October 1989 LIDS-P-1919

*On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators**

by

Jonathan ECKSTEIN
Graduate School of Business, Harvard University
Boston, MA 02163

Dimitri P. BERTSEKAS
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

This paper shows, by means of a new type of operator called a *splitting operator*, that the Douglas-Rachford splitting method for finding a zero of the sum of two monotone operators is a special case of the proximal point algorithm. Therefore, applications of Douglas-Rachford splitting, such as the alternating direction method of multipliers for convex programming decomposition, are also special cases of the proximal point algorithm. The approach taken here also essentially subsumes the theory of partial inverses developed by Spingarn. We show the usefulness of the connection between Douglas-Rachford splitting and the proximal point algorithm by deriving a new, *generalized* alternating direction method of multipliers for convex programming.

Running Heads: Operator Splitting and the Proximal Point Algorithm

Key Words: Monotone Operators, Decomposition, Alternating Direction Methods, Method of Multipliers

---

* This paper consists mainly of dissertation research results of the first author. The dissertation was performed at M.I.T. under the supervision of the second author, and was supported in part by the Army Research Office under grant number DAAL03-86-K-0171, and by the National Science Foundation under grant number ECS-8519058.

1. Introduction

The theory of monotone set-valued monotone operators (see, for example, Brézis 1973) provides a powerful general framework for the study of convex programming and variational inequalities. A fundamental algorithm for finding a root of a monotone operator is the *proximal point algorithm* (Rockafellar 1976a). The well-known *method of multipliers* (Hestenes 1969, Powell 1969) for constrained convex programming is known to be a special case of the proximal point algorithm (Rockafellar 1976b).

The proximal point algorithm requires evaluation of *resolvent* operators of the form $(I+\lambda T)^{-1}$, where T is monotone and set-valued, λ is a positive scalar, and I denotes the identity mapping. The main difficulty with the method is that $I+\lambda T$ may be hard to invert, depending on the nature of T. One alternative is to find maximal monotone operators A and B such that $A+B=T$, but $I+\lambda A$ and $I+\lambda B$ are easier to invert that $I+\lambda T$. One can then devise an algorithm that uses only operators of the form $(I+\lambda A)^{-1}$ and $(I+\lambda B)^{-1}$, rather than $(I+\lambda(A+B))^{-1}=(I+\lambda T)^{-1}$. Such an approach is called a *splitting method*, and is inspired by well-established techniques from numerical linear algebra (for example, see Marchuk 1975).

A number of authors, mainly in the French mathematical community, have extensively studied such methods, which fall into four principle classes: forward-backward (Passty 1979, Gabay 1983, Tseng 1988), double-backward (Lions 1978, Passty 1979), Peaceman-Rachford (Lions and Mercier 1979), and Douglas-Rachford (Lions and Mercier 1979). For a survey, readers may wish to refer to Eckstein (1989). We will focus on the "Douglas-Rachford" class, which appears to have the most general convergence properties. Gabay (1983) has shown that the *alternating direction* method of multipliers, a variation on the method of multipliers designed to be more conducive to decomposition, is a special case of Douglas-Rachford splitting.

A principle contribution of this paper is a demonstration that Douglas-Rachford splitting is an application of the proximal point algorithm. As a consequence, much of the theory of the proximal point algorithm may be carried over to the context of Douglas-Rachford splitting and its special cases, including the alternating direction method of multipliers. As one example of this carryover, we present a *generalized* form of the proximal point algorithm — created by synthesizing the work of Rockafellar (1976a) with that of Gol'shtein and Tret'yakov (1979) — and show how it gives rise to a new method, *generalized* Douglas-Rachford splitting. We give some further examples of the application of this theory, namely Spingarn's (1983, 1985b) *method of partial inverses* (with a minor extension), and a new augmented Lagrangian method for convex programming, the *generalized* alternating direction method of multipliers.

Most of the results presented here are taken from the recent thesis by Eckstein (1989), which contains more detailed development, and also relates the theory to the work of Gol'shtein (1985, 1986, 1987). Some preliminary versions of our results have also appeared somewhat earlier in Eckstein (1988).

This paper is organized a follows: Section 2 introduces the basic theory of monotone operators in Hilbert space, while Section 3 proves the convergence of a generalized form of the proximal point algorithm. Section 4 discusses Douglas-Rachford splitting, showing it to be a special case of the proximal point algorithm by means of a specially-constructed *splitting operator*. This notion is combined with the result of Section 3 to yield *generalized* Douglas-Rachford splitting. Section 5 demonstrates applications of this theory in the method of partial inverses and in generalizing the alternating direction method of multipliers.

2. *Monotone Operators*

An *operator* T on a Hilbert space $\mathcal{H}$ is a (possibly null-valued) point-to-set map $T\colon \mathcal{H} \to 2^{\mathcal{H}}$. We will make no distinction between an operator T and its graph, that is, the set

$\{(x, y) \mid y \in T(x)\}$. Thus, we may simply say that an operator is any subset T of $\mathcal{H} \times \mathcal{H}$, and define $T(x) = Tx = \{y \mid (x, y) \in T\}$.

If T is single-valued, that is, the cardinality of Tx is at most 1 for all $x \in \mathcal{H}$, we will by slight abuse of notation allow Tx and $T(x)$ to stand for the unique $y \in Y$ such that $(x, y) \in T$, rather than the singleton set $\{y\}$. The intended meaning should be clear from the context.

The *domain* of a mapping T is its "projection" onto the first coordinate,

$$\operatorname{dom} T = \{x \in \mathcal{H} \mid \exists\, y \in \mathcal{H} : (x, y) \in T\} = \{x \in \mathcal{H} \mid Tx \neq \varnothing\}.$$

We say that T has *full domain* if $\operatorname{dom} T = \mathcal{H}$. The *range* or *image* of T is similarly defined as its projection onto the second coordinate,

$$\operatorname{im} T = \{ y \in \mathcal{H} \mid \exists\, x \in \mathcal{H} : (x, y) \in T\} .$$

The *inverse* T^{-1} of T is $\{(y, x) \mid (x, y) \in T\}$.

For any real number c and operator T, we let cT be the operator $\{(x, cy) \mid (x, y) \in T\}$, and if A and B are any operators, we let

$$A+B = \{(x, y+z) \mid (x, y) \in A, (x, z) \in B\} .$$

We will use the symbol I to denote the *identity* operator $\{(x, x) \mid x \in \mathcal{H}\}$. Let $\langle \cdot , \cdot \rangle$ denote the inner product on $\mathcal{H}$. Then an operator T is *monotone* if

$$\langle x' - x, y' - y \rangle \geq 0 \qquad \forall\, (x, y), (x', y') \in T .$$

A monotone operator is *maximal* if (considered as a graph) it is not strictly contained in any other monotone operator on $\mathcal{H}$. Note that an operator is (maximal) monotone if and only if its inverse is (maximal) monotone. The best-known example of maximal monotone operator is the subgradient mapping ∂f of a closed proper convex function $f: \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ (see Rockafellar 1970a, 1970b). The following theorem, originally due to Minty (1962), provides a crucial characterization of maximal monotone operators:

Theorem 1. If T is a monotone operator on $\mathcal{H}$, T is maximal if and only if $\mathrm{im}(I+T) = \mathcal{H}$.

For alternate proofs of Theorem 1, or stronger related theorems, see Rockafellar (1970b), Brézis (1973), Doležal (1979), or Joshi and Bose (1985). All proofs of the theorem require Zorn's Lemma, or, equivalently, the axiom of choice.

Given any operator A, let J_A denote the operator $(I+A)^{-1}$. Given any positive scalar c and operator T, $J_{cT} = (I + cT)^{-1}$ is called a *resolvent* of T. An operator C on $\mathcal{H}$ is said to be *nonexpansive* if

$$\|y' - y\| \leq \|x' - x\| \qquad \forall\ (x, y), (x', y') \in C \quad .$$

Note that nonexpansive operators are necessarily single-valued and Lipschitz continuous. An operator J on $\mathcal{H}$ is said to be *firmly nonexpansive* if

$$\|y' - y\|^2 \leq \langle x' - x, y' - y\rangle \qquad \forall\ (x, y), (x', y') \in J \quad .$$

The following lemma summarizes some well-known properties of firmly nonexpansive operators.

Lemma 1. (i) All firmly nonexpansive operators are nonexpansive. (ii) An operator J is firmly nonexpansive if and only if $2J - I$ is nonexpansive. (iii) An operator is firmly nonexpansive if and only if it is of the form $\frac{1}{2}(C + I)$, where C is nonexpansive. (iv) An operator J is firmly nonexpansive if and only if $I - J$ is firmly nonexpansive.

Proof. Statement (i) follows directly from the Cauchy-Schwartz inequality. To prove (ii), first let J be firmly nonexpansive. Then for any $x, y \in \mathcal{H}$,

$$\|(2J - I)x \ - \ (2J - I)y\|^2 \ = \ 4\|Jx - Jy\|^2 \ - \ 4\langle Jx - Jy, x - y\rangle \ + \ \|x - y\|^2 \ .$$

Since J is firmly nonexpansive, $4(\|Jx - Jy\|^2 - \langle Jx - Jy, x - y\rangle) \leq 0$, and one deduces that

$$\|(2J - I)x \ - \ (2J - I)y\|^2 \ \leq \ \|x - y\|^2 \ ,$$

and so $2J - I$ is nonexpansive. Conversely, now suppose $C = 2J - I$ is nonexpansive. Then $J = \frac{1}{2}(C + I)$, and for any $x, y \in \mathcal{H}$,

$$
\begin{aligned}
\|Jx - Jy\|^2 &= \tfrac{1}{4}\|Cx - Cy\|^2 + \tfrac{1}{2}\langle Cx - Cy, x - y\rangle + \tfrac{1}{4}\|x - y\|^2 \\
&\le \tfrac{1}{2}\|x - y\|^2 + \tfrac{1}{2}\langle Cx - Cy, x - y\rangle \\
&= \langle \tfrac{1}{2}(C + I)x - \tfrac{1}{2}(C + I)y, x - y\rangle \\
&= \langle Jx - Jy, x - y\rangle \quad .
\end{aligned}
$$

Therefore, J is firmly nonexpansive. This proves (ii); claim (iii) is simply a reformulation of (ii). Finally, consider (iv). From (ii), we have

$$
\begin{aligned}
& I - J \text{ is firmly nonexpansive} \\
\Leftrightarrow \quad & 2(I - J) - I \text{ is nonexpansive} \\
\Leftrightarrow \quad & -2J + I = -(2J - I) \text{ is nonexpansive} \\
\Leftrightarrow \quad & 2J - I \text{ is nonexpansive} \\
\Leftrightarrow \quad & J \text{ is firmly nonexpansive} \quad . \quad \blacksquare
\end{aligned}
$$

Figure 1 illustrates the lemma.

