April 28, 1989 LIDS-P-1866

# Further Applications of a Splitting Algorithm to Decomposition in Variational Inequalities and Convex Programming*

Paul Tseng†

**Abstract**

A classical method for solving the variational inequality problem is the projection algorithm. We show that existing convergence results for this algorithm follow from one given by Gabay for a splitting algorithm for finding a zero of the sum of two maximal monotone operators. Moreover, we extend the projection algorithm to solve <u>any</u> monotone affine variational inequality problem. When applied to linear complementarity problems, we obtain a matrix splitting algorithm that is simple and, for linear/quadratic programs, massively parallelizable. Unlike existing matrix splitting algorithms, this algorithm converges under no additional assumption on the problem. When applied to generalized linear/quadratic programs, we obtain a decomposition method that, unlike existing decomposition methods, can simultaneously dualize the linear constraints and diagonalize the cost function. This method gives rise to highly parallelizable algorithms for solving a problem of deterministic control in discrete time and for computing the orthogonal projection onto the intersection of convex sets.

KEY WORDS: variational inequality, operator splitting, decomposition, linear complementarity, linear/quadratic programming.

---

* This research is partially supported by the U.S. Army Research Office, contract DAAL03-86-K-0171 (Center for Intelligent Control Systems), and by the National Science Foundation under grant NSF-ECS-8519058.

† The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

## 1. Introduction

Let X be a nonempty closed convex set in $\Re^n$ and let $f:X \rightarrow \Re^n$ be a continuous function. Consider the following problem:

$$\text{Find an } x^* \in X \text{ satisfying } \quad \langle f(x^*), x - x^* \rangle \geq 0, \quad \forall\, x \in X. \qquad \mathrm{VI}(X,f)$$

This problem, called the variational inequality problem, has numerous applications to optimization, including the solution of systems of equations, constrained and unconstrained optimization, traffic assignment, and saddlepoint point problems. [See for example [Aus76], [BeT89], [CGL80], [GLT81], [KiS80].]

We make the following standing assumptions regarding f and X:

**Assumption A:**

(a) The function f is monotone, i.e. $\langle f(y) - f(x), y - x \rangle \geq 0, \quad \forall\, x \in X, \forall\, y \in X.$

(b) The problem VI(X,f) has a solution.

Let D be an n×n positive definite matrix D. Consider the following algorithm for solving VI(X,f) whereby the original variational inequality is approximated by a sequence of affine variational inequalities:

**Asymmetric Projection (AP) Algorithm**

Iter. 0 Start with any $x^0 \in X$.

Iter. r+1 Given an $x^r \in X$, compute a new iterate $x^{r+1} \in X$ satisfying

$$\langle D(x^{r+1} - x^r) + f(x^r), x - x^{r+1} \rangle \geq 0, \qquad \forall\, x \in X. \tag{1.1}$$

[The iteration (1.1) is well defined because D is positive definite [BeT89, §3.5], [KiS80, §2].] We have called the above algorithm the asymmetric projection (AP) algorithm because if D is symmetric, then it reduces to the well-known projection algorithm [Sib70] (also see [BeT89], [Daf83], [KiS80], [PaC82])

$$x^{r+1} = \operatorname{argmin}_{x \in X} \{ \|x - x^r + D^{-1} f(x^r)\|_D \}, \qquad r = 0, 1, 2, \ldots,$$

where $\|\cdot\|_D$ denotes the norm $\|x\|_D = \langle x, Dx \rangle^{1/2}$.

It has been shown that if D and f satisfy a certain contraction condition [PaC82], [Daf83], then $\{x^r\}$ generated by the AP iteration (1.1) converges to a solution of VI(X,f). Unfortunately, this condition implies that f is strictly monotone, which excludes from consideration important special cases of VI(X,f) such as linear complementarity problems and linear/quadratic programs. The goal of this paper is two-fold: First we show that the existing convergence conditions for the AP algorithm follow as a corollary of a general convergence condition given by Gabay [Gab83] for a forward-backward splitting algorithm. This leads to a unified and a much simpler characterization of the convergence conditions. Second, we show that the convergence condition for the AP algorithm can be broadened such that it is applicable to all monotone (not necessarily strictly monotone) affine variational inequality problems. In particular, we apply this algorithm to linear complementarity problems (for which X is the non-negative orthant) to obtain a matrix splitting algorithm that is simple and, for linear/quadratic programs, massively parallelizable. Unlike existing matrix splitting algorithms [Man77], [Pan84], [LiP87], this algorithm requires no additional assumption (such as symmetry) on the problem data for convergence. We also apply this algorithm to generalized linear/quadratic programming problems to obtain a new decomposition method for solving these problems. This method has the important advantage that it can simultaneously dualize any subset of the constraints and diagonalize the cost function; hence it is highly parallelizable. We describe applications of this method to a problem of deterministic control in discrete time and to computing the orthogonal projection onto the intersection of convex sets.

This paper proceeds as follows: In §2 we describe the forward-backward splitting algorithm and a convergence result of Gabay for this algorithm. In §3 we show that the AP algorithm is a special case of this splitting algorithm and that Gabay's result contains as special cases existing convergence results for the AP algorithm. In §4 we show that the AP algorithm can be applied to solve any monotone affine variational inequality problem. In §5 we further specialize the AP algorithm to a decomposition method for generalized linear/quadratic programming. In §6 we further specialize the AP algorithm to a matrix splitting algorithm for solving linear complementarity problems.

In our notation, all vectors are column vectors and superscript T denotes transpose. We denote by $\langle \cdot,\cdot \rangle$ the usual Euclidean inner product and by $\|\cdot\|$ its induced norm. [The argument of $\|\cdot\|$ can be either a matrix or a vector.] For any n×n matrix D, we denote its symmetric part by

$$\bar{D} = (D + D^T)/2.$$

For any set $Y \subseteq \Re^n$, we denote by $\delta_Y(\cdot)$ the indicator function for Y, i.e. $\delta_Y(y)$ is zero if $y \in Y$ and is $\infty$ otherwise. For any closed convex function $h:\Re^n \to (-\infty,\infty]$ and any $x \in \Re^n$, we denote by $\partial h(x)$ the subdifferential of h at x. For any closed set Y in $\Re^n$ we say a function $h:Y \to \Re^n$ is co-coercive with modulus $\sigma > 0$ if

$$\langle h(y) - h(x), y - x \rangle \geq \sigma \|h(y) - h(x)\|^2, \qquad \forall\, x \in Y,\ \forall\, y \in Y.$$

[Note that a co-coercive function is Lipschitz continuous and monotone.]

