EORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 9/16/88

roject No. G-37-614/R5952-0A0 School/~~Lab~~X Mathematics

ncludes Subproject No.(s) N/A

roject Director(s) J. E. Spingarn GTRC/~~GIT~~XXX

ponsor NSF

itle The Proxomal Point Algorithm in Mathematical Programming

ffective Completion Date: 11/30/87 (Performance) 11/30/87 (Reports)

rant/Contract Closeout Actions Remaining:

- [ ] None
- [ ] Final Invoice or Copy of Last Invoice Serving as Final
- [ ] Release and Assignment
- [x] Final Report of Inventions and/or Subcontract: Patent and Subcontract Questionnaire sent to Project Director [x]
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other

ontinues Project No. Continued by Project No.

OPIES TO:

roject Director
esearch Administrative Network
esearch Property Management
ccounting
rocurement/~~GTRI Supply Services~~
~~esearch Security Services~~
eports Coordinator (OCA)
~~rogram Administration Division~~
ontract Support Division (2)

~~Facilities Management~~
Library
GTRC
Project File
Other

GEORGIA INSTITUTE OF TECHNOLOGY — OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL [ ] REVISION NO. ______

Project No. G-37-614 (R5952-0A0) GTRC/~~GIT~~ DATE 6 / 7 / 85

Project Director: J. E. Spingarn School/~~Lab~~ Math

Sponsor: National Science Foundation

Type Agreement: Grant NO. DMS-8506712

Award Period: From 6/15/85 To ~~11/30/86~~ 11/30/87 (Performance) ~~2/28/87~~ 11/30/87 (Reports)

Sponsor Amount: This Change / Total to Date

Estimated: $ ______ $ 33,189

Funded: $ ______ $ 16,089

Cost Sharing Amount: $ 12,938 (Includes year 2) Cost Sharing No: G-37-320

Title: The Proximal Point Algorithm in Mathematical Programming

ADMINISTRATIVE DATA

OCA Contact John Schonk x4820

1) Sponsor Technical Contact:	2) Sponsor Admin/Contractual Matters:
John E. Lagnese	Stephen G. Burnisky
National Science Foundation	National Science Foundation
MPS/DMS	DGS/MPS
Washington, DC 20550	Washington, DC 20550
202/357-9765	202/357-9671

Defense Priority Rating: N/A Military Security Classification: N/A

(or) Company/Industrial Proprietary: N/A

RESTRICTIONS

See Attached NSF Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval – Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with GIT

COMMENTS:

No Funds may be expended after 11/30/86.

RECEIVED Research Reports

COPIES TO:

SPONSOR'S I. D. NO. 02.107.000.85.073

Project Director	Procurement/EES Supply Services	GTRC
Research Administrative Network	~~Research Security Services~~	Library
Research Property Management	~~Reports Coordinator~~ (OCA)	Project File
Accounting	Research Communications (2)	Other A. Jones

PROGRESS REPORT

Working with Professor Jim Lawrence (George Mason University), we have been studying the question of finite termination of the proximal point algorithm. Briefly, the question we are addressing is this: if q is a nonexpansive mapping and

$$x_{n+1} = (x_n+q(x_n))/2$$

then what conditions will guarantee that $x_k = x_{k+1} = \dots$ for some k.

We have found that the class of nonexpansive piecewise isometries behaves especially nicely under the above iteration. For such q, we have demonstrated that there exists K such that for all $k \geq K$, the set $\{x_k, x_{k+1}, \dots\}$ positively spans a subspace on which q is linear. This demonstrates linear convergence to a fixed point if one exists. If, furthermore, some additional condition is satisfied, such as if the subspace on which q is linear is $\{0\}$, or if $((I+q)/2)^{-1} - I$ is a subdifferential mapping on the subspace, then finite termination can be shown to be a consequence.

We have applied these results to show finite termination occurs when solving systems of linear inequalities, and for the problem of finding a point in the intersection of a polyhedron

<u>STATEMENT OF SUPPORT</u>

The principal investigator is currently supported by this NSF grant number D[illegible]S-8[illegible]06712. He has not submitted nor has he pending any proposals with any other organization or agency.

with a subspace. These results are currently being written up in a paper [1] that should be ready for submission shortly.

The current research was motivated by previous work in which we studied iterative methods for solving systems of linear inequalities via the proximal point algorithm. We have recently extensively revised and resubmitted one paper on that subject [2]. A copy is enclosed.

CURRENT PAPERS

1. J. Lawrence and J. E. Spingarn, paper in progress on finite termination in the proximal point algorithm.

2. J. E. Spingarn, *A projection method for least-square solutions to overdetermined systems of linear inequalities*, resubmitted in August, 1984, for publication.

Revised September 1985

A PROJECTION METHOD FOR LEAST-SQUARE SOLUTIONS TO OVERDETERMINED SYSTEMS OF LINEAR INEQUALITIES

by

Jonathan E. Spingarn
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

An algorithm previously introduced by the author for finding a feasible point of a system of linear inequalities is further investigated. For inconsistent systems, it is shown to generate a sequence converging at a linear rate to the set of least-square solutions. The algorithm is a projection type method, and is a manifestation of the proximal point algorithm.

Key Words: linear inequalities, feasibility, relaxation method, projection method, proximal point algorithm, monotone operator.

This research was supported, in part, by the Air Force Office of Scientific Research, under grant 80-0195C, and by the National Science Foundation, under grant DMS-8506712.

Introduction

Let a system

$$\langle x,u_i\rangle \le b_i, \quad i = 1,\dots,n \tag{0.1}$$

of linear inequalities be given ($0 \ne u_i \in R^d$, $b_i \in R$). Let $C_i = \{x: \langle x,u_i\rangle \le b_i\}$ and $C = C_1 \cap \dots \cap C_n$. In [37], the author introduced the following primal-dual algorithm for finding a point $x \in C$:

Start: Choose arbitrary $x_0, y_{01}, \dots, y_{0n} \in R^d$ (0.2)
with $y_{01} + \dots + y_{0n} = 0$.

Step k ($k = 0,1,\dots$): Compute

$$x'_{ki} = \mathrm{proj}_{C_i}(x_k + y_{ki}), \quad i = 1,\dots,n$$

$$y'_{ki} = x_k + y_{ki} - x'_{ki}, \quad i = 1,\dots,n$$

and update

$$x_{k+1} = \frac{1}{n}\Sigma_{i=1}^n x'_{ki}$$

$$y_{k+1,i} = y'_{ki} - \frac{1}{n}\Sigma_{j=1}^n y'_{kj}, \quad i = 1,\dots,n.$$

It was shown in [37] that regardless of the choice of starting values, either $x_k \to x$ and $y_{ki} \to y_i$ with $x \in C$, y_i an outward normal to C_i at x, and $y_1 + \dots + y_n = 0$; or $|(x_k+y_{k1},\dots,x_k+y_{kn})| \to \infty$ and the system is inconsistent. The algorithm was investigated further in [38], where it was shown that termination occurs after a finite number of iterations if the interior of C is nonempty. More precisely, $\mathrm{int}(C) \ne \emptyset$ implies for some k that $x_k = x_{k+1} = \dots$ with

$x_k \in C$.

In this paper, we investigate the behavior of the algorithm in the cases where $\text{int}(C) = \emptyset$ or where $C = \emptyset$. If $\text{int}(C) = \emptyset$ but $C \neq \emptyset$, we show that the sequence (x_k) converges to a solution and that the distance to the solution set approaches zero at a linear rate. In the case where $C = \emptyset$, we demonstrate convergence at a linear rate to the set of least-square solutions and convergence of (x_k) to one particular least-square solution.

The algorithm is actually a manifestation of the known proximal point algorithm and the results of Rockafellar [36], Bruck and Reich [5], [34], and Luque [24] on the behavior of the proximal point algorithm will play a key role here.

Our method is closely related to similar projection methods for solving systems of linear equations or inequalities. The Kaczmarz method [22] of cyclic projection solves a system of linear equations by the procedure

$$x_{k+1} = P_n(P_{n-1}(\ldots(P_1(x_k)))),$$

where the P_j are projections onto the defining hyperplanes. If the system is consistent then x_n converges to a minimum norm solution if x_0 belongs to the linear span of the normal vectors to the hyperplanes [21]. Tanabe [39] showed that for each j the subsequence $x_j, x_{n+j}, x_{2n+j}, \ldots$ converges even if the system is inconsistent. Censor, Eggermont, and Gordon [8] showed that when strong underrelaxation is applied in

Kaczmarz's method to an inconsistent system of linear equations, the limits of the cyclic subsequences approach a least squares solution of the system. Strong underrelaxation is undesirable since it slows down convergence. Plotnikov [31] has obtained a similar result for more general systems of convex sets with empty intersection. Other methods related to that of Kaczmarz for systems of equations are discussed in [6], [7], [11], [12]. The method of successive projections was also noticed by von Neumann [30] (cf. also [10], [20]). The cyclic projection method was extended to inequalities in the "relaxation method" of Agmon [1] and Motzkin and Schoenberg [29]. This method and its relationship to subgradient optimization was studied extensively by Goffin [13]-[19]. A survey of methods, including some projection methods, for computing least-square solutions to systems of linear equations can be found in [41].

Some projection methods rely, as does ours, on an averaging process. Cimmino's method [9] for systems of linear equations is

$$x_{k+1} = (1-\lambda)x_k + \frac{\lambda}{n} \Sigma_{j=1}^{n} \mathrm{Proj}_{C_j}(x_k)$$

with $\lambda = 2$. This is the gradient method $x_{k+1} = x_k - \lambda\nabla f(x_k)$ for the function $f(x) = \frac{1}{2n} \Sigma_{i=1}^{n} |x - \mathrm{Proj}_{C_j}(x)|^2$ and when $0 < \lambda < 2$, convergence to a least square solution can be deduced from the results in [32], [33]. Cimmino's method was generalized by Auslender [3], [4] to the inequality case.

The reliance on dual variables in algorithm (0.2) is less of a handicap than it might seem since it is not actually necessary to keep track of the vectors y_k and y_k' (see remarks in [38]). An unusual feature of the algorithm is that, unlike other projection methods, it can be observed to accelerate when caught in small solid angles.

After reviewing preliminaries and establishing needed properties of the proximal point algorithm in §I, we will prove convergence to a least-square solution in §II, and establish that the convergence is linear in §III.

I. Background

Let H be a Euclidean space equipped with the standard inner project $\langle x,y\rangle = \Sigma x_i y_i$. A multifunction $M: H \rightrightarrows H$ is monotone if $\langle x-x', y-y'\rangle \geq 0$ whenever $y \in M(x)$ and $y' \in M(x')$. The graph of M is the set $Gr(M) = \{(x,y) \in H \times H: y \in M(x)\}$. If $Gr(M) \subset Gr(S)$ implies $M = S$ whenever S is monotone then M is maximal monotone.

If $M: H \rightrightarrows H$ is maximal monotone, then for each $x \in H$ there exists a unique $P(x) \in H$ such that $x - P(x) \in M(P(x))$ [27]. This function $P = (I+M)^{-1}$ is the proximal mapping associated with M. Its fixed points are the zeros of M. The simplest example, due to Moreau [28], is the case where $M = \partial f$ is the subdifferential of a lower semicontinuous convex function f. In this case,

$$P(x) = \arg\min_{y \in H}\{f(y) + \frac{1}{2}|y-x|^2\}.$$

The property of $P: H \to H$ being the proximal mapping $(I+M)^{-1}$ for some maximal monotone M is equivalent to P having the property (letting $Q = I-P$)

$$|z-z'|^2 \geq |P(z)-P(z')|^2 + |Q(z)-Q(z')|^2 \quad \forall\, z,z' \in H \tag{1.1}$$

and also to the existence of a nonexpansive mapping $N: H \to H$ such that $P = \frac{1}{2}(I+N)$. The convergence to a fixed point of P (if one exists) of the sequence

$$x_{k+1} = P(x_k) \tag{1.2}$$

(x_0 being arbitrary) for functions P satisfying (1.1) was observed by Martinet, who also observed that the proximal mapping of Moreau has this property. For functions of the form $P = \frac{1}{2}(I+N)$, convergence to a fixed point (if one exists) follows from a result of Krasnoselski [23]. In [36], Rockafellar proposed the iteration (1.2) with $P = (I+M)^{-1}$ as an algorithm for finding a zero of a maximal monotone multifunction M, observing that its convergence was a consequence of earlier results. Taking his cue from the *application prox* of Moreau, he named it the "proximal point algorithm" and established several of its properties useful for applications to convex programming.

As an immediate consequence of (1.1), P is nonexpansive:

$$|z-z'| \geq |P(z)-P(z')| \quad \forall\, z,z' \in H. \tag{1.3}$$

Taking $z = x_k$ and $z' = x_{k+1}$, we see that

$$|Q(z_k)| \geq |Q(x_{k+1})|, \quad k = 0,1,2,... \tag{1.4}$$

The following lemma of Bruck and Reich [5], [34] gives useful information in the event that P is fixed-point free. In the more general result they prove, v is taken to be the element of least norm in the *closure* of range(M).

Lemma 1 ([5], [34]). Let M be maximal monotone and assume the range of M has an element v of minimum norm. For any sequence $x_{k+1} = P(x_k)$ generated by the proximal point algorithm, $Q(x_k)$ converges to v.

Proof. By [26], range (M) $\supset$ relint(conv(range(M))). By [35, Theorem 6.3], this implies cl(range(M)) is convex, so v is unique. Noting that $v \in$ range(M) = range (Q), pick $z_0 \in Q^{-1}(v)$. Defining $z_{k+1} = P(z_k)$, we get

$$|v| = |Q(z_0)| \geq |Q(z_1)| \geq \ldots$$

by (1.4). The uniqueness of v thus implies $v = Q(z_0) = Q(z_1) = \ldots$ By (1.1),

$$|z_k - x_k|^2 - |z_{k+1} - x_{k+1}|^2 \geq |v - Q(x_k)|^2.$$

But $|z_k - x_k|$ is nonincreasing (1.3) so $Q(x_k) \to v$. □

The convergence of (x_k) for arbitrary $x_0 \in H$ to a fixed point of P (or a zero of M), if one exists, is an immediate consequence of Lemma 1. For if P has a fixed point, (x_k) is clearly bounded. By Lemma 1, $Q(x_k) \to 0$ so if x_∞ is a cluster point, $Q(x_\infty) = 0$ by the continuity of Q (1.1). Hence $0 \in M(x_\infty)$. But $|x_k - x_\infty|$ is nonincreasing (1.3), so $x_k \to x_\infty$, as desired. Note, furthermore, that

$$x_k = x_0 - \Sigma_{i=0}^{k-1} Q(x_i), \quad k = 1,2,3,\ldots$$

so by Lemma 1,

$$\frac{x_k}{k} \to -v, \tag{1.5}$$

another result of Reich [34]. In particular, $|x_k| \to \infty$ if $M^{-1}(0) = \emptyset$.

