November 5, 1988 LIDS-P-1836

Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities[1]

by

Paul Tseng[2]

Abstract

Recently Han and Lou [18] proposed a highly parallelizable decomposition algorithm for convex programming involving strongly convex costs. We show in this paper that their algorithm, as well as the method of multipliers [17, 19, 34] and the dual gradient method [8, 40], are special cases of a certain multiplier method for separable convex programming. This multiplier method is similar to the alternating direction method of multipliers [10, 15] but uses both Lagrangian and augmented Lagrangian functions. We also apply this method to symmetric linear complementarity problems to obtain a new class of matrix splitting algorithms. Finally, we show that this method is itself a dual application of an algorithm of Gabay [12] for finding a zero of the sum of two maximal monotone operators. We give an extension of Gabay's algorithm that allows dynamic stepsizes and show that, under certain conditions, it has a linear rate of convergence. We also apply this algorithm to variational inequalities.

KEY WORDS: maximal monotone operator, augmented Lagrangian, alternating minimization

---

[1] 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, grant NSF-ECS-8519058.

[2] Laboratory for Information and Decision Systems, Room 35-205, Massachusetts Institute of Technology, Cambridge, MA 02139.

Acknowledgement: Thanks are due to D. P. Bertsekas and J. Eckstein for their many helpful comments.

1. Introduction

One of the most important applications of convex duality theory is in decomposition algorithms for solving problems with special structure. A canonical example is the following separable convex programming problem

$$\text{Minimize} \quad f(x) + g(z) \tag{1.1}$$
$$\text{Subject to} \quad Ax + Bz = b, \tag{1.2}$$

where $f:\Re^n \to (-\infty,\infty]$ and $g:\Re^m \to (-\infty,\infty]$ are given convex functions, A is a given rxn matrix, B is a given rxm matrix, and b is a given vector in $\Re^r$. In our notation, all vectors are column vectors and superscript T denotes the transpose. We will denote by $\langle\cdot,\cdot\rangle$ the usual Euclidean inner product and $\|\cdot\|$ its induced norm, i.e. $\|x\|^2 = \langle x,x\rangle$.

By attaching a Lagrange multiplier vector $p \in \Re^r$ to the constraints (1.2), the problem (1.1) can be decomposed into two independent problems involving, respectively, x and z. One algorithm based on this dual approach, proposed by Uzawa [40] and others, operates by successively minimizing the Lagrangian function

$$L(x,z,p) \;=\; f(x) + g(z) + \langle p, b-Ax-Bz\rangle.$$

with respect to x and z (with p fixed) and then updating the multiplier by the iteration

$$p \;:=\; p + c(b - Ax - Bz),$$

where c is a positive stepsize. [We assume for the sake of discussion that the minimum above is attained.] It can be shown that this algorithm is convergent if both f and g are strictly convex and c is chosen to be sufficiently small. [In this case the dual functional defined by $q(p) = \min_{x,z} L(x,z,p)$ is differentiable and this algorithm can be viewed as a gradient method for maximizing q.]

Unfortunately, for many problems of interest, the function f may be strictly convex but not g. This is particularly the case when a problem is transformed in a way to bring about a structure that is favorable for decomposition (see §4 for an example). A solution to this difficulty is suggested by a recent work of Han and Lou. In [18] they proposed a decomposition algorithm for minimizing a strongly convex function over the intersection of a finite number of closed convex sets. It can be shown, by introducing auxiliary variables, that this convex program is a special case of (1.1). Moreover, it can be shown (see §4) that their algorithm is similar to the dual gradient method above, except for the key difference that the Lagrangian function is replaced by an augmented Lagrangian function when the minimization is taken with respect to z.

In this paper we generalize the Han and Lou algorithm to solve the general problem (1.1). [The main interest here is in problems where f is strongly convex and separable but g is not strictly convex.] At each iteration of our algorithm, the Lagrangian L(x,z,p) is first minimized with respect to x (with z and p held fixed), and then the augmented Lagrangian

$$L_c(x,z,p) = L(x,z,p) + c\|Ax+Bz-b\|^2/2$$

is minimized with respect to z (with x and p held fixed). Finally the multiplier is updated according to the usual augmented Lagrangian iteration

$$p := p + c(b - Ax - Bz)$$

and the process is repeated. This algorithm, which we call alternating minimization algorithm, has the nice feature that, if B has full column rank, then both minimizations involve strongly convex objective functions. Moreover, if f is separable (in addition to being strongly convex), the first minimization is also separable – a feature that makes this algorithm particularly suitable for problems where f is separable and g is such that the minimization of the augmented Lagrangian with respect to z is easily carried out. The alternating minimization algorithm is a very useful method for decomposition. Indeed, as we shall see, it contains as special cases (in addition to the algorithm of Han and Lou) the dual gradient method, the method of multipliers, and a class of matrix splitting algorithms for symmetric linear complementarity problems.

Our method should be contrasted with the alternating direction method of multipliers, proposed by Gabay-Mercier [10], Glowinski-Marrocco [15] and extended by Gabay [11] (see also [2, 7, 9, 14, 39]), which is another multiplier method that alternates between minimization with respect to x and minimization with respect to z. The only difference between the two methods is that at each iteration of the alternating direction method of multipliers, x is updated by minimizing the augmented Lagrangian rather than the Lagrangian function as in our method. The quadratic term of the augmented Lagrangian affects adversely the decomposition of the minimization with respect to x based on separability properties of f, and this is an advantage for our method. On the other hand, in contrast with the alternating direction method of multipliers, the penalty parameter c in our method must be chosen from a restricted range (as will be seen later), usually through trial and error.

It turns out however that the alternating minimization algorithm is itself a special case of an algorithm analyzed by Gabay [12] for finding a zero of the sum of two maximal monotone operators. [Such operators have been studied extensively because of their role in convex analysis and certain partial differential equations. Finding a zero of the sum of these operators is a fundamental problem (see also [3, 6, 38, 23]).] Let $\Pi:\Re^r\to\Re^r$ and $\Psi:\Re^r\to\Re^r$ denote two arbitrary maximal monotone operators and suppose that Π^{-1} is also strongly monotone. The algorithm of Gabay computes a zero of $\Pi + \Psi$ by successively applying the iteration

$$p := [I+c\Psi]^{-1}[I-c\Pi]p,$$

where c is some fixed, sufficiently small stepsize. We will give a proof of convergence for the above algorithm – different from the one given by Gabay – that also provides an estimate of the rate of convergence and does not require the stepsizes to be fixed. Gabay also considered applications of his algorithm to decomposition, but limited his applications to the case where either Ψ or Ψ^{-1} is the subdifferential of the indicator function for a convex set (an example is the gradient projection method of Goldstein [16]).

This paper is organized as follows: in §2 we describe the general algorithm for finding a zero of the sum of two maximal monotone operators and analyze its convergence properties. In §3 we apply this algorithm to the separable convex program (1.1) to derive the alternating minimization algorithm. In §4 and §5 we show that the algorithm of Han and Lou, the method of multipliers, and the dual gradient method can be obtained as special cases of the alternating minimization algorithm. In §6 we apply the alternating minimization algorithm to the symmetric linear complementarity problem to obtain a new class of matrix splitting algorithms. In §7 we apply the general algorithm of §2 to variational inequalities.