## 2. A Splitting Algorithm

A multifunction $T:\Re^n \to \Re^n$ is said to be a monotone operator if

$$\langle y - y', x - x' \rangle \geq 0 \qquad \text{whenever} \qquad y \in T(x),\ y' \in T(x').$$

It is said to be maximal monotone if, in addition, the graph $\{ (x,y) \in \Re^n \times \Re^n \mid y \in T(x) \}$ is not properly contained in the graph of any other monotone operator $T':\Re^n \to \Re^n$. [A classical example of a maximal monotone operator is the subdifferential of a closed proper convex function (see [Min64] or [Mor65]). General discussion of maximal monotone operators can be found in [Bré73], [Dei85], [Roc76].] We denote by $T^{-1}$ the inverse of T, i.e.

$$(T^{-1})(y) = \{ x \in \Re^n \mid y \in T(x) \}, \quad \forall\, y \in \Re^n.$$

It is easily seen from symmetry that the inverse of a maximal monotone operator is also a maximal monotone operator.

Let Y be any closed convex set in $\Re^n$. Let $T:\Re^n \to \Re^n$ be a maximal monotone operator whose effective domain is Y (i.e. $Y = \{ y \in \Re^n \mid T(y) \neq \emptyset \}$) and let $h:Y \to \Re^n$ be any function. Consider the following problem

$$\text{Find an } y^* \in Y \text{ satisfying} \qquad 0 \in T(y^*) + h(y^*). \tag{2.1}$$

Suppose that (2.1) has a solution and consider an algorithm for solving (2.1) that uses the following iteration

$$y^{r+1} = [\, I + T \,]^{-1} [\, I - h \,](y^r), \qquad r = 0, 1, 2, \ldots. \tag{2.2}$$

[It has been shown [Min62] that the proximal mapping $[\, I + T \,]^{-1}$ is a single-valued mapping and its effective domain is all of $\Re^n$. Furthermore the range of $[\, I + T \,]^{-1}$ is Y so that the iteration (2.2) is well defined.] The iteration (2.2), in the terminology of [LiM79], is a splitting iteration that alternates between a forward step with respect to h and a backward step with respect to T. Convergence of this iteration has been extensively studied [Bru75], [Gab83], [Pas79], [Lem88], [Tse88]. The result that is most useful to us is the following given by Gabay:

**Proposition 1** ([Gab83, §6]) If h is a co-coercive function with modulus greater than 1/2, then the sequence $\{y^r\}$ generated by the iteration (2.2) converges to a solution of (2.1) from any starting point $y^0$ in Y.

The iteration (2.2), despite its simplicity, is a very powerful tool for the development of decomposition methods. Both a projection algorithm of Goldstein [Gol64] and an algorithm of Han and Lou [HaL88] can be shown to be special cases of this iteration (see [Tse88]). A number of new decomposition algorithms for solving variational inequality problems and convex programs can also be derived from it [Gab83], [Tse88]. We will presently derive many more algorithms of this kind.

It should be noted that a function is co-coercive if and only if its inverse is a coercive (i.e. strongly monotone) operator. Hence it would appear that the splitting iteration (2.2) is restricted to problems having some strongly monotone component. For example, the applications given in [HaL88], [Lem88, Theorem 1], [Tse88] for convex programming are restricted to problems having some strongly convex component in the cost function. However, we shall see that this is not the case.

## 3. Relation to Splitting Algorithm

Let D be an n×n positive definite matrix. To simplify the notation, let

$$L = \bar{D}^{-1/2}(D-\bar{D})\bar{D}^{-1/2}, \tag{3.1}$$

and $Y = \{ \bar{D}^{1/2}x \mid x \in X \}$. Also let $\tilde{f}:Y \to \Re^n$ be the function

$$\tilde{f}(y) = \bar{D}^{-1/2}f(\bar{D}^{-1/2}y), \qquad \forall\, y \in Y. \tag{3.2}$$

We have the following result:

**Proposition 2** If $\tilde{f} - L$ is co-coercive with modulus greater than 1/2, then the AP iteration (1.1) is a special case of (2.2) and $\{x^r\}$ generated by (1.1) converges to a solution of VI(X,f).

**Proof:** From (1.1) we have

$$0 \in Dx^{r+1} + \partial\delta_X(x^{r+1}) - Dx^r + f(x^r),$$

or equivalently,

$$0 \in \bar{D}x^{r+1} + (D-\bar{D})x^{r+1} + \partial\delta_X(x^{r+1}) - Dx^r + f(x^r).$$

By making the substitution $y^r = \bar{D}^{1/2}x^r$, we can express the above equation as

$$0 \in y^{r+1} + Ly^{r+1} + \partial\delta_Y(y^{r+1}) - (I+L)y^r + \tilde{f}(y^r).$$

This in turn is equivalent to

$$y^{r+1} = [I + L + \partial\delta_Y]^{-1}[\, I + L - \tilde{f}\,](y^r).$$

Now we have that $\langle Lz\,,\, z\rangle = \langle (D-\bar{D})\bar{D}^{-1/2}z\,,\, \bar{D}^{-1/2}z\rangle/2 = 0$ for all $z \in \Re^n$. Hence (cf. [Roc70b]) $L + \partial\delta_Y$ is a maximal monotone operator with effective domain Y, and the above iteration is a special case of (2.2). Since by assumption $\tilde{f} - L$ is co-coercive with modulus greater than 1/2, it follows from Proposition 1 that the sequence $\{y^r\}$ converges to an $y^\infty \in Y$ satisfying

$$0 \in \partial\delta_Y(y^\infty) + \tilde{f}(y^\infty).$$

Since $x^\infty = \bar{D}^{-1/2} y^\infty$ is easily seen to be a solution of VI(X,f), this completes the proof.
Q.E.D.

