

On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators

Jonathan Eckstein

Mathematical Sciences Research Group, Thinking Machines Corporation, Cambridge, MA 02142, USA

Dimitri P. Bertsekas

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 20 November 1989

Revised manuscript received 9 July 1990

This paper shows, by means of an 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. This observation allows the unification and generalization of a variety of convex programming algorithms. By introducing a modified version of the proximal point algorithm, we derive a new, *generalized* alternating direction method of multipliers for convex programming. Advances of this sort illustrate the power and generality gained by adopting monotone operator theory as a conceptual framework.

Key words: Monotone operators, proximal point algorithm, decomposition.

1. Introduction

The theory of maximal set-valued monotone operators (see, for example, [4]) 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* [48]. The well-known *method of multipliers* [23, 41] for constrained convex programming is known to be a special case of the proximal point algorithm [49]. This paper will reemphasize the power and generality of the monotone operator framework in the analysis and derivation of convex optimization algorithms, with an emphasis on decomposition algorithms.

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

This paper is drawn largely from the dissertation research 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.

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 than $I + \lambda T$. 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 [33]).

A number of authors, mainly in the French mathematical community, have extensively studied monotone operator splitting methods, which fall into four principal classes: forward-backward [40, 13, 56], double-backward [30, 40], Peaceman–Rachford [31], and Douglas–Rachford [31]. For a survey, readers may wish to refer to [11, Chapter 3]. We will focus on the “Douglas–Rachford” class, which appears to have the most general convergence properties. Gabay [13] 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. The alternating direction method of multipliers was first introduced in [16] and [14]; additional contributions appear in [12]. An interesting presentation can be found in [15], and [3] provides a relative accessible exposition. Despite Gabay’s result, most developments of the alternating direction method multipliers rely on a lengthy analysis from first principles. Here, we seek to demonstrate the benefit of using the operator-theoretic approach.

This paper hinges on 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 and related algorithms 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 [48] with that of Gol’shtein and Tret’yakov [22] — and show how it gives rise to a new method, *generalized* Douglas–Rachford splitting. This in turn allows the derivation of a new augmented Lagrangian method for convex programming, the *generalized* alternating direction method of multipliers. This result illustrates the benefits of adopting the monotone operator analytic approach. Because it allows over-relaxation factors, which are often found to accelerate proximal point-based methods in practice, the generalized alternating direction method of multipliers may prove to be faster than the alternating direction method of multipliers in some applications. Because it permits approximate computation, it may also be more widely applicable.

While the current paper was under review, [28] was brought to our attention. There, Lawrence and Spingarn briefly draw the connection between the proximal point algorithm and Douglas–Rachford splitting in a somewhat different — and very elegant — manner. However, the implications for extensions to the Douglas–Rachford splitting methodology and for convex programming decomposition theory were not pursued.

Most of the results presented here are refinements of those in the recent thesis by Eckstein [11], which contains more detailed development, and also relates the theory to the work of Gol'shtein [17, 18, 19, 20, 21, 22]. Some preliminary versions of our results have also appeared in [10]. Subsequent papers will introduce applications of the development given here to parallel optimization algorithms, again capitalizing on the underpinnings provided by monotone operator theory.

This paper is organized as 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 applies this theory, generalizing the alternating direction method of multipliers. It also discusses Spingarn's [52, 54] *method of partial inverses*, with a minor extension. Section 6 briefly presents a negative result concerning finite termination of Douglas–Rachford splitting methods.

2. Monotone operators

An *operator* T on a Hilbert space \mathcal{H} is a (possibly null-valued) point-to-set map $T: \mathcal{H} \rightarrow 2^{\mathcal{H}}$. We will make no distinction between an operator T and its graph, that is, the set $\{(x, y) | y \in T(x)\}$. Thus, we may simply say that an operator is any subset T of $\mathcal{H} \times \mathcal{H}$, and define $T(x) = Tx = \{y | (x, y) \in T\}$.