In [36], Rockafellar established linear convergence of the proximal point algorithm under the assumption that M^{-1} satisfy a Lipschitz condition at 0. The Lipschitz condition requires in particular that $M^{-1}(0)$ be single-valued. Luque [24] showed that linear convergence to the set $M^{-1}(0)$ (though not necessarily to a particular point in $M^{-1}(0)$) still holds under a weaker condition not requiring that $M^{-1}(0)$ be single-valued. In the following lemma, we generalize Luque's result, showing that useful convergence information can be obtained even in the case where M has no zeros.

<u>Lemma 2</u>. Let M be maximal monotone. Assume range(M) has an element v of least norm and that there is $\kappa > 0$ and a neighborhood V of v such that for all $y \in V$, $Y \in M(x)$ implies

$$\operatorname{dist}(x, M^{-1}(v)) \leq \kappa |y-v|. \tag{1.6}$$

If $x_{k+1} = P(x_k)$ $(P = (I+M)^{-1})$, then $\operatorname{dist}(x_k, M^{-1}(v)) \to 0$ at a linear rate with modulus $\kappa/\sqrt{1+\kappa^2}$.

Remark. In the case where $v = 0$ this follows from [24, Theorem 2.1].

Proof. M^{-1} is closed- and convex-valued [2, Proposition 6.7.3] so $M^{-1}(v)$ is closed and convex. For each k, let u_k denote the unique nearest point to x_k in $M^{-1}(v)$. For all $x \in M^{-1}(v)$, $|v| = |(x+v)-x| \geq |x-P(x)| = |Q(x)|$ since P is nonexpansive. By choice of v, this implies $Q(x) = v$, or equivalently $x-v \in M^{-1}(v)$. Thus

$$M^{-1}(v) - v \subset M^{-1}(v). \tag{1.7}$$

By Lemma 1, $Q(x_k) \to v$ so $Q(x_k) \in V$ for all k sufficiently large. Thus (1.6) implies

$$\operatorname{dist}(x_{k+1}, M^{-1}(v)) \leq \kappa|Q(x_k)-v| \tag{1.8}$$

for all k sufficiently large. Then

$$\begin{aligned}
|x_{k+1}-u_{k+1}|^2 &\leq |x_{k+1} - (u_k-v)|^2 && \text{(by 1.7)} \\
&\leq |x_k-u_k|^2 - |Q(x_k)-v|^2 && \text{(by 1.1)} \\
&\leq |x_k-u_k|^2 - \frac{1}{\kappa^2}|x_{k+1}-u_{k+1}|^2 && \text{(by 1.8)}
\end{aligned}$$

and so

$$|x_{k+1}-u_{k+1}| \leq \frac{\kappa}{\sqrt{1+\kappa^2}}|x_k-u_k|. \qquad \square$$

Next, we relate Algorithm (0.2) to the proximal point algorithm. The notation introduced below will be observed throughout the paper. Henceforth, let $H = R^d \times \ldots \times R^d$ (n times, where n is the number of inequalities in the system (0.1)). Define subspaces A and B of H by

$$A = \{(x_1,\ldots,x_n) \in H\colon x_1 = \ldots = x_n\}$$

$$B = \{(y_1,\ldots,y_n) \in H\colon y_1 + \ldots + y_n = 0\}.$$

With H endowed with the standard inner product, A and B are orthogonal complements; i.e., $A = B^{\perp}$ and $B = A^{\perp}$. For any $x \in H$, x_A and x_B shall denote the orthogonal projections of x onto A and B, respectively. Hence for $x = (x_1,\ldots,x_n)$,

$$x_A = \left(\frac{1}{n}\Sigma_{i=1}^n x_i,\ldots,\frac{1}{n}\Sigma_{i=1}^n x_i\right) \tag{1.9}$$

$$x_B = \left(x_1 - \frac{1}{n}\Sigma_{i=1}^n x_i,\ldots,x_n - \frac{1}{n}\Sigma_{i=1}^n x_i\right).$$

For $i = 1,\ldots,n$, define $T_i\colon R^d \rightrightarrows R^d$ by

$$T_i(x) = \begin{cases} \{0\} & \text{if } \langle x,u_i\rangle < b_i \\ \{tu_i\colon t \ge 0\} & \text{if } \langle x,u_i\rangle = b_i \\ \emptyset & \text{if } \langle x,u_i\rangle > b_i. \end{cases}$$

T_i is just the *normal cone* mapping for the set C_i [35]. Define $T = T_1 \times \ldots \times T_n$, i.e., $(y_1,\ldots,y_n) \in T(x_1,\ldots,x_n)$ iff $y_i \in T_i(x_i)$, $i = 1,\ldots,n$. T is the normal cone mapping for $\hat{C} = C_1 \times \ldots \times C_n$. The T_i and hence also T are all maximal monotone. A new multifunction $T_A\colon H \rightrightarrows H$ is defined by declaring its graph to be

$$\text{graph}(T_A) = \{(x_A + y_B, x_B + y_A) : y \in T(x)\}. \tag{1.10}$$

T_A is the partial inverse of T with respect to A, a concept introduced in [37]. T_A is again maximal monotone. The significance of T_A lies in the following easy consequence of its definition: If $\bar{x} = (x,\ldots,x) \in A$ and $\bar{y} = (y_1,\ldots,y_n) \in B$, then the following are equivalent and each implies $x \in C$:

i. $0 \in T_A(\bar{x} + \bar{y})$

ii. $\bar{y} \in T(\bar{x})$.

Thus, the A-component of any zero of T_A solves (0.1); conversely, if x solves (0.1), then $0 \in T_A(\bar{x})$, where $\bar{x} = (x,\ldots,x)$. The problem of solving (0.1) thus reduces to the problem of finding a zero of T_A. This can be done by applying the proximal point algorithm to T_A. The resulting procedure is Algorithm (0.2). To make this precise, following the notation used in (0.2), let us define

$$\begin{aligned}\hat{x}_k &= (x_k,\ldots,x_k) \\ \hat{y}_k &= (y_{k1},\ldots,y_{kn}) \\ \hat{x}'_k &= (x'_{k1},\ldots,x'_{kn}) \\ \hat{y}'_k &= (y'_{k1},\ldots,y'_{kn}).\end{aligned} \tag{1.11}$$

Note that

$$\begin{aligned}&\text{i.} && \hat{x}_{k+1} = (\hat{x}'_k)_A \quad \text{and} \quad \hat{y}_{k+1} = (\hat{y}'_k)_B\,, \\ &\text{ii.} && \hat{x}_k + \hat{y}_k = \hat{x}'_k + \hat{y}'_k\,, \quad \text{and} \\ &\text{iii.} && \hat{y}'_k \in T(\hat{x}'_k).\end{aligned} \tag{1.12}$$

From the definition of T_A, (1.12iii), and then (1.12i),

$$(\hat{x}_k')_B + (\hat{y}_k')_A \in T_A((\hat{x}_k')_A + (\hat{y}_k')_B) = T_A(\hat{x}_{k+1} + \hat{y}_{k+1}).$$

But then by (1.12ii) and the definition of T_A,

$$\begin{aligned}\hat{x}_k + \hat{y}_k &= \hat{x}_k' + \hat{y}_k' \\ &= ((\hat{x}_k')_A + (\hat{y}_k')_B) + ((\hat{x}_k')_B + (\hat{y}_k')_A) \\ &\in (I+T_A)(\hat{x}_{k+1} + \hat{y}_{k+1})\end{aligned}$$

or

$$\hat{x}_{k+1} + \hat{y}_{k+1} = P_A(\hat{x}_k+\hat{y}_k), \quad \text{where } P_A = (I+T_A)^{-1}. \tag{1.13}$$

Thus, the passage from $\hat{x}_k + \hat{y}_k$ to $\hat{x}_{k+1} + \hat{y}_{k+1}$ is the outcome of one iteration of the proximal point algorithm, applied to the maximal monotone multifunction T_A.

II. Asymptotic Behavior

We begin in this section to study the behavior of the algorithm when the system of inequalities is inconsistent. We will stick with the notation ($T,A,B,T_A,T_i,\hat{x}_k,P_A$, etc.) introduced at the end of §I.

If the system (0.1) is inconsistent, T_A has no zeros. The proximal point algorithm applied to T_A will thus yield a divergent sequence, but the lemmas of the previous section can be applied to gain information about the asymptotic behavior of the algorithm. The principal result of this section is that for any sequence (x_k) generated by the algorithm (0.2), x_k converges to a least-square solution.

As a first step, we determine the element of least norm in the range of T_A. From the definition of T_A,

$$\text{range}(T_A) = \{\hat{y}_A + \hat{w}_B \colon \hat{y} \in T(\hat{w})\}$$

(in this and in the following section, "^" indicates an element or subset of $H = R^d \times \ldots \times R^d$). If we define $\hat{C} = C_1 \times \ldots \times C_n$, then $\hat{C} = \text{domain}(T)$ and $0 \in T(\hat{w})$ for all $\hat{w} \in \hat{C}$. Thus the element of least norm in range(T_A) must also be the element of least norm in the subset $\{\hat{w}_B \colon \hat{w} \in \hat{C}\}$. But this subset is the projection onto B of the polyhedral convex set $\hat{C}$; it is hence polyhedral (closed) convex and has a unique element $\hat{v}$ of minimum norm which equals $\hat{w}_B$ for at least one $\hat{w} \in \hat{C}$. Any $\hat{w} \in \hat{C}$ for which $\hat{w}_B = \hat{v}$ will be called a <u>minimally dispersed system of representatives</u> for the system (0.1). This terminology is motivated by the fact that such a $\hat{w} = (w_1, \ldots, w_n)$ minimizes

$$|\hat{w}_B|^2 = |w_1 - \tfrac{1}{n}\Sigma w_i|^2 + \ldots + |w_n - \tfrac{1}{n}\Sigma w_i|^2$$

subject to $w_i \in C_i$, $i = 1, \ldots, n$.

<u>Lemma 3</u>. The following are equivalent and are satisfied for at least one $\hat{w} = (w_1, \ldots, w_n) \in H$:

i. $\hat{w}$ is minimally dispersed (2.1)

ii. $0 \in \hat{w}_B + T(\hat{w})$ (equivalently,
$0 \in w_j - \frac{1}{n}\Sigma_{i=1}^n w_i + T_j(w_j)$, $j = 1, \ldots, n$).

iii. $\hat{w} = \text{proj}_{\hat{C}}(\hat{w}_A)$ (equivalently,
$w_j = \text{proj}_{C_j}(\frac{1}{n}\Sigma_{i=1}^n w_i)$, $j = 1, \ldots, n$).

<u>Proof</u>. $\hat{w}$ is minimally dispersed iff $\hat{w}$ minimizes $|\hat{w}_B|^2$ subject to $\hat{w} \in \hat{C}$. This i $\leftrightarrow$ ii follows by [35, Theorem 27.4]. The equivalence ii $\leftrightarrow$ iii is obvious. □

The set χ of <u>least-square solutions</u> to (0.1) consists of all $x \in R^d$ which, together with some $\hat{w} = (w_1,\ldots,w_n) \in H$, solve the problem

$$\text{to minimize } \tfrac{1}{2}\, \Sigma_{i=1}^{n} |x-w_i|^2 \text{ subject to } \langle w_j,u_j\rangle \le b_j, \quad j = 1,\ldots,n. \tag{2.2}$$

By [35, Corollary 28.3.1], x and $\hat{w}$ solve this problem if, and only if, there exist Lagrange multipliers $\lambda_1,\ldots,\lambda_n$ satisfying, together with x and $\hat{w}$, the Karush-Kuhn-Tucker conditions. For the problem (2.2), these conditions assert that

$$\begin{aligned} &\text{i.} \quad x = \tfrac{1}{n}\,\Sigma_{i=1}^{n} w_i \quad \text{and} \quad x-w_j = \lambda_j u_j, \quad j = 1,\ldots,n. \\ &\text{ii.} \quad \lambda_j \ge 0 \quad \text{and} \quad \langle w_j,u_j\rangle \le b_j, \quad j = 1,\ldots,n. \qquad (2.3) \\ &\text{iii.} \quad \langle w_j,u_j\rangle < b_j \text{ implies } \lambda_j = 0, \quad j = 1,\ldots,n. \end{aligned}$$

Observe that (2.3) implies $w_j = \text{proj}_{C_j}(x)$, and hence that $\hat{w}$ is minimally dispersed.

The next lemma describes the least-square solutions to (0.1).

<u>Lemma 4</u>. $x \in \chi$ if, and only if, there exists $\hat{w} = (w_1,\ldots,w_n) \in H$ such that

i. $\hat{w}$ is minimally dispersed (2.4)

ii. $x = \frac{1}{n} \Sigma_{i=1}^{n} w_i$.

If x and $\hat{w}$ satisfy (2.4) then

$$w_j = \text{proj}_{C_j}(x), \quad j = 1,\ldots,n.$$

Furthermore, χ is nonempty.

Proof. $x \in \chi$ then there exists $\hat{w} \in H$ and $\lambda \in R^n$ satisfying the Kuhn-Tucker conditions (2.3). This implies that (2.4) holds.

Conversely, suppose (2.4) holds. By (2.1iii), $w_j = \text{proj}_{C_j}(x)$, $j = 1,\ldots,n$. Thus, it is possible to pick $\lambda_j \geq 0$ such that $\lambda_j u_j = x - w_j$, and the Kuhn-Tucker conditions (2.3) hold with this choice of λ. In particular, x and $\hat{w}$ solve the problem (2.2) and $x \in \chi$.