We show below that if f is Lipschitz continuous, then the hypothesis of Proposition 2 are implied by the convergence conditions given by Pang and Chan [PaC82, Theorem 2.9] (also see [BeT89], [Daf83]).

**Proposition 3** If f is Lipschitz continuous and there exists a $\beta \in (0,1)$ such that

$$\|\tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\| \le \beta\|y-x\|, \qquad \forall\, x \in Y, \forall\, y \in Y,$$

then $\tilde{f} - L$ is co-coercive with modulus greater than 1/2.

**Proof:** It suffices to show that

$$\langle \tilde{f}(y) - Ly - \tilde{f}(x) + Lx\,,\, y - x\rangle \ge \sigma\|\tilde{f}(y) - Ly - \tilde{f}(x) + Lx\|^2, \;\; \forall\, x \in Y, \forall\, y \in Y,$$

for some $\sigma > 1/2$. Consider any $x, y \in Y$. Since $\|\tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\| \le \beta\|y-x\|$, we have (also using the definition of L (3.1))

$$\begin{aligned}
\|\tilde{f}(y) - Ly - \tilde{f}(x) + Lx\|^2 &= \|(y-x) + \tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\|^2 \\
&= \|y-x\|^2 + 2\langle y-x\,,\; \tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\rangle \\
&\quad + \|\tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\|^2 \\
&\le (1+\beta)\|y-x\|^2 + 2\langle y - x\,, \tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2} D \bar{D}^{-1/2}(y-x)\rangle \\
&= 2\langle y - x\,, \tilde{f}(y) - Ly - \tilde{f}(x) + Lx\rangle - (1-\beta)\|y-x\|^2.
\end{aligned}$$

Since f is Lipschitz continuous (with modulus say $\rho$), we have that $\|\tilde{f}(y) - Ly - \tilde{f}(x) + Lx\| \le (\|L\|+\rho\|\bar{D}^{-1}\|)\|y-x\|$ and the above inequality implies

$$[1 + (1-\beta)/(\|L\|+\rho\|\bar{D}^{-1}\|)^2]\cdot\|\tilde{f}(y) - Ly - \tilde{f}(x) + Lx\|^2 \le 2\langle y - x, \tilde{f}(y) - Ly - \tilde{f}(x) + Lx\rangle.$$

Q.E.D.

## 4. Application to Affine Variational Inequalities

Consider the special case of VI(X,f) where f is affine, i.e.

$$f(x) = Qx + q, \qquad \forall\, x \in X, \tag{4.1}$$

where Q is an $n \times n$ positive semi-definite matrix and q is an n-vector. We choose a positive definite matrix D as before and denote

$$E = Q - D, \qquad H = \bar{D}^{-1/2} E \bar{D}^{-1/2}. \tag{4.2}$$

Let L and $\tilde{f}$ be given by, respectively, (3.1) and (3.2). We strengthen Proposition 3 as follows:

**Proposition 4** If $\|I + H\| < 2$ and E is symmetric, then $\tilde{f} - L$ is co-coercive with modulus greater than 1/2.

**Proof:** Let $P = I + H$. Then (by (3.1)-(3.2), (4.1)-(4.2)) $\tilde{f}(y) - Ly - \tilde{f}(x) + Lx = P(y-x)$ and P is symmetric positive semi-definite (since H is symmetric and $\langle z , Pz\rangle = \langle \bar{D}^{-1/2}z , Q\bar{D}^{-1/2}z\rangle \ge 0$ for all $z \in \Re^n$), so that $P^{1/2}$ exists. Hence

$$\begin{aligned} \|Pz\|^2 &\le \|P\|\cdot\|P^{1/2}z\|^2 \\ &= \|P\|\cdot\langle z , Pz\rangle, \quad \forall\, z \in \Re^n, \end{aligned}$$

and $\tilde{f} - L$ is co-coercive with modulus $1/\|P\|$. Q.E.D.

[Notice that if both D and Q are symmetric, then $I + H = D^{-1/2}QD^{-1/2}$.] We show below that the condition $\|I + H\| < 2$ can always be satisfied by choosing D appropriately. This is a crucial step towards the development of useful algorithms.

**Proposition 5** Let $D = \gamma I - F$, where F is any n×n matrix. Then, for all $\gamma$ sufficiently large ($\gamma$ depends on F and Q only), $\|I + H\| < 2$.

**Proof:** First note that $\bar{D} = \gamma I - \bar{F}$. Hence, for all $\gamma > \|\bar{F}\|$, we can expand $\bar{D}^{-1/2}$ as

$$\begin{aligned} \bar{D}^{-1/2} &= \gamma^{-1/2}\cdot[\, I - \bar{F}/\gamma\, ]^{-1/2} \\ &= \gamma^{-1/2}\cdot[\, I + \alpha_1\cdot\bar{F}/\gamma + \alpha_2\cdot(\bar{F}/\gamma)^2 + \alpha_3\cdot(\bar{F}/\gamma)^3 + \ldots\, ], \end{aligned}$$

where $\alpha_k = (1\cdot3\cdot\ldots\cdot(2k-1))/(k!\cdot2^k)$, $k = 1, 2, \ldots$. Since the $\alpha_k$'s are bounded (in fact between 0 and 1/2), we can write

$$\bar{D}^{-1/2} = \gamma^{-1/2}\cdot[\, I + \alpha_1\cdot\bar{F}/\gamma + O(1/\gamma^2)\, ],$$

where $O(\varepsilon) \to 0$ as $\varepsilon \to 0$ (the constant in the $O(\cdot)$ notation depends on F and Q only). Direct multiplication yields (cf. (4.2))

$$\begin{aligned} \bar{D}^{-1/2}E\bar{D}^{-1/2} &= \gamma^{-1}\cdot[\, I + \alpha_1\cdot\bar{F}/\gamma + O(1/\gamma^2)\, ]\cdot[Q + F - \gamma I]\cdot[\, I + \alpha_1\cdot\bar{F}/\gamma + O(1/\gamma^2)\, ] \\ &= -I + O(1/\gamma). \end{aligned}$$

Hence, for all $\gamma$ sufficiently large, $\|I + \bar{D}^{-1/2}E\bar{D}^{-1/2}\| < 2$. Q.E.D.