We now give a critical theorem. The "only if" part of the following theorem has been well known for some time, but the "if" part, just as easily obtained, appears to have been obscure or unknown. The purpose here is to stress the complete symmetry that exists between (maximal) monotone operators and (full-domained) firmly nonexpansive operators over any Hilbert space.

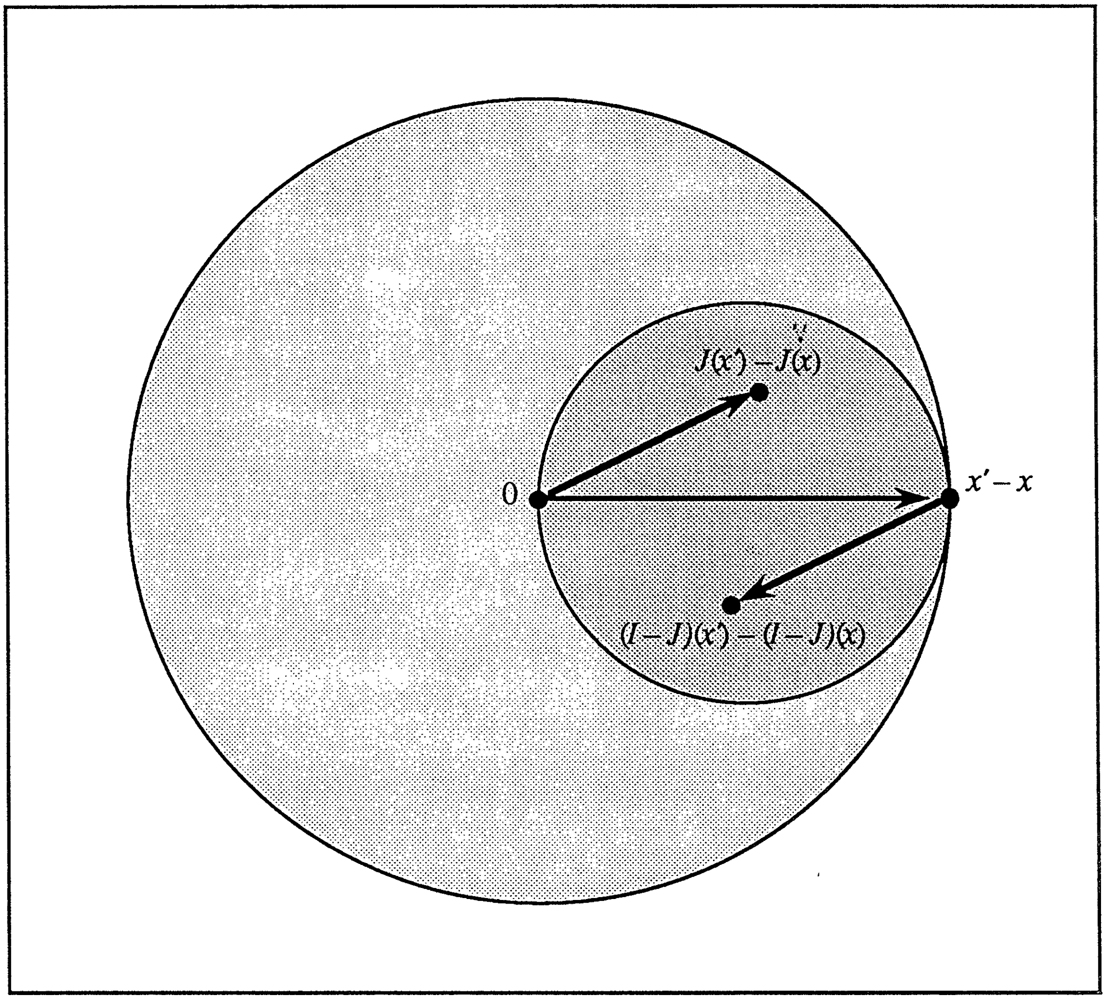

Figure 1. Illustration of the action of firmly nonexpansive operators in Hilbert space. If J is nonexpansive, then $J(x') - J(x)$ must lie in the larger sphere, which has radius $\|x' - x\|$ and is centered at 0. If J is *firmly* nonexpansive, then $J(x') - J(x)$ must lie in the smaller sphere, which has radius $(1/2)\|x' - x\|$ and is centered at $(1/2)(x' - x)$. This characterization follows directly from J being of the form $(1/2)I + (1/2)C$, where C is nonexpansive. Note that if $J(x') - J(x)$ lies in the smaller sphere, so must $(I - J)(x') - (I - J)(x)$, illustrating Lemma 1(iv).

Theorem 2. Let c be any positive scalar. An operator T on $\mathcal{H}$ is monotone if and only if its resolvent $J_{cT} = (I + cT)^{-1}$ is firmly nonexpansive. Furthermore, T is maximal monotone if and only if J_{cT} is firmly nonexpansive and $\text{dom}(J_{cT}) = \mathcal{H}$.

Proof. By the definition of the scaling, addition, and inversion operations,

$$(x, y) \in T \Leftrightarrow (x + cy, x) \in (I + cT)^{-1} .$$

Therefore,

$$\begin{aligned}
T \text{ monotone} &\Leftrightarrow \langle x' - x, y' - y \rangle \geq 0 && \forall\, (x, y), (x', y') \in T \\
&\Leftrightarrow \langle x' - x, cy' - cy \rangle \geq 0 && \forall\, (x, y), (x', y') \in T \\
&\Leftrightarrow \langle x' - x + cy' - cy, x' - x \rangle \geq \|x' - x\|^2 && \forall\, (x, y), (x', y') \in T \\
&\Leftrightarrow (I + cT)^{-1} \text{ firmly nonexpansive.}
\end{aligned}$$

The first claim is established. Clearly, T is maximal if and only if cT is maximal. So, by Theorem 1, T is maximal if and only if $\mathrm{im}(I + cT) = \mathcal{H}$. This is in turn true if and only if $(I + cT)^{-1}$ has domain $\mathcal{H}$, establishing the second statement. ■

Corollary 2.1. An operator K is firmly nonexpansive if and only if $K^{-1} - I$ is monotone. K is firmly nonexpansive with full domain if and only if $K^{-1} - I$ is maximal monotone.

Corollary 2.2. For any $c > 0$, the resolvent J_{cT} of a monotone operator T is single-valued. If T is also maximal, then J_{cT} has full domain.

Corollary 2.3 (The Representation Lemma). Let $c > 0$ and let T be monotone on $\mathcal{H}$. Then every element z of $\mathcal{H}$ can be written in at most one way as $x + cy$, where $y \in Tx$. If T is maximal, then every element z of $\mathcal{H}$ can be written in *exactly* one way as $x + cy$, where $y \in Tx$.

Corollary 2.4. The functional taking each operator T to $(I+T)^{-1}$ is a bijection between the collection of maximal monotone operators on $\mathcal{H}$ and the collection of firmly nonexpansive operators on $\mathcal{H}$.

Corollary 2.1 simply restates the $c = 1$ case of the theorem, while Corollary 2.2 follows because firmly nonexpansive operators are single-valued. Corollary 2.3 is essentially a restatement of Corollary 2.2. Corollary 2.4 resembles a result of Minty (1962), but is not identical (Minty did not use the concept of *firm* nonexpansiveness).

A *root* or *zero* of an operator T is a point x such that $Tx \ni 0$. We let $\mathrm{zer}(T) = T^{-1}(0)$ denote the set of all such points. In the case that T is the subdifferential map ∂f of a convex function f, $\mathrm{zer}(T)$ is the set of all global minima of f. The zeroes of a monotone operator precisely coincide with the fixed points of its resolvents:

Lemma 2. Given any maximal monotone operator T, real number $c > 0$, and $x \in \mathcal{H}$, we have $0 \in Tx$ if and only if $J_{cT}(x) = x$.

Proof. By direct calculation, $J_{cT} = \{(x + cy, x) \mid (x, y) \in T\}$. Hence,

$$0 \in Tx \quad \Leftrightarrow \quad (x, 0) \in T \quad \Leftrightarrow \quad (x, x) \in J_{cT} \ .$$

Since J_{cT} is single-valued, the proof is complete. ■

3. *A Generalized Proximal Point Algorithm*

Lemma 1 suggests that one way of finding a zero of a maximal monotone operator T might be to perform the iteration $z^{k+1} = J_{cT}(z^k)$, starting from some arbitrary point z^0. This procedure is the essence of the *proximal point algorithm*, as named by Rockafellar (1976a). Specialized versions of this method were known earlier to Martinet (1970, 1972). Rockafellar's analysis allows c to vary from one iteration to the next: given a maximal monotone operator T and a sequence of positive scalars $\{c_k\}$, called *stepsizes*, we say that $\{z^k\}$ is generated by the *proximal point algorithm* if $z^{k+1} = J_{c_kT}(z^k)$ for all $k \geq 0$. Rockafellar's convergence theorem also allows the resolvents J_{c_kT} to be evaluated approximately, so long as the sum of all errors is finite. A related result due to Gol'shtein and Tret'yakov (1979) considers iterations of the form

$$z^{k+1} = (1 - \rho_k)z^k + \rho_k J_{cT}(z^k) \quad ,$$

where $\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2)$ is a sequence of *over-* or *under-relaxation* factors. In at least one important application of the proximal point algorithm, the method of multipliers for convex programming, using relaxation factors ρ_k greater than 1 is known to accelerate convergence (Bertsekas 1982, p. 129). Gol'shtein and Tret'yakov also allow resolvents to be evaluated approximately, but, unlike Rockafellar, do not allow the stepsize c to vary with k, restrict $\mathcal{H}$ to be finite-dimensional, and do not consider the case in which $\operatorname{zer}(T) = \emptyset$. The following theorem effectively combines the results of Rockafellar and Gol'shtein-Tret'yakov.