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,

$$\text{dom } T = \{x \in \mathcal{H} | \exists y \in \mathcal{H}: (x, y) \in T\} = \{x \in \mathcal{H} | Tx \neq \emptyset\}.$$

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

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

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

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

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

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

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

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 $f: \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ [42, 44, 45]. The following theorem, originally due to Minty [36, 37], provides a crucial characterization of maximal monotone operators:

Theorem 1. *A monotone operator T on \mathcal{H} is maximal if and only if $\text{im}(I + T) = \mathcal{H}$. \square*

For alternative proofs of Theorem 1, or stronger related theorems, see [45, 4, 6, or 24]. 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

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

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

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

The following lemma summarizes some well-known properties of firmly nonexpansive operators. The proof is straightforward and is omitted (or see, for example, [48] or [11, Section 3.2.4]). Figure 1 illustrates the lemma.

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. \square*

We now give a critical theorem. The “only if” part of the following theorem has been well known for some time [48], but the “if” part, just as easily obtained, appears to have been obscure. 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.

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,

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

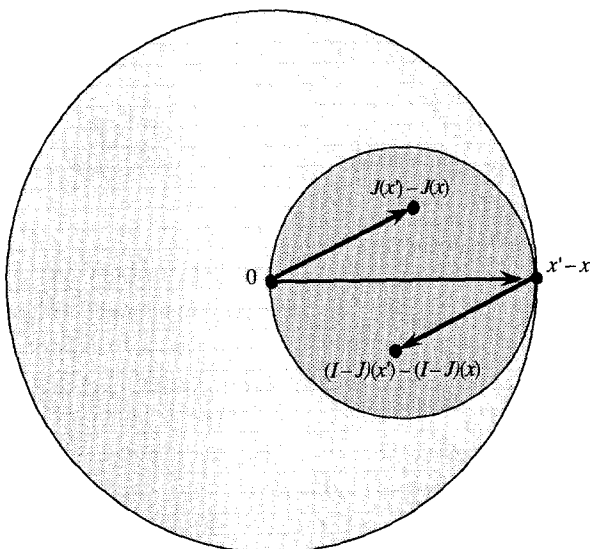


Fig. 1. Illustration of the action of firmly nonexpansive operators in Hilbert space. If J is nonexpansive, then $J(x') - J(x)$ must lie in the larger sphere, which has radius $\|x' - x\|$ and is centered at 0. If J is firmly nonexpansive, then $J(x') - J(x)$ must lie in the smaller sphere, which has radius $\frac{1}{2}\|x' - x\|$ and is centered at $\frac{1}{2}(x' - x)$. This characterization follows directly from J being of the form $\frac{1}{2}I + \frac{1}{2}C$, 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).

Therefore,

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

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 $\text{im}(I + cT) = \mathcal{H}$. This is in turn true if and only if $(I + cT)^{-1}$ has domain \mathcal{H} , establishing the second statement. \square

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. \square*

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. \square*

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 $y \in Tx$. \square

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} . \square*

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 [37], but is not identical (Minty did not use the concept of *firm* nonexpansiveness; but see also [28]). A *root* or *zero* of an operator T is a point x such that $Tx \ni 0$. We let $\text{zer}(T) = T^{-1}(0)$ denote the set of all such points. In the case that T is the subdifferential map ∂f of a convex function f , $\text{zer}(T)$ 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,

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

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

3. A generalized proximal point algorithm

Lemma 2 suggests that one way of finding a zero of a maximal monotone operator T might be to perform the iteration $z^{k+1} = J_{c_k T}(z^k)$, starting from some arbitrary point z^0 . This procedure is the essence of the *proximal point algorithm*, as named by Rockafellar [48]. Specialized versions of this method were known earlier to Martinet [34, 35], and another development appears in [5]. Rockafellar's analysis allows c to vary from one iteration to the next: give 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 $z^{k+1} = J_{c_k T}(z^k)$ for all $k \geq 0$. Rockafellar's convergence theorem also allows the resolvents $J_{c_k T}$ to be evaluated approximately, so long as the sum of all errors is finite. A related result due to Gol'shtein and Tret'yakov [22] considers iterations of the form

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

where $\{\rho_k\}_{k=0}^\infty \subseteq (0, 2)$ is a sequence of *over-* or *under-relaxation* factors. 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 $\text{zer}(T) = \emptyset$. The following theorem effectively combines the results of Rockafellar and Gol'shtein–Tret'yakov. The notation “ \xrightarrow{w} ” denotes convergence in the weak topology on \mathcal{H} , where “ \rightarrow ” denotes convergence in the strong topology induced by the usual norm $\langle x, x \rangle^{1/2}$.