[A more refined analysis shows that $\gamma > 4(\|Q\|+\|F\|+\|\bar{F}\|)$ is sufficient.] Also, more generally, we can choose $D = \gamma M - F$, where M is any n×n symmetric positive definite matrix, and the conclusion of Proposition 5 can be seen to hold still.

Propositions 2, 4 and 5 motivate the following strategy for choosing the matrix D in the AP iteration (1.1): First choose an F for which Q + F is symmetric; then choose D =

$\gamma I - F$, where $\gamma$ is any positive scalar for which $\|I + H\| < 2$ (H is given by (4.2)). [In practice, $\gamma$ must be estimated, perhaps dynamically.] Since we are free to choose F, we may choose F to simplify the AP iteration. In particular, we may choose F to be either upper or lower triangular (e.g. $F = R^T - S$, where R (S) denotes the upper (lower) triangular part of Q). For linear complementarity problems, this gives rise to a method that is simple and, for quadratic programs, highly parallelizable (see §6).

How does Proposition 3 compare with Proposition 4 in this affine case? It can be seen from (3.1), (4.1)-(4.2) that $\tilde{f}(y) - \tilde{f}(x) - \bar{D}^{-1/2}D\bar{D}^{-1/2}(y-x) = H(y-x)$, so that the convergence condition in Proposition 3 is equivalent to $\|H\| < 1$, which is stronger than the condition $\|I + H\| < 2$. In fact, $\|H\| < 1$ implies

$$\begin{aligned}\langle Qz\,,z\rangle &= \|\bar{D}^{1/2}z\|^2 + \langle H\bar{D}^{1/2}z\,,\bar{D}^{1/2}z\rangle \\ &\geq \|\bar{D}^{1/2}z\|^2 - \|H\|\cdot\|\bar{D}^{1/2}z\|^2 \\ &> 0, \quad \forall\, z \in \Re^n,\end{aligned}$$

so that Q is <u>positive definite</u>. [However, the contraction condition $\|H\| < 1$ can be extended to general linearization methods (see [PaC82], [Daf83]) while an analogous extension does not seem possible with the condition $\|I + H\| < 2$.]

The condition $\|H\| < 1$ is not always easy to verify, especially if the size of H is large. Below we show that if Q is also positive definite and E is symmetric, then $\|H\| < 1$ is equivalent to the simpler condition that $D - E$ be positive definite. This result is a slight generalization of that stated in [LiP87, §3.2] for the case where D is symmetric and diagonal. For completeness we include it here.

**Proposition 6** For any $n\times n$ symmetric matrix E and any $n\times n$ positive definite matrix D such that $D + E$ is positive definite, $\|\bar{D}^{-1/2}E\bar{D}^{-1/2}\| < 1$ if and only if $D - E$ is positive definite.

**Proof:** Suppose that $\|\bar{D}^{-1/2}E\bar{D}^{-1/2}\| < 1$. Then

$$\begin{aligned}\langle (D-E)z\,,z\rangle &= \langle \bar{D}z\,,z\rangle - \langle Ez\,,z\rangle \\ &\geq \|\bar{D}^{1/2}z\|^2 - \|\bar{D}^{-1/2}E\bar{D}^{-1/2}\|\cdot\|\bar{D}^{1/2}z\|^2\end{aligned}$$

$$\geq (1 - \|\bar{D}^{-1/2}E\bar{D}^{-1/2}\|)\cdot\rho\cdot\|z\|^2, \qquad \forall\, z\in\Re^n,$$

where $\rho$ denotes the smallest eigenvalue of $\bar{D}$ and the first inequality follows from the Cauchy-Schwartz inequality.

Now suppose that $D - E$ is positive definite. Since $D + E$ is also positive definite, there exists some $\sigma > 0$ such that

$$\begin{aligned} \langle \bar{D}z , z\rangle &\geq \langle Ez , z\rangle + \sigma\|z\|^2, && \forall\, z\in\Re^n, \\ \langle \bar{D}z , z\rangle &\geq -\langle Ez , z\rangle + \sigma\|z\|^2, && \forall\, z\in\Re^n. \end{aligned}$$

Letting $w = \bar{D}^{1/2}z$ and using the inequality $\|\bar{D}^{-1/2}w\|^2 \geq \|w\|^2/\|\bar{D}\|$, we obtain

$$(1-\sigma/\|\bar{D}\|)\cdot\|w\|^2 \geq |\langle\bar{D}^{-1/2}E\bar{D}^{-1/2}w, w\rangle|, \qquad \forall\, w\in\Re^n.$$

Since for any $n\times n$ symmetric matrix $A$, $\|A\| = \max_{\|w\|=1} |\langle Aw , w\rangle|$, this implies

$$(1-\sigma/\|\bar{D}\|) \geq \|\bar{D}^{-1/2}E\bar{D}^{-1/2}\|.$$

Q.E.D.

## 5. Decomposition in Generalized Linear/Quadratic Programming

Consider the the affine variational inequality problem (cf. (4.1))

$$\text{Find an } x^*\in X \text{ satisfying } \langle Qx^* + q , x - x^*\rangle \geq 0, \ \forall\, x\in X, \tag{5.1}$$

and suppose that $Q$, $q$ and $X$ have the following special form

$$Q = \begin{bmatrix} G & A \\ -A^T & H \end{bmatrix}, \qquad q = \begin{bmatrix} -b \\ c \end{bmatrix}, \qquad X = V\times W, \tag{5.2}$$

where A is an $k \times m$ matrix, G and H are, respectively, $k \times k$ and $m \times m$ symmetric positive semi-definite matrices, b is an k-vector, c is an m-vector, and V and W are closed convex set in, respectively, $\Re^k$ and $\Re^m$. The matrix Q can be seen to be positive semi-definite (but not positive definite). This problem contains a number of important problems as special cases. For example, if both V and W are polyhedral sets, then this problem reduces to the generalized linear-quadratic programming problem considered by Rockafellar [Roc87] (see Application 3 below). As another example, if $G = 0$ and $V = \Re^k$, then this problem reduces to the following convex program with quadratic costs

$$\begin{aligned} &\text{Minimize} && \langle w , Hw\rangle/2 + \langle c , w\rangle \\ &\text{subject to} && Aw = b, \ w \in W. \end{aligned} \tag{5.3}$$