Before preceding to the next section, let us familiarize ourselves with the notation that is used throughout this paper. For any real symmetric matrix E, we denote by $\rho(E)$ the <u>spectral radius</u> of E, i.e. $\rho(E)$ is the square root of the largest eigenvalue of E^TE. For any set Ω, we denote by $\delta(\cdot|\Omega)$ the <u>indicator</u> function for Ω, i.e. $\delta(x|\Omega)$ is zero if $x\in\Omega$ and is ∞ otherwise. For any convex function $h:\Re^h\to(-\infty,\infty]$ and any $x\in\Re^h$, we denote by $\partial h(x)$ the <u>subdifferential</u> of h at x. A multifunction $T:\Re^h\to\Re^h$ is said to be a <u>monotone operator</u> 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

$$\{\ (x,y)\in \Re^h\times\Re^h \mid y\in T(x)\ \}$$

is not properly contained in the graph of any other monotone operator $T':\Re^h\to\Re^h$. We denote by T^{-1} the inverse of T, i.e.

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

It is easily seen from symmetry that the inverse of a maximal monotone operator is also a maximal monotone operator. For any monotone operator T, we will mean by the modulus of T the largest (nonnegative) scalar σ such that

$$\langle y-y',x-x'\rangle \geq \sigma\|x-x'\|^2 \qquad \text{whenever} \qquad y\in T(x),\ y'\in T(x').$$

We say that T is strongly monotone if its modulus is positive.

2. A Splitting Algorithm for the Sum of Two Maximal Monotone Operators

In this section we consider the general problem of finding a zero of the sum of two maximal monotone operators, with the inverse of one of them being strongly monotone. We describe an extension of the algorithm by Gabay [12] for solving this problem and analyze its convergence. A number of applications of this algorithm will be given in subsequent sections.

Let $\Phi:\Re^n\to\Re^n$ and $\Gamma:\Re^m\to\Re^m$ be two arbitrary maximal monotone operators. Let A be a $r\times n$ matrix, B be a $r\times m$ matrix, and b be a vector in $\Re^r$. Consider the problem of finding a $p^*\in\Re^r$ satisfying

$$b\in A\Phi(A^Tp^*) + B\Gamma(B^Tp^*). \tag{2.1}$$

This problem can be shown to contain as a special case the convex program (1.1) (see discussion in §3). We make the following standing assumption:

Assumption A:

(a) Eq. (2.1) has a solution.

(b) Φ^{-1} is strongly monotone with modulus σ.

Notice that Assumption A (b) implies that $\Phi^{-1}-\sigma I$ is a maximal monotone operator. Hence a result of Minty [27] says that Φ is single valued and defined on all of $\Re^r$. Furthermore, the value of $\Phi(A^T p^*)$ is the same for all solutions p^* of (2.1). To see the latter, note that if both $\tilde{p}$ and p^* are solutions of (2.1), then there exists $\tilde{\gamma} \in \Gamma(B^T\tilde{p})$ such that $b = A\Phi(A^T\tilde{p}) + B\tilde{\gamma}$ and there exists $\gamma^* \in \Gamma(B^T p^*)$ such that $b = A\Phi(A^T p^*) + B\gamma^*$. Hence

$$
\begin{aligned}
0 &= \langle \Phi(A^T\tilde{p}) - \Phi(A^T p^*), A^T\tilde{p} - A^T p^* \rangle + \langle \tilde{\gamma} - \gamma^*, B^T\tilde{p} - B^T p^* \rangle \\
&\geq \sigma \| \Phi(A^T\tilde{p}) - \Phi(A^T p^*) \|^2,
\end{aligned}
$$

where the inequality follows from the monotonicity of Γ and the fact that Φ^{-1} has modulus σ. Since $\sigma > 0$, we have $\Phi(A^T\tilde{p}) = \Phi(A^T p^*)$. We will denote by x^* the vector $\Phi(A^T p^*)$.

We describe below our algorithm for solving (2.1). This algorithm, for any starting multiplier $p(0) \in \Re^r$, generates a sequence of three-tuples $\{(x(t), z(t), p(t))\}$ using the following iteration:

$$x(t) = \Phi(A^T p(t)), \tag{2.2a}$$

$$z(t) \in \Gamma(B^T[p(t) - c(t)(Ax(t) + Bz(t) - b)]), \tag{2.2b}$$

$$p(t+1) = p(t) + c(t)(b - Ax(t) - Bz(t)). \tag{2.2c}$$

The stepsizes $\{c(t)\}_{t=0,1,\ldots}$ is any sequence of scalars satisfying

$$\varepsilon \leq c(t) \leq 2\sigma/\rho(A^T A) - \varepsilon, \quad t = 0, 1, \ldots, \tag{2.2d}$$

where ε is any fixed positive scalar not exceeding $\sigma/\rho(A^T A)$. We will show that $z(t)$ is well defined below.

In [12] Gabay proposed an algorithm for finding a zero of the sum of two maximal monotone operators Π and Ψ, with Π^{-1} being strongly monotone. In his algorithm, a sequence

$\{p(t)\}$ is generated by applying a forward Euler step for Π followed by a backward Euler step for Ψ at each iteration, i.e.

$$p(t+1) = [I+c\Psi]^{-1}[I-c\Pi]p(t). \tag{2.3}$$

Gabay showed that, for any fixed positive c less than twice the modulus of Π^{-1}, the sequence $\{p(t)\}$ generated by (2.3) converges to a zero of $\Pi + \Psi$. We claim that the algorithm (2.2a)-(2.2d) is in fact an extension of Gabay's algorithm. To see this, we use (2.2c) to replace (2.2b) by

$$[p(t)-p(t+1)]/c(t) + b - Ax(t) \in B\Gamma(B^T p(t+1)).$$

Combining this with (2.2a), we obtain that

$$p(t+1) = [I+c(t)B\Gamma B^T]^{-1}(b + [I-c(t)A\Phi A^T]p(t)), \tag{2.4}$$

which, for c(t) fixed, $b = 0$, and A and B both being $r \times r$ identity matrices, is identical to the iteration (2.3). Algorithms such as this, where a step for $A\Phi A^T$ alternates with a step for $B\Gamma B^T$, are called (in the terminology of Lions-Mercier [23]) splitting algorithms.

To see that z(t) given by (2.2b) is well defined, note first that $B\Gamma B^T$ is itself a maximal monotone operator. Hence, by a result of Minty [27], the proximal mapping $[I+c(t)B\Gamma B^T]^{-1}$ is single valued and defined on all of $\Re^r$, and the iteration (2.4) is well defined. This in turn implies that $\Gamma(B^T p(t+1))$ is nonempty and therefore z(t) is well defined.

The main difference between Gabay's algorithm and the iteration (2.2a)-(2.2d) is that the latter allows the stepsize c(t) to vary with t. Below we present our convergence results for the iteration (2.2a)-(2.2d) (whose proof we give in Appendix A). These results sharpen those given by Gabay (cf. [12], Theorem 6.1).

Proposition 1 The sequences $\{x(t)\}$, $\{z(t)\}$, $\{p(t)\}$ generated by (2.2a)-(2.2d) satisfy:
(a) $x(t) \rightarrow x^*$.
(b) $Bz(t) \rightarrow b-Ax^*$.
(c) $p(t) \rightarrow$ a solution of (2.1).
(d) Let δ and η denote the modulus of $A\Phi A^T$ and $B\Gamma B^T$ respectively. Then, for $t = 1, 2, \ldots$,

$$\varepsilon^2\rho(A^TA)\|x(t)-x^*\|^2 + \varepsilon^2\|Bz(t)+Ax^*-b\|^2 \le \|p(t)-p^\infty\|^2$$
$$\le [(1-\delta^2\varepsilon^2)/(1+\eta^2\varepsilon^2)]\|p(t-1)-p^\infty\|^2,$$

where (cf. part (c)) p^∞ denotes the unique limit point of $\{p(t)\}$.