Theorem 3. Let T be a maximal monotone operator on a Hilbert space $\mathcal{H}$, and let $\{z^k\}$ be such that

$$z^{k+1} = (1-\rho_k)z^k + (1-\rho_k)w^k \qquad \forall\, k \geq 0,$$

where

$$\| w^k - (I + cT)^{-1}(z^k) \| \leq \varepsilon_k \qquad \forall\, k \geq 0$$

and $\{\varepsilon_k\}_{k=0}^{\infty}, \{\rho_k\}_{k=0}^{\infty}, \{c_k\} \subseteq [0, \infty)$ are sequences such that

$$E_1 = \sum_{k=0}^{\infty} \varepsilon_k < \infty$$

$$\Delta_1 = \inf_{k \geq 0} \rho_k > 0$$

$$\Delta_2 = \sup_{k \geq 0} \rho_k < 2$$

$$\overline{c} = \inf_{k \geq 0} c_k > 0 \ .$$

Such a sequence $\{z^k\}$ is said to conform to the *generalized proximal point algorithm.* Then if T possesses any zero, $\{z^k\}$ converges weakly to a zero of T. If T has no zeroes, then $\{z^k\}$ is an unbounded sequence.

Proof. Suppose first that T has some zero. For all k, define

$$Q_k = I - J_{c_kT} = I - (I + c_kT)^{-1} .$$

We know that Q_k is firmly nonexpansive from Lemma 1(iv). Note also that any zero of T is a fixed point of $(I + c_kT)^{-1}$ by Lemma 2, and hence a zero of Q_k for any k. For all k, define

$$\bar{z}^{k+1} = (1 - \rho_k)z^k + \rho_k J_{c_kT}(z^k) = (I - \rho_k Q_k)(z^k) .$$

For any zero z^* of T,

$$\begin{aligned} \| \bar{z}^{k+1} - z^* \|^2 &= \| z^k - \rho_k Q_k(z^k) - z^* \|^2 \\ &= \| z^k - z^* \|^2 - 2\rho_k \langle z^k - z^*, Q_k(z^k)\rangle + \rho_k^2 \| Q_k(z^k) \|^2 . \end{aligned}$$

Since $0 \in Q_k(z^*)$ and Q_k is firmly nonexpansive, we have

$$\begin{aligned} \| \bar{z}^{k+1} - z^* \|^2 &\le \| z^k - z^* \|^2 - \rho_k(2 - \rho_k)\| Q_k(z^k) \|^2 \\ &\le \| z^k - z^* \|^2 - \Delta_1(2 - \Delta_2)\| Q_k(z^k) \|^2 . \end{aligned}$$

As $\Delta_1(2 - \Delta_2) > 0$, we have that $\| \bar{z}^{k+1} - z^* \| \le \| z^k - z^* \|$. Now, $\| z^{k+1} - \bar{z}^{k+1} \| \le \rho_k\varepsilon_k$, so

$$\begin{aligned} \| z^{k+1} - z^* \| &\le \| \bar{z}^{k+1} - z^* \| + \| z^{k+1} - \bar{z}^{k+1} \| \\ &\le \| \bar{z}^k - z^* \| + \rho_k\varepsilon_k . \end{aligned}$$

Combining this inequality for all k,

$$\begin{aligned} \| z^{k+1} - z^* \| &\le \| z^0 - z^* \| + \sum_{i=0}^{k} \rho_i\varepsilon_i \\ &\le \| z^0 - z^* \| + 2E_1 , \end{aligned}$$

and $\{z^k\}$ is bounded. Furthermore,

$$\begin{aligned} \| z^{k+1} - z^* \|^2 &= \| \bar{z}^{k+1} - z^* + (z^{k+1} - \bar{z}^{k+1}) \|^2 \\ &= \| \bar{z}^{k+1} - z^* \|^2 + 2\langle \bar{z}^{k+1} - z^*, z^{k+1} - \bar{z}^{k+1} \rangle + \|z^{k+1} - \bar{z}^{k+1} \|^2 \\ &\le \| \bar{z}^{k+1} - z^* \|^2 + 2 \| \bar{z}^{k+1} - z^* \| \|z^{k+1} - \bar{z}^{k+1} \| + \|z^{k+1} - \bar{z}^{k+1} \|^2 \\ &\le \| z^k - z^* \|^2 - \Delta_1(2 - \Delta_2)\| Q_k(z^k) \|^2 + 2\rho_k\varepsilon_k(\| z^0 - z^* \| + 2E_1) + \rho_k^2\varepsilon_k^2 . \end{aligned}$$

Since $\{\varepsilon_k\}$ is summable, so is $\{\varepsilon_k{}^2\}$, hence $E_2 = \sum_{k=0}^{\infty} \varepsilon_k^2 < \infty$. It follows that for all k,

$$\| z^{k+1} - z^* \|^2 \le \| z^0 - z^* \|^2 + 4E_1(\| z^0 - z^* \| + 2E_1) + 4E_2 - \Delta_1(2 - \Delta_2) \sum_{i=0}^{k} \| Q_i(z^i) \|^2 .$$

Letting $k \to \infty$, we have that

$$\sum_{i=0}^{\infty} \| Q_i(z^i) \|^2 < \infty \quad \Rightarrow \quad \lim_{k \to \infty} Q_k(z^k) = 0 .$$

For all k, define (x^k, y^k) to be the unique point in T such that $x^k + c_k y^k = z^k$. Then $Q_k(z^k) \to 0$ implies that $z^k - x^k \to 0$. Furthermore, since $\{c_k\}$ is bounded away from zero, we also have $c_k^{-1} Q_k(z^k) = y_k \to 0$.

Now, $\{z^k\}$ is bounded, and so possesses at least one weak cluster point. Let z^∞ be any weak cluster point of $\{z^k\}$. Let " $\xrightarrow{w}$ " denote weak convergence, and let $\{z^{k(j)}\}_{j=0}^{\infty}$ be a subsequence such that $z^{k(j)} \xrightarrow{w} z^\infty$. Since $z^k - x^k \to 0$, we also have $x^{k(j)} \xrightarrow{w} z^\infty$.

Let (x, y) be any point in T. By the monotonicity of T, we have that $\langle x - x^k, y - y^k \rangle \ge 0$ for all k. Taking the limit over the subsequence $k(j)$ and using that $x^{k(j)} \xrightarrow{w} z^\infty$ and $y_k \to 0$, one obtains $\langle x - z^\infty, y - 0 \rangle \ge 0$. Since (x, y) was chosen arbitrarily, we conclude from the assumed maximality of T that $(z^\infty, 0) \in T$, that is $z^\infty \in \text{zer}(T)$.

It remains to show that $\{z^k\}$ has only one weak cluster point. Consider any zero z^* of T. Then $\| z^{k+1} - z^* \| \le \| z^k - z^* \| + 2\varepsilon_k$, and thus $\| z^k - z^* \| \le \| z^0 - z^* \| + 2E_1$ for all k. Therefore,

$$\alpha^* = \liminf_{k \to \infty} \|z^k - z^*\|$$

is finite and nonnegative, and one may show that $\| z^k - z^* \| \to \alpha^*$. Now take any two weak cluster points z_1^∞ and z_2^∞ of $\{z^k\}$. By the reasoning above, both are zeroes of T, and hence

$$\alpha_1 = \lim_{k \to \infty} \|z^k - z_1^\infty\|$$

$$\alpha_2 = \lim_{k \to \infty} \|z^k - z_2^\infty\|$$

both exist and are finite. Writing

$$\|z^k - z_2^\infty\|^2 = \|z^k - z_1^\infty\|^2 + 2\langle z^k - z_1^\infty, z_1^\infty - z_2^\infty\rangle + \|z_1^\infty - z_2^\infty\|^2 ,$$

one concludes that

$$\lim_{k \to \infty} \langle z^k - z_1^\infty, z_1^\infty - z_2^\infty\rangle = \tfrac{1}{2}\left(\alpha_2^2 - \alpha_1^2 - \|z_1^\infty - z_2^\infty\|^2\right) .$$

Since z_1^∞ is a weak cluster point of $\{z^k\}$, this limit must be zero. Hence,

$$\alpha_2^2 = \alpha_1^2 + \|z_1^\infty - z_2^\infty\|^2 .$$

Reversing the roles of z_1^∞ and z_2^∞, we also obtain that

$$\alpha_1^2 = \alpha_2^2 + \|z_1^\infty - z_2^\infty\|^2 .$$

We then are forced to conclude that $\| z_1^\infty - z_2^\infty \| = 0$, that is, $z_1^\infty = z_2^\infty$. Thus, $\{z^k\}$ has exactly one weak cluster point. This concludes the proof in the case that T possesses at least one zero.