[This can be seen by attaching a Lagrange multiplier vector to the constraints $Aw = b$ and writing down the corresponding Kuhn-Tucker conditions for (5.3).] If $G = 0$ and V is the non-negative orthant in $\Re^k$, then this problem reduces to (5.3) with the equality constraints $Aw = b$ replaced by the inequality constraints $Aw \geq b$. The convex program (5.3) is an important problem in optimization, including as special cases ordinary linear/quadratic programs as well as the problems of projecting on to a convex set [Han88] and of finding a point in the intersection of convex sets [GPR67] (see Applications 1 and 2 below).

The special structures of Q and X motivate a number of very interesting applications of the AP algorithm. Suppose that we choose $D = \gamma I - F$ (cf. Proposition 5), where $\gamma$ is some yet to be determined scalar and F is the $(k+m)\times(k+m)$ matrix

$$F = \begin{bmatrix} 0 & 0 \\ 2A^T & 0 \end{bmatrix}.$$

Then $Q - D$ is symmetric, and the AP iteration (1.1) applied to solve (5.1)-(5.2) reduces to the following:

$$p^{r+1} = \operatorname{argmin}_{p \in V}\{\ \gamma\|p - p^r\|^2/2 + \langle Gp^r - b + Aw^r , p\rangle\ \}, \tag{5.4a}$$

$$w^{r+1} = \operatorname{argmin}_{w \in W}\{\ \gamma\|w - w^r\|^2/2 + \langle Hw^r + c - 2A^Tp^{r+1} + A^Tp^r , w\rangle\ \}. \tag{5.4b}$$

From Propositions 2, 4 and 5 we have the following result:

**Proposition 7** Suppose that (5.1)-(5.2) has a solution. Then, for all $\gamma$ sufficiently large ($\gamma$ depends on A, G and H only), the sequence $\{(w^r,p^r)\}$ generated by (5.4a)-(5.4b) converges to a solution of (5.1)-(5.2).

Notice that the iteration (5.4a)-(5.4b) involves the minimization of a separable, strictly convex quadratic function over, respectively, V and W. The separability of the cost function is an important feature of this iteration. For example, if W has special structures, say, W is a Cartesian product of sets in lower dimension spaces, i.e.,

$$W = W_1 \times W_2 \times \ldots \times W_h,$$

where each $W_i$ is a closed convex set in $\Re^{m_i}$ ($m_1 + m_2 + \ldots + m_h = m$), then (5.4b) decomposes into h separate problems. As another example, if W is a polyhedral set given by $W = \{ w \in \Re^m \mid Bw = d, 0 \le w \le u \}$ for some matrix B and vectors d and u of appropriate dimensions, then (5.4b) is a special case of <u>monotropic programming</u> problem [Roc84] and can be solved by one of many methods.

Some alternative choices of F also lead to interesting methods. For example, if we choose F to be the (k+m)×(k+m) matrix comprising $-2A$ in its top right corner and zero entries everywhere else, then the AP iteration (1.1) reduces to the following:

$$w^{r+1} = \mathrm{argmin}_{w \in W}\{ \gamma\|w - w^r\|^2/2 + \langle Hw^r + c - A^Tp^r , w\rangle \}, \tag{5.5a}$$
$$p^{r+1} = \mathrm{argmin}_{p \in V}\{ \gamma\|p - p^r\|^2/2 + \langle Gp^r - b + 2Aw^{r+1} - Aw^r , w\rangle \}. \tag{5.5a}$$

As another example, we can choose F to have the form

$$\begin{bmatrix} -F' & 0 \\ 2A^T & -F'' \end{bmatrix},$$

where F′ and F″ are, respectively, some k×k and m×m symmetric positive semi-definite matrices. For certain choices of F′ and F″ (e.g. F′ = G, F″ = H), this may significantly improve the rate of convergence. Problem decomposition can still be achieved if we choose $F'$ and $F''$ to be block diagonal according to the Cartesian product structure of, respectively, V and W.

We can also alternate between the two iterations (5.4a)-(5.4b) and (5.5a)-(5.5b) similar to Aitken's double sweep method. This typically accelerates the convergence in practice (although we do not have a convergence proof for this mixed method).

**Application 1** (Linear/Quadratic Programming)

Consider the convex program (5.3). As we noted earlier, this program can be formulated in the form of (5.1)-(5.2) with $G = 0$ and $V = \Re^k$. Hence by applying (5.4a)-(5.4b) we obtain the following iteration for solving (5.3):

$$p^{r+1} = p^r + (b - Aw^r)/\gamma, \tag{5.6a}$$
$$w^{r+1} = \mathrm{argmin}_{w \in W}\{ \gamma\|w - w^r\|^2/2 + \langle Hw^r + c - 2A^Tp^{r+1} + A^Tp^r , w\rangle \}. \tag{5.6b}$$

By Proposition 7, if (5.3) has an optimal solution and there exists an optimal Lagrange multiplier vector associated with the constraints $Aw = b$, then, for all $\gamma$ sufficiently large, the sequence $\{(w^r,p^r)\}$ generated by (5.6a)-(5.6b) converges to an optimal primal dual pair. The iteration (5.6a)-(5.6b), in constrast to existing decomposition methods for solving (5.3), has the important feature that it simultaneously dualizes the constraints and diagonalizes the cost function. This allows the problem to be decomposed according to the structure of W only, independent of the structures of H and A. [This feature of diagonalizing the cost function is reminiscent of an algorithm of Feijoo and Meyer [FeM88], but the Feijoo-Meyer algorithm cannot dualize any constraint and must use an additional line search at each iteration to ensure convergence.]