For any minimally dispersed $\hat{w}$, (2.4ii) defines a least-square solution x. Hence $\chi \neq \emptyset$. □

If $x \in \chi$, then for $\hat{w}$ as in Lemma 4, $\hat{w}_B = \hat{v}$. In other words, $v_j = w_j - \frac{1}{n} \Sigma_{i=1}^{n} w_i = w_j - x$, $j = 1,\ldots,n$. By (2.1iii),

$$x = \text{proj}_{C_j}(x) - v_j, \quad j = 1,\ldots,n, \quad (x \in \chi) \tag{2.5}$$

Thus

i. If $x \in \chi$ and $v_j = 0$, then $x \in C_j$. (2.6)

ii. If $x \in \chi$ and $v_j \neq 0$, then
$x \notin C_j$, $\text{Proj}_{C_j}(x) \in \text{bd}(C_j)$, and v_j is a positive multiple of u_j.

The next lemma describes the structure of χ more completely. It asserts that χ is the intersection of an affine flat consisting of the least-square solutions to a certain subsystem of (0.1) with a polyhedral convex set consisting of the feasible solutions of the complementary subsystem of (0.1). These two complementary subsystems correspond to a partition of the constraint indices into the two sets:

$$I = \{i: v_i \neq 0\} \quad \text{and} \quad J = \{j: v_j = 0\}.$$

Defining

$$\chi_I = \{x: x \text{ is a least-square solution to the subsystem } \langle x,u_i\rangle \le b_i,\ i \in I\}$$

and

$$\chi_J = \bigcap_{j \in J} C_j$$

(with $\chi_I = R^d$ if $I = \emptyset$, and $\chi_J = R^d$ if $J = \emptyset$), we have

<u>Lemma 5</u>. $\chi = \chi_I \cap \chi_J$. Moreover, χ_I is a translate of the subspace

$$N := \{\zeta \in R^d: \langle \zeta, v_i\rangle = 0,\ i = 1,\dots,n\}.$$

<u>Proof</u>. Let $\phi_j(x) = \text{dist}^2(x,C_j)$, $\phi = \Sigma_{j=1}^n \phi_j$. Suppose that $x \in \chi_I \cap \chi_J$. Using first the fact that $x \in \chi_I$ and then that $x \in \chi_J$, we have for any $x' \in R^d$

$$\phi(x') \ge \sum_{i \in I} \phi_i(x') \ge \sum_{i \in I} \phi_i(x) = \phi(x)$$

so $x \in \chi$. Thus $\chi_I \cap \chi_J \subset \chi$.

Suppose next that $x \in \chi$ and define $w_j = \mathrm{proj}_{C_j}(x)$. By (2.6i), $x \in \chi_J$. To show $x \in \chi_I$, we may clearly assume $I \neq \emptyset$. Since $x = \frac{1}{n} \Sigma_{i=1}^{n} w_i$ (Lemma 4) and $w_j = x$ for $j \in J$, it follows that $x = \frac{1}{|I|} \sum_{i \in I} w_i$. Then $w_i = \mathrm{Proj}_{C_i}(\frac{1}{|I|} \sum_{j \in I} w_j)$ for $i \in I$ and the w_i $(i \in I)$ are minimally dispersed for the subsystem indexed by I (2.1iii). By Lemma 4, $x \in \chi_I$. Thus $\chi \subset \chi_I \cap \chi_J$.

Next, we show that χ_I is a translate of N. It is obvious that $\chi_I + N \subset \chi_I$, so it is enough to show that $x^1-x^2 \in N$ whenever $x^1,x^2 \in \chi_I$. Recall that v is an n-dimensional vector depending on the system (0.1). To the I-subsystem of (0.1) there corresponds likewise an $|I|$-dimensional vector v^I. But since $\chi \subset \chi_I$, (2.5) shows that v^I is the restriction to I of v, i.e., $v_i^I = v_i$ for all $i \in I$. By (2.6ii), for all $i \in I$, $x^1,x^2 \in \chi_I$,

$$x^1-x^2 = (\mathrm{Proj}_{C_i}(x^1)-v_i) - (\mathrm{Proj}_{C_i}(x^2)-v_i) \in \mathrm{bd}(C_i) - \mathrm{bd}(C_i).$$

Thus $\langle x^1-x^2,v_i\rangle = 0$ for all $i \in I$, and $x^1-x^2 \in N$. □

For an inconsistent system (0.1), the proximal mapping P_A for T_A has no fixed points. However, if x is a least-square solution, $\hat{x} = (x,\ldots,x)$, and $t \geq 0$, then according to the following lemma, the algorithm (0.2) initiated with $\hat{x}_0 = \hat{x}$, $\hat{y}_0 = 0$ will generate $\hat{x}_n = \hat{x}$, $\hat{y}_n = -n\hat{v}$. Since P_A is nonexpansive, this shows for <u>any</u> sequences $\hat{x}_n,\hat{y}_n$ generated by the algorithm, that

$$|(\hat{x}_n+\hat{y}_n) - (\hat{x}-n\hat{v})| \text{ is nonincreasing.} \tag{2.7}$$

Lemma 6. If $x \in \chi$, $\hat{x} = (x,\ldots,x)$, and $t \geq 0$, then $P_A(\hat{x}-t\hat{v}) = \hat{x} - (t+1)\hat{v}$.

Proof. From (1.12) and (1.13), $P_A(\hat{x}-t\hat{v}) = \hat{x}'_A + \hat{y}'_B$, where $\hat{x}' = \mathrm{Proj}_{\hat{C}}(\hat{x}-t\hat{v})$ and $\hat{y}' = \hat{x}-t\hat{v}-\hat{x}'$. Let $w_j = \mathrm{Proj}_{C_j}(x)$, $\hat{w} = (w_1,\ldots,w_n)$. Then $\hat{v} = \hat{w}_B = \hat{w}-\hat{w}_A = \hat{w}-\hat{x}$, so $\hat{x}-t\hat{v} = \hat{w}-(t+1)\hat{v} \in \hat{w} + T(\hat{w})$ (2.1i,ii). Thus, $\hat{x}' = \hat{w}$ and $\hat{y}' = \hat{x}-t\hat{v}-\hat{w} = -(t+1)\hat{v}$. Finally, $P_A(\hat{x}-t\hat{v}) = \hat{x}'_A + \hat{y}'_B = \hat{w}_A - (t+1)\hat{v} = \hat{x} - (t+1)\hat{v}$. (The last equality uses (2.4ii)). □

Lemma 7. Any sequence $\hat{x}'_k$ generated by the algorithm is bounded and its cluster points are minimally dispersed.

Proof. Recall the notation (1.11) for sequences generated by the algorithm. As in (1.13), let P_A denote the proximal mapping associated with T_A and $Q_A = I-P_A$. Then

$$\begin{aligned} Q_A(\hat{x}_k+\hat{y}_k) &= (\hat{x}_k+\hat{y}_k) - (\hat{x}_{k+1}+\hat{y}_{k+1}) && \text{(by (1.13)} \\ &= (\hat{x}'_k)_B + (\hat{y}'_k)_A. && \text{(by 1.12i, ii)} \end{aligned} \tag{2.8}$$

By Lemma 1, we know that $Q_A(\hat{x}_k+\hat{y}_k)$ converges to the element $\hat{v}$ of minimum norm in the range of T_A. Since $\hat{v} \in B$, this is equivalent to

$$(\hat{x}'_k)_B \to \hat{v} \quad \text{and} \quad (\hat{y}'_k)_A \to 0. \tag{2.9}$$

Fix $x \in \chi$ and let $\hat{x} = (x,\ldots,x)$. By (2.7), $|(\hat{x}_{k+1}+\hat{y}_{k+1})-(\hat{x}-k\hat{v})|$ is nonincreasing. But

$$|(\hat{x}_{k+1}+\hat{y}_{k+1})-(\hat{x}-k\hat{v})| = |((x_k')_A + (y_k')_B) - (\hat{x}-k\hat{v})| \quad \text{(by 1.12i)}$$

$$\geq |(\hat{x}_k')_A - \hat{x}|$$

since the norm of a vector dominates the norm of its A-component. Thus $(\hat{x}_k')_A$ is bounded and by (2.9) $(\hat{x}_k')$ is also bounded. The sequence $(\hat{x}_k')$ thus has cluster points and by (2.9), all its cluster points have B-component equal to $\hat{v}$. From the algorithm (0.2), we know that $\hat{x}_k' \in \hat{C}$ for all k, so the cluster points also lie in $\hat{C}$ and are therefore minimally dispersed (2.1i). □

In light of Lemma 7, the cluster points of $(\hat{x}_k')$ are of the form $\hat{a}+\hat{v}$ $(\hat{a} \in A)$. Suppose $\hat{a}_1+\hat{v}$ and $\hat{a}_2+\hat{v}$ to be two distinct such cluster points ($\hat{a}_1 = (a_1,\ldots,a_1)$, $\hat{a}_2 = (a_2,\ldots,a_2)$). It is impossible that κ exist such that

$$\langle x_\kappa - x_{\kappa+1}, a_1-a_2\rangle < 0 \quad \text{and} \quad \langle y_{\kappa j}', a_1-a_2\rangle \geq 0. \qquad (2.10)$$

This is impossible because (2.8) implies $\hat{x}_\kappa - \hat{x}_{\kappa+1} = (\hat{y}_\kappa')_A$, or

$$x_\kappa - x_{\kappa+1} = \frac{1}{n} \Sigma_{j=1}^n y_{\kappa j}'.$$

In the proof of the following lemma, it will be shown that if two distinct cluster points did exist then it would be possible to produce a contradiction by demonstrating the existence of κ satisfying (2.10).

<u>Lemma 8</u>. Any sequence $\hat{x}_k'$ generated by the algorithm has at most one cluster point, and hence converges.

Proof. Suppose the sequence $(\hat{x}_k')$ has two distinct cluster points, $\hat{a}_1 + \hat{v}$ and $\hat{a}_2 + \hat{v}$ ($\hat{a}_1 = (a_1,\ldots,a_1)$ and $\hat{a}_2 = (a_2,\ldots,a_2)$). Since the set of cluster points is compact (Lemma 7), we may assume a_1 and a_2 to be picked so as to maximize $|a_1-a_2|$. By Lemma 4, a_1 and a_2 belong to χ. Of course, $\hat{x}_{k+1} = (\hat{x}_k')_A$ by (1.12i), so the A-components of the cluster points of $(\hat{x}_k')$ are cluster points of $(\hat{x}_k)$. In particular, a_1 and a_2 are cluster points of the sequence (x_k).

By Lemma 5, $a_1, a_2 \in \chi_J \subset C_j$ for all $j \in J$. Let us partition J into subsets $J = J_- \cup J_0$, where

$$J_- = \{j \in J: \langle a_1, u_j \rangle < b_j\}$$

$$J_0 = \{j \in J: \langle a_1, u_j \rangle = b_j\}.$$

Define $t = |a_1-a_2|$, $w = (a_1-a_2)/t$, and for each $\varepsilon > 0$,

$$U_\varepsilon = B(a_2;t+\varepsilon) \cap \{x: \langle a_1-x, w\rangle < \varepsilon\}$$

(where $B(\cdot;\cdot)$ denote the open ball with given center and radius). The sets U_ε are open neighborhoods of a_1 whose diameters shrink to zero as $\varepsilon \downarrow 0$.

By (2.6), v_j is a positive multiple of u_j for all $j \in I$. Now, $x_{kj}' = \mathrm{Proj}_{C_j}(x_k + y_{kj})$ (0.2) so $y_{kj}' = x_k + y_{kj} - x_{kj}'$ is a nonnegative multiple of u_j. Since $v_j \neq 0$, this implies that y_{kj}' is also a multiple of v_j. In particular,

$$y_{kj}' \in N^\perp \quad (j \in I,\ k = 0,1,2,\ldots) \tag{2.11}$$

Choose ε small enough so

$$U_\varepsilon + B(0;\varepsilon) \subset \mathrm{int}(C_j) \text{ for all } j \in J_- \text{ and } 0 < \varepsilon < t. \quad (2.12)$$

Choose K large enough so that

$$\text{i.} \quad |(\hat{x}_k')_B - \hat{v}| < \varepsilon \quad \text{for } k \geq K \quad (2.13)$$

$$\text{ii.} \quad x_k \in B(a_2;t+\varepsilon) \quad \text{for } k \geq K$$

$$\text{iii.} \quad x_K \notin U_\varepsilon$$

It is possible to satisfy (i) by (2.9). If it were not possible to satisfy (ii) there would be cluster points of x_k outside of $B(a_2;t+\varepsilon)$, contradicting the maximality of $|a_1-a_2|$. For (iii), simply note that a_2 is a cluster point of (x_k) and $a_2 \notin U_\varepsilon$.