Now consider the case in which T has no zero. We show by contradiction that $\{z^k\}$ is unbounded. Suppose that $\{z^k\}$ is bounded, that is, there is some finite S such that $\| z^k \| < S$ for all k. Let

$$\bar{\varepsilon} = \sup_{k \geq 0} \{\varepsilon_k\} .$$

Then let

$$r = \frac{2S}{\min\{1, \Delta_1\}} + \bar{\varepsilon} + 1 \ .$$

We claim that for all k, one has $\|z^k\|, \|w^k\|, \|J_{c_kT}(z^k)\| < r - 1$. Clearly, $\|z^k\| < S < r - 1$, so the claim holds for z^k. Now, $w^k = \rho_k^{-1}(z^{k+1} - (1 - \rho_k)z^k)$, so

$$\|w^k\| \le \frac{1}{\rho_k}\left(\|z^{k+1}\| - (1 - \rho_k)z^k\right) < \frac{1}{\Delta_1}\left(S + S\right) = \frac{2S}{\Delta_1} \le r - 1 \ .$$

Finally,

$$\|w^k - J_{c_kT}(z^k)\| \le \varepsilon_k \quad \Rightarrow \quad \|J_{c_kT}(z^k)\| \le \|w^k\| + \varepsilon_k < \frac{2S}{\Delta_1} + \bar{\varepsilon} \le r - 1 \ .$$

Now, let $h: \mathbb{R}^n \to [0, \infty]$ be the convex function

$$h(x) = \begin{cases} 0\,, & \|x\| \le r \\ +\infty\,, & \|x\| > r \end{cases} \ ,$$

and let $T = T' + \partial h$, so that

$$T'(\mathrm{x}) = \begin{cases} T(x)\,, & \|x\| < r \\ \{\, y + ax \mid y \in T(x), a \ge 0 \}\,, & \|x\| = r \\ \varnothing\,, & \|x\| > r \end{cases} \ .$$

Since $\operatorname{dom} T \cap \operatorname{int}(\operatorname{dom} \partial h) = \operatorname{dom} T \cap \{x \mid \|x\| < r\} \ne \varnothing$, T' is maximal monotone (Rockafellar 1970c). Further, $\operatorname{dom} T'$ is bounded, so $\operatorname{zer}(T') \ne \varnothing$ (Rockafellar 1969). Since $\|z^k\|, \|w^k\|$, and $\|J_{c_kT}(z^k)\|$ are all less than r for all k, the sequence $\{z^k\}$ obeys the generalized proximal point iteration for T', as well as for T. That is,

$$z^{k+1} = (1 - \rho_k)z^k + \rho_k w^k \qquad \forall\, k \ge 0,$$

where

$$\|w^k - (I + cT')^{-1}(z^k)\| \le \varepsilon_k \ .$$

By the logic of the first part of the theorem, $\{z^k\}$ converges weakly to some zero z^∞ of T'. Furthermore, as $\|z^k\| \le r - 1$ for all k, $\|z^\infty\| \le r - 1 < r$, and so $T'(z^\infty) = T(z^\infty)$, and z^∞ is also a zero of T. This is a contradiction; hence, we conclude that $\{z^k\}$ cannot be bounded.

■

4. *Decomposition: Douglas-Rachford Splitting Methods*

The main difficulty in applying the proximal point algorithm and related methods is the evaluation of *inverses* of operators of the form $I + \lambda T$, where $\lambda > 0$. For many maximal monotone operators T, such inversion operations may be prohibitively difficult. Now suppose that we can choose two maximal monotone operators A and B such that $A+B = T$, but $J_{\lambda A}$ and $J_{\lambda B}$ are easier to evaluate that $J_{\lambda T}$. A *splitting algorithm* is a method that employs the resolvents $J_{\lambda A}$ and $J_{\lambda B}$ of A and B, but does not use the resolvent $J_{\lambda T}$ of the original operator T. Here, we will consider only one kind of splitting algorithm, the Douglas-Rachford scheme of Lions and Mercier (1979). It is patterned after an alternating direction method for the discretized heat equation that dates back to the mid-1950's (Douglas and Rachford 1956).

Let us fix some $\lambda > 0$ and two maximal monotone operators A and B. The sequence $\{z^k\}_{k=0}^{\infty}$ is said to obey the *Douglas-Rachford recursion* for λ, A, and B if

$$z^{k+1} = J_{\lambda A}((2J_{\lambda B} - I)(z^k)) + (I - J_{\lambda B})(z^k) \quad .$$

Given any sequence obeying this recurrence, let (x^k, b^k) be, for all $k \geq 0$, the unique element of B such that $x^k + \lambda b^k = z^k$ (see the Representation Lemma, Corollary 2.3). Then, for all k, one has

$$(I - J_{\lambda B})(z^k) = x^k + \lambda b^k - x^k = \lambda b^k$$

$$(2J_{\lambda B} - I)(z^k) = 2x^k - (x^k + \lambda b^k) = x^k - \lambda b^k \quad .$$

Similarly, if $(y^k, a^k) \in A$, then $J_{\lambda A}(y^k + \lambda a^k) = y^k$. In view of these identities, one may give the following alternative prescription for finding z^{k+1} from z^k:

(a) Find the unique $(y^{k+1}, a^{k+1}) \in A$ such that $y^{k+1} + \lambda a^{k+1} = x^k - \lambda b^k$

(b) Find the unique $(x^{k+1}, b^{k+1}) \in B$ such that $x^{k+1} + \lambda b^{k+1} = y^{k+1} + \lambda b^k$.

Lions' and Mercier's original analysis of Douglas-Rachford splitting (1979) centered on the operator

$$G_{\lambda,A,B} = J_{\lambda A} \circ (2J_{\lambda B} - I) + (I - J_{\lambda B}) \ ,$$

where $\circ$ denotes functional composition; the Douglas-Rachford recursion can be written $z^{k+1} = G_{\lambda,A,B}(z^k)$. Lions and Mercier showed that $G_{\lambda,A,B}$ is firmly nonexpansive, from which they obtained convergence of $\{z^k\}$. Our aim is to broaden their analysis by exploiting the connection between firm nonexpansiveness and maximal monotonicity.

Consider the operator

$$S_{\lambda,A,B} = (G_{\lambda,A,B})^{-1} - I \ .$$

We first seek a set-theoretical expression for $S_{\lambda,A,B}$. Following the algorithmic description (a)-(b) above, we arrive at the following expression for $G_{\lambda,A,B}$:

$$G_{\lambda,A,B} = \{(u + \lambda b, v + \lambda b) \mid (u, b) \in B, (v, a) \in A, v + \lambda a = u - \lambda b\} \ .$$

A simple manipulation provides an expression for $S_{\lambda,A,B} = (G_{\lambda,A,B})^{-1} - I$:

$$S_{\lambda,A,B} = (G_{\lambda,A,B})^{-1} - I = \{(v + \lambda b, u - v) \mid (u, b) \in B, (v, a) \in A, v + \lambda a = u - \lambda b\} \ .$$

Given any Hilbert space $\mathcal{H}$, $\lambda > 0$, and operators A and B on $\mathcal{H}$, we define $S_{\lambda,A,B}$ to be the *splitting operator* of A and B with respect to λ. We now directly establish the maximal monotonicity of $S_{\lambda,A,B}$.

Theorem 4. If A and B are monotone then $S_{\lambda,A,B}$ is monotone. If A and B are maximal monotone, then $S_{\lambda,A,B}$ is maximal monotone.

Proof. First we show that $S_{\lambda,A,B}$ is monotone. Let $u, b, v, a, u', b', v', a' \in \mathcal{H}$ be such that $(u, b), (u', b') \in B$, $(v, a), (v', a') \in A$, $v + \lambda a = u - \lambda b$, and $v' + \lambda a' = u' - \lambda b'$. Then

$$a = \frac{1}{\lambda}(u - v) - b \qquad a' = \frac{1}{\lambda}(u' - v') - b',$$

and

$$\begin{aligned}
&\langle (v'+\lambda b') - (v+\lambda b), (u' - v') - (u - v) \rangle \\
&= \lambda \langle (v'+\lambda b') - (v+\lambda b), \lambda^{-1}(u' - v') - b' - \lambda^{-1}(u - v) + b \rangle \\
&\quad + \lambda \langle (v'+\lambda b') - (v+\lambda b), b' - b \rangle \\
&= \lambda \langle v' - v, \lambda^{-1}(u' - v') - b' - \lambda^{-1}(u - v) + b \rangle \\
&\quad + \lambda^2 \langle b' - b, \lambda^{-1}(u' - v') - b' - \lambda^{-1}(u - v) + b \rangle \\
&\quad + \lambda \langle v' - v, b' - b \rangle + \lambda^2 \langle b' - b, b' - b \rangle \\
&= \lambda \langle v' - v, a' - a \rangle + \lambda \langle b' - b, u' - u \rangle - \lambda \langle b' - b, v' - v \rangle - \lambda^2 \langle b' - b, b' - b \rangle \\
&\quad + \lambda \langle v' - v, b' - b \rangle + \lambda^2 \langle b' - b, b' - b \rangle \\
&= \lambda \langle v' - v, a' - a \rangle + \lambda \langle b' - b, u' - u \rangle .
\end{aligned}$$

By the monotonicity of A and B, the two terms in the final line are nonnegative, so we obtain that $\langle (v'+\lambda b') - (v+\lambda b), (u' - v') - (u - v) \rangle \geq 0$, and $S_{\lambda,A,B}$ is monotone. It remains to show that $S_{\lambda,A,B}$ is maximal in the case that A and B are. By Theorem 1, we only need to show that $(I + S_{\lambda,A,B})^{-1} = G_{\lambda,A,B} = J_{\lambda A} \circ (2J_{\lambda B} - I) + (I - J_{\lambda B})$ has full domain. This is indeed the case, as $J_{\lambda A}$ and $J_{\lambda B}$ are defined everywhere. ■

Combining Theorems 4 and 2, we have the key Lions-Mercier result:

Corollary 4.1. If A and B are maximal monotone, then $G_{\lambda,A,B} = (I + S_{\lambda,A,B})^{-1}$ is firmly nonexpansive and has full domain.

There is also an important relationship between the zeroes of $S_{\lambda,A,B}$ and those of $A+B$:

Theorem 5. Given $\lambda > 0$ and operators A and B on $\mathcal{H}$,

$$\mathrm{zer}(S_{\lambda,A,B}) = Z_\lambda^* \triangleq \{ u+\lambda b \mid b \in Bu, \ -b \in Au \}$$

$$\subseteq \{\ u+\lambda b \mid u \in \text{zer}(A+B),\ b \in Bu\}\ .$$

Proof. Let $S = S_{\lambda,A,B}$. We wish to show that zer(S) is equal to Z_λ^*. Let $z \in$ zer(S). Then there exist some $u, b, v, a \in \mathcal{H}$ such that $v+\lambda b = z$, $u - v = 0$, $(u, b) \in B$, and $(v, a) \in A$. So,

$$u - v = 0 \ \Rightarrow\ u = v \ \Rightarrow\ \lambda a = -\lambda b \ \Rightarrow\ a = -b\ ,$$

and we have $u+\lambda b = z$, $(u, b) \in B$, and $(u, -b) \in A$, hence $z \in Z_\lambda^*$. Conversely, if $z \in Z_\lambda^*$, then $z = u+\lambda b$, $b \in Bu$, and $-b \in Au$. Setting $u = v$ and $a = -b$, we see that $(z, 0) \in S$. Finally, the inclusion $Z_\lambda^* \subseteq \{\ u+\lambda b \mid u \in \text{zer}(A+B),\ b \in Bu\}$ follows because $b \in Bu$ and $-b \in Au$ imply that $u \in$ zer($A+B$). ■

Thus, given any zero z of $S_{\lambda,A,B}$, $J_{\lambda B}(z)$ is a zero of $A+B$. Thus one may imagine finding a zero of $A+B$ by using the proximal point algorithm on $S_{\lambda,A,B}$, and then applying the operator $J_{\lambda B}$ to the result. In fact, this is precisely what the Douglas-Rachford splitting method does:

Theorem 6. The Douglas-Rachford iteration $z^{k+1} = [J_{\lambda A} \circ (2J_{\lambda B} - I) + (I - J_{\lambda B})]z^k$ is equivalent to applying the proximal point algorithm to the maximal monotone operator $S_{\lambda,A,B}$, with the proximal point stepsizes c_k fixed at 1, and exact evaluation of resolvents.