Notice that Proposition 1 (b) implies that if B has full row rank, then $\{z(t)\}$ converges. Proposition 1 (d) implies that if either $A\Phi A^T$ or $B\Gamma B^T$ is strongly monotone, then the rate of convergence of the sequence $\{(x(t),z(t),p(t))\}$ is linear. The proof of Proposition 1 is based on an argument used by Glowinski and Le Tallec [14] for the alternating direction method of multipliers (also see [2], §3.4.4). Also, in practice, exact solutions of the Eq. (2.2a) and (2.2b) are difficult to obtain. It can be seen from (3.3a)-(3.3b) and (3.4) that Proposition 1 holds even if the solutions of (2.2a) and (2.2b) are computed inexactly. Unfortunately the amount of inexactness allowable cannot be easily estimated. As a final remark, the results in this section also extend directly to problems defined on a Hilbert space.

3. Application to Separable Convex Programming: the Alternating Minimization Algorithm

Consider again the separable convex program (1.1)

$$\begin{array}{ll} \text{Minimize} & f(x) + g(z) \\ \text{Subject to} & Ax + Bz = b, \end{array} \tag{3.1}$$

where $f:\Re^n \to (-\infty,\infty]$, $g:\Re^m \to (-\infty,\infty]$ are given convex functions, A is a r×n matrix, B is a r×m matrix, and b is a vector in $\Re^r$. In this section we will derive the alternating minimization algorithm for solving (3.1) by applying the iteration (2.2a)-(2.2d). We make the following assumptions regarding (3.1):

Assumption B:

(a) f and g are convex lower semicontinuous functions.

(b) f is strongly convex with modulus $\alpha > 0$, i.e., for any $\lambda \in (0,1)$,

$$\lambda f(x) + (1-\lambda)f(y) - f(\lambda x+(1-\lambda)y) \ \ge\ \alpha\lambda(1-\lambda)\|x-y\|^2 \quad \forall\, x\in\Re^n,\ \forall\, y\in\Re^n. \tag{3.2}$$

(c) Problem (3.1) is feasible, i.e. there exists $x \in \Re^n$, $z \in \Re^m$ such that $f(x) + g(z) < \infty$ and $Ax+Bz = b$.

(d) The function $g(z)+\|Bz\|^2$ has a minimum.

Assumption B implies that problem (3.1) has an optimal solution. To see this, note that because f and g are lower semicontinuous and f is strongly convex, if (3.1) does not have an optimal solution, there must exist a $z \in \Re^m$ and a $w \in \Re^m$ such that $Bw = 0$ and $g(z+\lambda w)$ is strictly decreasing with $\lambda \geq 0$ – contradicting Assumption B (d). Moreover, the strict convexity of f implies that (3.1) has a unique optimal solution in x, which we denote by x^*.

Notice that Assumption B (d) holds if either g has a minimum or B has full column rank. If Assumption B (d) does not hold, but (3.1) has an optimal solution, we can define the perturbation function $h(w) = \inf\{ g(z) \mid w = Bz \}$, which is convex and proper. Then, if h is lower semicontinuous, we can instead solve the reduced problem $\min\{ f(x) + h(w) \mid Ax + w = b \}$, which can be seen to satisfy Assumption B. Upon obtaining x^*, we then solve $\min\{ g(z) \mid Bz = b-Ax^* \}$. For various properties of strongly convex functions see pp. 83 of [30].

By assigning a Lagrange multiplier vector $p \in \Re^r$ to the constraints $Ax + Bz = b$, we obtain the dual program (see [35], §28) of (3.1) to be

$$\begin{array}{ll} \text{Minimize} & \phi(A^T p) + \gamma(B^T p) - \langle b,p \rangle \\ \text{subject to} & p \in \Re^r, \end{array} \tag{3.3}$$

where $\phi:\Re^r \to (-\infty,\infty]$ and $\gamma:\Re^r \to (-\infty,\infty]$ are respectively the conjugate function of f and g, i.e.,

$$\phi(y) = \sup\{ \langle y,x \rangle - f(x) \},$$
$$\gamma(w) = \sup\{ \langle w,z \rangle - g(z) \}.$$

Both ϕ and γ are lower semicontinuous convex (see [35], §12) and, because f is strongly convex, ϕ is in addition real valued and differentiable (see [35], Corollary 13.3.1 and Theorem 26.3). We make the following assumption regarding (3.3):

Assumption C: The program (3.3) has an optimal solution, i.e., (3.1) has an optimal Lagrange multiplier vector corresponding to the constraints $Ax + Bz = b$.

Since the optimal objective value of (3.3) is not $+\infty$ by Assumption C, the function γ cannot be $+\infty$ everywhere. This, together with the fact that γ is lower semicontinuous convex, implies that $\partial\gamma$ is a maximal monotone operator (see Minty [28] or Moreau [29]). Because ϕ is convex and real valued, $\nabla\phi$ is also a maximal monotone operator. This, together with the observation that p^* is a solution of (3.3) if and only if p^* satisfies

$$0 \in A\nabla\phi(A^T p^*) + B\partial\gamma(B^T p^*) - b,$$

then implies that the dual program (3.3) is a special case of the general problem (2.1) with $\Phi = \nabla\phi$ and $\Gamma = \partial\gamma$. Furthermore, the strong convexity condition (3.2) implies that $\partial f = (\nabla\phi)^{-1}$ is strongly monotone with modulus 2α. Hence Assumption A holds for the above choice of Φ and Γ (with $\sigma = 2\alpha$) and we can apply the splitting algorithm (2.2a)-(2.2d) to solve this special case of (2.1). This produces the following algorithm, which we have named the alternating minimization algorithm earlier, for solving (3.1) and its dual (3.3):

$$x(t) = \operatorname{argmin}_x\{ f(x) - \langle p(t),Ax\rangle \}, \tag{3.4a}$$
$$z(t) = \operatorname{argmin}_z\{ g(z) - \langle p(t),Bz\rangle + c(t)\|Ax(t)+Bz-b\|^2/2 \}, \tag{3.4b}$$
$$p(t+1) = p(t) + c(t)(b-Ax(t)-Bz(t)), \tag{3.4c}$$

where $p(0)$ is any element of $\Re^r$, and $\{c(t)\}$ is any sequence of scalars satisfying

$$\varepsilon \le c(t) \le 4\alpha/\rho(A^TA) - \varepsilon, \quad t = 0,1,\ldots, \tag{3.4d}$$

and ε is any fixed positive scalar not exceeding $2\alpha/\rho(A^TA)$. In practice, the threshold $4\alpha/\rho(A^TA)$ will typically be unknown, and some trial and error may be required to select the sequence $c(t)$. This is a drawback of the method.

Convergence of the alternating minimization algorithm follows from Proposition 1:

Proposition 2 The sequences $\{x(t)\}$, $\{z(t)\}$, $(p(t)\}$ generated by (3.4a)-(3.4d) satisfy the following:

(a) $x(t) \to x^*$.

(b) $Bz(t) \to b-Ax^*$.