To illustrate the computational advantages of the iteration (5.6a)-(5.6b), consider the special case of (5.3) where W is a box in $\Re^m$ (this is the convex quadratic program in standard form). In this case the iteration (5.6a)-(5.6b) reduces to:

$$p^{r+1} = p^r + (b - Aw^r)/\gamma,$$
$$w^{r+1} = [\, w^r + (-Hw^r - c + 2A^Tp^{r+1} - A^Tp^r)/\gamma \,]^+,$$

where $[\cdot]^+$ denotes the orthogonal projection onto W. The above iteration is simple and massively parallelizable.

**Application 2** (Projection onto the Intersection of Convex Sets)

Consider the problem of computing the orthogonal projection of a vector $\bar{z}$ in $\Re^m$ onto the intersection of a finite number of closed convex sets $Z_1, Z_2, \ldots, Z_h$ in $\Re^m$ ($m \geq 1$, $h \geq 1$) [Han88]. This problem can be written in the form of (5.3) as

$$\begin{array}{llll} \text{Minimize} & \|z_1 - \bar{z}\|^2 \; + & \|z_2 - \bar{z}\|^2 \; + \; \ldots \; + & \|z_h - \bar{z}\|^2 \\ \text{subject to} & z_1 - z_2 = 0, & z_1 - z_3 = 0, \; \ldots, & z_1 - z_h = 0, \\ & z_1 \in Z_1, & z_2 \in Z_2, \quad\quad \ldots, & z_h \in Z_h, \end{array}$$

where each $z_i$ is an auxiliary vector in $\Re^m$. Direct application of (5.6a)-(5.6b) to this problem produces the following iteration:

$$\begin{aligned} p_i^{r+1} &= p_i^r + (z_i^r - z_1^r)/\gamma, && i = 2, 3, \ldots, h, \\ z_i^{r+1} &= [\, z_i^r - (\, z_i^r - \bar{z} + 2p_i^{r+1} - p_i^r \,)/\gamma \,]_i^+, && i = 2, 3, \ldots, h, \\ z_1^{r+1} &= [\, z_1^r - (\, z_1^r - \bar{z} - \sum_{i=2}^{h} (2p_i^{r+1} + p_i^r \,))/\gamma \,]_1^+, \end{aligned}$$

where $[\cdot]_i^+$ denotes the orthogonal projection onto $Z_i$. This iteration is highly parallelizable and, by Proposition 7, the sequence $\{z_1^r\}$ converges to the orthogonal projection of $\bar{z}$ onto $Z_1 \cap \ldots \cap Z_h$ if (i) $\gamma$ is sufficiently large, (ii) $Z_1 \cap \ldots \cap Z_h \neq \varnothing$, and (iii) there exists an optimal Lagrange multiplier vector associated with the constraints $z_1 - z_i = 0$ for all i. [The latter holds under a certain constraint qualification [Roc70, Theorem 30.4] (also see [Han88]).]

We can alternatively use an iteration based on (5.5a)-(5.5b) or use a different problem formulation (e.g., change the cost function or replace the constraints $z_1 - z_i = 0$ by $z_{i-1} - z_i = 0$ for all $i = 2, 3, \ldots, h$).

**Application 3** (Deterministic Control in Discrete Time)

Consider the following optimal control problem considered in [RoW87]

$$\text{Minimize} \quad \sum_{\tau=0}^{T} \langle u_\tau , P_\tau u_\tau\rangle/2 + \langle p_\tau , u_\tau\rangle + \langle c_{\tau+1} , x_\tau\rangle \tag{5.7}$$
$$+ \sum_{\tau=1}^{T} \rho_{V_\tau,Q_\tau}(q_\tau - C_\tau x_{\tau-1} - D_\tau u_\tau) + \rho_{V_{T+1},Q_{T+1}}(q_{T+1} - C_{T+1}x_T)$$

$$\begin{aligned} \text{subject to} \quad x_\tau &= A_\tau x_{\tau-1} + B_\tau u_\tau + b_\tau, && \tau = 1, \ldots, T, && (5.8a)\\ x_0 &= B_0 u_0 + b_0, && && (5.8b)\\ u_\tau &\in U_\tau, && \tau = 0, \ldots, T, \end{aligned}$$

where each $P_\tau$ is an $m_\tau \times m_\tau$ symmetric positive semi-definite matrix, each $Q_\tau$ is an $l_\tau \times l_\tau$ symmetric positive semi-definite matrix, each $A_\tau$ is an $k_\tau \times n_\tau$ matrix, each $B_\tau$ is an $k_\tau \times m_\tau$ matrix, each $C_\tau$ is an $l_\tau \times n_\tau$ matrix, each $D_\tau$ is an $l_\tau \times m_\tau$ matrix, each $U_\tau$ is a polyhedral set in $\Re^{m_\tau}$, each $V_\tau$ is a polyhedral set in $\Re^{l_\tau}$, and $b_\tau$, $c_\tau$, $p_\tau$, $q_\tau$ are vectors of appropriate dimensions. The function $\rho_{V_\tau,Q_\tau}:\Re^{l_\tau} \to (-\infty,\infty]$ denotes the polyhedral function

$$\rho_{V_\tau,Q_\tau}(\xi) \;=\; \max_{\eta \in V_\tau}\{ \langle \xi , \eta\rangle - \langle \eta , Q_\tau \eta\rangle/2 \}.$$

The vectors $x_0, x_1, \ldots, x_T$ are the states and the vectors $u_0, u_1, \ldots, u_T$ are the controls.