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

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

where

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

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

$$E_1 = \sum_{k=0}^\infty \varepsilon_k < \infty, \quad \Delta_1 = \inf_{k \geq 0} \rho_k > 0, \quad \Delta_2 = \sup_{k \geq 0} \rho_k < 2, \quad \bar{c} = \inf_{k \geq 0} c_k > 0.$$

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

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

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

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

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

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

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

$$\|z^{k+1} - z^*\| \leq \|\bar{z}^{k+1} - z^*\| + \|z^{k+1} - \bar{z}^{k+1}\| \leq \|\bar{z}^k - z^*\| + \rho_k \varepsilon_k.$$

Combining this inequality for all k ,

$$\|z^{k+1} - z^*\| \leq \|z^0 - z^*\| + \sum_{i=0}^k \rho_i \varepsilon_i \leq \|z^0 - z^*\| + 2E_1,$$

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

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

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 ,

$$\begin{aligned}
 \|z^{k+1} - z^*\|^2 &\leq \|z^0 - z^*\|^2 + 4E_1(\|z^0 - z^*\| + 2E_1) \\
 &\quad + 4E_2 - \Delta_1(2 - \Delta_2) \sum_{i=0}^k \|Q_i(z^i)\|^2.
 \end{aligned}$$

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

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

For all k , define (x^k, y^k) to be the unique point in T such that $x^k + c_k y^k = z^k$ (here we use the Representation Lemma, Corollary 2.3). Then $Q_k(z^k) \rightarrow 0$ implies that $z^k - x^k \rightarrow 0$. Furthermore, since $\{c_k\}$ is bounded away from zero, we also have $c_k^{-1} Q_k(z^k) = y_k \rightarrow 0$.

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 $\{z^{k(j)}\}_{j=0}^{\infty}$ be a subsequence such that $z^{k(j)} \xrightarrow{w} z^\infty$. Since $z^k - x^k \rightarrow 0$, 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 $\langle x - x^k, y - y^k \rangle \geq 0$ for all k . Taking the limit over the subsequence $k(j)$ and using that $x^{k(j)} \xrightarrow{w} z^\infty$ and $y_k \rightarrow 0$, one obtains $\langle x - z^\infty, y - 0 \rangle \geq 0$. 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 . Since $\|z^k - z^*\| \leq \|z^0 - z^*\| + 2E_1$ for all k ,

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

is finite and nonnegative, and one may show that $\|z^k - z^*\| \rightarrow \alpha^*$. 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

$$\alpha_1 = \lim_{k \rightarrow \infty} \|z^k - z_1^\infty\|, \quad \alpha_2 = \lim_{k \rightarrow \infty} \|z^k - z_2^\infty\|$$

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

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

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 are then 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

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

We claim that for all k , one has $\|z^k\|, \|w^k\|, \|J_{c_k T}(z^k)\| < r - 1$. 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

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

Finally,

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

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

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

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

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

Since $\text{dom } T \cap \text{int}(\text{dom } \partial h) = \text{dom } T \cap \{x \mid \|x\| < r\} \neq \emptyset$, T' is maximal monotone [46]. Further, $\text{dom } T'$ is bounded, so $\text{zer}(T') \neq \emptyset$ [43]. Since $\|z^k\|, \|w^k\|$, and

$\|J_{c_k T}(z^k)\|$ 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,

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

where

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

By the logic of the first part of the theorem, $\{z^k\}$ converges weakly to some zero z^∞ of T' . Furthermore, as $\|z^k\| \leq r - 1$ for all k , $\|z^\infty\| \leq r - 1 < r$, 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. \square