(c) $p(t) \to$ an optimal solution of the dual program (3.3).

(d) If either $A\nabla\phi A^T$ or $B\partial\gamma B^T$ is strongly monotone, then the rate of convergence of $\{(x(t),Bz(t),p(t))\}$ is linear.

(e) If the convex function $g(z)+\|Bz\|^2$ has bounded level sets, then $\{z(t)\}$ is bounded and, for any of its limit points z^∞, (x^*,z^∞) is an optimal solution of (3.1).

Proof: Parts (a)-(d) follow directly from Proposition 1. To prove part (e), let z^* denote an m-vector that, together with x^*, forms an optimal solution of (3.1). Then from (3.4b)-(3.4c) we have that

$$g(z(t)) - \langle p(t+1),B(z(t)-z^*)\rangle \quad \le \quad g(z^*), \qquad t=0,1,\ldots.$$

Since (cf. parts (b) and (c)) $Bz(t) \to b-Ax^* = Bz^*$ and $\{p(t)\}$ is bounded, this implies that

$$\limsup_{t\to+\infty}\{g(z(t))\} \quad \le \quad g(z^*). \tag{3.5}$$

Hence $g(z(t))+\|Bz(t)\|^2$ is bounded and, by hypothesis, $\{z(t)\}$ is bounded. Since g is lower semicontinuous, each limit point of $\{z(t)\}$, say z^∞, satisfies $g(z^\infty) \le g(z^*)$ (cf. (3.5)). Since (cf. part (b)) $Bz^\infty = b-Ax^*$, (x^*,z^∞) is feasible for (3.1) and its cost $f(x^*)+g(z^\infty)$ does not exceed $f(x^*)+g(z^*)$. Hence (x^*,z^∞) is an optimal solution of (3.1). Q.E.D.

We remark that the hypothesis in Proposition 2 (e) holds if B has full column rank or if g has bounded level sets. In practice, the latter can always be enforced by constraining z to be inside the ball $\{ z\in\Re^m \mid \|z\| \le \mu \}$ with μ a sufficiently large scalar. An example for which Proposition 2 (e) applies is when $f(x) = \|x-d\|^2/2$ for some $d\in\Re^n$ and A has full row rank. Straightforward calculation finds that $A\nabla\phi(A^Tp) = (AA^T)p-Ad$ and hence $A\nabla\phi A^T$ is strongly monotone with modulus being the smallest eigenvalue of AA^T.

4. The Algorithm of Han and Lou is a Special Case

In this section we show that the Han and Lou algorithm [18] is a special case of the alternating minimization algorithm (3.4a)-(3.4d). We also improve upon the results in [18] by applying Proposition 2. Consider the following problem studied by Han and Lou

$$\text{Minimize} \quad f(x) \qquad (4.1)$$
$$\text{Subject to} \quad x \in X_1 \cap \ldots \cap X_k,$$

where $f:\Re^n \to \Re$ is a strongly convex differentiable function (with modulus α) and each X_i is a convex closed set in $\Re^n$. Let X_0 denote the effective domain of f, i.e., $X_0 = \{ x \in \Re^n \mid f(x) < \infty \}$. We make the following assumption regarding (4.1):

Assumption D: Either (a) $ri(X_0) \cap ri(X_1) \cap \ldots \cap ri(X_k) \neq \varnothing$ or (b) $X_0 \cap X_1 \cap \ldots \cap X_k \neq \varnothing$ and all X_i's are polyhedral sets.

We can transform the problem (4.1) into the following form:

$$\text{Minimize} \quad f(x) + g(z_1, \ldots, z_k) \qquad (4.2)$$
$$\text{Subject to} \quad x = z_i, \quad i = 1, \ldots, k,$$

where $z_1, z_2, \ldots, z_k$ are auxiliary vectors and $g:\Re^{nk} \to (-\infty,\infty]$ is the indicator function for $X_1 \times \ldots \times X_k$, i.e.

$$g(z_1, \ldots, z_k) = \textstyle\sum_i \delta(z_i|X_i).$$

The problem (4.2) is clearly a special case of (3.1), where f and g are as above, b = 0, B is the negative of the kn×kn identity matrix, and A is the kn×n matrix composed of k n×n identity matrices stacked one on top of the next.

Assumption D implies that (4.2) is feasible. Since it is easily seen that g is convex lower semicontinuous and that the function $g(z_1, \ldots, z_k) + \sum_i \|z_i\|^2$ has a minimum, Assumption B holds. Hence (4.2) has an optimal solution. Moreover, it can be seen from the strict convexity of f that (4.1) has a unique optimal solution, which we denote by x^*, and that $(x^*,\ldots,x^*) \in \Re^{nk+n}$ is the unique optimal solution of (4.2). Also, by Theorem 28.2 in [35], the problem (4.2) has an optimal Lagrange multiplier vector associated with the constraints $x = z_i$, $i = 1,\ldots,k$. Hence Assumption C holds, and we can apply the alternating minimization algorithm to solve problem (4.1). This produces the following iteration:

$$x(t) = \text{argmin}_x \{ f(x) - \textstyle\sum_i \langle p_i(t), x \rangle \}, \qquad (4.3a)$$
$$z_i(t) = \text{argmin} \{ \langle p_i(t), z_i \rangle + c(t)\|x(t) - z_i\|^2/2 \mid z_i \in X_i \}, \qquad i = 1,\ldots,k, \qquad (4.3b)$$

$$p_i(t+1) = p_i(t) + c(t)(z_i(t)-x(t)), \qquad i = 1,\ldots,k, \tag{4.3c}$$

where p_i $(i = 1,\ldots,k)$ is a Lagrange multiplier vector associated with the constraints $x = z_i$ and $\{c(t)\}$ is any sequence of scalars bounded strictly between zero and $4\alpha/k$. [The initial multiplier $p_i(0)$ is any element of $\Re^n$.] Notice that the iterations (4.3b)-(4.3c) are highly parallelizable, and the same is true for iteration (4.3a) if f is separable.

To see the connection between the above algorithm and the Han and Lou algorithm, note from the strict convexity of f that the conjugate function of f, denote by ϕ, is differentiable; hence (4.3a) is equivalent to

$$x(t) = \nabla\phi(\textstyle\sum_i p_i(t)). \tag{4.4a}$$

Also (4.3b) can be written as

$$z_i(t) = \operatorname{argmin}\{ \|z_i+p_i(t)/c(t)-x(t)\|^2 \mid z_i\in X_i \}, \qquad i = 1,\ldots,k. \tag{4.4b}$$

The iteration (4.4a)-(4.4b), (4.3c) is identical to the Han and Lou algorithm, except that the latter algorithm further restricts $p_i(0)$ to be zero for all i and c(t) to be a fixed scalar inside $(0,2\alpha/k]$ for all t.

Convergence of the algorithm (4.3a)-(4.3d) follows from Proposition 2:

Proposition 3 The sequences $\{x(t)\}$, $\{z(t)\}$, $\{p(t)\}$ generated by (4.3a)-(4.3c) satisfy the following:

(a) $x(t) \to x^*$.

(b) $z_i(t) \to x^*$, $i = 1,\ldots,k$.

(c) $p_i(t) \to$ an optimal Lagrange multiplier vector for (4.2) corresponding to $x = z_i$, $i = 1,\ldots,k$,.

In the case where the X_i's are not all polyhedral sets, Proposition 3 further improves upon the results in [18] (since it asserts convergence without requiring that $X_1\cap\ldots\cap X_k$ has a nonempty interior).