Since $x_K \notin U_\varepsilon$ and a_1 is a cluster point of (x_k), there is a smallest $\kappa \geq K$ such that

$$\text{i.} \quad \langle a_1 - x_{\kappa+1}, w\rangle < \varepsilon \quad (2.14)$$

$$\text{ii.} \quad \langle a_1 - x_\kappa, w\rangle \geq \varepsilon.$$

Subtracting, we get

$$\langle x_\kappa - x_{\kappa+1}, w\rangle < 0. \quad (2.15)$$

By (2.13ii) and (2.14i),

$$x_{\kappa+1} \in U_\varepsilon. \quad (2.16)$$

Now, $(\hat{x}_\kappa')_B = \hat{x}_\kappa' - (\hat{x}_\kappa')_A = \hat{x}_\kappa' - \hat{x}_{\kappa+1}$ (i.12i) so from (2.13i), we have

$$\varepsilon > |\hat{x}'_\kappa - \hat{x}_{\kappa+1} - \hat{v}| \geq |x'_{\kappa j} - x_{\kappa+1} - v_j| \qquad (j = 1,\ldots,n).$$

In particular,

$$|x'_{\kappa j} - x_{\kappa+1}| < \varepsilon \quad \text{for all } j \in J,$$

whence, by (2.12) and (2.16),

$$x'_{\kappa j} \in \operatorname{int}(C_j) \quad \text{for all } j \in J_-. \tag{2.17}$$

But $y'_{\kappa j} \in T_j(x'_{\kappa j})$ $(j = 1,\ldots,n)$ by (0.2), so (2.17) implies

$$y'_{\kappa j} = 0, \quad \text{for all } j \in J_-. \tag{2.18}$$

For $j \in J_0$, we have

$$\langle a_1, u_j\rangle = b_j \quad (\text{by definition of } J_0)$$

$$\langle a_2, u_j\rangle \leq b_j \quad (a_2 \in \chi_J \text{ by Lemma 5}).$$

Subtracting,

$$\langle w, u_j\rangle \geq 0 \quad \text{for all } j \in J_0.$$

Now, $y'_{\kappa j}$ is a nonnegative multiple of u_j (0.2) so

$$\langle w, y'_{\kappa j}\rangle \geq 0 \quad \text{for all } j \in J_0. \tag{2.19}$$

By Lemma 5, $w = (a_1 - a_2)/t \in N$ so (2.11) gives

$$\langle w, y'_{\kappa j}\rangle = 0 \quad \text{for all } j \in I. \tag{2.20}$$

Combining (2.18), (2.19), and (2.20), we have

$$\langle w, y'_{\kappa j}\rangle \geq 0 \quad \text{for all } j.$$

Combining this with (2.15), we see that (2.10) holds, giving the desired contradiction. □

We now summarize our observations so far about convergence of the algorithm in

Theorem 1. Let sequences (x_k), (y_{ki}), (x'_{ki}), (y'_{ki}) be produced by the algorithm (0.2). The sequences (x_k) and (x'_{ki}) $(i = 1,\ldots,n)$ converge to limits

$$\text{i.} \qquad x_k \to x_\infty \tag{2.21}$$

$$\text{ii.} \qquad x'_{ki} \to x'_{\infty i}, \quad i = 1,\ldots,n,$$

where

$$\text{i. } \hat{x}'_\infty = (x'_{\infty 1},\ldots,x'_{\infty n}) \text{ is minimally dispersed} \tag{2.22}$$

$$\text{ii.} \qquad x_\infty = \frac{1}{n}\,\Sigma_{i=1}^{n}\, x'_{\infty i}.$$

Thus also, x_∞ is a least square solution, $x'_{\infty i} = \mathrm{proj}_{C_i}(x_\infty)$, and $(x'_{\infty 1}-x_\infty,\ldots,x'_{\infty n}-x_\infty) = \hat{v} =$ the element of smallest norm in range(T_A). For the sequences (y_{ki}), (y'_{ki}) one has

$$\lim_{k\to\infty} (y_{k1},\ldots,y_{kn})/k = \lim_{k\to\infty} (y'_{k1},\ldots,y'_{kn})/k = -\hat{v}. \tag{2.23}$$

Proof. By Lemma 8, (2.12ii) holds. But $x_{k+1} = \frac{1}{n}\,\Sigma_{i=1}^{n}\, x'_{ki}$ and $x'_{ki} \in C_i$ by (0.2) so (2.21ii) implies (2.21i) and (2.22ii). By Lemma 7, $\hat{x}'_\infty$ is minimally dispersed. From Lemmas 3 and 4, $x_\infty \in \chi$, $x'_{\infty i} = \mathrm{Proj}_{C_i}(x_\infty)$, and $(\hat{x}'_\infty)_B = \hat{v}$.

Regarding the sequences $(\hat{y}_k)$ and $(\hat{y}'_k)$, we have

$$\lim_{k\to\infty} \hat{y}_k' - \hat{y}_{k+1} = \lim_{k\to\infty} \hat{y}_k' - (\hat{y}_k')_B \qquad \text{(by 1.12i)}$$

$$= \lim_{k\to\infty} (\hat{y}_k')_A = 0 \qquad \text{(by 2.9)}.$$

Thus, if either of the two sequences $(\hat{y}_k')/k$ and $(\hat{y}_k/k)$ has a limit so does the other and the two limits must coincide. By (1.5),

$$\lim_{k\to\infty} (\hat{x}_k + \hat{y}_k)/k = -\hat{v}.$$

Since $\hat{x}_k$ converges,

$$\lim_{k\to\infty} \hat{y}_k/k = -\hat{v}$$

and (2.23) holds. □

III. Convergence Rate

According to Theorem 1, any sequence (x_k) generated by the algorithm converges to a least-square solution. In this section, it will be shown that $\mathrm{dist}(x_k,\chi) \le c\mu^k$ for some $c > 0$ and $0 \le \mu < 1$.

If we let π_A denote orthogonal projection onto A, then since π_A is nonexpansive,

$$\mathrm{dist}(\hat{x}_k+\hat{y}_k, T_A^{-1}(\hat{v})) \ge \mathrm{dist}(\hat{x}_k, \pi_A(T_A^{-1}(\hat{v})))$$

$$= \mathrm{dist}(\hat{x}_k, \{\hat{w}_A : \hat{w} \in \hat{C},\ \hat{w}_B = \hat{v}\}) \qquad \text{(by 1.10)}$$

$$= \mathrm{dist}(\hat{x}_k, \{(w,\ldots,w) : w \in \chi\}) \qquad \text{(by Lemma 4)}$$

$$= \sqrt{n}\ \mathrm{dist}(x_k,\chi), \qquad (\hat{x}_k = (x_k,\ldots,x_k))$$

so it is enough to show that

$$\operatorname{dist}(\hat{x}_k+\hat{y}_k, T_A^{-1}(\hat{v})) \to 0, \text{ linearly} \qquad (3.1)$$
$$\text{with modulus } \mu \ (0 \le \mu < 1).$$

The sequence $(\hat{x}_k+\hat{y}_k)$ is generated by applying the proximal point algorithm to T_A. So to establish (3.1) it suffices, by Lemma 2, to show that a constant κ exists such that for all $\hat{w}$ in a neighborhood of $\hat{v}$, $\hat{w} \in T_A(\hat{z})$ implies $\operatorname{dist}(\hat{z}, T_A^{-1}(\hat{v})) \le \kappa|\hat{w}-\hat{v}|$. By definition of T_A, this is equivalent to:

$$\text{there is } \delta > 0 \text{ such that if} \qquad (3.2)$$
$$\hat{y} \in T(\hat{x}) \quad \text{and} \quad |\hat{x}_B+\hat{y}_A-\hat{v}| \le \delta$$
$$\text{then}$$
$$\operatorname{dist}(\hat{x}_A+\hat{y}_B, T_A^{-1}(\hat{v})) \le \kappa\varepsilon.$$

Our strategy to establish (3.2) is to show $\kappa \ge 0$ exists such that for any $\hat{x}$ and $\hat{y}$ with $\hat{y} \in T(\hat{x})$ and $|\hat{x}_B+\hat{y}_A-\hat{v}|$ sufficiently small, there exist $\hat{x}^*$ and $\hat{y}^*$ such that

i. $\qquad \hat{y}^* \in T(\hat{x}^*) \qquad (3.3)$

ii. $\qquad |\hat{y}-\hat{y}^*| \le \frac{\kappa}{2}\,|\hat{x}_B+\hat{y}_A-\hat{v}|$

iii. $\qquad \hat{y}_A^* = 0$

iv. $\qquad \hat{x}^*$ is minimally dispersed and

$$|\hat{x}^*-\hat{x}| \le \frac{\kappa}{2}\,|\hat{x}_B+\hat{y}_A-\hat{v}|.$$

Once this is done, $\hat{v} = \hat{x}_B^* + \hat{y}_A^* \in T_A(\hat{x}_A^* + \hat{y}_B^*)$ so

$$\operatorname{dist}(\hat{x}_A+\hat{y}_B, T_A^{-1}(\hat{v})) \le |(\hat{x}_A+\hat{y}_B)-(\hat{x}_A^*+\hat{y}_B^*)|$$

$$\le |\hat{x}-\hat{x}^*| + |\hat{y}-\hat{y}^*| \le \kappa|\hat{x}_B+\hat{y}_A-\hat{v}|,$$

as desired.

The following fact, derived easily from [40, Theorem 1], will play an essential role:

Lemma 9 [40]. Let ϕ: $R^d \rightrightarrows R^e$ be a polyhedral convex multifunction (i.e., one whose graph is a polyhedral convex set). There is a constant $c \ge 0$ such that for all $y, y' \in R^e$, if $\phi^{-1}(y') \ne \emptyset$,

$$\phi^{-1}(y) \subset \phi^{-1}(y') + c|y-y'|\bar{B},$$

where $\bar{B}$ is the closed unit ball in R^d.

If ϕ: $R^d \to R \cup \{\infty\}$ is a polyhedral convex function (i.e., one whose epigraph is a polyhedral convex set), let $\phi_\alpha = \{x \in R^d: \phi(x) \le \alpha\}$. Lemma 9 may be applied to the multifunction $x \mapsto \{y \in R: y \ge \phi(x)\}$ to obtain

Lemma 10 (from [40, Theorem 1]). Let ϕ be a polyhedral convex function on R^d. There exists a constant $c \ge 0$ such that for all α, $\alpha' \in R$, if $\phi_{\alpha'} \ne \emptyset$,

$$\phi_\alpha \subset \phi_{\alpha'} + c|\alpha-\alpha'|\bar{B}.$$

For $\hat{x} = (x_1,\dots,x_n) \in H$, consider the function

$$\phi(\hat{x}) := \max_j\{|x_j - \tfrac{1}{n}\Sigma x_i - v_j|\} + \psi_{\hat{C}}(\hat{x})$$

(where $\psi_{\hat{C}}(\hat{x}) = 0$ if $\hat{x} \in \hat{C}$, $\psi_{\hat{C}}(\hat{x}) = \infty$ otherwise). ϕ achieves its minimum value of zero on the set $\phi^{-1}(0) = \{\hat{x} \in \hat{C}: \hat{x}_B = \hat{v}\}$ of all $\hat{x} \in H$ that are minimally dispersed. ϕ is also polyhedral convex, so Lemma 10 (applied with $\alpha' = 0$, $\alpha = \phi(\hat{x})$) implies the existence of $\kappa_1 > 0$ such that

$$\hat{x} \in \phi^{-1}(0) + \kappa_1 \phi(\hat{x}) \bar{B} \quad \text{for all } \hat{x} \in H,$$

where $\bar{B}$ is the closed unit ball in H. Note that the constant κ_1 depends only on the function ϕ, which is determined by the sets C_i, $i = 1,\ldots,n$. For later reference, we summarize our observations as:

<u>Lemma 11</u>. There is a constant $\kappa_1 > 0$ (depending only on $C_1,\ldots,C_n$) such that if $\hat{x} \in \hat{C}$ then there exists $\hat{x}^* = (x_1^*,\ldots,x_n^*) \in H$ which is minimally dispersed and which satisfies $|\hat{x}^*-\hat{x}| \le \kappa_1 |\hat{x}_B - \hat{v}|$.

<u>Proof</u>. By the above choice of κ_1, there exists $\hat{x}^* \in \phi^{-1}(0)$ such that $|\hat{x}^*-\hat{x}| \le \kappa_1 \phi(\hat{x})$. The conclusion follows since $\phi(\hat{x}) = \max_j \{|x_j - \frac{1}{n} \Sigma x_i - v_j|\} \le |\hat{x}_B - \hat{v}|$. □

A convex set $K \subset R^d$ is a <u>cone</u> if $tK \subset K$ for all $t \ge 0$. K is <u>pointed</u> if for some ℓ, $\langle k,\ell\rangle > 0$ for all $0 \ne k \in K$. (Under this definition, note that $K = \{0\}$ is pointed.)

<u>Lemma 12</u>. Let $K \subset R^d$ be a closed pointed convex cone. There is a constant c such that for any $k_1,\ldots,k_s \in K$,

$$|k_i| \le c|k_1 + \ldots + k_s|, \quad i = 1,\ldots,s.$$

Proof. If $K = \{0\}$, we can take $c = 0$. Otherwise choose $\ell \in R^d$, $|\ell| = 1$, such that $\langle k,\ell\rangle > 0$ for all $0 \neq k \in K$. Let $m = \min\{\langle k,\ell\rangle: |k| = 1, k \in K\}$. Then $|k_1 + \ldots + k_s| \geq \langle \ell, k_1 + \ldots + k_s\rangle \geq \langle \ell, k_i\rangle \geq m|k_i|$, $i = 1,\ldots,s$, so we can take $c = 1/m$. □

By Lemma 5, $\chi \subset C_j$ for all $j \in J$. Let us partition J into three subsets: $J = J_1 \cup J_2 \cup J_3$, where

$$J_1 = \{j: \chi \subset \mathrm{bd}(C_j)\}$$

$$J_2 = \{j: \chi \cap \mathrm{bd}(C_j) \neq \emptyset \text{ and } \chi \cap \mathrm{int}(C_j) \neq \emptyset\}$$

$$J_3 = \{j: \chi \subset \mathrm{int}(C_j)\}.$$

Lemma 13. There are constants $\kappa_2 \geq 0$ and $\delta > 0$ (depending only on $C_1,\ldots,C_n$) such that if $\hat{y} \in T(\hat{x})$ and $|\hat{x}_B+\hat{y}_A-\hat{v}| \leq \delta$ then

$$|y_j| \leq \kappa_2|\hat{x}_B+\hat{y}_A-\hat{v}| \quad \text{for all } j \in J_2.$$

Proof. If $J_2 = \emptyset$ there is nothing to show, so assume $J_2 \neq \emptyset$. Choose $z \in \chi$ and $\rho > 0$ sufficiently small so $\bar{B}(z;\rho) \subset C_j$ for all $j \in J_2$ and so $\chi + \bar{B}(0;\rho) \subset C_j$ for all $j \in J_3$ (the latter is possible because χ is polyhedral by Lemma 5).

Let $\delta = \frac{\rho}{2\kappa_1}$ and suppose that $\hat{y} \in T(\hat{x})$ and $|\hat{x}_B+\hat{y}_A-\hat{v}| \leq \delta$. By Lemma 11, there is a minimally dispersed $\hat{x}^* = (x_1^*,\ldots,x_n^*)$ such that $|\hat{x}^*-\hat{x}| \leq \kappa_1|\hat{x}_B+\hat{y}_A-\hat{v}|$.

