to establish existence and uniqueness results. Despite this, it has long been recognized that it has a serious drawback; in order to formulate as an extremal problem, the operators that describe supply and demand must be integrable. This imposes severe restrictions on the functions one can use to represent supply and demand, and it rules out some of those which are most popular in economics [3]. For this reason, there is strong motivation to search for different lines of attack. Naturally, the issue only arises in multicommodity models -- since integrability is not a serious restriction for real-valued functions of one variable, and only when systems of demand or supply functions are interdependent -- since otherwise individual commodities can be handled separately.

The integrability issue has been confronted by many researchers, and numerous lines of attack have been suggested. These include reformulation as a complementarity problem or variational inequality for which solution techniques are available. Surveys of such techniques can be found in [3], [11], [24]. Jacobi or diagonalization methods have also been studied [1], [5], [8], [9], [33]. However, global convergence results are scarce (excepting the linear case), and those that are available are complex and the required assumptions are difficult to verify [11], [24]. Also, methods based on Newton or secant method approaches require differentiability assumptions that seem unnatural in an economic setting. Practical success has been achieved with the Jacobi iteration, but here again, meaningful global convergence results are not to be found.

The approach we will present here does not require formulating an equivalent optimization problem, and converges globally under only monotonicity assumptions. It requires no differentiability or linearity, and hence provides much modelling freedom. This flexibility will be illustrated on several equilibrium models in this and the following section.

To place the simplest d-commodity spatial price equilibrium problem in a suitable framework, we consider a connected network with node set $N = \{0, ..., n\}$, arc set A, a circulation $x: A \to R^d$, and potential function $p: N \to R^d$ such that $p_0=0$. The nodes 1, ..., n represent the markets through which all traded units pass; node 0 is a dummy node. The vector $p_i \in R^d$ ($i \neq 0$) gives the unit prices of all d commodities at market i. The arcs are divided into two classes: $A = A_\tau \cup A_\sigma$ ($A_\tau \cap A_\sigma = \emptyset$). The set of all supplementary arcs is denoted by A_σ. Node 0 is connected to each market j by just one supplementary arc $a=(0,j)$, and the flow x_a in that arc represents the excess supply (supply minus demand) at market j; in other words, x_a is whatever it needs to be in order that x be conservative at node j. Arcs in A_τ are transport arcs which join markets. The flow x_a in a transport arc $a \sim (i,j)$ represents units shipped from market i to market j through arc a. Shipping costs in transport arc $a \sim (i,j)$ are given by a constant vector $c_a \in R^d_+$ whose k^{th} component is the unit cost of shipping commodity k from market i to market j through arc a. The economic behavior of market j is described by a maximal monotone multifunction $P_a(x_a)$ ($a=(0,j)$) which gives the set of possible prices compatible with the excess supply x_a.

It is possible to provide economic justification for the assumption that P_a is monotone. The excess supply $x_a = \sigma - \delta$ is the difference between supply σ and demand δ at market j ($a=(0,j)$). Let us write $\sigma \in S_a(p_j)$ and $\delta \in D_a(p_j)$ to indicate that σ and δ are compatible with the prices p_j. Then $p_j \in P_a(x_a)$ means there exist $\delta \in D_a(p_j)$ and $\sigma \in S_a(p_j)$ such that $x_a = \sigma - \delta$, or equivalently, $p_j \in (S_a - D_a)^{-1}(x_a)$. Thus $P_a = (S_a - D_a)^{-1}$, and P_a will be monotone provided S_a is monotone and D_a is antitone ($-D_a$ is monotone).

To conclude that supply S_a is monotone, suppose a firm acts to maximize its profits $\langle p,\sigma \rangle - \phi(\sigma)$, where $\phi(\sigma)$ is the cost of producing σ. This maximum is achieved at σ_0 provided $\phi(\sigma) \geq \phi(\sigma_0) + \langle \sigma - \sigma_0, p \rangle \; \forall \sigma$, which says that $p \in \partial(\mathrm{cl}\, \phi)(\sigma_0)$, where $\mathrm{cl}\, \phi$ is the closure of ϕ in the sense of [25]. Equivalently, $\sigma_0 \in \partial\phi^*(p)$ [25,31.5.2] (where ϕ^* is the convex conjugate of ϕ). Thus $S_a = \partial\phi^*$. Assuming $\phi^* \quad +\infty$ (i.e., ϕ dominates some linear function), $\partial\phi^*$ is maximal monotone [25,31.5.2]. Note that it is not necessary to assume that ϕ is convex.

To justify the antitonicity of D_a, we follow Ahn [1] in distinguishing two classes of consumers: industrial and household. An industrial consumer seeks to maximize $\Psi(\delta) - \langle p,\delta\rangle$, where $\Psi(\delta)$ is the revenue generated by using the input δ. As in the supply case, this behavior implies that D_a is the subdifferential of a closed concave function, hence antitone. For household consumers, we consider the so called "compensated demand" problem [34, p. 80] whereby expendidure $\langle p,\delta\rangle$ is minimized while maintaining a desired utility, that is subjected to a constraint of the form $U(\delta) \geq c$, for some prescribed c. Since the constraint set is convex [34, p.79], δ achieves the minimum of $\langle p,\delta\rangle$ if, and only if, $p \in -N(\delta)$ where N is the normal cone mapping for the constraint set. Thus $D_a = (-N)^{-1}$ is maximal antitone. In using the compensated demand problem, we are neglecting the "fixed income effect" whereby a consumer maximizes utility subject to his limited income. Quoting Ahn [1, p.36]: "If the fixed income effect is not negligible and no income compensation is allowed, the demand function might not have the antitonicity property. If the income effect is not strong or the industrial demand dominates the household demand, however, the corresponding aggregate demand function is likely to be antitone. For example, the actual demand data used at FEA in the first application of the PIES model satisfy the strict antitonicity condition"

To phrase the problem in the form (1.1), we introduce the space C of circulations, and the space T of differentials. Since the network is assumed to be connected, for every $t \in T$ there is a ***unique*** potential $p \in U$ such that $t=\Delta p$ and $p_0=0$. The standard spatial price equilibrium problem is

(2.1) to find $x \in C$ and $t=\Delta p \in T$ $(p_0=0)$ such that

i. for all transport arcs $a \in A_\tau$ $(a\sim(i,j))$,
$\langle x_a, c_a-t_a\rangle = 0$, $x_a \geq 0$, and $c_a-t_a \geq 0$.

ii. for all supplementary arcs $a \in A_\sigma$ $(a=(0,j))$,
$p_j = t_a \in P_a(x_a)$.

The first condition says that no trade takes place at a loss and there is no further incentive to trade. (The condition $p_0=0$ creates no real restriction since the value of p can be specified arbitrarily at any one node. It is only included to force $p_j=t_a$ for $a=(0,j)$). The second says that the excess supply at market j is compatible with the prices at market j.

For all $a \in A_\tau$ define $P_a(x_a) = c_a + N_{R^d_+}(x_a)$. Then P_a is maximal monotone and (2.1i) is equivalent to $t_a \in P_a(x_a)$. Hence, if we let $P = \Pi_{a\in A} P_a$, then (2.1) is equivalent to the simpler

(2.2) $x \in C$, $t=\Delta p \in T$ $(p_0=0)$, and $t \in P(x)$.