Proof. The Douglas-Rachford iteration is $z^{k+1} = G_{\lambda,A,B}(z^k)$, which is just $z^{k+1} = (I + S_{\lambda,A,B})^{-1}(z^k)$. ■

Theorem 6 appears to be a new characterization of Douglas-Rachford splitting. In view of Theorem 3, Theorem 5, and the Lipschitz continuity of $J_{\lambda B}$, we immediately obtain the following Lions-Mercier convergence result:

Corollary 6.1 (Lions and Mercier 1979) If $A+B$ has a zero, then the Douglas-Rachford splitting method produces a sequence $\{z^k\}$ weakly convergent to a limit z of the form $u + \lambda b$, where $u \in$ zer($A+B$), $b \in Bu$, and $-b \in Au$. If procedure (a)-(b) is used to implement the Douglas-Rachford iteration, then $\{x^k\} = \{J_{\lambda B}(z^k)\}$ converges to some zero of $A+B$.

Theorem 3 also states that, in general Hilbert space, the proximal point algorithm produces an unbounded sequence when applied to a maximal monotone operator that has no zeroes. Thus, one obtains a further result apparently unknown to Lions and Mercier:

Corollary 6.2. Suppose A and B are maximal monotone and $\text{zer}(A+B) = \varnothing$. Then the sequence $\{z^k\}$ produced by the Douglas-Rachford splitting is unbounded. If procedure (a)-(b) is used, then at least one of the sequences $\{x^k\}$ or $\{b^k\}$ is unbounded.

Note that it is not necessary to assume that $A+B$ is maximal; only A and B need be maximal.

Because the Douglas-Rachford splitting method is a special case of the proximal point algorithm as applied to the splitting operator $S_{\lambda,A,B}$, a number of generalizations of Douglas-Rachford splitting now suggest themselves: one can imagine applying the *generalized* proximal point algorithm to $S_{\lambda,A,B}$, with stepsizes c_k other than 1, with relaxation factors ρ_k other than 1, or with approximate evaluation of the resolvent $G_{\lambda,A,B}$. We will show that while the first of these options is not practical, the last two are.

Consider, for any $c > 0$, trying to compute $(I + cS_{\lambda,A,B})^{-1}(z)$. Now,

$$(I + cS_{\lambda,A,B})^{-1} = \{ \, ((1-c)v + cu + \lambda b,\ v + \lambda b) \mid (u, b) \in B,\ (v, a) \in A,\ v + \lambda a = u - \lambda b \, \} .$$

Thus, to calculate $(I + cS_{\lambda,A,B})^{-1}(z)$, one must find $(u, b) \in B$ and $(v, a) \in A$ such that

$$(1-c)v + cu + \lambda b = z \qquad\qquad a = \tfrac{1}{\lambda}(u - v) - b \ .$$

Alternatively, we may state the problem as that of finding $u, v \in \mathcal{H}$ such that

$$z - (cu + (1-c)v) \in \lambda Bu \qquad\qquad -z + ((1+c)u - cv) \in \lambda Av \, .$$

This does not appear to be a particularly easy problem. Specifically, it does not appear to be any less difficult than the calculation of $J_{\lambda(A+B)}$ at an arbitrary point z, which, when using a splitting algorithm, we are expressly trying to avoid. By comparison, that calculation involves finding $(u, b) \in B$ such that $(u, \lambda^{-1}(z-u) - b) \in A$.

Consider, however, what happens when one fixes c at 1. Then one has only to find

$$(u, b) \in B \quad \text{such that} \quad u+\lambda b = z$$

$$(v, a) \in A \quad \text{such that} \quad v+\lambda a = u - \lambda b.$$

The conditions $(u, b) \in B$, $u+\lambda b = z$ uniquely determine $u = J_{\lambda,B}(z)$ and $b = \frac{1}{\lambda}(z-u)$ independently of v. Once u is known, then v is likewise uniquely determined by $u = J_{\lambda,A}(u - \lambda b)$. We have thus achieved a *decomposition* in which the calculation of $J_{S_{\lambda,A,B}} = (I + S_{\lambda,A,B})^{-1}$ is replaced by separate, sequential evaluations of $J_{\lambda A} = (I+\lambda A)^{-1}$ and $J_{\lambda B} = (I+\lambda B)^{-1}$. This procedure is essentially the procedure (a)-(b) given above. It seems that keeping $c = 1$ at all times is critical to the decomposition. Spingarn (1985) has already recognized this phenomenon, but in the more restrictive context of his method of partial inverses. The next section will show that Spingarn's method is a special case of Douglas-Rachford splitting.

The formulation of the splitting operator $S_{\lambda,A,B}$ is a way of combining A and B having the special property that evaluating the resolvent $G_{\lambda,A,B} = (I + S_{\lambda,A,B})^{-1}$ decomposes into sequential evaluations of $J_{\lambda A}$ and $J_{\lambda B}$. Simple addition of operators does not have such a decomposition property. Furthermore, the close relationship between $\text{zer}(S_{\lambda,A,B})$ and $\text{zer}(A+B)$ makes $S_{\lambda,A,B}$ useful in finding zeroes of $A+B$.

Despite the impracticality of using stepsizes other than 1, it *is* possible to use varying relaxation factors, and to evaluate $G_{\lambda,A,B} = (I + S_{\lambda,A,B})^{-1}$ approximately, obtaining a *generalized Douglas-Rachford splitting method.* The properties of this (new) method are summarized by the following theorem:

Theorem 7. Given a Hilbert space $\mathcal{H}$, some $z^0 \in \mathcal{H}$, $\lambda > 0$, and maximal monotone operators A and B on $\mathcal{H}$, let $\{z^k\}_{k=0}^{\infty} \subseteq \mathbb{R}^n$, $\{u^k\}_{k=0}^{\infty} \subseteq \mathbb{R}^n$, $\{v^k\}_{k=1}^{\infty} \subseteq \mathbb{R}^n$, $\{\alpha_k\}_{k=0}^{\infty} \subseteq [0, \infty)$, $\{\beta_k\}_{k=0}^{\infty} \subseteq [0, \infty)$, and $\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2)$ conform to the following conditions:

$$\| u^k - J_{\lambda B}(z^k) \| \leq \beta_k \qquad \forall\, k \geq 0 \qquad \text{(T1)}$$

$$\| v^{k+1} - J_{\lambda A}(2u^k - z^k) \| \leq \alpha_k \qquad \forall\, k \geq 0 \qquad \text{(T2)}$$

$$z^{k+1} = z^k + \rho_k(v^{k+1} - u^k) \qquad \forall\, k \geq 0 \qquad \text{(T3)}$$

$$\sum_{k=0}^{\infty} \alpha_k < \infty$$

$$\sum_{k=0}^{\infty} \beta_k < \infty$$

$$0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 \quad .$$

Then if $\operatorname{zer}(A+B) \neq \varnothing$, $\{z^k\}$ converges weakly to some element of $Z_\lambda^* = \{\, u+\lambda b \mid b \in Bu, -b \in Au \,\}$. If $\operatorname{zer}(A+B) = \varnothing$, then $\{z^k\}$ is unbounded.

Proof. Fix any k. Then $\| u^k - J_{\lambda B}(z^k) \| \leq \beta_k$ implies that

$$\| (2u^k - z^k) - (2J_{\lambda B} - I)(z^k) \| \leq 2\beta_k \quad .$$

Since $J_{\lambda A}$ is nonexpansive,

$$\| J_{\lambda A}(2u^k - z^k) - J_{\lambda A}(2J_{\lambda B} - I)(z^k) \| \leq 2\beta_k \quad ,$$

and so

$$\| v^{k+1} - J_{\lambda A}(2J_{\lambda B} - I)(z^k) \| \leq 2\beta_k + \alpha_k$$

$$\| (v^{k+1} + z^k - u^k) - [J_{\lambda A}(2J_{\lambda B} - I) + (I - J_{\lambda B})](z^k) \| \leq 3\beta_k + \alpha_k \quad .$$

Let $\varepsilon_k = 3\beta_k + \alpha_k$ for all k. Then

$$\sum_{k=0}^{\infty} \varepsilon_k = 3\sum_{k=0}^{\infty} \beta_k + \sum_{k=0}^{\infty} \alpha_k < \infty \quad .$$

We also have

$$z^{k+1} = z^k + \rho_k(v^{k+1} - u^k) = (1 - \rho_k)z^k + \rho_k(v^{k+1} + z^k - u^k) \ .$$

Thus, letting $y^k = v^{k+1} + z^k - u^k$, we have

$$\begin{aligned} 0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 \\ \sum_{k=0}^{\infty} \varepsilon_k < +\infty \ , \\ \| y^k - G_{\lambda,A,B}(z^k) \| \leq \varepsilon_k & \qquad \forall\, k \geq 0 \\ z^{k+1} = (1 - \rho_k)z^k + \rho_k y^k & \qquad \forall\, k \geq 0 \ . \end{aligned}$$

The conclusion then follows from Theorems 3 and 5. ■

In at least one real example (Eckstein 1989, Chapter 7), using the generalized Douglas-Rachford splitting method with relaxation factors ρ_k other than 1 has been shown to converge faster than regular Douglas-Rachford splitting. Thus, the above convergence result is of some practical significance.

5. *Some Interesting Special Cases*

We now consider some interesting applications of splitting operator theory, namely the method of partial inverses (Spingarn 1983, 1985b) and the generalized alternating direction method of multipliers. We begin with the method of partial inverses.