For $w \in R^d$, denote by w_χ the orthogonal projection of w onto the linear subspace spanned by $\chi-\chi = \{x-x': x,x' \in \chi\}$. Apply Lemma 12 to each pointed convex cone that is generated by some subset of $\{(u_j)_\chi : j \in J_2\}$ and let c_1 be the maximum of the constants so obtained. Since $(u_j)_\chi \neq 0$ for $j \in J_2$ we may define

$$c_2 = \max\{\frac{|u_j|}{|(u_j)_\chi|} : j \in J_2\}.$$

Let $x^* = \frac{1}{n}\Sigma_{i=1}^n x_i^*$. For $j = 1,\ldots,n$, $x_j^* = x^* + v_j$ because $\hat{x}^*$ is minimally dispersed. Thus $x_j^* = x^*$ for $j \in J$. For $j \in J$ we then have

$$|x_j - x^*| = |x_j - x_j^*| \leq |\hat{x} - \hat{x}^*|$$

$$\leq \kappa_1 |\hat{x}_B + \hat{y}_A - \hat{v}| \leq \frac{\rho}{2} \quad (j \in J)$$

Since $\hat{y} \in T(\hat{x})$, $y_j \neq 0$ implies $x_j \in bd(C_j)$. Thus

$$\langle x_j - z, y_j \rangle = |y_j| \mathrm{dist}(z, bd(C_j)) \quad (\text{for } j = 1,\ldots,n)$$

$$\geq |y_j|\rho \quad (\text{for } j \in J_2).$$

For all $j \in J_2$, we then have

$$\langle x^* - z, (y_j)_\chi \rangle = \langle x^* - z, y_j \rangle \quad (\text{since } x^*, z \in \chi)$$

$$= \langle x_j - z, y_j \rangle + \langle x^* - x_j, y_j \rangle$$

$$\geq |y_j|\rho - \frac{1}{2}|y_j|\rho = \frac{1}{2}|y_j|\rho$$

$$> 0 \text{ if } y_j \neq 0.$$

For $j \in J_2$, $y_j = 0$ iff $(y_j)_\chi = 0$. Thus $\langle x^* - z, (y_j)_\chi \rangle > 0$ for all $j \in J_2$ for which $(y_j)_\chi \neq 0$, showing that $\{(y_j)_\chi : j \in J_2\}$ generates a pointed cone. By choice of c_1,

$$|(y_j)_\chi| \le c_1 |\sum_{i\in J_2} (y_i)_\chi| \quad (\text{for } j \in J_2)$$

Next, we show that

$$(y_j)_\chi = 0 \quad \text{for } j \in I \cup J_1 \cup J_3.$$

The case $j \in J_1$ is evident. For $j \in I$, $v_j \neq 0$ so y_j is a multiple of v_j. By Lemma 5, this implies $(y_j)_\chi = 0$. For the remaining case of $j \in J_3$, we have $\mathrm{dist}(x_j, \chi) \le |x_j - x^*| \le \frac{\rho}{2}$. By choice of ρ, this implies

$$x_j \in \mathrm{int}(C_j), \quad \text{and hence } y_j = 0 \quad (\text{for } j \in J_3) \quad (3.4)$$

Finally, for all $j \in J_2$,

$$\begin{aligned} |y_j| &\le c_2 |(y_j)_\chi| \\ &\le c_1 c_2 |\sum_{i\in J_2} (y_i)_\chi| \\ &= c_1 c_2 |\Sigma_{i=1}^n (y_i)_\chi| = c_1 c_2 |(\Sigma_{i=1}^n y_i)_\chi| \\ &\le c_1 c_2 |\Sigma_{i=1}^n y_i|. \end{aligned}$$

Since $|\Sigma_{i=1}^n y_i| \le |\hat{y}_A| \le |\hat{x}_B + \hat{y}_A - \hat{v}|$, the conclusion of Lemma 13 holds with $\kappa_2 = c_1 c_2$. □

<u>Lemma 14</u>. Let $u_1, \ldots, u_r \in R^d$ be nonzero. There is a constant $c > 0$ such that for any $t_1, \ldots, t_r \ge 0$, there exist $s_1, \ldots, s_r \ge 0$ such that $s_1 u_1 + \ldots + s_r u_r = 0$ and $|t_i - s_i| \le c|t_1 u_1 + \ldots + t_r u_r|$, $i = 1, \ldots, r$.

<u>Proof</u>. Consider the polyhedral convex multifunction

ϕ: $R^r \rightrightarrows R^d$ defined by

$$\phi(t) = \begin{cases} \{t_1u_1 + \dots + t_ru_r\} & \text{if } t \geq 0 \\ \emptyset & \text{otherwise.} \end{cases}$$

Apply Lemma 9 to ϕ to obtain a Lipschitz constant c such that for all $t \geq 0$,

$$t \in \phi^{-1}(t_1u_1 + \dots + t_ru_r) \subset \phi^{-1}(0) + c|t_1u_1 + \dots + t_ru_r|\bar{B}.$$

Then, for any $t \geq 0$, there exists $s \in \phi^{-1}(0)$ with

$$|t-s| \leq c|t_1u_1 + \dots + t_ru_r|. \qquad \square$$

Lemma 15. There are $\kappa \geq 0$ and $\delta > 0$ (depending only on $C_1,\dots,C_n$ such that if $\hat{y} \in T(\hat{x})$ and $|\hat{x}_B+\hat{y}_A-\hat{v}| \leq \delta$ then there exist $\hat{x}^*$ and $\hat{y}^*$ such that

$$\begin{array}{ll} \text{i.} & \hat{y}^* \in T(\hat{x}^*) \qquad (3.5) \\ \text{ii.} & |\hat{y}-\hat{y}^*| \leq \kappa_3|\hat{x}_B+\hat{y}_A-\hat{v}| \\ \text{iii.} & y_1^* + \dots + y_n^* = 0 \\ \text{iv.} & \hat{x}^* \text{ is minimally dispersed and} \\ & |\hat{x}^*-\hat{x}| \leq \kappa_1|\hat{x}_B+\hat{y}_A-\hat{v}|. \end{array}$$

Proof. Let δ be as in Lemma 13 and suppose that $\hat{y} \in T(\hat{x})$ and $|\hat{x}_B+\hat{y}_A-\hat{v}| \leq \delta$. Choose $\hat{x}^*$ by Lemma 11 to satisfy (iv).

If $I \cup J_1 = \emptyset$, then set $\hat{y}^* = 0$. Then (i) and (iii) hold trivially. By Lemma 13 and (3.4), $|\hat{y}| \leq \sum_{j \in J_2} |y_j| \leq n\kappa_2|\hat{x}_B+\hat{y}_A-\hat{v}|$ so (ii) holds with $\kappa_3 = n\kappa_2$, and we are done.

Suppose $I \cup J_1 \neq \emptyset$. Let c be chosen as in Lemma 14 for the vectors $u_i/|u_i|$ $(i \in I \cup J_1)$. Let $t_i = |y_i|$ and let s_i be chosen as in Lemma 14, and define $y_i^* = s_i u_i/|u_i|$ $(i \in I \cup J_1)$. Then

$$\sum_{I\cup J_1} y_j^* = 0 \text{ and } |y_i - y_i^*| = |t_i - s_i| \le c\left|\sum_{I\cup J_1} y_j\right| \text{ (for } i \in I \cup J_1).$$

For $i \in J_2 \cup J_3$, define $y_j^* = 0$. Clearly (iii) holds.

For $i \in I \cup J_1$,

$$\begin{aligned} |y_i - y_i^*| &\le c\left|\sum_{I\cup J_1} y_j\right| \le c\left|\Sigma_{i=1}^n y_i\right| + c\left|\sum_{J_2} y_i\right| + c\left|\sum_{J_3} y_i\right| \\ &\le cn|\hat{y}_A| + cn\kappa_2|\hat{x}_B + \hat{y}_A - \hat{v}| + 0 \qquad \text{(by Lemma 13 and 3.4)} \\ &\le (cn + cn\kappa_2)|\hat{x}_B + \hat{y}_A - \hat{v}|. \end{aligned}$$

Also, for $i \in J_2$,

$$|y_i - y_i^*| = |y_i| \le \kappa_2|\hat{x}_B + \hat{y}_A - \hat{v}|$$

by Lemma 13. For $i \in J_3$ we have by (3.4),

$$|y_i - y_i^*| = |y_i| = 0.$$

Combining the above, we have $|y_i - y_i^*| \le \kappa_3|\hat{x}_B + \hat{y}_A - \hat{v}|$ for $i = 1,\ldots,n$ with $\kappa_3 = \max(\kappa_2, cn + cn\kappa_2)$ and (ii) holds.

It remains to show (i). For $i \in J_2 \cup J_3$, $y_i^* = 0$ so it is trivial that $y_i^* \in T_i(x_i^*)$. To show the same for $i \in I \cup J_1$, it suffices to show $x_i^* \in \mathrm{bd}(C_i)$ since y_i^* is by definition a nonnegative multiple of u_i. For $i \in I$, $0 \ne -v_i \in T_i(x_i^*)$ since $\hat{x}^*$ is minimally dispersed, so $x_i^* \in \mathrm{bd}(C_i)$. For all $i \in J_1$, $x_i^* = x^* + v_i = x^* \in \chi \subset \mathrm{bd}(C_i)$ since $\hat{x}^*$ is minimally dispersed and by definition of J_1. Thus (i) holds. □

Choosing $\kappa = \min\{2\kappa_1, 2\kappa_3\}$, (3.5) implies (3.3), which is all that was needed to prove linear convergence. In summary:

Theorem 2. Any sequence (x_k) generated by the algorithm (0.2) converges to the set of least square solutions at a linear rate, in the sense that $\text{dist}(x_k, \chi) \le c\mu^k$ (for some $c > 0$, $0 \le \mu < 1$) for all k sufficiently large.

Proof. By Lemma 2, (3.1) holds with $\mu = \kappa/\sqrt{1 + \kappa^2}$. $\square$

BIBLIOGRAPHY

1. S. Agmon, The relaxation method for linear inequalities, Canadian J. Math. 6:382-392 (1954).

2. J.-P. Aubin, I. Ekeland, Applied Nonlinear Analysis, John Wiley and Sons, N.Y. (1984).

3. A. Auslender, Méthodes numériques pour la résolution des problemes d'optimisation avec contraintes, Thèse, Faculté des Sciences de Grenoble, June 1969, Ch. VI.

4. A. Auslender, Optimisation, Méthodes Numériques, Masson, Paris, 1976.

5. R. E. Bruck and S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, Houston J. Math. 3:459-470 (1977).

6. Y. Censor, Row-action methods for huge and sparse systems and their applications, SIAM Review 23 (1981) 444-466.

7. Y. Censor, Finite series-expansion reconstruction methods, Proc. of the IEEE 71, no. 3 (1983) 409-419.

8. Y. Censor, P. P. B. Eggermont, D. Gordon, Numerische Mathematik 41 (1983) 83-92.

9. G. Cimmino, Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari, La Ricerca Scientifica XVI, Serie II, Anno IX, Vol. 1 (1938) 326-333.

10. F. Deutsch, Rate of convergence of the method of alternating projections, Proceedings of the Conference on "Parametric Optimization and Approximation," Oberwolfach, Germany, Oct. 1983, Birkhäuser, Boston, to appear.

11. P. Eggermont, G. Herman, A. Lent, Iterative algorithms for large partitioned systems, with applications to image reconstruction, Linear Algebra and its Applications 40 (1981) 37-67.

12. T. Elfving, Block-iterative methods for consistent and inconsistent linear equations, Numer. Math. 35, 1-12 (1980).

13. J. L. Goffin, "On the finite convergence of the relaxation method for solving systems of inequalities," Operations Research Center Report ORC 71-36, University of California at Berkeley, 1971.

14. J. L. Goffin, "On convergence rates of subgradient optimization methods," Mathematical Programming, 13, 1977, 329-347.

15. J. L. Goffin, "Nondifferentiable optimization and the relaxation method," in Nonsmooth Optimization, Editors C. Lemaréchal and R. Mifflin, IIASA Proceedings Series, Pergamon Press, 1978, 31-49.

16. J. L. Goffin, "Acceleration in the relaxation method for linear inequalities and subgradient optimization," Progress in nondifferentiable optimization, 29-59, IIASA Collaborative Proc. Ser. CP-82, 8, Internat. Inst. Appl. Systems Anal., Laxenburg, 1982

17. J. L. Goffin, "Chaos and order in relaxation methods with maximizing controls," Proceedings of the 6th Conference on Probability Theory, held at Brasov, Romania, September 1979, 67-76, Ed. Acad. R. S. Romania, Bucharest, 1981.

18. J. L. Goffin, "On the nonpolynomiality of the relaxation method for systems of linear inequalities," Mathematical Programming, 22 (1982), 93-103.

19. J. L. Goffin, "The relaxation method for solving systems of linear inequalities," Mathematics of Operations Research, 5, 1980, 388-414.

20. I. Halperin, The product of projection operators, Acta Sci. Math. (Szeged) 23 (1962) 96-99.

21. G. T. Herman, A. Lent, P. H. Lutz, Relaxation methods for image reconstruction, Comm. ACM 21 (1978) 152-158.

22. S. Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen, Bull. Acad. Polon. Sciences et Lettres, A, 355-357 (1937).

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

24. F. J. Luque, Asymptotic convergence analysis of the proximal point algorithm, to appear in SIAM J. Control Optim.

25. B. Martinet, Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox, C.R. Acad. Sci. Paris (Ser. A), 274:163-165 (1972).

26. G. J. Minty, On the maximal domain of a "monotone" function, Mich. Math. J. 8:135-137 (1961).

27. G. J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29:341-346 (1962).

28. J.-J. Moreau, Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France 93:273-299 (1965).

29. T. S. Motzkin and I. J. Schoenberg, The relaxation method for linear inequalities, Canadian J. Math. 6:393-404 (1954).

30. J. von Neumann, Functional Operators -- Vol. II. The Geometry of Orthogonal Spaces, Annals of Math. Studies #22, Princeton University Press, 1950.

31. S. V. Plotnikov, Cyclic projection on a system of convex sets with empty intersection (in Russian; English abstract in MR 85g:90094) Improper optimization problems, 60-66, Akad. Nauk SSSR, Ural. Nauchn. Tsentr. Inst. Mat. i Mekh., Sverdlovsk, 1982.