Since $C = T^\perp$, this fits into the partial inverse framework (1.1) and lends itself to solution by algorithm (1.2). Hence the partial-inverse procedure for solving (2.2) is:

(2.3) Given $x \in C$ and $t=\Delta p \in T$ $(p_0=0)$, compute for every $a \in A$, $x_a' = \text{prox}_a(x_a+t_a)$ and $t_a' = x_a+t_a-x_a'$. The next iterates are $x^+ = (x')_C$ and $t^+ = \Delta p^+ = (t')_T$.

For $a \in A_\tau$, the proximal mapping for P_a is the function $\text{prox}_a(u) = \max\{0,u-c_a\}$. Let E_0 denote the node-arc incidence matrix with row 0 (the row corresponding to node 0) deleted. Since any flow conservative at nodes 1,...,n is automatically conservative at node 0, we know that $C = \{x : E_0x=0\}$. That node 0 is connected by a supplementary arc to every other node ensures that $E_0E_0^t$ is invertible. Hence we can compute the required projections as follows: set $p_0^+=0$ and compute $(p_1^+,\ldots,p_n^+)^t = (E_0E_0^t)^{-1}E_0t'$. Then $t^+ = \Delta p^+$, and $x^+ = x'-(x')_T = x'-(x+t-t')_T = x'-t+(t')_T = x-t'+t^+$. The sequences $t^{k+1}=(t^k)^+$, $x^{k+1}=(x^k)^+$, and $p^{k+1}=(p^k)^+$ generated in this manner converge to a solution to (2.2) (if one exists). The ability to execute this algorithm depends on having available an efficient procedure for calculating the functions prox_a $(a\in A_\sigma)$.

There are two approaches one can use to devise a stopping criterion for the partial inverse algorithm: an approximate fixed point approach, and one based on duality. The first halts execution after finding an approximate fixed point of the underlying nonexpansive mapping [18]. It has the advantage that it can always be used, since it relies on information that must be generated by the algorithm. Applied to the spatial economic equilibrium problem, it works as follows. For a chosen $\varepsilon>0$, one halts when $|x-x^+| < \varepsilon$ and $|t-t^+| < \varepsilon$ (as must happen if a solution exists). Noting that $x-x^+ = x-x'_C = x-(x+t-t')_C = t'_C$ and $t-t^+ = t-t'_T = t-(t+x-x')_T = x'_T$, we see that when the termination criterion is satisfied, x' and t' provide an approximate solution in the sense that

$$x' \in x'_T + C,\ \ t' \in t'_C + T,\ \text{ and }\ t'\in P(x'),\ \text{ with }\ |x'_T| < \varepsilon \text{ and } |t'_C| < \varepsilon.$$

This says that x' is "almost" a circulation and t' is "almost" a differential. More precisely, we have $|\mathrm{div}\ x'| = |\mathrm{div}\ x'_T| \le \varepsilon v$, where v is the largest valence of any node in the network. Likewise, if we assign the prices $p_0 = 0$ and $p_j = (t')_a$ $(\forall a\in A_\sigma,\ a=(0,j))$, we obtain a potential whose differential is "almost" t'. To interpret this statement, note that the sum $\Sigma_{a\in\Gamma} \pm(t')_a$ taken around any circuit Γ of the network (using + for arcs in the direction of the circuit, and - for those opposed) equals the sum $\Sigma_{a\in\Gamma} \pm(t'_C)_a$ around the same circuit. Any arc $b\in A_\tau$ $(b\sim(i,j))$ forms, together with two supplementary arcs, a circuit of three arcs, and for such a circuit Γ, we thus have $|p_j-p_i-(t')_b| = |\Sigma_{a\in\Gamma} \pm(t')_a| = |\Sigma_{a\in\Gamma} \pm(t'_C)_a| \le 3\varepsilon$, showing that the differential of p is close to t'.

A second approach to devising stopping criteria will be discussed later in our discussion of location problems. It depends on the existence of an underlying optimization problem. Using duality, it may be possible to sandwich the optimal value into an arbitrarily small interval. While this approach is more satisfying, it has the disadvantages that it cannot always be used and that much additional work inessential to the progress of the algorithm may have to be done to compute the needed function values.

Example 1.

To fix ideas and notation used in problem (2.1) and algorithm (2.3), let us first consider the linear one-commodity case which is the one most commonly treated in the literature. For every supplementary arc a=(0,j), a relation $x_a=d_jp_j-b_j$ $(d_j>0,\ b_j>0)$ is assumed to hold between the excess supply x_a and price p_j when $p_j>0$, while for $p_j=0$, demand can be considered to be unlimited. This leads to the monotone function $P_a(x_a) = \max\{0,(x_a+b_j)/d_j\}$ whose proximal mapping is $\mathrm{prox}_a(x_a) = \max\{x_a, \frac{d_jx_a-b_j}{1+d_j}\}$. For transport arcs $a\in A_\tau$, the monotone relation is

$$P_a(x_a) = \begin{cases} c_a & \text{if } x_a > 0 \\ \{r : r \le c_a\} & \text{if } x_a \le 0 \end{cases},$$

whose proximal mapping is $\mathrm{prox}_a(x_a) = \max\{0, x_a - c_a\}$.

We have assumed constant transportation costs only because that is the most common case considered in the literature. But any monotone relation will do, so long as its proximal mapping can be easily computed. In particular, congestion effects (whereby increasing traffic causes increased unit transportation costs) can be incorporated in a straightforward manner.

Example 2.

Let us now see what happens if we try to modify example 1 to allow for $d>1$ commodities. Assume that at each market $j=1,\ldots,n$, the excess supply s_j is related to price p_j by $s_j \in D_j p_j - b_j + N_{R_+^d}(p_j)$, where D_j is a $d \times d$ positive semidefinite (but not necessarily symmetric) matrix. The inclusion here of the normal cone mapping $N_{R_+^d}$ has the effect of saying that demand is unlimited for any commodity whose price is zero, and that no commodity has a negative price. In the case $d=1$, this model is exactly the same as the previous one, and could theoretically be solved in exactly the same way. But unfortunately, the multifunction $p_j \to D_j p_j - b_j + N_{R_+^d}(p_j)$ has a proximal mapping that cannot easily be computed. It is therefore advisable to modify the network in a way that makes it unnecessary to compute this mapping.

For this reason, we join node 0 to each market j by two arcs $a_{j1} \sim (0,j)$ and $a_{j2} \sim (0,j)$. To these arcs we assign the multifunctions

$$\begin{aligned} p_j &\in P_{a_{j1}}(x_{a_{j1}}) \text{ if and only if } x_{a_{j1}} = D_j p_j - b_j, \\ p_j &\in P_{a_{j2}}(x_{a_{j2}}) \text{ if and only if } x_{a_{j2}} \in N_{R_+^d}(p_j). \end{aligned} \tag{2.4}$$

For a circulation $x: A \to R^d$ and potential $p: N \to R^d$ with $p_0 = 0$ and $t = \Delta p$, the excess supply at market j is now represented by $s_j = x_{a_{j1}} + x_{a_{j2}}$. Since $t_{a_{j1}} = t_{a_{j2}} = p_j$, the relations (2.4) are equivalent to

$$t_{a_{j1}} \in P_{a_{j1}}(x_{a_{j1}}) \text{ and } t_{a_{j2}} \in P_{a_{j2}}(x_{a_{j2}})$$

and these conditions hold if and only if the price p_j is compatible with the excess supply $x_{a_{j1}} + x_{a_{j2}}$. An equilibrium can now be solved for much as in the first example. But now, the proximal mappings for $P_{a_{j1}}$ and $P_{a_{j2}}$ can be computed in a straightforward manner: $\mathrm{prox}_{a_{j1}}(u) = u - (I + D_j)^{-1}(u + b_j)$, and $\mathrm{prox}_{a_{j2}}(u) = u - \max\{0, u\}$.