5. The Method of Multipliers and the Dual Gradient Method are Special Cases

Consider the following convex program

$$\begin{array}{ll} \text{Minimize} & q(z) \\ \text{Subject to} & Ez = d, \end{array} \tag{5.1}$$

where $q:\Re^m \to \Re \cup \{+\infty\}$ is a convex, lower semicontinuous function, E is an $n \times m$ real matrix having no zero row, and d is a vector in $\Re^n$. We assume that problem (5.1) has a nonempty, bounded optimal solution set and an optimal Lagrange multiplier vector associated with the constraints $Ez = d$.

We can rewrite (5.1), after introducing an auxiliary variable x, as the following convex program:

$$\begin{array}{ll} \text{Minimize} & \delta(x|\{0\}) + q(z) \\ \text{subject to} & -x + Ez = d. \end{array} \tag{5.2}$$

The preceding problem is clearly a special case of problem (3.1) if we choose $f(\cdot) = \delta(\cdot|\{0\})$, $g(\cdot) = q(\cdot)$, $A = -I$, $B = E$ and $b = d$. With this choice, f satisfies the strong convexity condition (3.2) for any $\alpha > 0$ and Assumptions B and C hold. The alternating minimization algorithm (3.4a)-(3.4d) in this case reduces to the <u>method of multipliers</u> proposed in [17, 19, 34] (see also [1, 2, 24, 33, 37]):

$$\begin{aligned} z(t) &= \operatorname{argmin}_z\{\, q(z) - \langle p(t), Ez\rangle + c(t)\|d-Ez\|^2/2 \,\}, \\ p(t+1) &= p(t) + c(t)(d-Ez(t)), \end{aligned}$$

where $\{c(t)\}$ is any sequence of scalars bounded away from zero.

Now consider the program (5.1) again, but this time we further assume that q satisfies the strong convexity condition (2.2) for some $\alpha > 0$ and we choose $f(\cdot) = q(\cdot)$, $g(\cdot) = \delta(\cdot|\{0\})$, $A = E$, $B = -I$ and $b = d$. With this choice, Assumptions B and C hold. The alternating minimization algorithm (3.4a)-(3.4d) in this case reduces to the dual gradient method discussed in §1:

$$\begin{aligned} x(t) &= \operatorname{argmin}_x\{\, q(x) - \langle p(t), Ex\rangle \,\}, \\ p(t+1) &= p(t) + c(t)(d-Ex(t)), \end{aligned}$$

where $\{c(t)\}$ is any sequence of scalars bounded strictly between zero and $4\alpha/\rho(E^TE)$. This algorithm was first proposed by Uzawa [40] for the more general case where q is strictly convex, but no explicit bound on the stepsizes was given. Other discussion of this algorithm can be found in Ch. 2.6 of [1] and in [8, 21, 24, 33].

6. Application to Symmetric Linear Complementarity Problems

Let M be a given rxr symmetric positive-semidefinite matrix and let w be a given vector in $\Re^r$. Consider the symmetric linear complementarity problem of finding a vector $p \in \Re^r$ satisfying

$$Mp + w \geq 0, \quad p \geq 0, \quad \langle Mp + w, p\rangle \geq 0, \tag{6.1}$$

where (6.1) is assumed to have a solution. The above is a fundamental problem in optimization. One method for solving (6.1) is based on matrix splitting (see [22, 25, 31]). In this method the matrix M is decomposed into the sum of two matrices

$$M = K + L,$$

and, given the current iterate p(t), the next iterate p(t+1) is computed to be a solution of the following linear complementarity problem

$$(\omega(t)I + L)p - (\omega(t)I - K)p(t) + w \geq 0, \quad p \geq 0, \tag{6.2a}$$

$$\langle(\omega(t)I + L)p - (\omega(t)I - K)p(t) + w, p\rangle = 0, \tag{6.2b}$$

where $\omega(t)$ is a relaxation parameter. [p(0) is assumed given.] It has been shown (see for example [25]) that if the above iteration is well defined (i.e. the problem (6.2a)-(6.2b) has a solution) and $\omega(t) = \omega$ for all t, where ω is a nonnegative scalar for which $2\omega I + L - K$ is positive-definite, then the sequence $\{Mp(t)\}$ converges.

In this section we will use the alternating minimization algorithm to obtain a matrix splitting algorithm based on a choice of K and L different from the one above. In particular, we have the following main result of this section:

Proposition 4 If both K and L are symmetric positive-semidefinite matrices and $\{\omega(t)^{-1}\}$ is bounded strictly between zero and $2/\rho(K)$, then the sequence $\{p(t)\}$ generated by (6.2a)-(6.2b) converges to a solution of (6.1). If K is also positive-definite, then the rate of convergence is linear.

Proof: Since K is symmetric positive-semidefinite, it can be expressed as

$$K = AQA^T, \tag{6.3a}$$

where Q is an $n \times n$ positive-definite diagonal matrix ($n \le r$) and A is a $r \times n$ matrix whose columns form a set of orthonormal vectors (see [13]). Similarly, we can express L as

$$L = CRC^T, \tag{6.3b}$$

where R is an $m \times m$ positive diagonal matrix ($m \le r$) and C is a $r \times m$ matrix whose columns form a set of orthonormal vectors. Now consider the convex quadratic program

$$\begin{aligned} &\text{Minimize } \langle x, Q^{-1}x \rangle/2 + g(z_1,z_2) \\ &\text{subject to } Ax + Cz_1 + z_2 = -w, \end{aligned} \tag{6.4}$$

where $g:\Re^{r+m} \to (-\infty,\infty]$ is the convex lower semicontinuous function

$$g(z_1,z_2) = \begin{cases} \langle z_1, R^{-1}z_1 \rangle/2 & \text{if } z_2 \le 0, \\ \infty & \text{otherwise.} \end{cases}$$

It is easily seen (using (6.3a)-(6.3b)) that any optimal Lagrange multiplier vector for (6.4) corresponding to the constraints $Ax + Cz_1 + z_2 = -w$ is a solution of the symmetric linear complementarity problem (6.1) and conversely.

The problem (6.4) is a special case of (3.1) with A as above and with $B = [\ C\ \ I\ ]$, $b = -w$, $f(x) = \langle x, Q^{-1}x \rangle/2$, and g as defined above. Furthermore it has an optimal solution (since its dual has an optimal solution). This, together with the observation that f is strongly convex with modulus $1/(2\rho(Q)) = 1/(2\rho(K))$ and $g(z_1,z_2) + \|Cz_1 + z_2\|^2$ has a minimum, implies that

Assumptions B and C hold. Hence we can apply the alternating minimization algorithm (3.4a)-(3.4d) to the quadratic program (6.4). This produces the iteration (also using the observation that $\rho(A^TA) = 1$)

$$x(t) = \operatorname{argmin}_x\{ \langle x,Q^{-1}x\rangle/2 - \langle p(t),Ax\rangle \}, \tag{6.5a}$$
$$(u(t),s(t)) = \operatorname{argmin}_{s\le 0,u}\{ \langle u,R^{-1}u\rangle/2 + \langle p(t),Bu+s\rangle + c(t)\|Ax(t)+Bu+s-b\|^2/2 \}, \tag{6.5b}$$
$$p(t+1) = p(t) + c(t)(b-Ax(t)-Bu(t)-s(t)). \tag{6.5c}$$

where $\{c(t)\}$ is any sequence of scalars bounded strictly between zero and $2/\rho(K)$.