Let T be an operator on a Hilbert space $\mathcal{H}$, and let V be any linear subspace of $\mathcal{H}$, $V^{\perp}$ denoting its orthogonal complement. Then the *partial inverse* T_V of T with respect to V is the operator obtained by swapping the $V^{\perp}$ components of each pair in T, thus (Spingarn 1983, 1985b):

$$T_V = \{(x_V + y_{V^{\perp}}, y_V + x_{V^{\perp}}) \mid (x, y) \in T\} \ .$$

Here, we use the notation that for any vector z, z_V denotes the projection of z on V, and $z_{V\perp}$ its projection onto $V^\perp$.

Spingarn has suggested applying the proximal point algorithm to T_V to solve the problem

$$\text{Find } (x, y) \in T \text{ such that } x \in V \text{ and } y \in V^\perp , \qquad \text{(ZV)}$$

where T is maximal monotone. In particular, if $T = \partial f$, where f is a closed proper convex function, this problem reduces to that of minimizing f over V. One application of this method is the "progressive hedging" stochastic programming method of Rockafellar and Wets (1987).

Consider now the operator

$$N_V = V \times V^\perp = \{(x, y) \mid x \in V, y \in V^\perp\} .$$

It is easily seen that N_V is the subdifferential $\partial(\delta_V)$ of the closed proper convex function

$$\delta_V(x) = \begin{cases} 0 , & x \in V \\ +\infty , & x \notin V \end{cases} ,$$

and hence that N_V is maximal monotone. Now consider the problem

$$\text{Find } x \text{ such that } 0 \in (T + N_V)x , \qquad \text{(ZV')}$$

which is equivalent to (ZV).

If one forms the splitting operator $S_{\lambda,A,B}$ with $\lambda = 1$, $A = N_V = V \times V^\perp$, and $B = T$, one obtains

$$\begin{aligned} S_{1,V\times V^\perp,T} &= \{(v + b, u - v) \mid (u, b) \in T, v \in V, a \in V^\perp, v + a = u - b\} . \\ &= \{((u - b)_V + b, u - (u - b)_V) \mid (u, b) \in T\} \\ &= \{(u_V + b_{V\perp}, b_V + u_{V\perp}) \mid (u, b) \in T\} \\ &= T_V . \end{aligned}$$

Thus, the partial inverse T_V is a special kind of splitting operator, and applying the proximal point algorithm to T_V is a specialized form of Douglas-Rachford splitting. Naturally, one can apply the generalized proximal point algorithm to T_V just as easily one can apply the regular proximal point algorithm, and one can allow values of λ (but not c_k) other than 1. Following a derivation similar to Spingarn's (1985b), one obtains the following algorithm for (ZV):

> Start with any $x^0 \in V$, $y^0 \in V^\perp$.
>
> At iteration k:
>
> Find $\tilde{y}^k \in \mathcal{H}$ such that $\|\tilde{y}^k - J_{\lambda T}(x^k + y^k)\| \leq \beta^k$.
>
> Let $\tilde{x}^k = (x^k + y^k) - \tilde{y}^k$.
>
> Let $x^{k+1} = (1 - \rho_k)x^k + \rho_k(\tilde{x}^k)_V$.
>
> Let $y^{k+1} = (1 - \rho_k)y^k + \rho_k(\tilde{x}^k)_{V^\perp}$.

Here $\{\rho_k\}_{k=0}^{\infty}$ and $\{\beta_k\}_{k=0}^{\infty}$ are sequences meeting the restrictions of Theorem 7. It is interesting to compare this method to Algorithm 1 of Spingarn (1985b). In cases where $T = \partial f$, the computation of $\tilde{y}^k$ reduces to an approximate, unconstrained minimization of f plus a quadratic term.

In addition to partial-inverse-based methods, the class of Douglas-Rachford splitting algorithms also includes the general monotone operator method of Gol'shtein (1987), and related convex programming methods (Gol'shtein 1985, 1986). Demonstrating this relationship is rather laborious, however, and interested readers should refer to Eckstein (1989).

We now turn to our second example application of splitting operator theory, the derivation of a new augmented Lagrangian method called the *generalized alternating direction method of multipliers*.

Consider a general finite-dimensional optimization problem of the form

$$\text{minimize}_{\ \mathbf{x}\in\mathbb{R}^n}\ f(\mathbf{x}) + g(\mathbf{M}\mathbf{x}) \ , \tag{P}$$

where $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ and $g\colon \mathbb{R}^n \to (-\infty, +\infty]$ are closed proper convex, and $\mathbf{M}$ is some $m \times n$ matrix. By writing (P) in the form

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) + g(\mathbf{w}) \\ \text{subject to} & \mathbf{M}\mathbf{x} = \mathbf{w} \ , \end{array} \tag{P'}$$

and attaching a multiplier vector $\mathbf{p} \in \mathbb{R}^m$ to the constraints $\mathbf{M}\mathbf{x} = \mathbf{w}$, one obtains an equivalent dual problem

$$\text{maximize}_{\ \mathbf{p}\in\mathbb{R}^m}\ -\left(f^*(-\mathbf{M}^{\mathsf{T}}\mathbf{p}) + g^*(\mathbf{p})\right) \ , \tag{D}$$

where * denotes the convex conjugacy operation. One way of solving the problem (P)-(D) is to let $A = \partial[f^* \circ (-\mathbf{M}^{\mathsf{T}})]$ and $B = \partial g^*$, and apply Douglas-Rachford splitting to A and B. This approach was shown by Gabay (1983) to yield the *alternating direction method of multipliers* (Glowinski and Marroco 1975, Gabay and Mercier 1976, Fortin and Glowinski 1983, Gabay 1983, Glowinski and Le Tallec 1987),

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg\min_{\mathbf{x}} \left\{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x}\rangle + \tfrac{\lambda}{2}\|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2\right\} \\ \mathbf{w}^{k+1} &= \arg\min_{\mathbf{w}} \left\{g(\mathbf{w}) - \langle \mathbf{p}^k, \mathbf{w}\rangle + \tfrac{\lambda}{2}\|\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}\|^2\right\} \\ \mathbf{p}^{k+1} &= \mathbf{p}^k + \lambda_k(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^{k+1}) \quad . \end{aligned}$$

This method resembles the conventional Hestenes-Powell method of multipliers for (P'), except that it minimizes the augmented Lagrangian function

$$L_\lambda(\mathbf{x}, \mathbf{w}, \mathbf{p}) = f(\mathbf{x}) + g(\mathbf{w}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} - \mathbf{w}\rangle + \tfrac{\lambda}{2}\|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2$$

first with respect to $\mathbf{x}$, and then with respect to $\mathbf{w}$, rather than with respect to both $\mathbf{x}$ and $\mathbf{w}$ simultaneously. Notice also that the penalty parameter λ is not permitted to vary with k.

We now show how Theorem 7 yields a generalized version of this algorithm. Let the maximal monotone operators $A = \partial[f^* \circ (-\mathbf{M}^{\mathrm{T}})]$ and $B = \partial g^*$ be defined as above.

A pair $(\mathbf{x}, \mathbf{p}) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be a *Kuhn-Tucker pair* for (P) if $(\mathbf{x}, -\mathbf{M}^{\mathrm{T}}\mathbf{p}) \in \partial f$ and $(\mathbf{Mx}, \mathbf{p}) \in \partial g$. It is a basic exercise in convex analysis to show that if $(\mathbf{x}, \mathbf{p})$ is a Kuhn-Tucker pair, then $\mathbf{x}$ is optimal for (P) and $\mathbf{p}$ is optimal for (D), and also that if $\mathbf{p} \in \operatorname{zer}(A+B)$, then $\mathbf{p}$ is optimal for (D). We can now state a new variation on the alternating direction method of multipliers for (P):

Theorem 8 (The generalized alternating direction method of multipliers). Consider a convex program in the form (P), minimize$_{\mathbf{x} \in \mathbb{R}^n}$ $f(\mathbf{x}) + g(\mathbf{Mx})$, where $\mathbf{M}$ has full column rank. Let $\mathbf{p}^0$, $\mathbf{z}^0 \in \mathbb{R}^m$, and suppose we are given $\lambda > 0$ and

$$\{\mu_k\}_{k=0}^{\infty} \subseteq [0, \infty), \qquad \sum_{k=0}^{\infty} \mu_k < \infty$$

$$\{\nu_k\}_{k=0}^{\infty} \subseteq [0, \infty), \qquad \sum_{k=0}^{\infty} \nu_k < \infty$$

$$\{\rho_k\}_{k=0}^{\infty} \subseteq (0, 2), \qquad 0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2 .$$

Suppose $\{\mathbf{x}^k\}_{k=1}^{\infty}$, $\{\mathbf{w}^k\}_{k=0}^{\infty}$, and $\{\mathbf{p}^k\}_{k=0}^{\infty}$ conform, for all k, to

$$\left\| \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{Mx}\rangle + \tfrac{\lambda}{2}\|\mathbf{Mx} - \mathbf{w}^k\|^2\} \right\| \leq \mu_k$$

$$\left\| \mathbf{w}^{k+1} - \arg\min_{\mathbf{w}} \{g(\mathbf{w}) - \langle \mathbf{p}^k, \mathbf{w}\rangle + \tfrac{\lambda}{2}\|\rho_k\mathbf{Mx}^{k+1} + (1-\rho_k)\mathbf{w}^k - \mathbf{w}\|^2\} \right\| \leq \nu_k$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \lambda(\rho_k\mathbf{Mx}^{k+1} + (1-\rho_k)\mathbf{w}^k - \mathbf{w}^{k+1}) .$$

Then if (P) has a Kuhn-Tucker pair, $\{\mathbf{x}^k\}$ converges to a solution of (P) and $\{\mathbf{p}^k\}$ converges to a solution of the dual problem (D). Furthermore, $\{\mathbf{w}^k\}$ converges to $\mathbf{Mx}^*$, where $\mathbf{x}^*$ is the limit of $\{\mathbf{x}^k\}$. If (D) has no optimal solution, then at least one of the sequences $\{\mathbf{p}^k\}$ or $\{\mathbf{w}^k\}$

is unbounded.