Example 3.

Next, we discuss a variation of the spatial price equilibrium model in which a global competitive market exists for transportation services. This type of model has been studied in [9], [12], [15], [20]. As before, we have a node set $N = \{0,\ldots,n\}$, with nodes $1,\ldots,n$ representing the n markets. The arc set $A = A_\tau \cup A_\sigma$ is partitioned into two subsets: A_τ contains an arbitrary number of transport arcs $a \sim (i,j)$, and A_σ contains the supplementary arcs $a = (0,j)$, precisely one for each market j. To each $a \in A_\tau$ ($a \sim (i,j)$), there is associated a scalar $w_a > 0$ which measures the effort required to traverse the link a. We will refer to w_a as the "distance" between i and j, but it could equally well be time or some other unit. The flow $x_a \in R^d$ in arc $a \in A_\tau$ $(a \sim (i,j))$ represents units shipped from market i to market j through a. The vector $\bar{x} := \sum_{a \in A_\tau} w_a x_a$ represents the total amount of shipping used. The cost of shipping is given by a vector $\bar{p} \in R^d$ whose k^{th} component gives the unit cost per unit of distance to ship commodity k. The supply-price relationship for shipping is described by a multifunction $\bar{P} : R^d \rightrightarrows R^d$; that is, the set of price vectors $\bar{p}$ at which the shipping vector $\bar{x}$ is available is $\bar{P}(\bar{x})$. $\bar{P}$ is assumed to be maximal monotone. The flow x_a in a supplementary arc $a=(0,j)$ represents the excess supply at market j. The behavior of market j is described by a maximal monotone multifunction $P_a(x_a)$ which gives the set of possible price vectors compatible with the excess supply x_a.

We can now state the <u>spatial price equilibrium problem with a global market for transportation</u> as:

(2.5) to find $x \in C$, $p \in U$ $(p_0=0)$, and $\bar{p} \in R^d$, such that

i. for all transport arcs $a \in A_\tau$ $(a \sim (i,j))$,

$\langle x_a, w_a\bar{p}-p_j+p_i \rangle = 0$, $x_a \geq 0$, and $w_a\bar{p}-p_j+p_i \geq 0$

ii. for all supplementary arcs $a \in A_\sigma$ $(a=(0,j))$,

$p_j \in P_a(x_a)$

iii. $\bar{p} \in \bar{P}(\bar{x})$, where $\bar{x} := \sum_{a \in A_\tau} w_a x_a$.

The first two conditions are the same as (2.1) except that in (2.5i) the transportation cost in arc a is $w_a\bar{p}$ rather than c_a. The third condition states that the price of shipping is compatible with the supply. To rephrase the problem in the form of (1.1), define the spaces

(2.6)
$$X = \{(x,\bar{x}) : x \in C \text{ and } \bar{x} := \textstyle\sum_{a \in A_\tau} w_a x_a\}$$
$$Y = \{(t,\bar{t}) : \exists\, p \in U,\ \bar{p} \in R^d, \text{ such that (i) } t_a = p_j - p_i - w_a\bar{p},\ \forall a \in A_\tau\ (a \sim (i,j)),$$
$$\text{(ii) } t_a = p_j,\ \forall a \in A_\sigma\ (a=(0,j)), \text{ and (iii) } \bar{t} = \bar{p}\ \}.$$

Letting $w_a=0$ for all $a \in A_\sigma$, define the matrix $\bar{E} = \begin{pmatrix} E & 0 \\ -w^t & 1 \end{pmatrix}$. Then $X = \left\{\begin{pmatrix} x \\ \bar{x} \end{pmatrix} : \bar{E}\begin{pmatrix} x \\ \bar{x} \end{pmatrix} = 0\right\}$, and $Y = \left\{ \bar{E}^t \begin{pmatrix} p \\ \bar{p} \end{pmatrix} : p \in U, \bar{p} \in R^d \right\}$, so that $X = Y^\perp$. For every $a \in A_\tau$, define $P_a = N_{R^d_+}$. Then P_a is maximal monotone, its proximal mapping is $prox_a(u) = \max\{0,u\}$, and (2.5i) is equivalent to $p_j-p_i-w_a\bar{p} \in P_a(x_a)$. If we then define $P = \left(\prod_{a \in A} P_a\right) \times \bar{P}$, then (2.5) is equivalent to

(2.7) to find $(x,\bar{x}) \in X$ and $(t,\bar{t}) \in X^\perp$ with $(t,\bar{t}) \in P(x,\bar{x})$.

To recognize this equivalence, note that if $(x,\bar{x})$ and $(t,\bar{t})$ solve (2.7), then there is a unique $p \in U$

(with $p_0=0$) and $\bar{p}\in R^d$ such that $\begin{pmatrix} t \\ \bar{t} \end{pmatrix} = \bar{E}^t \begin{pmatrix} p \\ \bar{p} \end{pmatrix}$ and this gives a solution to (2.5). Conversely, if $x, p, \bar{p}$ solve (2.5), then by defining $\bar{x} := \sum_{a\in A_\tau} w_a x_a$ and $\begin{pmatrix} t \\ \bar{t} \end{pmatrix} = \bar{E}^t \begin{pmatrix} p \\ \bar{p} \end{pmatrix}$ we obtain a solution to (2.7). The algorithm (1.2) thus gives a globally convergent procedure solve (2.5):

(2.8) given $x\in C$, $p\in U$, and $\bar{p}\in R^d$,

i. for every $a \in A_\tau$ $(a\sim(i,j))$, compute $x_a' = \max\{0,\ x_a+p_j-p_i-w_a\bar{p}\}$,
$t_a' = x_a+p_j-p_i-w_a\bar{p}-x_a'$.

ii. for every $a \in A_\sigma$ $(a\sim(0,j))$, compute $x_a' = \text{prox}_a\{0,\ x_a+p_j\}$,
$t_a' = x_a+p_j-x_a'$.

iii. compute $\bar{x} := \sum_{a\in A_\tau} w_a x_a$, $\bar{x}' = \text{prox}_p(\bar{p} + \bar{x})$, $\bar{t}' = \bar{x} + \bar{p} - \bar{x}'$.