From the Karush-Kuhn-Tucker conditions for the minimization problem in (6.5a) and in (6.5b) (also using (6.5c)) we obtain that

$$x(t) = QA^Tp(t),$$
$$u(t) = RB^Tp(t+1),$$
$$p(t+1) \ge 0, \; s(t) \le 0, \; \langle p(t+1),s(t)\rangle = 0.$$

Substituting for x(t) and u(t) in (6.5c) gives

$$(I + c(t)BRB^T)p(t+1) - (I - c(t)AQA^T)p(t) - c(t)b = -c(t)s(t) \ge 0,$$
$$p(t+1) \ge 0, \; \langle p(t+1),s(t)\rangle = 0.$$

It then follows from (6.3a)-(6.3b) that the iteration (6.2a)-(6.2b) with $\omega(t) = c(t)^{-1}$ is identical to the iteration (6.5a)-(6.5c). Hence, by Proposition 2 (c), the sequence $\{p(t)\}$ generated by (6.2a)-(6.2b), with $\{\omega(t)^{-1}\}$ bounded strictly between zero and $2/\rho(K)$, converges to a solution of (6.1). Moreover, by Proposition 2 (d), if $K = AQA^T$ is positive-definite, then the rate of convergence is linear. Q.E.D.

Notice that Proposition 4 asserts convergence of the sequence $\{p(t)\}$ eventhough (6.1) may have many solutions. To the best of our knowledge, this is the first such result for a matrix splitting algorithm. Also notice that since L is symmetric positive-semidefinite, the iteration (6.2a)-(6.2b) may be carried out by minimizing a convex quadratic function over the nonnegative orthant. There exist a number of efficient methods for this minimization (see for example §1.5 in [1]). If L is diagonal or tridiagonal, a direct method such as that given in [4] may be used.

Although the restriction of both K and L to symmetric and positive-semidefinite matrices excludes a number of well-known choices for K and L – such as those associated with the Gauss-Seidel and the Jacobi methods – it permits other choices that are very suitable when M is specially structured. One example is when M is of the form

$$M = DQD^T + ERE^T,$$

where Q and R are positive-definite diagonal matrices and D and E are matrices of appropriate dimension (such form arises in, for example, quadratic programs with strictly convex separable costs and linear inequality constraints). Suppose that we choose $K = DQD^T$ and $L = ERE^T$. Then if the matrix [D E] has the staircase structure shown in Figure 1a, the matrices K and L would have respectively the upper and lower block diagonal form shown in Figure 1b. In this case the problem (6.2a)-(6.2b) is significantly smaller than the original problem (6.1).

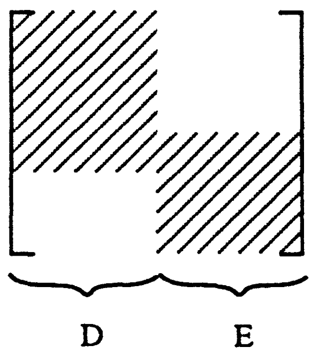

Figure 1a. The matrix [D E] has a staircase structure.

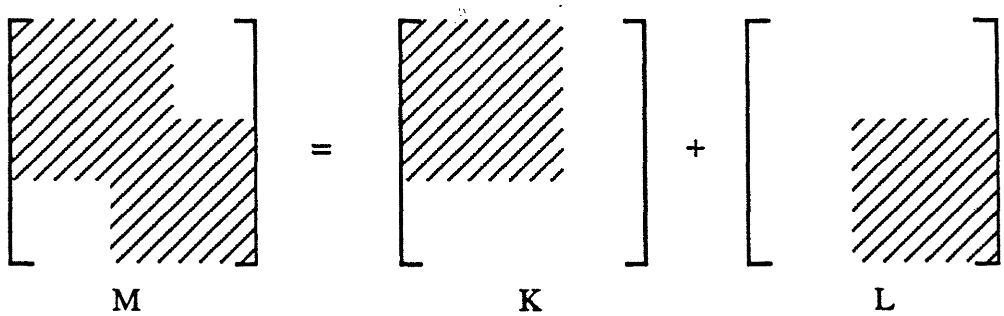

Figure 1b. The matrix M decomposes into a upper block diagonal matrix K and a lower block diagonal matrix L.

7. Application to Variational Inequalities

Consider the following separable variational inequality problem. We are given two closed convex sets $X \subset \Re^n$ and $Z \subset \Re^m$, two continuous functions $R:\Re^n \to \Re^n$ and $S:\Re^m \to \Re^m$, an $r \times n$ matrix A, an $r \times m$ matrix B and a vector $b \in \Re^r$. Our objective is to find a vector $(x^*,z^*) \in X \times Z$ satisfying $Ax^* + Bz^* = b$ and

$$\langle x-x^*, R(x^*)\rangle + \langle z-z^*, S(z^*)\rangle \geq 0, \qquad \forall\ (x,z) \in X \times Z \text{ satisfying } Ax + Bz = b. \tag{7.1}$$

This problem has numerous applications to numerical computation – including the solution of a systems of equations, constrained and unconstrained optimization, traffic assignment problems, game theory, and saddle point problems (see [2], §3.5; [12, 20]). For example, the convex program (3.1) is a special case of (7.1) if its objective function is the sum of the indicator function for a closed convex set and a differentiable convex function. We make the following assumptions regarding (7.1):

Assumption E:

(a) The problem (7.1) has a solution.

(b) R is strongly monotone (with modulus σ) and S is monotone.

(c) Either both X and Z are polyhedral sets or there exist $x \in \mathrm{ri}(X)$ and $z \in \mathrm{ri}(Z)$ satisfying $Ax + Bz = b$.

In this section we will derive a decomposition algorithm for (7.1) by applying the splitting algorithm (2.2a)-(2.2d). First we claim that (7.1) is a special case of the problem (2.1). To see this, note that (x^*,z^*) solves the variational inequality (7.1) if and only if it solves the convex program

$$\begin{aligned} &\text{Minimize} \quad \langle R(x^*),x\rangle + \langle S(z^*),z\rangle \\ &\text{subject to} \quad x \in X,\ z \in Z,\ Ax + Bz = b. \end{aligned} \tag{7.2}$$

Let p^* be an optimal Lagrange multiplier for (7.2) corresponding to the equality constraints $Ax + Bz = b$ (such p^* exists by Theorem 28.2 of [35]). Let (x^*,z^*) be an optimal solution of (7.2). The Karush-Kuhn-Tucker conditions for (7.2) then imply that

$$A^T p^* \in N(x^* | X) + R(x^*), \tag{7.3a}$$

$$B^T p^* \in N(z^*|Z) + S(z^*), \tag{7.3b}$$
$$Ax^* + Bz^* = b. \tag{7.3c}$$

where $N(\cdot|X)$ and $N(\cdot|Z)$ denote the subdifferential of $\delta(\cdot|X)$ and $\delta(\cdot|Z)$ respectively. Consider the multifunctions $F:\Re^n \to \Re^n$ and $G:\Re^m \to \Re^m$ defined by

$$F(x) = R(x) + N(x|X),$$
$$G(z) = S(z) + N(z|Z),$$

Because R and S are monotone and continuous, they are maximal monotone operators (see Minty [27]). Hence, by a result of Rockafellar [36], both F and G are also maximal monotone operators. Let us rewrite (7.3a)-(7.3c) equivalently as

$$AF^{-1}(A^T p^*) + BG^{-1}(B^T p^*) = b. \tag{7.4}$$

Since F is easily seen to be strongly monotone (with modulus σ), (7.4) is a special case of (2.1) and Assumption A holds.