32. B. T. Polyak, Gradient methods for the minimization of functionals, U.S.S.R. Computational Mathematics and Mathematical Physics 3, no. 4 (1963) 864-878.

33. B. T. Polyak, Constrained minimization methods, U.S.S.R. Computational Mathematics and Mathematical Physics, 6, no. 5, 1-50 (1966).

34. S. Reich, On infinite products of resolvents, Rendiconti della Classe di Scienze fisiche, matematiche e naturali, Serie VIII, Vol. LXIII, fasc. 5, Accademia Nazionale dei Lincei, 1977.

35. R. T. Rockafellar, Convex Analysis, Princeton U.P., Princeton, N. J., 1970.

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

37. J. E. Spingarn, Partial inverse of a monotone operator, Appl. Math. Optim. 10:247-265 (1983).

38. J. E. Spingarn, A primal-dual projection method for solving systems of linear inequalities, Linear Algebra and its Applications 65:45-62 (1985).

39. K. Tanabe, Projection method for solving a singular system of linear equations and its applications, Numer. Math. 17 (1971) 203-214.

40. D. W. Walkup and R. J.-B. Wets, A Lipschitzian characterization of convex polyhedra, Proc. Amer. Math. Soc. 23 (1969) 167-173.

41. A. Björck, Methods for sparse linear least squares problems, Sparse Matrix Computations, J. Bunch and D. Rose (eds.), Academic Press, New York (1976).

G-37-614

FINAL REPORT

PROXIMAL POINT ALGORITHMS IN MATHEMATICAL PROGRAMMING

Submitted by

Jon Spingarn
School of Mathematics

Submitted to

National Science Foundation
1800 G Street, NW
Washington, DC 20550

Contracting through

Georgia Tech Research Corporation

February 1987

GEORGIA INSTITUTE OF TECHNOLOGY
A UNIT OF THE UNIVERSITY SYSTEM OF GEORGIA
SCHOOL OF MATHEMATICS
ATLANTA, GEORGIA 30332

FINAL REPORT

Proximal Point Algorithm in Mathematical Programming

Submitted by

Jon Spingarn
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332

Submitted to

National Science Foundation
1800 G Street, NW
Washington, DC 20550

Contracted through

Georgia Tech Research Corporation
Georgia Institute of Technology
Atlanta, Georgia 30332

February 1987

EXHIBIT VI-1

NATIONAL SCIENCE FOUNDATION
Washington, D.C. 20550

FINAL PROJECT REPORT
NSF FORM 98A

PLEASE READ INSTRUCTIONS ON REVERSE BEFORE COMPLETING

PART I–PROJECT IDENTIFICATION INFORMATION

1. Institution and Address	2. NSF Program	3. NSF Award Number
Georgia Institute of Tech. School of Mathematics Atlanta, GA 30332	Math. Sci. Applied Math.	DMS-8506712
	4. Award Period From 6/15/85 To 11/30/87	5. Cumulative Award Amount $33,089

6. Project Title

Proximal Point Algorithm in Mathematical Programming

PART II–SUMMARY OF COMPLETED PROJECT *(FOR PUBLIC USE)*

For a nonexpansive piecewise isometric mapping q on R^d having a fixed point, we have studied the iterates of $(I+q)/2$. Such iterates were shown to behave, eventually, as if q were actually an isometry. Under certain circumstances, a finite number of iterations are guaranteed to produce a fixed point. Examples dealing with systems of linear inequalities and with network flows have been examined. A new constructive proof has been presented that for any real anti-symmetric matrix A, there exists $x \geq 0$ such that $Ax \geq 0$ and $x+Ax > 0$.

Some properties of iterates of nonexpansive mappings on R^d and especially of nonexpansive piecewise isometries (foldings) have been explored. For a nonexpansive q, the set S if all cluster points of q-sequences $x, q(x), q^2(x), \ldots$ is nonempty only if q has a fixed point. S is closed, convex, $q|S$ is an isometry, and all q-sequences converge to S. A characterization was presented for the class of all foldings having the property that the set S absorbs every q-sequence after only finitely many iterations.

A solution approach has been proposed for a certain class of network equilibrium problems. The method of solution is a specialization to a network setting of the method of partial inverses. Application is discussed to the computation of economic equilibria.

PART III–TECHNICAL INFORMATION *(FOR PROGRAM MANAGEMENT USES)*

1. ITEM *(Check appropriate blocks)*	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM: Check (✓)	Approx. Date
a. Abstracts of Theses					
b. Publication Citations					
c. Data on Scientific Collaborators					
d. Information on Inventions					
e. Technical Description of Project and Results					
f. Other *(specify)*					

2. Principal Investigator/Project Director Name *(Typed)*	3. Principal Investigator/Project Director Signature	4. Date
Jon Spingarn		2/13/87

TABLE OF CONTENTS

I. Iterates of Nonexpansive Mappings

A. Summary

Working with Dr. Jim Lawrence (George Mason University), we have studied nonexpansive piecewise isometric mappings (foldings) and the behavior of sequences of iterates of such mappings. The results of our research to date are contained in [1] and [2].

If $q\colon R^d \to R^d$ is a nonexpansive mapping having a fixed point, it is known that any sequence

(1) $$x_{k+1} = (q(x_k) + x_k)/2$$

converges to a fixed point of q [8]. In [1], the principal goal was to exhibit a class of mappings, the nonexpansive piecewise isometric mappings, or "foldings", for which the behavior of the iteration (1) is especially favorable.

The iteration (1) is equivalent to the "proximal point algorithm" for finding a zero of a maximal monotone multifunction [9], [10], [11]. If T is maximal monotone on R^d, the proximal point algorithm generates a sequence

(2) $$x_{k+1} = p(x_k), \quad p = (I + T)^{-1}.$$

Furthermore, the mapping $q = 2p - I$ is nonexpansive so (1) and (2) are equivalent.

In [1], the iteration (1) was studied in detail for the case where q is a folding. It was shown that the

sequence (1) behaves, eventually, as if q were actually an isometry.

It was also shown in [1] that this fact has many surprising consequences, especially the finite termination of several algorithms. As examples, we described algorithms for solving systems of linear inequalities and network flow problems. We also obtained a constructive and simple new proof of the fact that for any antisymmetric real $n \times n$ matrix A, there exists $x \geq 0$ such that $Ax \geq 0$ and $x + Ax > 0$.

The paper [2] was motivated by a desire to determine whether or not the iterates

$$x_{k+1} = q(x_k) \tag{3}$$

of a folding q coincide, as do the iterates (1), after finitely many steps, with the iterates of some isometry. The answer turned out to be negative. This outcome is consistent with the fact that for a general nonexpansive mapping q (not necessarily a folding), the iterates of $(I+q)/2$ have long been known to behave more nicely than the iterates of q; for instance, any sequence (1) converges to a fixed point of q (if one exists) [8], but the same is not in general true for (3).

It was shown in [2] that much can still be said about the iterates of (3) when q is nonexpansive, and still more can be said if q is a folding. In [2], we first analyzed the structure of foldings. We showed that a folding induces a decomposition of the underlying space into finitely many polyhedral convex sets (the "folds" of q). If q has a fixed point, we showed that there is a unique fold whose interior contains a fixed point, and that the fixed point set is the intersection of this fold with an affine set.

In [2], we defined the "cluster set" consisting of all cluster points of q-sequences, and showed this set to be nonempty only if q possesses a fixed point. If q has a fixed point, we demonstrated that all q-sequences converge to S (though not necessarily to a particular point in S). Further, S is closed and convex, $q(S) = S$, and the restriction of q to S is an isometry.

We then turned to the question of characterizing all foldings having the property: for every x there exists k such that $q^k(x) \in S$. As a particularly illuminating example of a folding that fails to have this property, consider the folding $q: R^2 \to R^2$ defined by $q(x,y) = s(r(x,y))$, where r is a rotation about the origin by an angle that is an irrational multiple of pi and

$$s(x,y) = \begin{cases} (x,y) & \text{if } x < 1 \\ (2-x,y) & \text{if } x \geq 1. \end{cases}$$

It is easy to see that for any $(x,y) \in R^2$, the iterates $q^k(x,y)$ move arbitrarily close to the cluster set $S = \{(x,y): x^2 + y^2 \leq 1\}$ and that the restriction of q to S coincides with the isometry r. But it followed from our results in [2] that there do exist points (x,y) for which $q^k(x,y) \notin S$ for all k. Further, we presented in [2] a characterization of the class of foldings q whose cluster set absorbs all q-sequences.

B. The Proximal Iterates of a Folding -- Summary of the Main Results from [1]

A multifunction $r: R^d \rightrightarrows R^d$ assigns a subset r(x) of R^d to each $x \in R^d$. The set of x such that r(x) is nonempty is dom(r), the domain of r. We may identify r with its graph, namely the set $\{(x,y): y \in r(x)\}$. For the sake of simplicity, we will use the same symbol to represent a multifunction and its graph. For example, the notations $y \in r(x)$ and $(x,y) \in r$ will be used interchangeably. If for all (x,y) and (x',y') in r

i. $|y-y'| \le |x-x'|$, then r is nonexpansive

ii. $|y-y'|^2 \le |x-x'|^2 - |(x-y) - (x'-y')|^2$, then r is proximal

iii. $\langle x-x', y-y'\rangle \ge 0$, then r is monotone.

If r is nonexpansive, then r is maximal nonexpansive if (the graph of) r is not properly contained in (the graph of) another nonexpansive r'. In an identical manner, we can define maximal proximal and maximal monotone.

The mapping $(x,y) \overset{\alpha}{\mapsto} (x,2y-x)$ induces a one-to-one correspondence of the class of proximal multifunctions onto the class of nonexpansive multifunctions. By this we mean that the image under α of (the graph of) a proximal multifunction is (the graph of) a nonexpansive multifunction and the image under α^{-1} of a nonexpansive multifunction is proximal. Likewise, $(x,y) \overset{\beta}{\mapsto} (x+y,x-y)$ carries monotone onto nonexpansive multifunctions. Of course, the composition $(x,y) \overset{\alpha^{-1}\beta}{\mapsto} (x+y,x)$ carries monotone onto proximal multifunctions.

In the following, we take q, p, and T to be corresponding nonexpansive, proximal, and monotone multifunctions i.e., $q = \alpha(p) = \beta(T)$. From the definitions, it is apparent that

$$p = (I+q)/2 = (I+T)^{-1}.$$

The correspondences α and β clearly preserve maximality. Thus q is maximal nonexpansive iff p is maximal proximal iff T is maximal monotone. Since $\beta(x,0) = \alpha(x,x) = (x,x)$, the fixed point set for p, the fixed point set for q, and the set of zeros of T coincide.

A nonexpansive multifunction q is obviously single-valued on its domain. A theorem of Kirszbraun [12] states that q is maximal if, and only if, $\mathrm{dom}(q) = R^d$. Thus "maximal nonexpansive" and "nonexpansive function on R^d" have the same meaning. Since $\mathrm{dom}(p) = \mathrm{dom}(q)$, the following are equivalent:

i. q is a nonexpansive function: $R^d \to R^d$

ii. p is a proximal function: $R^d \to R^d$

iii. T is a maximal monotone multifunction: $R^d \rightrightarrows R^d$.

If these equivalent conditions hold, it follows by a result of Krasnoselski [8] that for any x, the sequence

(4) $((I+q)/2)^n(x) = p^n(x) = ((I+T)^{-1})^n, \quad n = 0,1,2,\ldots$

converges to a fixed point of p and q (a zero of T) if one exists. Specifically, Krasnoselski's result asserts that the iterates of $(I+q)/2$ converge to a fixed point of q for any nonexpansive function q having a fixed point.

Before continuing, some historical remarks are in order. In the literature, "proximal" mappings have often been called "firmly nonexpansive" or the "resolvent of the monotone operator T". The correspondence between nonexpansive and monotone multifunctions was observed by Minty [13] who exploited it to show that for T maximal monotone on a Hilbert space, $(I+T)^{-1}$ is a function defined on the whole space. The term "proximal mapping" was first introduced by Moreau [10] to describe the function

$$x \mapsto \arg_z \min\{f(z) + \tfrac{1}{2}|z-x|^2\},$$

where f is lower semicontinuous convex. Moreau's "proximal mapping" is actually $(I+\partial f)^{-1}$, ∂f being the subdifferential of f (∂f is maximal monotone [14]). It was observed by Martinet [9] that Moreau's "proximal mapping" is in fact a "proximal function" as we have defined the term here, and that this implies convergence of its iterates. The correspondence between the class of proximal multifunctions and the class of nonexpansive multifunctions has also been observed before.

The iteration (4) has also been called the proximal point algorithm, especially when the emphasis has been placed on convergence to a zero of T. This name was introduced by Rockafellar [11], [15], in his study of algorithms for the solution of convex programming problems.

In [1], we investigated the behavior of the proximal point algorithm on nonexpansive mappings that are piecewise isometries. We call these mappings "foldings". To define this notion precisely, suppose that q is a nonexpansive function on R^d and let I denote the collection of all convex sets $K \subset R^d$ such that the restriction $q|K$ is an isometry. Let I be partially ordered by set inclusion. Every singleton trivially belongs to I, and I is closed under chain unions, so every point of R^d is contained in a maximal element of I. If $K \in I$ and $L \in I$ and $\mathrm{int}(K) \cap \mathrm{int}(L) \neq \emptyset$ then $\mathrm{conv}(K \cup L) \in I$. Thus, if K and L are distinct maximal elements of I, then $\mathrm{int}(K) \cap \mathrm{int}(L) = \emptyset$. If q has only a finite number of folds, then q is a folding and the maximal elements of I are the folds of q. If q is a folding, it is not hard to see that each fold is closed, polyhedral, and has nonempty interior.

The class of foldings is closed under composition.

For suppose q and q' are foldings with (locally finite) folds K_i $(i \in I)$ and K'_j $(j \in J)$, respectively. Then $q' \circ q$ is an isometry on each of the convex sets

$$L_{ij} := K_i \cap q^{-1}(q(K_i) \cap K'_j)$$

and these sets L_{ij} are locally finite. Thus $q' \circ q$ is a folding (although the sets L_{ij} are not necessarily its folds).