Let E_0 denote the node-arc incidence matrix for the network with row 0 deleted, and define $\bar{E}_0 = \begin{pmatrix} E_0 & 0 \\ -w^t & 1 \end{pmatrix}$. Then X is the null space of $\bar{E}_0$. That node 0 is connected by a supplementary arc to every other node ensures that $\bar{E}_0\bar{E}_0^t$ is invertible. Hence we can compute the next iterates x^+, p^+, $\bar{p}^+$ as follows: set $p_0^+=0$ and compute $(p_1^+,\ldots,p_n^+,\bar{p}^+)^t = (\bar{E}_0\bar{E}_0{}^t)^{-1}\bar{E}_0(t',\bar{t}')^t$. Then $t^+ = E^t p^+ - \bar{p}^+ w$ and $x^+ = x-t'+t^+$.

D. Economic Equilibrium with Implicit Supply Curve.

A single-location equilibrium model known as the PIES model has been an impetus for much research into solution methods for equilibrium problems [1], [13]. It differs from those of section II in one important aspect: the supply-price relationship is only *implicitly* known. The consumers' behavior is described by a (nonintegrable) multifunction $D : R^d \rightrightarrows R^d$ such that $-D$ is maximal monotone; for each $u\in R^d$, $D(u)$ is the set of prices at which consumers are willing to purchase the commodity bundle u. However, no corresponding relationship is given which describes the supplier's behavior. Rather, in response to a perceived demand vector $u\in R^d$, the supplier is assumed to choose a production plan $x\in R^r$ by solving an optimization problem:

P(u) to minimize $f_0(x)$ subject to $f_1(x)+u_1\le 0,\ldots, f_d(x)+u_d\le 0$, and $x\in K$.

In the original PIES model, P(u) is a linear programming problem, so we are being less restrictive here, assuming only that the functions $f_1,\ldots,f_d$ be convex real-valued, and $K\subset R^r$ closed convex. The function f_0 can be interpreted as cost, and $-f_1, \ldots, -f_m$ as quantities of goods produced. A vector $y\in R^d$ of <u>shadow prices</u> or a <u>Kuhn-Tucker vector</u> [25, p.274] for the problem P(u) is one having the properties: $y \ge 0$ and $\inf_{v\in K} \{f_0(v)+\sum y_i(f_i(v)+u_i)\} = \inf P(u)$. The set of all such y is implicitly determined by u (although it may be empty). If a solution to P(u) and a shadow price vector y exist such that the price vector y is also acceptable to the consumers, then an <u>equilibrium</u> is said to exist. Thus we can summarize the problem as

(PIES) to find $u \in R^d$, $x \in R^r$, and $y \in R^d$ such that $y \in D(u)$, x solves $P(u)$, and y is a Kuhn-Tucker vector for $P(u)$.

It will be helpful to introduce the closed proper convex function

$$F(x,u) = \begin{cases} f_0(x) & \text{if } f_1(x)+u_1 \le 0, \ldots, f_d(x)+u_d \le 0, \text{ and } x \in K \\ +\infty & \text{otherwise} \end{cases}$$

We will need the following

LEMMA . For $u \in R^d$, $x \in R^r$, and $y \in R^d$, we have $(0,y) \in \partial F(x,u)$ if and only if x solves $P(u)$ and y is a Kuhn-Tucker vector for $P(u)$.

Proof: Suppose $(0,y) \in \partial F(x,u)$. Then $F(x,u)$ is finite, implying $F(x,u)=f_0(x)$, or equivalently, $x \in K$ and $f_i(x)+u_i \le 0$ for all i (x is feasible for $P(u)$). By the definition of ∂F [25 ,p.214],

(+.1) if $x' \in K$ and $f_i(x')+u_i' \le 0$ $(i=1,\ldots,d)$, then
$f_0(x') \ge f_0(x) + \Sigma\, y_i(u_i'-u_i)$.

Setting $u_i'=u_i$ in (+.1), we see that x solves $P(u)$. Setting $x'=x$ in (+.1) and using the fact that $f_i(x)+u_i \le 0$ $(\forall i)$, we obtain $y \ge 0$ and $y_i(f_i(x)+u_i) = 0$ $(\forall i)$. Setting $u_i' = -f_i(x')$ in (+.1), we see that

$$f_0(x')+\Sigma\, y_i(f_i(x')+u_i) \ge f_0(x) \qquad (\forall x' \in K)$$

and that equality holds for $x'=x$. Hence y is a Kuhn-Tucker vector for $P(u)$.

To prove the converse, suppose x solves $P(u)$ and y is a Kuhn-Tucker vector for $P(u)$. Fix $x' \in K$ and u' such that $f_i(x')+u_i' \le 0$ $(i=1,\ldots,d)$. Since $x' \in K$ and y is a Kuhn-Tucker vector, we know that $f(x) \le f_0(x')+\Sigma\, y_i(f_i(x')+u_i)$. Using this, $y \ge 0$, and $f_i(x')+u_i' \le 0$, we get

$$f_0(x)+\Sigma\, y_i(u_i'-u_i) \;\le\; f_0(x')+\Sigma\, y_i(f_i(x')+u_i') \;\le\; f_0(x')$$

so that (+.1) holds, i.e. $(0,y) \in \partial F(x,u)$.

To place the problem in a framework where it can be solved by algorithm (1.2), we introduce the spaces $X = \{(x,u,u) : x \in R^r, u \in R^d\}$ and $Y = \{(0,y,-y) : y \in R^d\}$. It is clear that $X = Y^{\perp}$ and (PIES) is equivalent to the problem

(3.1) to find $(x,u,u) \in X$ and $(0,y,-y) \in X^{\perp}$ such that $(0,y,-y) \in (\partial F \times -D)(x,u,u)$.

To interpret the partial inverse algorithm (1.2) with these choices, we need to know how to compute projections $(x,y,z)_X$ and how to proximate ∂F. (It is merely *assumed* that the operator -D has a known proximal mapping prox_{-D} whose computation is supplied as a subroutine.) The projection of (x,y,z) onto X is $(x,\frac{y+z}{2},\frac{y+z}{2})$. As for the proximal mapping prox_F of ∂F, we have

$$\mathrm{prox}_F(\bar{x},\bar{u}) = \mathrm{argmin}_{x,u}\ \frac{1}{2}\,|(x,u)-(\bar{x},\bar{u})|^2 + F(x,u).$$

To compute this, note that for fixed x, the minimum in u can be computed explicitly -- it occurs when $u_i=\min\{\bar{u}_i,-f_i(x)\}$, $i=1,...,d$. The correct value of x can thus be found by minimizing the strictly convex function

$$f_0(x) + \frac{1}{2}\,|x-\bar{x}| + \frac{1}{2}\,\Sigma_i \max^2\{0,f_i(x)+\bar{u}_i\}$$

subject to the single constraint $x \in K$. The procedure (1.2) can therefore be stated concisely as

(3.2) Given $u \in R^d$, $x \in R^r$, and $y \in R^d$, compute the unique minimizer x' of the function $f_0(x') + \frac{1}{2}\,|x'-x| + \frac{1}{2}\,\Sigma_i \max^2\{0,f_i(x')+u_i+y_i\}$ subject to $x' \in K$. Then let $u_i' := \min\{u_i+y_i,-f_i(x')\}$, $u'' := \mathrm{prox}_{-D}(u-y)$, $y' := u+y-u'$, and $y'' := u-y-u''$. Begin the next iteration with $x^+=x'$, $u^+ = \frac{u'+u''}{2}$, and $y^+ = \frac{y'-y''}{2}$.