We can then apply the splitting algorithm (2.2a)-(2.2d) to solve (7.4). This produces the iteration whereby we first compute an $x(t) \in X$ satisfying

$$\langle x - x(t), R(x(t)) - A^T p(t) \rangle \geq 0, \qquad \forall\ x \in X,$$

then compute a $z(t) \in Z$ satisfying

$$\langle z - z(t), S(z(t)) - B^T(p(t) - c(t)(Ax(t) + Bz(t) - b)) \rangle \geq 0, \qquad \forall\ z \in Z,$$

and finally update the multipliers by

$$p(t+1) = p(t) + c(t)(b - Ax(t) - Bz(t)),$$

where $\{c(t)\}$ is any sequence of scalars bounded strictly between zero and $2\sigma/\rho(A^TA)$. Convergence of the sequences $\{x(t)\}$, $\{Bz(t)\}$, $\{p(t)\}$ generated by this iteration follows from Proposition 1. We leave the issue of computing x(t) and z(t) open (see [2, 5, 20, 32] for solution methods).

Appendix A:

In this appendix we prove Proposition 1. Let p^* denote any solution of (2.1). Then

$$x^* \in \Phi(A^T p^*), \tag{A.1a}$$

and there exists $z^* \in \Re^m$ satisfying

$$z^* \in \Gamma(B^T p^*), \tag{A.1b}$$
$$Ax^* + Bz^* = b. \tag{A.1c}$$

From (2.2a)-(2.2c) we also have that, for $t = 0,1,\ldots,$

$$x(t) \in \Phi(A^T p(t)), \tag{A.2a}$$
$$z(t) \in \Gamma(B^T p(t+1)). \tag{A.2b}$$

Fix any integer $t \geq 0$ and, for convenience, let $c = c(t)$. Now, since (cf. (A.1a) and (A.2a)) $A^T p^* \in \Phi^{-1}(x^*)$ and $A^T p(t) \in \Phi^{-1}(x(t))$, we have by Assumption A (b)

$$\begin{aligned} 0 &= \langle A^T p(t) - A^T p^*, x(t) - x^* \rangle - \langle A^T p(t) - A^T p^*, x(t) - x^* \rangle \\ &\geq \sigma \|x(t) - x^*\|^2 - \langle p(t) - p^*, Ax(t) - Ax^* \rangle \\ &= c\langle Ax(t) - Ax^*, Ax(t) - Ax^* \rangle - \langle p(t) - p^*, Ax(t) - Ax^* \rangle - c\|Ax(t) - Ax^*\|2 + \sigma\|x(t) - x^*\|^2. \end{aligned}$$

Let $\theta = -c\|Ax(t) - Ax^*\|^2 + \sigma\|x(t) - x^*\|^2$. The above then implies that

$$\begin{aligned} 0 &\geq \langle -p(t) + c(Ax(t) + Bz(t) - b) + p^*, Ax(t) - Ax^* \rangle - c\langle Bz(t) - Bz^*, Ax(t) - Ax^* \rangle + \theta \\ &= \langle -\hat{p}(t+1), Ax(t) - Ax^* \rangle - c\langle B\hat{z}(t), A\hat{x}(t) \rangle + \theta, \end{aligned} \tag{A.3}$$

where we let $\hat{x}(t) = x(t) - x^*$, $\hat{z}(t) = z(t) - z^*$, $\hat{p}(t+1) = p(t+1) - p^*$, and the equality follows from (2.2c). Similarly, since (cf. (A.1b) and (A.2b)) $z^* \in \Gamma(B^T p^*)$ and $z(t) \in \Gamma(B^T p(t+1))$, we have from the monotonicity of Γ

$$\begin{aligned} 0 &= \langle B^T p(t+1) - B^T p^*, z(t) - z^* \rangle - \langle B^T p(t+1) - B^T p^*, z(t) - z^* \rangle \\ &\geq -\langle p(t+1) - p^*, Bz(t) - Bz^* \rangle \\ &= \langle -\hat{p}(t+1), Bz(t) - Bz^* \rangle. \end{aligned} \tag{A.4}$$

Summing (A.3)-(A.4) and using the fact (cf. (A.1c)) $Ax^* + Bz^* = b$, we obtain

$$\begin{aligned} 0 &\geq \langle -\hat{p}(t+1), Ax(t)+Bz(t)-b\rangle - c\langle B\hat{z}(t), A\hat{x}(t)\rangle + \theta \\ &= \langle \hat{p}(t+1), \hat{p}(t+1)-\hat{p}(t)\rangle/c - c\langle B\hat{z}(t), A\hat{x}(t)\rangle + \theta, \end{aligned}$$

where $\hat{p}(t) = p(t)-p^*$ and the equality follows from (2.2c). This, together with the identities (cf. (2.2c))

$$\begin{aligned} 2\cdot\langle \hat{p}(t+1), \hat{p}(t+1)-\hat{p}(t)\rangle &= \|\hat{p}(t+1)-\hat{p}(t)\|^2 + \|\hat{p}(t+1)\|^2 - \|\hat{p}(t)\|^2, \\ \|\hat{p}(t+1)-\hat{p}(t)\|^2/c^2 &= \|A\hat{x}(t)\|^2 + \|B\hat{z}(t)\|^2 + 2\cdot\langle B\hat{z}(t), A\hat{x}(t)\rangle, \end{aligned}$$

implies that

$$0 \geq \|\hat{p}(t+1)\|^2 - \|\hat{p}(t)\|^2 + c^2\cdot\|A\hat{x}(t)\|^2 + c^2\|B\hat{z}(t)\|^2 + 2c\theta.$$

Hence, by the definition of c and θ,

$$\begin{aligned} \|\hat{p}(t)\|^2 &\geq \|\hat{p}(t+1)\|^2 - c(t)^2\cdot\|A\hat{x}(t)\|^2 + 2c(t)\sigma\|\hat{x}(t)\|^2 + c(t)^2\|B\hat{z}(t)\|^2 \\ &\geq \|\hat{p}(t+1)\|^2 + c(t)(2\sigma - c(t)\rho(A^TA))\|\hat{x}(t)\|^2 + c(t)^2\|B\hat{z}(t)\|^2. \end{aligned}$$

Since the choice of t and p^* was arbitrary and (cf. (2.2d)) both $2\sigma/\rho(A^TA)-c(t)$ and $c(t)$ are bounded away from ε, we obtain that

$$\|p(t)-p^*\|^2 \geq \|p(t+1)-p^*\|^2 + \varepsilon^2\rho(A^TA)\|x(t)-x^*\|^2 + \varepsilon^2\|Bz(t)+Ax^*-b\|^2, \tag{A.5}$$

for all $t = 0,1,\ldots,$ for any solution p^* of (2.1). Eq. (A.5) implies that $\{p(t)\}$ is bounded and

$$x(t) \to x^*, \quad Bz(t) \to b-Ax^*. \tag{A.6}$$

Hence parts (a) and (b) are proven. To prove part (c), notice that since (cf. (A.2b))

$$Bz(t) \in B\Gamma(B^Tp(t+1)), \quad t = 0,1,\ldots,$$

it follows from (A.6) and the lower semicontinuity of $B\Gamma B^T$ (cf. Proposition 2.5 in [3]) that, for any limit point p^∞ of $\{p(t)\}$,

$$b - Ax^* \in B\Gamma(B^T p^\infty).$$

Similarly, we have from (A.2a) and the lower semicontinuity of $A\Phi A^T$ that

$$Ax^* \in A\Phi(A^T p^\infty).$$

Hence $b \in A\Phi(A^T p^\infty) + B\Gamma(B^T p^\infty)$ and therefore p^∞ solves (2.1). By replacing p^* in (A.5) by p^∞, we obtain that $p(t) \to p^\infty$.