Via the correspondences α and β, composition of non-expansive functions induces corresponding operators on the classes of proximal functions and maximal monotone multifunctions. Thus, if p_1 and p_2 are proximal functions and T_1 and T_2 are maximal monotone multifunctions, it is natural to define two new operations:

$$p_1 * p_2 = \alpha^{-1}(\alpha(p_1) \circ \alpha(p_2))$$

and

$$T_1 \circledast T_2 = \beta^{-1}(\beta(T_1) \circ \beta(T_2)).$$

Since the class of foldings is closed under composition, it follows that the class of proximal mappings that correspond to foldings is closed under *, and the class of maximal monotone multifunctions that correspond to foldings is closed under $\circledast$. Since composition is associative, the operations * and $\circledast$ are also associative.

In [1], we have investigated the behavior of sequences $x_{n+1} = p(x_n)$, where p is the proximal function corresponding to a folding $q = \alpha(p) = 2p-I$.

An observation will simplify the discussion. Suppose $\bar{x}$ is a fixed point for q (for p). Since folds are closed, there is $r > 0$ such that the open ball $B(\bar{x};r)$ is contained in the finite union of all the folds containing $\bar{x}$. For all x in $B(0;r)$, the line segment $[\bar{x},\bar{x}+x]$ is entirely contained in each fold that contains both $\bar{x}$ and $\bar{x}+x$ since folds are convex. It follows that for all x in $B(0;r)$ the restriction of q to $[\bar{x},\bar{x}+x]$ is an isometry. Defining $\hat{q}(x) = q(\bar{x}+x) - \bar{x}$ and $\hat{p}(x) = p(\bar{x}+x) - \bar{x}$, we then have $\hat{q}(0) = \hat{p}(0) = 0$, $\hat{q}$ and $\hat{p}$ are positively homogeneous on $B(0;r)$, and $\hat{q} = \alpha(\hat{p})$. Now, $\hat{q}$ and $\hat{p}$ are merely the translation of q and p for which the fixed point at $\bar{x}$ is displaced to the origin. Any statement about the behavior of q or p near $\bar{x}$ translates into a statement about $\hat{q}$ or $\hat{p}$ near 0. For this reason, we lose no generality if in discussing the behavior of q (or p) near a fixed point, we assume that fixed point to be 0 and q (or p) to be positively homogeneous on a neighborhood. And when we consider a sequence of iterates $x_{k+1} = p(x_k)$ converging to that fixed point, we may as well assume q (or p) to be positively homogeneous on the entire space since only local behavior is relevant.

The following three theorems were proved in [1] to describe the behavior of the proximal iterates of a folding.

Theorem 1 below can be interpreted as saying that the proximal iteration $x_{k+1} = p(x_k)$, when applied to a folding having a fixed point, spirals towards a fixed point $\bar{x}$. It states that for arbitrary x_0, there is a subspace $V(x_0)$ such that for all k sufficiently large, the set $\{x_k-\bar{x}, x_{k+1}-\bar{x}, \ldots\}$ positively spans $V(x_0)$. (If $V(x_0) = \{0\}$, this says that $x_k = \bar{x}$ and the iteration terminates.)

The second theorem states that the restriction of q to the subspace described in the first theorem is an isometry with a unique fixed point.

The subspace described in the first two theorems depends on the starting point x. The third theorem establishes the existence of a subspace not depending on x, but having some of the same properties:

Theorem 1. Let $q = \alpha(p)$ be a positively homogeneous folding, $p^k(x) \to 0$. Let $x_k = p^k(x)$, $C_k = \text{cone}\{x_k, x_{k+1}, \ldots\}$, and $L_k = \text{span}\{x_k, x_{k+1}, \ldots\}$. There is a subspace $V(x)$ of R^d and $K > 0$ such that $L_k = C_k = V(x)$ for all $k \geq K$.

Theorem 2. Let $q = \alpha(p)$ be a positively homogeneous folding on R^d. Suppose $x_k = p^k(x) \to 0$, and let $V = V(x)$. Then $q|V$ is an isometry, 0 is the only fixed point of q in V, and $p|V$ is a linear isomorphism.

Theorem 3. Let $q = \alpha(p)$ be a positively homogeneous folding on R^d. There is a subspace $\hat{V} \subset \text{lin}\, N_F$ such that q is an isometry on $\hat{V}$, 0 is the only fixed point of q in $\hat{V}$, and if $x \in R^d$ is such that $p^k(x) \to 0$ then there is $K > 0$ (possibly depending on x) such that $p^k(x) \in \hat{V}$ for all $k \geq K$.

C. Applications of the Results of [1]

1. Polyhedral Convex Functions. In [1], we established the following, which asserts that every polyhedral convex function gives rise, in a natural way, to a folding:

Theorem 4. If $g: R^d \to R \cup \{\infty\}$ is a proper polyhedral convex function, then $\beta(\partial g)$ is a folding on R^d.

Using our results on foldings, we were then able in [1] to prove that the proximal point algorithm can be used to minimize a convex polyhedral function in only finitely many steps:

Theorem 5. Let $g: R^d \to R \cup \{\infty\}$ be a proper polyhedral convex function that achieves a minimum value, and let p be the proximal mapping $(I+\partial g)^{-1}$. Then for aribtrary x, there is k such that $p^k(x)$ is a minimizer for g. In other words, the proximal point algorithm terminates after a finite number of iterations with a minimizer.

The following presents an important class of problems for which the proximal point algorithm always finds a solution in only finitely many iterations:

Theorem 6. Let A be a subspace of R^d, $B = A^\perp$, $g: R^d \to R \cup \{\infty\}$ a proper polyhedral convex function, $p = \pi_A * p_g$. If the fixed point set F for p is nonempty, then it has the form $F = C+D$ with $C \subset A$ and $D \subset B$. If C has nonempty interior relative to A, or if D has nonempty interior relative to B, then for each $z \in R^d$, there is some m such that $p^m(z) \in F$.

If $A = R^d$, then $p_g = \pi_A * p_g$ and $D = B = \{0\}$. Since D (trivially) has nonempty interior relative to B, Theorem 6 implies that the iteration $x_{k+1} = p_g(x_k)$ must yield a fixed point (provided one exists) after a finite number of iterations.

If $K \subset R^d$ is a polyhedral convex set, then ψ_K, the characteristic function of K, is a polyhedral convex function. Applying Theorem 6 to the case $g = \psi_K$, we obtain the following

Theorem 7. Let $K \subset R^d$ be a nonempty polyhedral convex set, $A \subset R^d$ a linear subspace, $B = A^\perp$, $p = \pi_A * \pi_K$, $z_{k+1} = p(z_k)$. The set of fixed points of p is nonempty if, and only if, $A \cap K \neq \emptyset$. If $A \cap K$ has nonempty interior with respect to A, or if $A \cap K \neq \emptyset$ and $B \cap K^o$ has nonempty interior with respect to B, then for some m, z_m is a fixed point for p, $\pi_A(z_m) \in A \cap K$, and $\pi_B(z_m) \in B \cap K^o$.

2. Systems of Inequalities. We will now describe two iterative schemes for finding a feasible point for a system of linear inequalities.

Consider a system

$$\langle x, u_i \rangle \le b_i \ , \qquad i = 1,\dots,n$$

of linear inequalities ($u_i \in R^d$, $b \in R$). For each i, let $C_i = \{x: \langle x,u_i\rangle \leq b_i\}$, p_i = projection onto C_i, and $q_i = \alpha(p_i)$. The q_i are clearly foldings. Also, define $C = C_1 \cap \ldots \cap C_n$, $p = p_1 * \ldots * p_n$, and $q = q_1 \circ \ldots \circ q_n$ $(= \alpha(p))$.

The first scheme we wish to suggest for solving the system is to iterate the proximal mapping p:

Theorem 8. If C has nonempty interior, then for each $x \in R^d$ there is $K > 0$ such that $p^K(x) = p^{K+1}(x) = \ldots \in C$.

This scheme for solving a system of linear inequalities is very closely related to the methods of Agmon [16] and Motzkin and Schoenberg [17]. In the method of "successive projection with relaxation parameter = 2" (also a finitely terminating algorithm [17, Theorem 1]), one takes $x_{k+1} = q(x_k)$ instead of $x_{k+1} = p(x_k)$.

We now describe a second method for solving the system. This latter method has been studied in some detail in [5], [6].

Define $K = C_1 \times \ldots \times C_n$. K is a polyhedral convex set in R^{dn}. Define the subspaces

$$A = \{(x_1,\ldots,x_n) \in R^{dn}: x_1 = \ldots = x_n\}$$

$$B = \{(y_1,\ldots,y_n) \in R^{dn}: y_1 + \ldots + y_n = 0\}$$

of R^{dn} and note that $B = A^\perp$. Clearly

$$A \cap K = \{(x,\ldots,x): x \in C\}$$

so solving the system is equivalent to finding a point in the intersection of A amd K. Let π_A and π_K denote the orthogonal projection mappings onto A and K, respectively. The algorithm is described in the following

Theorem 9 [5]. Suppose C has nonempty interior. If a sequence (z_k) is generated by the rule $z_{k+1} = (\pi_A * \pi_K)(z_k)$, then $\pi_A(z_m) \in A \cap K$ for some m.

3. Network Flows. Let G be a network (directed graph). Abstractly, G consists of a finite set N of "nodes", a finite set A of "arcs", and an incidence matrix $E = (e_{ij})$, where

$$e_{ij} = \begin{cases} 1 & \text{if } i \text{ is the initial node of arc } j \\ -1 & \text{if } i \text{ is the terminal node of arc } j \\ 0 & \text{otherwise} \end{cases}$$

Let us suppose for each arc $j \in A$, there is a polyhedral convex set $K_j \subset R^d$. A flow in G is a function $x\colon A \to R^d$. We will call x a circulation, and write $x \in C$, if it is conservative at each node, that is,

$$\sum_{j \in A} e_{ij} x(j) = 0 \quad \text{for all } i \in N.$$

The set C is a subspace of the vector space of all flows in G. Define $K = \prod_{j \in A} K_j$; in other words, K is the set of all flows x such that $x(j) \in K_j$ for all $j \in A$. The feasible circulation problem is

(P) $\qquad$ to find $x \in K \cap C$.

This is a problem of finding a point in the intersection of a polyhedral convex set and a subspace. Applying Theorem 7, we obtain

Theorem 10. If $\mathrm{int}(K) \cap C \neq \emptyset$, any sequence $x_{n+1} = (\pi_C * \pi_K)(x_n)$ terminates in a finite number of iterations with a flow x_m such that $\pi_C(x_m) \in K \cap C$.

The regularity condition $\mathrm{int}(K) \cap C \neq \emptyset$ means there exists a circulation x such that $x(j) \in \mathrm{int}\ K_j$ for all j. For each $a: A \to R^d$ (a may be regarded as a point in $R^{d|A|}$), define the perturbed problem

(P_a) $\qquad$ to find $x \in (K+a) \cap C$.

Thus $(P_0) = (P)$. The regularity condition can be expected to hold for most problems in the following sense:

Theorem 11. Suppose $\mathrm{int}\ K_j \neq \emptyset$ for all j. The set of all parameter values a which fail to satisfy one of the following conditions

i. $\mathrm{int}(K+a) \cap C \neq \emptyset$ (so (P_a) satisfies the hypothesis for Theorem 10;

ii. $(K+a) \cap C = \emptyset$ (so (P_a) is infeasible);

forms the boundary of a convex subset $R^{d|A|}$ and is thus a set of measure zero.

Theorem 11 means that, except for a belonging to a small set, either the perturbed problem (P_a) is solvable in a finite number of steps by the algorithm, or (P_a) has no solution and the iterates will diverge.

4. Linear Programming Duality. As a final application, we proved in [1] that if A is a real $d \times d$ antisymmetric matrix, there exists a vector $x \geq 0$ such that $Ax \geq 0$ and $x+Ax > 0$. While this is not a new result, our proof is novel, and provides an elegant application of our results on foldings.

D. The Simple Iterates of a Folding -- Summary of the Main Results from [2]

In [2], we proved several properties of foldings, the first being

Theorem 12. The folds of a folding q are closed, convex and have nonempty interior.

Folds are not necessarily disjoint, but their interiors cannot overlap, as the following shows:

Theorem 13. If P and Q are distinct folds then $P \cap \mathrm{int}(Q) = \emptyset$.

The fact that folds are convex, have nonoverlapping interiors, and their union is the whole space implies that they are convex polyhedral sets:

Theorem 14. Folds are polyhedral convex sets.

In the next theorem, it was shown that the fixed point set of a folding has a very special structure; it is the intersection of an affine flat and a fold.

Theorem 15. Let $q: R^d \to R^d$ be a folding with nonempty fixed point set F. Then there is an affine flat A and a fold Q such that $F = A \cap Q$.

The following is a characterization of linear isometries:

Theorem 16. Let $q: R^d \to R^d$ be positively homogeneous ($q(tx) = tq(x)$ for $t \geq 0$) and a folding. Then q is a (linear) isometry if, and only if, its fixed point set F is a subspace. If q is a linearly isometry, then q maps $F^{\perp}$ onto $F^{\perp}$.

We have seen already that the fixed point set of a folding is the intersection of a fold and a flat. The next result shows that the flat must meet the interior of the fold.

Theorem 17. Let $q: R^d \to R^d$ be a folding. If q has a fixed point, there is a unique fold P whose interior contains a fixed point.

If $q: R^d \to R^d$, we define the cluster set of q to be the set S of all cluster points of q-sequences $x, q(x), \ldots$

If q is nonexpansive and has a fixed point, then every q-sequence has a cluster point. On the other hand, if the cluster set of q is nonempty, Theorem 18 shows that q must have a fixed point. If q is nonexpansive and has a fixed point, Theorem 19 demonstrates that S is closed, convex, $q(S) = S$, and q is distance preserving on S. Also, all q-sequences converge to S in the sense that $\mathrm{dist}(q^k(x),S) \to 0$ for all $x \in R^d$. Thoerem 20 gives a sufficient condition guaranteeing that a folding q have the property: for every x there is some finite k such that $q^k(x) \in S$. Theorem 21 asserts the necessity of this same condition.