So far, no mention has been made of networks because the network interpretation of the PIES model is trivial, involving only two nodes and two arcs forming a circuit. Every circulation in such a network has the form (u,u) and every differential has the form (y,-y), so that it is possible, though not very useful, to describe the problem (3.1) using network terminology. However, it should be clear that there are a large number of potential variations of the spatial price equilibrium model that can be modelled using networks and solved by the algorithm (1.2). One can mix any of the modifications considered so far in any combination desired. For instance, one could consider a spatial model with n locations, specifying at each one a supply and/or demand "curve", or an excess supply curve, or implicitly defined supply and/or demand curves. Congestion and/or global transportation markets could be easily included and numerous other modifications are possible.

In comparison with other methods that have been proposed to solve the PIES model, the above trades an enormous theoretical advantage (global convergence) for many practical ones. Linearity of the supplier's optimization problem P(u) in the PIES model is completely destroyed by the above algorithm. While the original PIES method solves a linear programming problem at each iteration, the above demands that a nonlinear function be minimized (subject, however, only to the constraint $x \in K$). Of course, it also may be inconvenient to evaluate the proximal mapping for -D. This is the case, for example, with the log-linear demand model used in Hogan's PIES example [13], [16]. This being said, global convergence is more than a small advantage, and it is not found in the literature for this type of equilibrium model under such weak conditions. The original PIES algorithm, and Jacobi based methods like it, do not possess this strong property. Indeed, the original PIES algorithm was not guaranteed even to produce a sequence of iterates, let alone converge to a solution [1].

E. Multifacility Location Problem.

The object of the multifacility location problem is to locate n facilities at points $p_1, ..., p_n$ in R^d (usually d=2) in a manner that minimizes, subject to constraints, a weighted sum of distances between pairs of these facilities or between these and some previously existing facilities. The location problem is included in this article to emphasize its close connection with other network equilibrium models. Its resemblance to the basic spatial economic equilibrium model is especially striking, in spite of the distinct developments these two problems have had. The problem has an extensive history and many methods have been proposed for its solution. We refer to [14], [19] for further references.

Michelot, *et. al.* [14], [19], have realized that the problem readily lends itself to solution by the partial inverse approach. Their investigations, which include numerous computational experiments, have pointed to some advantages of the method: the nondifferentiability of the objective function (which is inherent to the problem) presents no impediment because the partial inverse algorithm works directly with the optimality conditions, avoiding any direct function minimization. Consequently, there is much flexibility in choosing the norms and types of constraints. Although we present the problem in a simplified setting, our method here is equivalent to Michelot, *et. al.*

Consider a connected network with node set $N = \{0,...,n\}$, nodes $1,...,n$ representing the facilities, and node 0 a dummy node. As in section II, it is assumed that no arc terminates at node 0; but here, we allow multiple arcs to connect node 0 to node j. Because the network is connected, each differential $t \in T$ corresponds to a *unique* $p=(p_0,...p_n)$ such that $t=\Delta p$ and $p_0=0$. In this manner, each $t \in T$ determines a unique choice $p_1,...p_n$ of locations for the n facilities. A <u>location problem</u> is

(4.1) to find $p \in U$ $(p_0=0)$ such that $t=\Delta p$ minimizes $f(t) = \Sigma_{a\in A} f_a(t_a)$

and where the arcs are partitioned into two classes $A = A_V \cup A_K$ with arcs in A_V representing penalties, and those in A_K representing constraints. Specifically,

$$\text{for each } a\in A_V,\ f_a(u) = c_a|u-b_a|_a \text{ for some } c_a\in R,\ b_a\in R^d, \text{ and norm } |\cdot|_a.$$

If $a\sim(0,j)$ (so that $t_a=p_j$, then such a term in the expression for f(t) penalizes the distance from p_j to b_a. Otherwise, if $a\sim(i,j)$ (so that $t_a=p_j-p_i$), then it penalizes the distance between p_i and p_j. Also,

$$\text{for each } a\in A_K,\ f_a = \Psi_{K_a} \text{ for some closed convex set } K_a \subset R^d.$$

If $a\sim(0,j)$, this choice of f_a imposes the constraint $p_j \in K_a$. Otherwise, if $a\sim(i,j)$, it forces $p_j-p_i \in K_a$.

The <u>dual problem</u> to (4.1) is

(4.2) to minimize $f^*(x) = \Sigma_{a\in A} (f_a)^*(x_a))$ over $x\in C$

where $(f_a)^*(u) = \sup\{\langle u,v\rangle - f_a(v)\}$ is the convex conjugate of f_a. Specifically, for $a\in A_K$,

$$(f_a)^*(u) = (\Psi_{K_a})^*(u) = \sup\{\langle u,v\rangle : v\in K_a\} \qquad (f_a(u) = \Psi_{K_a}(u))$$

and for $a\in A_V$,

$$(f_a)^*(u) = \langle u,b_a\rangle + \Psi_{B_a^*}\left(\frac{u}{c_a}\right) \qquad\qquad (f_a(u) = c_a|u-b_a|_a)$$

(where $B_a^* = \{v : \langle v,u\rangle \le 1$ for all $|u|_a \le 1\}$ is the unit ball for the dual norm $|\cdot|_a^*$). Let ∂f_a denote the subdifferential of f_a (whose proximal mapping will be denoted as $prox_a$). If t and x solve the primal-dual problem

(4.3) $\qquad t = \Delta p \in T\ (p_0=0),\ x \in C,$ and $x_a \in \partial f_a(t_a)$ for all $a \in A$

then it is trivial that t solves (4.1) and x solves (4.2). If the Slater condition

there exists $s \in T$ such that $s_a \in \text{relint}(\text{dom}(f_a))$ for all $a \in A$

holds, then it follows by [25, Theorem 27.4] that: if t solves (4.1) then there exists x such that (4.3) holds. Thus the problems (4.1) and (4.3) are equivalent, subject to the Slater condition. If (4.3) has a solution $\bar{t} = \Delta\bar{p}$, $\bar{x}$, then since $\bar{x}_a \in \partial f_a(\bar{t}_a)$, $t = \Delta p$ solves (4.1) if and only if $f_a(t_a) = f_a(\bar{t}_a)+\langle\bar{x}_a,t_a-\bar{t}_a\rangle$ for all a. This equality sometimes provides a means to completely describe the solution set to (4.1) and underscores the importance of solving for the dual variables, an observation that has been made by many authors.

Comparing (4.3) with (1.1), the partial-inverse algorithm (1.2) gives the following procedure to solve (4.3):

Given $x\in C$ and $t=\Delta p\in T$, compute for every $a\in A$, $t_a' = prox_a(x_a+t_a)$ and $x_a' = x_a+t_a-t_a'$. The next iterates are $x^+ = x_C'$ and $t^+ = \Delta p^+ = t_T'$.

For an arbitrary choice (t^0, x^0) of starting values, this generates sequences $t^{k+1}=(t^k)^+$ and $x^{k+1}=(x^k)^+$ converging to a solution to (4.3) (if one exists).