Proof. Let

$$\begin{aligned} \mathbf{z}^k &= \mathbf{p}^k + \lambda \mathbf{w}^k & \forall\, k \geq 0 \\ \mathbf{q}^k &= \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k) & \forall\, k \geq 0 \\ \alpha_k &= \lambda \|\mathbf{M}\| \mu_k & \forall\, k \geq 0 \\ \beta_0 &= \| \mathbf{p}^0 - J_{\lambda B}(\mathbf{p}^0 + \lambda \mathbf{w}^0) \| & \\ \beta_k &= \lambda v_k & \forall\, k \geq 1 \ , \end{aligned}$$

where $\|\mathbf{M}\|$ denotes the l_2-norm of the matrix $\mathbf{M}$,

$$\|\mathbf{M}\| = \sup_{\mathbf{x} \neq 0} \left\{ \frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|} \right\} \quad .$$

We wish to establish that the following hold for all $k \geq 0$:

$$\| \mathbf{p}^k - J_{\lambda B}(\mathbf{z}^k) \| \leq \beta_k \tag{Y1}$$

$$\| \mathbf{q}^k - J_{\lambda A}(2\mathbf{p}^k - \mathbf{z}^k) \| \leq \alpha_k \tag{Y2}$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + \rho_k(\mathbf{q}^k - \mathbf{p}^k) \ . \tag{Y3}$$

For $k = 0$, (Y1) is valid by the choice of β_0. Now suppose (Y1) holds for some k; we show that (Y2) also holds for k. Let

$$\begin{aligned} \bar{\mathbf{x}}^k &= \arg\min_{\mathbf{x}} \{ f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \tfrac{\lambda}{2} \| \mathbf{M}\mathbf{x} - \mathbf{w}^k \|^2 \} \\ \tilde{\mathbf{p}}^k &= (\mathbf{p}^k - \lambda \mathbf{w}^k) + \lambda \mathbf{M} \bar{\mathbf{x}}^k \quad . \end{aligned}$$

The existence of a unique $\bar{\mathbf{x}}^k$ is assured because f is proper and $\mathbf{M}$ has full column rank. Then

$$\begin{aligned} & \mathbf{0} \in \partial_{\mathbf{x}} [f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \tfrac{\lambda}{2} \| \mathbf{M}\mathbf{x} - \mathbf{w}^k \|^2]_{\mathbf{x} = \bar{\mathbf{x}}^k} \\ \Rightarrow \quad & \mathbf{0} \in \partial f(\bar{\mathbf{x}}^k) + \mathbf{M}^{\mathsf{T}} \mathbf{p}^k + \lambda \mathbf{M}^{\mathsf{T}} (\mathbf{M}\bar{\mathbf{x}}^k - \mathbf{w}^k) \\ \Rightarrow \quad & \mathbf{0} \in \partial f(\bar{\mathbf{x}}^k) + \mathbf{M}^{\mathsf{T}} \tilde{\mathbf{p}}^k \\ \Rightarrow \quad & -\mathbf{M}^{\mathsf{T}} \tilde{\mathbf{p}}^k \in \partial f(\bar{\mathbf{x}}^k) \end{aligned}$$

$$\Rightarrow \quad \bar{\mathbf{x}}^k \in \partial f^*(-\mathbf{M}^{\mathsf{T}}\tilde{\mathbf{p}}^k)$$

$$\Rightarrow \quad -\mathbf{M}\bar{\mathbf{x}}^k \in \partial[f \circ (-\mathbf{M}^{\mathsf{T}})](\tilde{\mathbf{p}}^k) = A\tilde{\mathbf{p}}^k \ .$$

Also

$$\tilde{\mathbf{p}}^k + \lambda(-\mathbf{M}\bar{\mathbf{x}}^k) = \mathbf{p}^k - \lambda\mathbf{w}^k \ ,$$

so

$$\tilde{\mathbf{p}}^k = (I + \lambda A)^{-1}(\mathbf{p}^k - \lambda\mathbf{w}^k) = J_{\lambda A}(2\mathbf{p}^k - \mathbf{z}^k) \ .$$

Thus, from

$$\begin{aligned} \left\| \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x}\rangle + \tfrac{\lambda}{2}\|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2\} \right\| &\le \mu_k \\ \mathbf{q}^k &= \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k) \ , \end{aligned}$$

we obtain

$$\|\mathbf{x}^k - \bar{\mathbf{x}}^k\| \le \mu_k$$

$$\|\mathbf{q}^k - \tilde{\mathbf{p}}^k\| \le \lambda\|\mathbf{M}\|\mu_k \ ,$$

establishing (Y2) for k.

Suppose that (Y1) and (Y2) hold for some k. We now show that (Y3) holds for k and (Y1) holds for k+1. Let

$$\begin{aligned} \mathbf{s}^k &= \mathbf{z}^k + \rho_k(\mathbf{q}^k - \mathbf{p}^k) \\ &= \mathbf{p}^k + \mathbf{w}^k + \lambda\rho_k(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k) \\ &= \mathbf{p}^k + \lambda(\rho_k\mathbf{M}\mathbf{x}^{k+1} + (1 - \rho_k)\mathbf{w}^k) \ . \end{aligned}$$

and also

$$\overline{\mathbf{w}}^k = \arg\min_{\mathbf{w}} \{g(\mathbf{w}) - \langle \mathbf{p}^k, \mathbf{w}\rangle + \tfrac{\lambda}{2}\|(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k) - \mathbf{w}\|^2\}$$
$$\tilde{\mathbf{s}}^k = \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k - \overline{\mathbf{w}}^k) \quad .$$

The existence of $\overline{\mathbf{w}}^k$ is guaranteed because g is proper. We then have

$$\begin{aligned}
& \mathbf{0} \in \partial_{\mathbf{w}}[g(\mathbf{w}) - \langle \mathbf{p}^k, \mathbf{w}\rangle + \tfrac{\lambda}{2}\|(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k) - \mathbf{w}\|^2]_{\mathbf{w}=\overline{\mathbf{w}}^k} \\
\Rightarrow\quad & \mathbf{0} \in \partial g(\overline{\mathbf{w}}^k) - \mathbf{p}^k + \lambda(\overline{\mathbf{w}}^k - (\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k)) \\
\Rightarrow\quad & \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k - \overline{\mathbf{w}}^k) = \tilde{\mathbf{s}}^k \in \partial g(\overline{\mathbf{w}}^k) \\
\Rightarrow\quad & \overline{\mathbf{w}}^k \in \partial g^*(\tilde{\mathbf{s}}^k) = B\tilde{\mathbf{s}}^k \quad .
\end{aligned}$$

As

$$\tilde{\mathbf{s}}^k + \lambda\overline{\mathbf{w}}^k = \mathbf{p}^k - \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k - \overline{\mathbf{w}}^k) = \mathbf{s}^k \quad ,$$

we have $\tilde{\mathbf{s}}^k = J_{\lambda B}(\mathbf{s}^k)$.

The condition on $\mathbf{w}^{k+1}$ is just $\|\mathbf{w}^{k+1} - \overline{\mathbf{w}}^k\| \le \nu_k$, so $\|\mathbf{p}^{k+1} - \tilde{\mathbf{s}}^k\| \le \lambda\nu_k$. We also have

$$\begin{aligned}
\mathbf{z}^{k+1} &= \mathbf{p}^{k+1} + \lambda\mathbf{w}^{k+1} \\
&= \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k - \mathbf{w}^{k+1}) + \lambda\mathbf{w}^{k+1} \\
&= \mathbf{p}^k + \lambda(\rho_k \mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k) \\
&= \mathbf{s}^k \quad .
\end{aligned}$$

Thus, (Y3) holds for k, and (Y1) holds for k+1 by $\|\mathbf{p}^{k+1} - \tilde{\mathbf{s}}^k\| \le \lambda\nu_k$. By induction, then, (Y1)-(Y3) hold for all k. The summability of $\{\mu_k\}$ and $\{\nu_k\}$ implies the summability of $\{\beta_k\}$ and $\{\alpha_k\}$. Suppose (P) has a Kuhn-Tucker pair. Then by Theorem 7, $\{\mathbf{z}^k\}$ converges to some element $\mathbf{z}^*$ of $\{\ \mathbf{p}+\lambda\mathbf{w} \mid \mathbf{w} \in B\mathbf{p},\ -\mathbf{w} \in A\mathbf{p}\}$. Applying the continuous operator $J_{\lambda B}$ to $\{\mathbf{z}^k\}$ and using (Y1), we obtain $\mathbf{p}^k \to \mathbf{p}^*$ and $\mathbf{w}^k \to \mathbf{w}^*$, where $(\mathbf{p}^*, \mathbf{w}^*) \in B$ and $\mathbf{p}^* + \lambda\mathbf{w}^* = \mathbf{z}^*$. By rearranging the multiplier update formula, we have

$$(\mathbf{p}^{k+1} - \mathbf{p}^k) + (\mathbf{w}^{k+1} - \mathbf{w}^k) = \lambda\rho_k(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k)$$

for all $k \geq 0$. Taking limits and using that ρ_k is bounded away from zero, we obtain that $(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k) \rightarrow \mathbf{0}$, hence $\mathbf{M}\mathbf{x}^k \rightarrow \mathbf{w}^*$. As $\mathbf{M}$ has full column rank, $\mathbf{x}^k \rightarrow \mathbf{x}^*$, where $\mathbf{x}^*$ is such that $\mathbf{M}\mathbf{x}^* = \mathbf{w}^*$. We thus have $(\mathbf{p}^*, \mathbf{w}^*) = (\mathbf{p}^*, \mathbf{M}\mathbf{x}^*) \in B = \partial g^*$, and so $(\mathbf{M}\mathbf{x}^*, \mathbf{p}^*) \in \partial g$. Now, we also have that $-\mathbf{M}^\mathsf{T}\tilde{\mathbf{p}}^k \in \partial f(\overline{\mathbf{x}}^k)$, or, equivalently, $(-\mathbf{M}^\mathsf{T}\tilde{\mathbf{p}}^k, \overline{\mathbf{x}}^k) \in \partial f$, for all k. Using

$$0 \leq \|\mathbf{q}^k - \tilde{\mathbf{p}}^k\| = \|\mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{z}^k) - \tilde{\mathbf{p}}^k\| \leq \lambda\|\mathbf{M}\|\mu_k \rightarrow 0 \quad ,$$

we have by taking limits that $\tilde{\mathbf{p}}^k \rightarrow \mathbf{p}^*$, and since $\|\mathbf{x}^k - \overline{\mathbf{x}}^k\| \leq \mu_k \rightarrow 0$, we also have $\overline{\mathbf{x}}^k \rightarrow \mathbf{x}^*$. Therefore, $(-\mathbf{M}^\mathsf{T}\mathbf{p}^*, \mathbf{x}^*) \in \partial f$ by the limit property for maximal monotone operators (*e.g.* Brézis 1973). We conclude that $(\mathbf{x}^*, \mathbf{p}^*)$ is a Kuhn-Tucker pair for (P), and we obtain the indicated convergence of $\{\mathbf{x}^k\}$, $\{\mathbf{p}^k\}$, and $\{\mathbf{w}^k\}$.