Figure 2 is intended to clarify the role of the relaxation factor ρ_k in the convergence of the method. It illustrates the case where $\varepsilon_k = 0$, forcing exact evaluation of the resolvent. For $\rho \in (0, 2)$, let

$$z^{k+1}(\rho) = (1 - \rho)z^k + \rho(I + c_k T)^{-1}(z^k) = (1 - \rho)z^k + \rho w^k.$$

From the figure, it is clear that

$$\|z^{k+1}(1) - z^*\| = \|w^k - z^*\| \leq \|z^{k+1}(\rho) - z^*\|$$

for all $\rho < 1$ and $z^* \in \text{zer } T$, so choosing $\rho < 1$ is unlikely to be beneficial. On the other hand, there may be an interval $(1, \bar{\rho}) \subseteq (1, 2)$ on which $\rho \in (1, \bar{\rho})$ implies

$$\text{dist}(z^{k+1}(\rho), \text{zer } T) < \text{dist}(z^{k+1}(1), \text{zer } T).$$

Thus, it should be possible for over-relaxation to accelerate convergence. In at least one important application of the proximal point algorithm, the method of multipliers for convex programming, such acceleration has been experimentally confirmed [2, pp. 129–131].

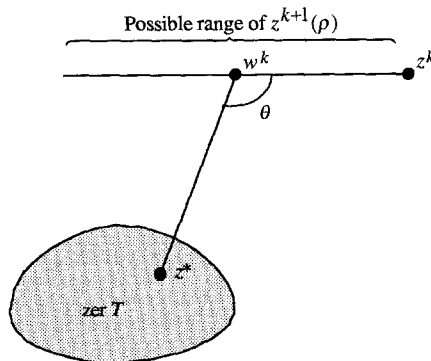


Fig. 2. Illustration of the use of the relaxation parameters. Here, $w^k = z^{k+1}(1) = (I + c_k T)^{-1}(z^k)$, and z^* is an arbitrary member of $\text{zer } T$. The angle θ must be at least 90 degrees by the monotonicity of T .

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 than $J_{\lambda T}$. 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 [31]. It is patterned after an alternating direction method for the discretized heat equation that dates back to the mid-1950's [7].

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

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

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

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

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

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 :

- Find the unique $(y^{k+1}, a^{k+1}) \in A$ such that $y^{k+1} + \lambda a^{k+1} = x^k - \lambda b^k$.
- 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 [31] 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 $S_{\lambda,A,B} = (G_{\lambda,A,B})^{-1} - I$:

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

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

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

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. \square

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. \square*

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} ,*

$$\begin{aligned} \text{zer}(S_{\lambda,A,B}) &= Z_{\lambda}^* \triangleq \{u + \lambda b \mid b \in Bu, -b \in Au\} \\ &\subseteq \{u + \lambda b \mid u \in \text{zer}(A + B), b \in Bu\}. \end{aligned}$$

Proof. Let $S = S_{\lambda,A,B}$. We wish to show that $\text{zer}(S)$ is equal to Z_{λ}^* . Let $z \in \text{zer}(S)$. 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 $u \in \text{zer}(A + B)$. \square

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 $z^{k+1} = (I + S_{\lambda,A,B})^{-1}(z^k)$. \square

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 [31]. *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 $u \in \text{zer}(A + B)$, $b \in Bu$, 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 $A + B$. \square*

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 $\text{zer}(A + B) = \emptyset$. 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. \square*

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

$$(1-c)v + cu + \lambda b = z, \quad a = \frac{1}{\lambda}(u - v) - b.$$

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

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

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. 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

$$\begin{aligned} (u, b) &\in B \quad \text{such that } u + \lambda b = z, \\ (v, a) &\in A \quad \text{such that } v + \lambda a = u - \lambda b. \end{aligned}$$

The conditions $(u, b) \in B, u + \lambda b = z$ uniquely determine $u = J_{\lambda,B}(z)$ and $b = (z - u)/\lambda$ 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 [54] has already commented on this phenomenon in the more restrictive context of the method of partial inverses.