To clarify this procedure, note that the proximal mappings are

$$prox_a(z) = Proj_{K_a}(z) \qquad\qquad (a\in A_K,\ f_a(u) = \Psi_{K_a}(u))$$

$$prox_a(z) = z - Proj_{c_aB_a^*}(z-b_a) \qquad\qquad (a\in A_V,\ f_a(u) = c_a|u-b_a|_a)$$

The required projections are computed most easily in the cases where the sets B^* and K_a are Euclidean balls or polyhedral sets. The network is connected, so an easy argument shows that the node-arc incidence E matrix has rank n. Since any flow conservative at nodes 1,...,n is automatically conservative at node 0, row 0 of E is a linear combination of the other rows. Let E_0 denote the matrix E with row 0 deleted. Since T = row space(E) = row space(E_0), E_0 also has rank n, $E_0E_0^t$ is invertible, and $x^+=\Delta p^+$ and t^+ can be computed as: $p^+ = (E_0E_0^t)^{-1}E_0t'$, $t^+ = E_0^tp^+$, and $x^+ = x-t'+t^+$.

In this, as in other integrable network models, duality can sometimes be used to devise a stopping criterion. Assume that we know a strictly feasible solution $\pi = (0,\pi_1,...,\pi_n)$ to (4.1) and that the sets K_a $(a\in A_K)$ are bounded. (To say that π is strictly feasible means that $\tau = \Delta\pi$ satisfies $\tau_a \in \text{int}(K_a)$ for all $a\in A_K$.) By definition of f^*, $f(t)+f^*(x) \ge \langle t,x\rangle = 0$ for all $t \in T$ and $x \in C$, with equality holding if, and only if (4.3) holds. Thus it seems reasonable to choose a small $\varepsilon>0$ and halt when $f(t)+f^*(x) < \varepsilon$, since then $f(t)-\varepsilon < -f^*(x) \le -\inf(4.2) \le \inf(4.1) \le f(t)$. After each iteration, we do in fact have $t \in T$ and $x \in C$ at hand, but there is unfortunately nothing to prevent either $f(t)$ or $f^*(x)$ from taking the value $+\infty$, even when t and x are close to a solution. We must therefore provide some rule to replace $t\in T$ and $x\in C$ with vectors $\tilde{t}\in T$ and $\tilde{x}\in C$ for which $f(\tilde{t})$ and $f^*(\tilde{x})$ are finite. If for every $t\in T$ we define $\lambda(t) := \max\{\lambda\le 1: \lambda t_a+(1-\lambda)\tau_a \in K_a$ for all $a\in A_\kappa\}$ and then let $\tilde{t}:=\lambda(t)t+(1-\lambda(t))\tau$, we have $\tilde{t}\in T$ and $f(\tilde{t}) = \sum_{a\in A_V} f_a(\tilde{t}_a) < \infty$. Similarly, for any $x\in C$, define $\mu(x) := \max\{\lambda\le 1: \lambda x_a \in c_a B_a^*$ for all $a \in A_V\}$ and $\tilde{x} := \mu(x)x$. Then $\tilde{x}\in C$ and since K_a is bounded, $f^*(\tilde{x}) = \sum_{a\in A_\kappa} (\Psi_{K_a})^*(\tilde{x}_a) + \sum_{a\in A_V} \langle \tilde{x}_a, b_a\rangle < \infty$. If (4.3) has a solution then the algorithm produces convergent sequences $t^k\to\bar{t}$ and $x^k\to\bar{x}$. By the choice of τ it is clear that $\lambda(t^k)\to 1$ and $\mu(x^k)\to 1$. Hence $\tilde{t}^k\to\bar{t}$ and $\tilde{x}^k\to\bar{x}$, and since f and f^* are continuous on their effective domains, it follows that $f(\tilde{t}^k)+f^*(\tilde{x}^k) \to 0$. Halting the algorithm when $f(\tilde{t}^k)+f^*(\tilde{x}^k) < \varepsilon$ ensures that $f(\tilde{t}^k)-\varepsilon \le -f^*(\tilde{x}^k) \le -\min(4.2) \le \min(4.1) \le f(\tilde{t}^k)$. This gives a stopping criterion similar to that developed in [14], [19].

F. Traffic Equilibrium.

In this section, we discuss a model of traffic equilibrium. It, and similar models, have served as a focal point in transportation research. The particular model we describe here is equivalent to that of [10]. For surveys on transportation models, consult [7], [21].

A transport network is given with a node set N and directed arc set A. Several commodities or travel types move through this network in the directions of the arcs. The arcs are partitioned into two classes: $A = A_\rho \cup A_\phi$ with $A_\phi=\{1,...,r\}$ representing <u>fictitious</u> <u>return</u> <u>links</u>, and $A_\rho=\{r+1,...,n\}$ representing <u>real</u> <u>roadway</u> <u>links</u>. There are r traffic types, corresponding to the r fictitious return links. A <u>possible</u> <u>circuit</u> is a sequence $\alpha = (a_1,...,a_k)$ of arcs such that $a_1,...,a_{k-1}\in A_\rho$, $a_k \in A_\phi$, $a_1\sim(i_1,i_2)$, $a_2\sim(i_2,i_3)$, ..., $a_k\sim(i_k,i_1)$, and the nodes $i_1, ..., i_k$ are distinct. The arc a_k is the fictitious return link, i_1 is the <u>origin</u>, and i_k the <u>destination</u> associated with the circuit α. Two circuits use the same return link when they correspond to the same traffic type. The flow $\delta^\alpha: A\to R$ is defined as $\delta^\alpha(a)=1$ if $a\in\alpha$ and $\delta^\alpha(a)=0$ if $a\notin\alpha$.

A nonempty collection **K** of possible circuits, called the <u>feasible</u> <u>circuits</u>, is given. A <u>feasible</u> <u>circuit</u> <u>flow</u> is a function $z : \mathbf{K} \to R_+$, where $z_\alpha = z(\alpha)$ is interpreted as the input flow on circuit α of the traffic type associated with the return link of α. Each such z generates a flow $\sum_{\alpha\in \mathbf{K}} z_\alpha\delta^\alpha$ whose value on arc a is the resulting <u>total</u> <u>flow</u> associated with the feasible circuit flow z. The cone of all total flows is denoted $L:= \{\sum_\alpha z_\alpha\delta^\alpha : z$ is a feasible circuit flow$\}$, and its <u>dual</u> <u>cone</u> is $M:= \{y: \langle x,y\rangle \ge 0$ for all $\alpha\in \mathbf{K}\} = \{y: \langle\delta^\alpha,y\rangle \ge 0$ for all $\alpha\in\mathbf{K}\}$.

To each arc $a\in A$, let there be assigned a maximal monotone $P_a : R \rightrightarrows R$. For $a\in A_\rho$,

$P_a(x_a) \subset R$ is an interval containing the possible values of the travel cost per unit of flow (or the time delay) in arc a when the total flow in that arc is x_a. Since P_a is monotone, increased traffic flow in an arc drives the unit cost upwards or leaves it unchanged. For $a \in A_\phi$, $y_a \in P_a(x_a)$ means that the demand x_a for that traffic type is compatible with origin to destination cost per unit flow $-y_a$. Since P_a is monotone, increased cost drives demand downwards or leaves it unchanged. A fixed demand d for that traffic type is modelled by setting $P_a(x_a) = R^1$ if $x_a = d$, $P_a(x_a) = \emptyset$ if $x_a \neq d$.