Theorem 18. A nonexpansive mapping $q\colon R^d \to R^d$ possesses a fixed point if and only if its cluster set is nonempty.

Theorem 19. Let $q\colon R^d \to R^d$ be a nonexpansive mapping with a nonempty fixed point set F. Then

i. for any $z \in S$, z is a cluster point of the sequence $z, q(z), q^2(z), \ldots$

ii. $q(S) = S$

iii. q is an isometry on S

iv. S is closed and convex

v. for any $x \in R^d$, and $\varepsilon > 0$, there exists K such that $\mathrm{dist}(q^k(x),S) < \varepsilon$ for all $k \geq K$.

If q is a folding, and P is the unique fold whose interior meets F, then also

vi. $F \cap \mathrm{int}(S)$ is nonempty

vii. $S \subset P$.

<u>Theorem 20</u>. Let $q: R^d \to R^d$ be a folding having a nonempty fixed point set F, P the fold whose interior meets F, i the isometry that agrees with q on P, S the set of all cluster points of q-sequences, and W the linear subspace parallel to the flat consisting of all points having periodic orbit under i. If $P = P + W^{\perp}$, then for every $x \in R^d$, there is some k such that $q^k(x) \in S$.

<u>Theorem 21</u>. Let q be a folding on R^d having a nonempty fixed point set F. Let P be the fold whose interior meets F, S the set of all cluster points of q-sequences. Let i denote the isometry that agrees with q on P, and W the subspace parallel to the flat consisting of all points having periodic orbit under i. If $P \neq P + W^{\perp}$, then there exists $y \notin S$ such that $q^k(y) \notin S$ for all k.

II. <u>Solution of Network Equilibrium Problems</u>

In a forthcoming paper [3], a new solution approach will be proposed for a certain class of network equilibrium problems. Applications to the computation of spatial economic equilibria will also be discussed.

The simplest problem to be considered will be the single-location supply-demand equilibrium problem. Here, the goal is to balance the supply and demand of a vector of consumable goods in a competitive market environment. Given a demand vector d, the suppliers are assumed to choose a production plan x so as to solve a certain convex programming problem

$$S(d) \quad \text{to minimize } f_0(x)$$
$$\text{subject to } f_i(x) + d_i \le 0, \quad i = 1,\dots,m,\ x \in C.$$

Solution of this problem also yields a vector $P(d)$ of shadow prices (Kuhn-Tucker vectors). The consumers are represented by a demand function $Q(p)$ which specifies the quantities of goods consumers are willing to purchase at unit prices p. The problem is then

$$\text{to determine } p \text{ and } d \text{ such that } p \in P(d) \text{ and } q \in Q(p).$$

This type of formulation is found in the PIES model [18] and others. The solution approach we will propose has some strong theoretical advantages over existing methods, the main one being global convergence under the assumption that the multifunction $-Q$ be maximal monotone. Explicit knowledge of the function $P(d)$ is not required; each iteration requires the solution of a convex programming problem for which the only constraint is $x \in C$. Not only are sequences produced converging to equilibrium values of p and d, but a sequence x_k is produced as well converging to an optimal value for x. Each iteration requires also one evaluation of the function $(I-Q)^{-1}$ which is known to be single-valued because of the fact that $-Q$ is maximal monotone.

More generally, our model will incorporate the spatial or multi-location equilibrium problem. Under this model, there are a finite number of countries or locations, each one having its own demand function $Q_j(p_j)$ and its own convex programming problem $S_j(d_j)$. Goods may be shipped between locations, and when this is done a certain transportation cost is incurred. It is required that the p_j are shadow prices for the associated convex programming problems $S_j(d_j)$, $q_j \in Q_j(p_j)$, and the markets are balanced in the sense that no goods are shipped in a manner that does not make sense for either producers or consumers.

The method of solution proposed here is a specialization to a network setting of the "method of partial inverses" introduced in [1]. Given a maximal monotone multifunction $T: H \rightrightarrows H$ on a Hilbert space H (which for our purposes will always be a finite dimensional Euclidean space equipped with the standard inner product) and a closed subspace $A \subset H$, the method of partial inverses is a procedure for solving the following problem (taking $B = A^{\perp}$)

(5) to find $x \in A$ and $y \in B$ such that $y \in T(x)$.

The projection of x onto the subspace A or B will be denoted as x_A or x_B, respectively. To solve (5), the method of partial inverses constructs sequences $x_n \in A$ and $y_n \in B$ in such a way that

(6) $$x_{n+1} = (x_n')_A \quad \text{and} \quad y_{n+1} = (y_n')_B,$$
where x_n' and y_n' are chosen so that
$$x_n' + y_n' = x_n + y_n \quad \text{and} \quad y_n' \in T(x_n').$$

The existence and uniqueness of x'_n and y'_n having the above property is established in [1]. The main result from [1] regarding the convergence of the algorithm is the following:

Theorem 22. Let x_k and y_k be sequences of iterates produced by the method of partial inverses. It will always happen either that

i. $x_k \to \bar{x}$ and $y_k \to \bar{y}$ for some solution $\bar{x},\bar{y}$, or that
ii. $|x_k + y_k| \to \infty$ and (5) has no solutions.

The distance from $x_k + y_k$ to the set $\{x+y\colon x,y \text{ solves } (5)\}$ is nonincreasing.

A network G is a triple (N,A,e). The finite sets N and A consist of the nodes and arcs of G, and the incidence function e maps $N \times A$ into $\{+1,-1,0\}$, where

$$e(i,j) = \begin{cases} +1 & \text{if } i \text{ is the "initial" node of arc } j \\ -1 & \text{if is the "terminal" node of arc } j \\ 0 & \text{otherwise} \end{cases}$$

Each arc is required to have exactly one initial and one terminal node, and these must be distinct. Let E be the $k \times n$ matrix whose (i,j)-entry is e(i,j) ($k = |N|$ and $n = |A|$).

A flow in G is a function $x\colon A \to R^m$. If $m > 1$, x is a "multicommodity" flow. The divergence $y = \operatorname{div} x$ of the flow x is the function $y\colon N \to H$ defined for each node i by

$$y(i) = \sum_{j \in A} e(i,j)x(j).$$

In matrix notation, we can write $y = Ex = \text{div } x$. If $\text{div } x = 0$, we say that x is a <u>circulation</u> in G. The set C of all circulations in G is a subspace of the vector space flows in G. A <u>potential</u> in G is a function $u: N \to H$. A potential determines, in a natural way, a function $v = \Delta u: A \to H$ called the <u>tension function</u> on A which is the <u>differential</u> of u. If i is the initial and i' the terminal node of arc j, then

$$\Delta u(j) = v(j) = u(i') - u(i),$$

or, in matrix form, $v = \Delta u = -E'u$. The tension v is called a <u>differential</u> if $v = \Delta u$ for some potential u. The set D of all differentials in G is also a subspace of the space of flows in G.

Since D is the range of E' and C is the kernel of E, we have the important relationship

$$C = D^{\perp}, \quad C^{\perp} = D.$$

An arbitrary flow x can thus be written in a unique way as the sum of a circulation and a differential:

$$x = x_C + x_D.$$

Both of the economic equilibrium problems mentioned above can be phrased in a network setting. Because the supply and demand functions arise in different manners, it turns out to be convenient to consider networks that contain two classes of arcs. Suppose that the set J of arcs of G is divided into two distinct classes J_1 and J_2. With each arc $j \in J_1$ let there be an associated maximal monotone operator $T_j: R^m \rightrightarrows R^m$. With each arc $j \in J_2$, let there be an associated family (P_{u_j}) of optimization problems, where, for each $u_j = (u_{1j},\ldots,u_{mj}) \in R^m$, the problem (P_{u_j}) is

(P_{u_j}) to minimize $f_{0j}(x_j)$ over $x_j \in R^{d_j}$
subject to the constraints $x_j \in C_j$ and
$f_{1j}(x_j) + u_{1j} \leq 0,\ldots,f_{mj}(x_j) + u_{mj} \leq 0$

where the functions $f_{ij}: R^{d_j} \to R$ are convex and $C_j \subset R^{d_j}$ is a nonempty closed convex set. Given such a network, it makes sense to consider the following problem (taking $d_j = 0$ for $j \in J_1$, and $d = \Sigma d_j$):

(Q) to find $u = \pi u_j \in C$, $y = \pi y_j \in D$, and $x = \pi x_j \in R^d$ such that

(a) for each $j \in J_1$, $y_j \in T_j(u_j)$ and

(b) for each $j \in J_2$, x_j solves the problem (P_{u_j}) and y_j is a Kuhn-Tucker vector.

To say that y_j is a Kuhn-Tucker vector for (P_{u_j}) means that

$y_{1j},\ldots,y_{mj} \geq 0$ and the infimum of the function
$f_{0j}(x_j') + \sum_i y_{ij}(f_{ij}(x_j') + u_{ij})$ over all $x_j' \in C_j$
is finite and equal to the optimal value in (P_{u_j}).

For each $j \in J_2$, define

$$F_j(x_j,u_j) = \begin{cases} f_{0j}(x_j) & \text{if } f_{1j}(x_j)+u_{1j} \le 0,\dots,f_{mj}(x_j)+u_{mj} \le 0 \\ & \quad\text{and } x_j \in C_j \\ +\infty & \text{otherwise} \end{cases}$$

and let $T_j = \partial F_j$.

Let $H = \{(x,u): x \in R^d$ and u is a flow in $G\}$. Let $A = \{(x,u) \in H: u \in C\}$ and $B = \{(v,y) \in H: v = 0$ and $y \in D\}$. It is clear that A and B are complementary subspaces of H. Define a maximal monotone multifunction $T: H \rightrightarrows H$ by declaring

$(v,y) \in T(x,u)$ if, and only if $y_j \in T_j(u_j)$ for all $j \in J_1$ and $(v_j,y_j) \in T_j(x_j,u_j)$ for all $j \in J_2$.

The network equilibrium problem (Q) is then equivalent to the problem

to find $(x,u) \in A$ and $(v,y) \in B$ such that $(v,y) \in T(x,u)$.

This problem can be solved by the method of partial inverses (6). In order that it be possible to implement this procedure, we know that it would be required that routines be available to perform each of the following tasks:

(1) given $(x,u) \in H$, to compute the projection of (x,u) onto A and B, and

(2) given $(x,u) \in A$ and $(0,y) \in B$, to determine (x',u')

and (v',y') such that

$$(x'+v',u'+y') = (x,u+y)$$

and

$$(v',y') \in T(x',u').$$

Procedures for performing each of these tasks will be investigated in [3]. The resulting algorithm for solving (Q) is the following:

Initialization: Start with an arbitrary flow x, $u \in C$, $y \in D$.

Step 1: (a) For each arc $j \in J_1$, find u_j' and y_j' such that

$$u_j' + y_j' = u_j + y_j \quad \text{and } y_j' \in T_j(u_j')$$

(b) For each arc $j \in J_2$, find x_j' to minimize the function

$$f_{0j}(x_j') + \frac{1}{2}\Sigma|x_{ij}'-x_{ij}|^2 + \frac{1}{2}\Sigma\min^2\{0,f_{ij}(x_j') + u_{ij} + y_{ij}\}$$

subject to the constraint $x_j \in C_j$, and let

$$u_{ij}' = \begin{cases} u_{ij} + y_{ij} & \text{if } f_{ij}(x_j') + u_{ij} + y_{ij} \geq 0 \\ -f_{ij}(x_j') & \text{if } f_{ij}(x_j') + u_{ij} + y_{ij} \leq 0 \end{cases}$$

and

$$y_{ij}' = u_{ij} + y_{ij} - u_{ij}'$$

Step 2: Update x, u, and y as follows:

$$x^+ = x', \quad u^+ = (u')_C, \quad y^+ = (y')_D$$

and repeat Step 1.

According to Theorem 22, we also have

<u>Theorem 23</u>. Let x_k, u_k, and y_k be iterates produced by the above algorithm. It will always happen either that

i. $x_k \to \bar{x}$, $u_k \to \bar{u}$, and $y_k \to \bar{y}$ for some solution $(\bar{x},\bar{u},\bar{y})$ to (Q),

or that

ii. $|(x_k,u_k+y_k)| \to \infty$ and (Q) has no solution.

References

Papers containing results of this project:

1. J. E. Spingarn and J. Lawrence, "On fixed points of nonexpansive piecewise isometric mappings," to appear in Proc. London Math. Soc.

2. J. E. Spingarn and J. Lawrence, "On iterates of nonexpansive mappings and foldings," submitted.

3. J. E. Spingarn, "On the computation of spatial economics equilibria," in progress.

Papers by the principal investigator closely related to this project:

4. "Partial inverse of a monotone operator," Applied Mathematics and Optimization 10 (1983) 247-265.

5. "A primal-dual projection method for solving systems of linear inequalities," Linear Algebra and Its Applications, 65: 45-62 (1985).

6. "A projection method for least square solutions to overdetermined systems of linear inequalities," accepted by Linear Algebra and Its Applications.

7. "Applications of the method of partial inverses to convex programming: decomposition," Mathematical Programming, 32 (1985) 199-223.

Other papers referenced in this report:

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

9. B. Martinet, Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox, C. R. Acad. Sci. Paris (Ser. A) 274 (1972) 163-165.

10. J.-J. Moreau, Proximité et dualité dans un espace Hilbertian, Bull. Soc. Math. France 93 (1965) 273-299.

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

12. M. D. Kirszbraun, Über die zusammenziehende and Lipschitzsche transformationen, Fundamenta Mathematicae 22 (1934) 77-108.

13. G. J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29 (1962) 341-346.

14. R. T. Rockafellar, On the maximal monotonicity of subdifferential mappings, Pacific J. Math. 33 (1970) 209-216.

15. R. T. Rockafellar, Augmented Lagrangians and the proximal point algorithm in convex programming, Math. of O. R. 1 (1976) 97-116.

16. S. Agmon, The relaxation method for linear inequalities, Canad. J. Math. 6 (1954) 382-392.

17. T. S. Motzkin, I. J. Schoenberg, The relaxation method for linear inequalities, Canad. J. Math. 6 (1954) 393-404.

18. B. Ahn, Computation of Market Equilibria for Policy Analysis: The Project Independence Evaluation System Approach, Garland Publ., New York 1979.