By attaching Lagrange multipliers to the system dynamic constraints (5.8a)-(5.8b) and writing each function $\rho_{V_\tau,Q_\tau}$ explicitly in terms of the Lagrangian that defines it, we can formulate the problem (5.7) in the form of (5.1)-(5.2), where

$$G = \begin{bmatrix} 0 &&&&&&& \\ & 0 &&&&&& \\ && 0 \ldots &&&&& \\ &&& 0 &&&& \\ &&&& Q_1 &&& \\ &&&&& Q_2 \ldots && \\ &&&&&& Q_T & \\ &&&&&&& Q_{T+1} \end{bmatrix}, \quad A = \begin{bmatrix} -I &&&& B_0 &&& \\ A_1 & -I &&&& B_1 && \\ & A_2 \ldots &&&&& B_2 \ldots & \\ && A_T & -I &&&& B_T \\ C_1 &&&& 0 & D_1 && \\ & C_2 \ldots &&&&& D_2 \ldots & \\ && C_T &&&&& D_T \\ &&& C_{T+1} &&&& 0 \end{bmatrix},$$

$$H = \begin{bmatrix} 0 & & & & & & & & \\ & 0 \ldots & & & & & & & \\ & & 0 & & & & & & \\ & & & 0 & & & & & \\ & & & & P_0 & & & & \\ & & & & & P_1 & & & \\ & & & & & & P_2 \ldots & & \\ & & & & & & & & P_T \end{bmatrix},$$

$$V = \Re^{k_0} \times \ldots \times \Re^{k_T} \times V_1 \times \ldots \times V_{T+1}, \qquad W = \Re^{n_0} \times \ldots \times \Re^{n_T} \times U_0 \times \ldots \times U_T,$$

and the vectors b and c are defined analogously using, respectively, the vectors $\{b_\tau\}$, $\{q_\tau\}$ and the vectors $\{c_\tau\}$, $\{p_\tau\}$. Let us apply the iteration (5.4a)-(5.4b) to solve this special case of (5.1)-(5.2). Then from the structures of V and W it can be seen that each (5.4a) and (5.4b) involves 2(T+1) separate subproblems, which can be solved in parallel. [For simplicity, we omit writing down this iteration. It suffices to notice that each subproblem involves the minimization of a separable, strictly convex quadratic function over either $\Re^{k_\tau}$ or $V_\tau$ or $\Re^{n_\tau}$ or $U_\tau$, for some $\tau$, which typically is easy to solve.] By Proposition 7, if (5.7) has an optimal solution and there exists optimal Lagrange multipliers associated with the constraints (5.8a)-(5.8b), then, for all $\gamma$ sufficiently large, the sequence of iterates generated by the iteration (5.4a)-(5.4b) (applied to solve this special case of (5.1)-(5.2)) converges. [Conditions for the existence of an optimal primal dual pair can be found in [Roc87].]

Notice that, in the above formulation, all of the system dynamic constraints are dualized (i.e., incorporated into the matrix A). Alternatively we can dualize only a subset of these constraints. For example, we can dualize only those constraints describing the system dynamic at <u>odd</u> time units. In this case, W becomes $W_0 \times W_2 \times \ldots \times W_{2h}$, where $h = \lfloor T/2 \rfloor$ and

$$W_0 = \{ (x_0, u_0) \mid x_0 = B_0 u_0 + b_0,\ u_0 \in U_0 \},$$
$$W_\tau = \{ (x_\tau, x_{\tau-1}, u_\tau) \mid x_\tau = A_\tau x_{\tau-1} + B_\tau u_\tau + b_\tau,\ u_\tau \in U_\tau \}, \quad \tau = 2, 4, \ldots, 2h.$$

[Of course, G, H, A, V, b and c also change accordingly.] Then the minimization problem in (5.4a) decomposes into h+1 separate subproblems. Although the number of separate subproblems is less in this formulation, the iteration will likely converge faster since only about half of the system dynamic constraints are dualized.

## 6. A Matrix Splitting Algorithm for Linear Complementarity Problems

Consider the problem

$$\text{Find an } x^* \in \Re^n \text{ satisfying} \quad x^* \geq 0, \;\; Qx^* + q \geq 0, \;\; \langle Qx^* + q \,, x^* \rangle = 0. \qquad \text{(LCP)}$$

where Q is an $n \times n$ positive semi-definite matrix and q is an n-vector. [The more general case where upper bound constraints are present may be treated analogously.] This problem, called the linear complementarity problem, is a classical problem in optimization (see [BaC78], [CGL80], [Man77], [Pan84]).

It is easily seen that (LCP) is a special case of VI(X,f) with $f(x) = Qx + q$ and X being the non-negative orthant in $\Re^n$. Let $D = \gamma I - F$ (cf. Proposition 5), where $\gamma$ is some yet to be determined scalar and $F = R^T - S$, where R (S) denotes the upper (lower) triangular part of Q. Then $Q - D = R + K + R^T - \gamma I$, where K denotes the $n \times n$ diagonal matrix whose j-th diagonal entry is the j-th diagonal entry of Q; hence $Q - D$ is symmetric. Consider applying the AP iteration (1.1) to solve this special case of VI(X,f) with the above choice of D. Then we obtain the following Gauss-Seidel iteration:

$$x_i^{r+1} = [\, x_i^r + (\Sigma_{j<i} (Q_{ji} - Q_{ij}) x_j^{r+1} - \Sigma_{j<i} Q_{ji} x_j^r - \Sigma_{j \geq i} Q_{ij} x_j^r - q_i)/\gamma \,]^+, \qquad i = 1, 2, \ldots, n,$$

where $Q_{ij}$ denotes the (i,j)-th entry of Q, $q_i$ denotes the i-th coordinate of q, and $[\cdot]^+$ denotes the projection onto the interval $[0,\infty)$. The above iteration is simple and, if Q has a certain special structure (e.g. Q given by (5.2) with $G = 0$ and q given by (5.2), which corresponds to the (LCP) formulation of the quadratic program $\min\{\langle w \,, Hw \rangle/2 + \langle c \,, w \rangle \mid Aw \geq b, w \geq 0 \}$), also highly parallelizable. Since $Q - D$ is symmetric, it follows from Propositions 2, 4 and 5 that if (LCP) has a solution, then, for all sufficiently large $\gamma$ ($\gamma$ depends on Q only), $\{x^r\}$ generated by the above iteration converges to a solution of

(LCP). The algorithm based on this iteration, in the terminology of [Pan84] (also see [LiP87]), is a matrix splitting algorithm. However, in contrast to existing matrix splitting algorithms, this algorithm converges without any additional assumption on the problem data (such as symmetry of Q).

We can alternatively choose F to be the matrix $F = S^T - R$. Depending on the structure of the matrix A, this choice may be more advantageous.

**References**

[AhH82] Ahn, B. H. and Hogan, W. W., "On Convergence of the PiES Algorithm for Computing Equilibria," *Oper. Res.*, 30 (1982), 281-300.