The <u>demand equilibrium problem</u> is

(5.1) to find x, y, and z such that

i. $x \in L$, with $x = \Sigma_{\alpha\in K}\, z_\alpha \delta^\alpha$, with z a feasible circuit flow,

ii. $y \in M$,

iii. $\langle x,y\rangle = 0$,

iv. $y_a \in P_a(x_a)$ for all $a \in A$,

Conditions (ii)-(iii) assert that $\langle\delta^\alpha,y\rangle \geq 0$ for all $\alpha \in K$ and $\langle\delta^\alpha,y\rangle = 0$ whenever $z_\alpha > 0$. In words, this says that for each $a \in A_\phi$, the total origin to destination cost equals $-y_a$ for all $\alpha \in K$ with $z_\alpha > 0$ and return link a, and that the total origin to destination cost is greater than or equal to $-y_a$ for all $\alpha \in K$ with $z_\alpha = 0$ and return link a, which is <u>Wardrop's user equilibrium law</u>. Conditions (i)-(iii) are equivalent to $-y \in N_L(x)$, where N_L is the normal cone mapping for L. So, writing $P = \Pi_{a\in A} P_a$, we see that (i)-(iv) are equivalent to the monotone variational inequality

(5.2) $0 \in P(x) + N_L(x), \quad y \in P(x), \quad -y \in N_L(x),$

which, in turn, can be written $(y,-y) \in (P \times N_L)(x,x)$. Letting $X = \{(x,x)$: x is a flow$\}$, we have $X^\perp = \{(y,-y)$: y is a flow$\}$, so (i)-(iv) can be expressed in the pattern of (1.1):

(5.3) to find $(x,x) \in X$ and $(y,-y) \in X^\perp$ such that $(y,-y) \in (P \times N_L)(x,x)$.

To describe the algorithm (1.2) with these choices, we need to know the proximal mappings $\mathrm{prox}_a : R \to R$ for P_a and for N_L. To evaluate prox_{N_L} requires that a quadratic programming problem be solved, since $\mathrm{prox}_{N_L}(u) = \mathrm{proj}_L(u) = \Sigma_{\alpha\in K}\, z_\alpha\delta^\alpha$, where $z = \mathrm{argmin}_{z\geq 0} |u - \Sigma_{\alpha\in K}\, z_\alpha \delta^\alpha|^2$. Equivalently, z solves the monotone linear complementarity problem: $z \geq 0$, $D^tDz - D^tu \geq 0$, $\langle z, D^tDz - D^tu\rangle = 0$. By making the appropriate substitutions in (1.2), we obtain a procedure for solving (5.3):

(5.4) Given $x : A \to R$ and $y : A \to R$, for every $a \in A$ compute $x'_a := \mathrm{prox}_a(x_a+y_a)$, and $y'_a := x_a+y_a-x'_a$. Then determine a feasible circuit flow z'' to minimize $|x-y-\Sigma_\alpha z''_\alpha\delta^\alpha|^2$ (subject to $z'' \geq 0$), and set $x'' := \Sigma_\alpha z''_\alpha \delta^\alpha$, $y'' := x''-x+y$.

The next iterates are $x^+ = \frac{x'+x''}{2}$, $y^+ = \frac{y'+y''}{2}$.

If (5.2) has a solution, this procedure yields sequences $x^k \rightarrow \bar{x}$, $y^k \rightarrow \bar{y}$ that converge to one. Also, $(x'')^k \in L$ and $(y'')^k \in M$ are sequences that converge to the same limits $\bar{x}$ and $\bar{y}$. However, it is not necessarily true that $(z'')^k \rightarrow \bar{z}$ for some feasible circuit flow $\bar{z}$, since the representation of a total flow as $\Sigma_\alpha z_\alpha \delta^\alpha$ may not be unique.

Each P_a is the subdifferential of some closed proper convex function $f_a : R \rightarrow R$, and these functions can be determined, up to an additive constant, by integration. Thus (5.2) are the optimality conditions for the optimization problem

(5.5) to minimize $f(x) = \Sigma_{a\in A} f_a(x_a)$ subject to $x \in L$.

The conditions (5.2) are easily seen to be completely equivalent to

(5.6) $0 \in P^{-1}(y) + N_M(y)$, $x \in P^{-1}(y)$, $-x \in N_M(y)$,

which are the optimality conditions for the dual optimization problem

(5.7) to minimize $f^*(y) = \Sigma_{a\in A} (f_a)^*(y_a)$ subject to $y \in M$.

By definition of f^*, $f(x)+f^*(y) \geq \langle x,y \rangle \geq 0$ whenever $x \in L$ and $y \in M$, and it is trivial to show that equality holds if and only if x and y solve the equivalent conditions (5.1), (5.2), (5.3), or (5.6). Thus $\inf(5.7) + \inf(5.5) \geq 0$, with equality and with the minima achieved when and only when (5.2) has a solution. This pair of dual minimization problems is a familiar one and it can be derived by other means [10].

Following the pattern of the location problem, this duality can be exploited to obtain a stopping criterion. After each iteration, we have at hand $x'' = \Sigma_{\alpha \in K} z''_\alpha \delta^\alpha \in L$ and $y'' \in M$. For some predetermined $\varepsilon > 0$ we halt if $f(x'')+f^*(y'') < \varepsilon$, in which case $f(x'')-\varepsilon \leq -f^*(y'') \leq -\inf(5.7) \leq \inf(5.5) \leq f(x'')$. Unfortunately, it is possible that $f(x'') = \infty$ or that $f^*(y'') = \infty$, even when x" and y" are close to a solution. This can be remedied by the same sort of device used in the previous section. Suppose the effective domain of each f_a is a closed finite interval $[u_a,v_a]$ (which implies, in particular that $\text{dom}(f_a^*) = R^1$) and that a feasible circuit flow ζ is known such that $\xi := \Sigma_{\alpha \in K} \zeta_\alpha \delta^\alpha$ satsifies $\xi_a \in (u_a,v_a)$ for all $a \in A$. For a given $x'' = \Sigma_\alpha z''_\alpha \delta^\alpha \in L$ with z" a feasible circuit flow, let $\tilde{z} = \lambda z'' + (1-\lambda)\zeta$, where $\lambda \in [0,1]$ is chosen as large as possible so $\tilde{x} := \lambda x''_a + (1-\lambda)\xi_a \in [u_a,v_a]$ for all $a \in A$. Then $\tilde{z}$ is a feasible circuit flow, $\tilde{x} = \Sigma_\alpha \tilde{z}_\alpha \delta^\alpha \in L$ and $f(\tilde{x}) < \infty$. For the stopping criterion, we replace x" with $\tilde{x}$. If a solution exists, continuity will force $f(\tilde{x})+f^*(y'')$ to converge towards zero, so we may halt when $f(\tilde{x})+f^*(y'') < \varepsilon$.

The above model is <u>completely separable</u> in the sense that the cost y_a in each arc depends only on the flow x_a in that arc. If we relax this assumption, we can model more general situations. Doing this will give us models that may no longer be integrable, which may thus deprive us of our stopping criterion based on duality. But, as we have seen in other examples, it is possible to find a (less satisfying) stopping criterion in any case, and we still get global convergence. For the partially separable model, we assume the arcs to be partitioned into disjoint classes: $A = A_1 \cup \ldots \cup A_s$ in such a way that the cost y_a in an arc a depends only on the flow for arcs *in the same class*. For instance, all the return links joining one origin to destination pair might form a class, or all links with a common origin, etc.