Now suppose that (D) has no optimal solution. Then $\text{zer}(A+B)$ must be empty, and by Theorem 7, $\{\mathbf{z}^k\}$ must be an unbounded sequence. By the definition of $\{\mathbf{z}^k\}$, either $\{\mathbf{p}^k\}$ or $\{\mathbf{w}^k\}$ must then be unbounded. ■

In a practical iterative optimization subroutine, it may be difficult to tell if the condition

$$\left\| \mathbf{x}^{k+1} - \arg\min_{\mathbf{x}} \{f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x}\rangle + \tfrac{\lambda}{2}\|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2\} \right\| \leq \mu_k$$

or

$$\left\| \mathbf{w}^{k+1} - \arg\min_{\mathbf{w}} \{g(\mathbf{w}) - \langle \mathbf{p}^k, \mathbf{w}\rangle + \tfrac{\lambda}{2}\|\rho_k\mathbf{M}\mathbf{x}^{k+1} + (1-\rho_k)\mathbf{w}^k - \mathbf{w}\|^2\} \right\| \leq \nu_k$$

has been satisfied. For more implementable stopping criteria, which, under appropriate assumptions, imply these kinds of conditions, we refer to Rockafellar (1976b).

References

BERTSEKAS, D. (1982), *Constrained Optimization and Lagrange Multiplier Methods* (New York: Academic Press).

BRÉZIS, H., (1973), *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert* (Amsterdam: North-Holland, in French).

BRÉZIS, H., LIONS, P.-L. (1978), "Produits Infinis de Resolvantes", *Israel Journal of Mathematics* **29**(4):329-345 (in French).

DOLEŽAL, V. (1979), *Monotone Operators and Applications in Control and Network Theory* (Amsterdam: Elsevier Scientific).

DOUGLAS, J., RACHFORD, H. H. (1956), "On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables". *Transactions of the American Mathematical Society* **82**:421-439.

ECKSTEIN, J. (1988), *The Lions-Mercier Algorithm and the Alternating Direction Method are Instances of the Proximal Point Algorithm.* Report LIDS-P-1769, Laboratory for Information and Decision Sciences, MIT.

ECKSTEIN, J. (1989), *Splitting Methods for Monotone Operators with Applications to Parallel Optimization.* Doctoral dissertation, department of civil engineering, Massachusetts Institute of Technology. Available as report LIDS-TH-1877, Laboratory for Information and Decision Sciences, MIT.

FORTIN, M., GLOWINSKI, R. (1983), "On Decomposition-Coordination Methods Using an Augmented Lagrangian". In M. FORTIN, R. GLOWINSKI, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-Holland).

GABAY, D. (1983), "Applications of the Method of Multipliers to Variational Inequalities". In M. FORTIN, R. GLOWINSKI, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (Amsterdam: North-Holland).

GABAY, D., MERCIER B. (1976), "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations". *Computers and Mathematics with Applications* **2**:17-40.

GLOWINSKI, R., LE TALLEC, P. (1987), *Augmented Lagrangian Methods for the Solution of Variational Problems.* MRC Technical Summary Report #2965, Mathematics Research Center, University of Wisconsin – Madison.

GLOWINSKI, R., MARROCO, A. (1975), "Sur L'Approximation, par Elements Finis d'Ordre Un, et la Resolution, par Penalisation-Dualité, d'une Classe de Problemes de Dirichlet non Lineares". *Revue Française d'Automatique, Informatique et Recherche Opérationelle* **9**(R-2):41-76 (in French).

GOL'SHTEIN, E. G. (1975), "Method for Modification of Monotone Mappings". *Ekonomika i Matemacheskaya Metody* **11**(6):1142-1159 (in Russian).

GOL'SHTEIN, E. G. (1985), "Decomposition Methods for Linear and Convex Programming Problems". *Matekon* **22**(4):75-101.

GOL'SHTEIN, E. G. (1986), "The Block Method of Convex Programming". *Soviet Mathematics Doklady* **33**(3):584-587.

GOL'SHTEIN, E. G. (1987), "A General Approach to Decomposition of Optimization Systems". *Soviet Journal of Computer and Systems Sciences* **25**(3):105-114.

GOL'SHTEIN, E. G., TRET'YAKOV, N. V. (1975), "The Gradient Method of Minimization and Algorithms of Convex Programming Based on Modified Lagrangians Functions". *Ekonomika i Matemacheskaya Metody* **11**(4):730-742 (in Russian).

GOL'SHTEIN, E. G., TRET'YAKOV, N. V. (1979), "Modified Lagrangians in Convex Programming and their Generalizations". *Mathematical Programming Study* **10**:86-97.

HESTENES, M. R. (1969), "Multiplier and Gradient Methods". *Journal of Optimization Theory and Applications* **4**:303-320.

JOSHI, M. C., BOSE, R. K. (1985), *Some Topics in Nonlinear Functional Analysis* (New Dehli: Halsted Press/John Wiley).

KACHUROVSKII, R. I. (1960), "On Monotone Operators and Convex Functionals". *Uspekhi Matemacheskaya Nauk* **15**(4):213-215 (in Russian).

KACHUROVSKII, R. I. (1968), "Nonlinear Monotone Operators in Banach Space". *Russian Mathematical Surveys* **23**(2):117-165.

LIONS, P.-L. (1978), "Une Méthode Itérative de Resolution d'une Inequation Variationnelle". *Israel Journal of Mathematics* **31**(2):204-208 (in French).

LIONS, P.-L., MERCIER, B. (1979), "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis* **16**(6):964-979.

MARCHUK, G. I. (1975), *Methods of Numerical Mathematics* (New York: Springer-Verlag).

MARTINET, B. (1970), "Regularisation d'Inequations Variationelles par Approximations Successives". *Revue Française d'Informatique et de Recherche Operationelle* **4**(R-3):154-158 (in French).

MARTINET, B. (1972), "Determination Approchée d'un Point Fixe d'une Application Pseudo-Contractante. Cas de l'Application prox". *Comptes Rendus de l'Academie des Sciences, Paris, Série A* **274**:163-165 (in French).

MINTY, G. J. (1961), "On the Maximal Domain of a 'Monotone' Function". *Michigan Mathematical Journal* **8**:135-137.

MINTY, G. J. (1962), "Monotone (Nonlinear) Operators in Hilbert Space". *Duke Mathematics Journal* **29**:341-346.

MINTY, G. J. (1964), "On the Monotonicity of the Gradient of a Convex Function". *Pacific Journal of Mathematics* **14**:243-247.

PASCALI, D., SBURLAN, S. (1978), *Nonlinear Mappings of Monotone Type* (Bucarest: Editura Academeie).

PASSTY, G. B. (1979), "Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Space", *Journal of Mathematical Analysis and Applications* **72**:383-390.

POWELL, M. J. D. (1969), "A Method for Nonlinear Constraints in Minimization Problems", in *Optimization*, R. FLETCHER, editor (New York: Academic Press).

ROCKAFELLAR, R. T. (1966), "Characterization of the Subdifferentials of Convex Functions". *Pacific Journal of Mathematics* **17**(3):497-510.

ROCKAFELLAR, R. T. (1969), "Local Boundedness of Nonlinear, Monotone Operators". *Michigan Mathematical Journal* **16**:397-407.

ROCKAFELLAR, R. T. (1970A), *Convex Analysis* (Princeton: Princeton University Press).

ROCKAFELLAR, R. T. (1970B), "On the Maximal Monotonicity of Subdifferential Mappings". *Pacific Journal of Mathematics* **33**(1):209-216.

ROCKAFELLAR, R. T. (1970C), "On the Maximality of Sums of Nonlinear Monotone Operators". *Transactions of the American Mathematical Society* **149**:75-88.

ROCKAFELLAR, R. T. (1970D), "On the Virtual Convexity of the Domain and Range of a Nonlinear Maximal Monotone Operator". *Mathematische Annalen* **185**:81-90.

ROCKAFELLAR, R. T. (1976A), "Monotone Operators and the Proximal Point Algorithm". *SIAM Journal on Control and Optimization* **14**(5):877-898.

ROCKAFELLAR, R. T. (1976B), "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming". *Mathematics of Operations Research* **1**(2):97-116.

ROCKAFELLAR, R. T. (1977), "Monotone Operators and and Augmented Lagrangian Methods in Nonlinear Programming", in *Nonlinear Programming 3*, O. L. MANGASARIAN, R. M. MEYER, S. M. ROBINSON, editors (New York: Academic Press).

ROCKAFELLAR, R. T., WETS R. J.-B. (1987), *Scenarios and Policy Aggregation in Optimization under Uncertainty*. Working paper WP-87-119, International Institute for Applied Systems Analysis (IIASA), Laxenberg, Austria.

SPINGARN, J. E. (1983), "Partial Inverse of a Monotone Operator". *Applied Mathematics and Optimization* **10**:247-265.

SPINGARN, J. E. (1985A), "A Primal-Dual Projection Method for Solving Systems of Linear Inequalities". *Linear Algebra and Its Applications* **65**:45-62.

SPINGARN, J. E. (1985B), "Application of the Method of Partial Inverses to Convex Programming: Decomposition". *Mathematical Programming* **32**:199-233.

TSENG, P. (1988), *Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities.* Report LIDS-P-1836, Laboratory for Information and Decision Sciences, MIT.