[Aus76] Auslender, A., *Optimization: Méthodes Numeriques*, Mason, Paris (1976).

[BaC78] Balinski, M. L. and Cottle, R. W. (eds.), *Mathematical Programming Study 7: Complementarity and Fixed Point Problems*, North-Holland, Amsterdam (1987).

[Ber82] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY (1982).

[BeT89] Bertsekas, D. P. and Tsitsiklis, J. N., *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ (1989).

[Bré73] Brézis, H., *Opérateurs Maximaux Monotones*, North-Holland, Amsterdam (1973).

[Bru75] Bruck, Jr., R. E., "An Iterative Solution of a Variational Inequality for Certain Monotone Operators in Hilbert Space," *Bull. A. M. S.*, 81 (1975), 890-892.

[CGL80] Cottle, R. W., Giannessi, F., and Lions, J-L. (eds.), *Variational Inequalities and Complementarity Problems: Theory and Applications*, Wiley, New York, NY (1980).

[Daf83] Dafermos, S., "An Iterative Scheme for Variational Inequalities," *Math. Prog.*, 26 (1983), 40-47.

[Dei85] Deimling, K., *Nonlinear Functional Analysis*, Springer-Verlag, Berlin (1985).

[FeM88] Feijoo, B. and Meyer, R. R., "Piecewise-Linear Approximation Methods for Nonseparable Convex Programming," *Management Science*, 34 (1988), 411-419.

[Gab83] Gabay, D., "Applications of the Method of Multipliers to Variational Inequalities," in (M. Fortin and R. Glowinski, eds.) *Augmented Lagrangian Methods:*

*Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam (1983).

[GLT81] Glowinski, R., Lions, J.L., and Tremolieres, R., *Numerical Analysis of Variational Inequalities*, North-Holland, Amsterdam (1981).

[Gol66] Goldstein, A. A., "Convex Programming in Hilbert Spaces," *Bull. A. M. S.*, 70 (1964), 709-710.

[GPR67] Gubin, L. G., Polyak, B. T. and Raik, E. V., "The Method of Projections for Finding the Common Point of Convex Sets," *Z. Vycisl. Mat. i Mat. Fiz.*, 7 (1967), 1211-1228; *USSR Comp. Math. and Math. Phys.*, 6 (1967), 1-24.

[Han88] Han, S. P., "A Successive Projection Method," *Math. Prog.*, 40 (1988), 1-14.

[HaL88] Han, S. P. and Lou, G., "A Parallel Algorithm for a Class of Convex Programs," *SIAM J. Contr. & Optim.*, 26 (1988), 345-355.

[KiS80] Kinderlehrer, D. and Stampacchia, G., *An Introduction to Variational Inequalities and Applications*, Academic Press, New York, NY (1980).

[Lem88] Lemaire, B., "Coupling Optimization Methods and Variational Convergence," in *Trends in Mathematical Optimization* (K.-H. Hoffman, J.-B. Hiriart-Urruty, J. Zowe, C. Lemarechal, eds.), Birkhäuser Verlag, Basel (1988), 163-179.

[LiP87] Lin, Y.Y. and Pang, J.-S., "Iterative Methods for Large Convex Quadratic Programs: A Survey," *SIAM J. Contr. & Optim.*, 25 (1987), 383-411.

[LiM79] Lions, P.L. and Mercier, B., "Splitting Algorithms for the Sum of Two Nonlinear Operators," *SIAM J. Num. Anal.*, 16 (1979), 964-979.

[Man77] Mangasarian, O. L., "Solution of Symmetric Linear Complementarity Problems by Iterative Methods," *J. Optim. Theory & Appl.*, 22 (1977), 465-485.

[Min62] Minty, G. J., "Monotone (Nonlinear) Operators in Hilbert Space," *Duke Math. J.*, 29 (1962), 341-346.

[Min64] Minty, G. J., "On the Monotonicity of the Gradient of a Convex Function," *Pacific J. Math.*, 14 (1964), 243-247.

[Mor65] Moreau, J. J., "Proximité et dualité dans un espace Hilbertien," *Bull. Soc. Math. France*, 93 (1965), 273-299.

[PaC82] Pang, J.-S. and Chan, D., "Iterative Methods for Variational Inequality and Complementarity Problems," *Math. Prog.*, 24 (1982), 284-313.

[Pan84] Pang, J.-S., "Necessary and Sufficient Conditions for the Convergence of Iterative Methods for the Linear Complementarity Problem," *J. Optim. Theory & Appl.*, 42 (1984), 1-17.

[Pan85] Pang, J.-S., "Asymmetric Variational Inequality Problems over Product Sets: Applications and Iterative Methods," *Math. Prog.*, 31 (1985), 206-219.

[Pas79] Passty, G.B., "Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Space," *J. Math. Anal. & Appl.*, 72 (1979), 383-390.

[Roc70a] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, NJ (1970).

[Roc70b] Rockafellar, R.T., "On the Maximality of Sums of Nonlinear Monotone Operators," *Trans. A. M. S.*, 149 (1970), 75-88.

[Roc76] Rockafellar, R.T., "Monotone Operators and the Proximal Point Algorithm," *SIAM J. Contr. & Optim.*, 14 (1976), 877-898.

[Roc84] Rockafellar R.T., *Network Flows and Monotropic Optimization*, Wiley-Interscience, N.Y., NY (1984).

[Roc87] Rockafellar, R.T., "Generalized Linear-Quadratic Programming and Optimal Control," *SIAM J. Contr. & Optim.*, 25 (1987), 781-814.

[RoW87] Rockafellar, R.T. and Wets, R.J-B., "Generalized Linear-Quadratic Problems of Deterministic and Stochastic Optimal Control in Discrete Time," Technical report, Dept. of Mathematics, University of Washington, Seattle, WA (1987), 63-93.

[Sib70] Sibony, M., "Méthodes itératives pour les équations et inéquations aux dérivées partielles nonlinéaires de type monotone," *Calcolo*, 2 (1970), 65-183.

[Tse88] Tseng, P., "Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities," LIDS Report P-1836, M.I.T., Cambridge, MA (November 1988).