The relaxed assumptions for the partially separable model are that for each class A_i, there is a maximal monotone $P_{A_i} : R^{A_i} \rightrightarrows R^{A_i}$ such that $y_{A_i} \in P_{A_i}(x_{A_i})$ whenever the flow vector $x_{A_i} \in R^{A_i}$ is compatible with the price vector $y_{A_i} \in R^{A_i}$. Let $P = \Pi P_{A_i}$. The equilibrium problem is exactly as before, except that (5.1iv) is replaced with

(5.1iv') $\qquad y_{A_i} \in P_{A_i}(x_{A_i}) \qquad (i=1,\ldots,s)$

And the procedure (5.4) is only modified to compute the proximal mappings for the P_{A_i}:

(5.4') Given $x : A \rightarrow R$ and $y : A \rightarrow R$, for $i=1,\ldots,s$, compute $x'_{A_i} := \mathrm{prox}_{A_i}(x_{A_i}+y_{A_i})$, and $y'_{A_i} := x_{A_i}+y_{A_i}-x'_{A_i}$. Then determine a feasible circuit flow z'' to minimize $|x-y-\Sigma_\alpha z''_\alpha \delta^\alpha|^2$ (subject to $z'' \geq 0$), and set $x'' := \Sigma_\alpha z''_\alpha \delta^\alpha$, $y'' := x''-x+y$. The next iterates are $x^+ = \frac{x'+x''}{2}$, $y^+ = \frac{y'+y''}{2}$.

To make this model usable, it is of course necessary to employ only multifunctions P_{A_i} whose proximal mappings are efficiently computable. As in section II (example 2) it may be necessary to further decompose the P_{A_i} in order to make this possible.

REFERENCES

1. Ahn, B.H., Computation of Market Equilibria for Policy Analysis: the Project Independence Evaluation System Approach, New York, Garland publishing.

2. Boughazi, M.A., Contribution à l'étude des algorithmes d'optimisation en analyse des données, thèse, Grenoble (1987).

3. Carey, M., Integrability and mathematical programming models: a survey and a parametric approach, Econometrica 45 (1977) 1957-1976.

4. Cournot, A.A., Mathematical Principles of the Theory of Wealth (1838).

5. Dafermos, S., The general multimodal network equilibrium problemwith elastic demand, Networks 12 (1982) 57-72.

6. Enke, S., Equilibrium among spatially separated markets: solution by electric analogue, Econometrica, 19 (1951) 40-47.

7. Florian, M., An introduction to models used in transportation planning, in M. Florian (ed.), Transportation Planning Models (1984) 137-152, North-Holland.

8. Florian, M., and H. Spiess, The convergence of diagonalization algorithms for asymmetric network equilibrium problems, Transportation Research 16B (1982) 447-483.

9. Florian, M., and M. Los, A new look at static spatial price equilibrium models, Regional Science and Urban Economics 12 (1982) 579-597.

10. Hall, M.A., and E.L. Peterson, Traffic equilibria analysed via geometric programming, in M.A. Florian (ed.), Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems (1976) 53-105, Springer-Verlag, New York.

11. Harker, P.T., and J.-S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms, and applications, preprint, U. Penn, December 1987.

12. Harker, P.T., and T.L. Friesz, Prediction of intercity freight flows, I: Theory, Transportation Research, 20B (1986) 139-153.

13. Hogan, W. W., Project independence evaluation system: structure and algorithms, Proceedings of Symposia in Applied Mathematics of the AMS 21 (1977) 121-137.

14. Idrissi, H., and O. Lefebvre, and C. Michelot, "Duality for constrained multifacility location problems with mixed norms and applications", working paper, March 1987.

15. Jones, P.C., and E.S. Theise, On the equivalence of competitive transportation markets and congestion in spatial price equilibrium models, Technical Report 87-02, Dept. I. E. and Mngt. Sci., Northwestern University.

16. Josephy, N. H., Hogan's PIES example and Lemke's algorithm, MRC Technical Report #1972, June 1979.

17. Krasnoselski, M.A., Two observations about the method of successive approximations, Uspekhi. Math. Nauk 10 (1955) 123-127.

18. Lawrence, J., and J.E. Spingarn, On fixed points of nonexpansive piecewise isometric mappings, Proceedings of the London Mathematical Society (3) 55 (1987) 605-624.

19. Lefebvre, O., and C. Michelot, "A primal-dual algorithm for the Fermat-Weber problem involving mixed gauges", working paper, April 1986.

20. MacKinnon, J.G., An algorithm for the generalized transportation problem, Regional Science and Urban Economics 5 (1975) 445-464.

21. Magnanti, T.L., Models and algorithms for predicting urban traffic equilibria, in M. Florian (ed.), Transportation Planning Models (1984) 153-185, North-Holland.

22. Minty, G.J., Monotone (nonlinear) operators in Hilbert space, Duke Mathematical Journal 29 (1962) 341-346.

23. Moreau, J.-J., Proximité et dualité dans un espace Hilbertien, Bull. Soc. math. France 93, 1965, 273-299.

24. Pang, J.-S., and D. Chan, Iterative methods for variational and complementarity problems, Mathematical Programming 24 (1982) 284-313.

25. Rockafellar, R.T., Convex Analysis (1970) Princeton University Press.

26. Rockafellar, R.T., Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976) 877-898.

27. Samuelson, P., Spatial price equilibrium and linear programming, American Economic Review 42 (1952) 283-303.

28. Spingarn, J.E., Approximate fixed points and the ellipsoid algorithm, Proceedings, Séminaire d'Analyse Numérique, Université Paul-Sabatier, Toulouse III (1986-1987).

29. Spingarn, J.E., Applications of the method of partial inverses to convex programming: decomposition, Mathematical Programming, 32 (1985) 199-223.

30. Spingarn, J.E., On computation of spatial economic equilibria, CORE discussion paper #8731, Université Catholique de Louvain, May 1987.

31. Spingarn, J.E., Partial inverse of a monotone operator, Applied Mathematics and Optimization 10 (1983) 247-265.

32. Takayama, T., and G.G. Judge, Spatial and Temporal Price and Allocation Models (1971) North-Holland, Amsterdam.

33. Theise, E.S., Alternative implementations of a diagonalization algorithm for multiple commodity spatial price equilibria, preprint, Northwestern University, April 1988.

34. Arrow, K.J., and F. H. Hahn, General Competitive Analysis (1971) Holden-Day, San Francisco.

35. T. Bonnesen, W. Fenchel, Theorie der Konvexen Körper, Chelsea Publishing Co., New York (1971). Published originally as Ergeb. Math. Grenzg. 3, No. 1, Berlin (1934).

36. A. S. Nemirovsky, D. B. Yudin, Problem Complexity and Method Efficiency in Optimization, english translation: John Wiley & Sons, New York (1983).

37. B. S. Mityagin, Two inequalities for the volume of convex bodies, Matem. Zametki 5, No. 1. English translation: Mathematical notes of the Academy of Sciences USSR 5, No. 1, 61-64. Plenum, New York.