Finally we prove part (d). From (A.1a)-(A.1b) and (A.2a)-(A.2b) we have that

$$Ax^* = A\Phi(A^T p^\infty), \quad b - Ax^* \in B\Gamma(B^T p^\infty),$$
$$Ax(t) = A\Phi(A^T p(t)), \quad Bz(t) \in B\Gamma(B^T p(t+1)), \qquad t = 0, 1, \ldots.$$

Fix any integer $t \geq 0$. Since δ and η are the modulus of $A\Phi A^T$ and $B\Gamma B^T$ respectively, the above implies that

$$\langle Ax(t) - Ax^*, p(t) - p^\infty \rangle \geq \delta \cdot \|p(t) - p^\infty\|^2,$$
$$\langle Bz(t) + Ax^* - b, p(t+1) - p^\infty \rangle \geq \eta \cdot \|p(t+1) - p^\infty\|^2,$$

and hence by the Cauchy-Schwarz inequality

$$\|Ax(t) - Ax^*\| \geq \delta \cdot \|p(t) - p^\infty\|,$$
$$\|Bz(t) + Ax^* - b\| \geq \eta \cdot \|p(t+1) - p^\infty\|.$$

This, together with (A.5), implies that

$$\|p(t) - p^\infty\|^2 \geq \|p(t+1) - p^\infty\|^2 + \varepsilon^2\delta^2 \cdot \|p(t) - p^\infty\|^2 + \varepsilon^2\eta^2 \cdot \|p(t+1) - p^\infty\|^2.$$

Also from (A.5) we have that

$$\|p(t) - p^\infty\|^2 \geq \varepsilon^2 \rho(A^T A) \cdot \|x(t) - x^*\|^2 + \varepsilon^2 \cdot \|Bz(t) + Ax^* - b\|^2.$$

Part (d) is then proven. Q.E.D.

References

[1] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, N. Y. (1982).

[2] Bertsekas, D.P. and Tsitsiklis, J.N., *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, N. J. (1988).

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

[4] Cottle, R. W. and Sacher, R. S., "On the Solution of Large, Structured, Linear Complementarity Problems: The Tridiagonal Case," *J. Appl. Math. and Optimization*, 3 (1977), 321-340.

[5] Dafermos, S., "An Iterative Scheme for Variational Inequalitites," *Mathematical Programming*, 26 (1983), 40-47.

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

[7] Eckstein, J., "The Lions-Mercier Splitting Algorithm and the Alternating Direction Method are Instances of the Proximal Point Algorithm," Report LIDS-P-1769, Laboratory for Information and Decision Systems, M.I.T., MA (August 1988).

[8] Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, 11 (1963), 399-417.

[9] Fortin, M. and Glowinski, R. (eds.), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems*, North-Holland, Amsterdam (1983).

[10] Gabay, D. and Mercier, B., "A Dual Algorithm for the Solution of Nonlinear Variational Problems Via Finite-Element Approximations," *Comp. and Math. with Applications*, 2 (1976), 17-40.

[11] Gabay, D., "Méthodes numériques pour l'optimisation non linéaire," Thèse de Doctorat d'Etat et Sciences Mathematiques, Université Pierre et Marie Curie, Paris VI, (1979).

[12] Gabay, D., "Applications of the Method of Multipliers to Variational Inequalities," in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems* (M. Fortin, R. Glowinski, eds.), North-Holland, Amsterdam (1983), 299-331.

[13] Gantmacher, F. R., *The Theory of Matrices*, Vols. 1-2, Chelsea, N. Y. (1959).

[14] Glowinski, R. and Le Tallec, P., "Augmented Lagrangian Methods for the Solution of Variational Problems," Mathematics Research Center Technical Summary Report #2965, University of Wisconsin-Madison, WI (January 1987).

[15] Glowinski, R. and Marrocco, A., "Sur l'approximation par éléments finis d'ordre un, et la résolution par penalisation-dualite d'une classe de problèmes de Dirichlet non linéaires," *Revue Française d'Automatique Informatique Recherche Opérationnelle, Analyse numérique*, R-2 (1975), 41-76.

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

[17] Haarhoff, P. C. and Buys, J. D., "A New Method for the Optimization of a Nonlinear Function Subject to Nonlinear Constraints," *The Computer Journal*, 13 (1970), 178-184.

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

[19] Hestenes, M. R., "Multiplier and Gradient Methods," *J. Optimization Theory and Applications*, 4 (1969), 303-320.

[20] Kinderlehrer, D. and Stampacchia, G., *An Introduction to Variational Inequalities and their Applications*, Academic Press, N.Y. (1980).

[21] Lasdon, L. S., *Optimization Theory for Large Systems*, MacMillan, N. Y. (1970).

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

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

[24] Luenberger, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA (1984).

[25] Mangasarian, O. L., "Solution of Symmetric Linear Complementarity Problems by Iterative Methods," *J. Optimization Theory and Applications*, 22 (1977), 465-485.

[26] Mangasarian, O. L. and De Leone, R., "Serial and Parallel Solution of Large Scale Linear Programs by Augmented Lagrangian Successive Overrelaxation," Computer Sciences Technical Report #701, University of Wisconsin-Madison, WI (June 1987).

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

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

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

[30] Ortega, J.M. and Rheinboldt, W.C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, N. Y. (1970).

[31] Pang, J. S., "On the Convergence of a Basic Iterative Method for the Implicit Complementarity Problem," *J. Optimization Theory and Applications*, 37 (1982), 149-162.

[32] Pang, J. S. and Chan, D., "Iterative Methods for Variational and Complementarity Problems," *Mathematical Programming*, 24 (1982), 284-313.

[33] Poljak, B. T., *Introduction to Optimization*, Optimization Software Inc., N. Y. (1987).

[34] Powell, M. J. D., "A Method for Nonlinear Constraints in Minimization Problems," In *Optimization* (R. Fletcher, ed.), Academic Press, N. Y. (1969), 283-298.

[35] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, N. J. (1970).

[36] Rockafellar, R.T., "On the Maximality of Sums of Nonlinear Monotone Operators," *Trans. Amer. Math. Soc. 149* (1970), 75-88.

[37] Rockafellar, R. T., "Augmented Lagrangians and Applications fo the Proximal Point Algorithm in Convex Programming," *Mathematics of Operations Research*, 1 (1976), 97-116.

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

[39] Springarn, J. E., "Applications of the Method of Partial Inverses to Convex Programming: Decomposition," *Mathematical Programming*, 32 (1985), 199-223.

[40] Uzawa, H., "Iterative Methods for Concave Programming," In *Studies in Linear and Non-linear Programming* (K. J. Arrow, L. Hurwicz, and H. Uzawa, eds.), Stanford University Press, Stanford (1958), 154-165.