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:*

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

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

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

$$\sum_{k=0}^{\infty} \alpha_k < \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad 0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2.$$

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

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

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

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

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

and so

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

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

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

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

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

$$0 < \inf_{k \geq 0} \rho_k \leq \sup_{k \geq 0} \rho_k < 2, \quad \sum_{k=0}^{\infty} \varepsilon_k < +\infty,$$

$$\|y^k - G_{\lambda, A, B}(z^k)\| \leq \varepsilon_k \quad \forall k \geq 0, \quad z^{k+1} = (1 - \rho_k)z^k + \rho_k y^k \quad \forall k \geq 0.$$

The conclusion then follows from Theorems 3 and 5. \square

In at least one real example [11, Section 7.2.3], 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. This example involved a highly parallel algorithm for linear programming which will be described in a later paper. There, a choice of $\rho_k = 1.5$ for all k appeared to converge to a given accuracy about 15% faster than the choice $\rho_k = 1$ for all k . Thus, the inclusion of over-relaxation factors is of some practical significance. In addition, the convergence of Douglas–Rachford splitting with approximate calculation of resolvents had not been formerly established.

5. Some interesting special cases

We now consider some interesting applications of splitting operator theory, namely the method of partial inverses [52, 54] and the generalized alternating direction method of multipliers. We begin with a brief discussion of 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 [52, 54]:

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

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

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

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 [51].

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

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

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 as easily as 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 $y^{k+1} = (1 - \rho_k)y^k + \rho_k(\tilde{y}^k)_{V^\perp}$.

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 [54]. 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 [20], and related convex programming methods [18, 19]. Demonstrating this relationship is rather laborious, however, and interested readers should refer to [11, Section 4.3].

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

$$(P) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + g(Mx),$$

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

$$(P') \quad \begin{aligned} &\underset{w}{\text{minimize}} \quad f(x) + g(w) \\ &\text{subject to} \quad Mx = w, \end{aligned}$$

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

$$(D) \quad \underset{p \in \mathbb{R}^m}{\text{maximize}} \quad -(f^*(-M^\top p) + g^*(r)),$$

where $*$ denotes the convex conjugacy operation. One way of solving the problem (P)-(D) is to let $A = \partial[f^* \circ (-M^\top)]$ and $B = \partial g^*$, and apply Douglas-Rachford splitting to A and B . This approach was shown by Gabay [13] to yield the *alternating direction method of multipliers* [16, 14, 12, 13, 15],

$$x^{k+1} = \arg \min_x \{f(x) + \langle p^k, Mx \rangle + \tfrac{1}{2}\lambda \|Mx - w^k\|^2\},$$

$$w^{k+1} = \arg \min_w \{g(w) - \langle p^k, w \rangle + \tfrac{1}{2}\lambda \|Mx^{k+1} - w\|^2\},$$

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

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

$$L_\lambda(x, w, p) = f(x) + g(w) + \langle p^k, Mx - w \rangle + \tfrac{1}{2}\lambda \|Mx - w\|^2,$$

first with respect to x , and then with respect to w , rather than with respect to both x and 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 $A = \partial[f^* \circ (-M^\top)]$ and $B = \partial g^*$ be defined as above.

A pair $(x, p) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be a *Kuhn-Tucker pair* for (P) if $(x, -M^\top p) \in \partial f$ and $(Mx, p) \in \partial g$. It is a basic exercise in convex analysis to show that if (x, p) is a Kuhn-Tucker pair, then x is optimal for (P) and p is optimal for (D), and also that if $p \in \text{zer}(A+B)$, then 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 $\underset{x \in \mathbb{R}^n}{f(x) + g(Mx)}$, where M has full*

column rank. Let $\mathbf{p}^0, \mathbf{z}^0 \in \mathbb{R}^m$, and suppose we are given $\lambda > 0$ and

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

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

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

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

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

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

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

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 $\mathbf{M}\mathbf{x}^*$, 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

$$\mathbf{z}^k = \mathbf{p}^k + \lambda \mathbf{w}^k \quad \forall k \geq 0,$$

$$\mathbf{q}^k = \mathbf{p}^k + \lambda(\mathbf{M}\mathbf{x}^{k+1} - \mathbf{w}^k) \quad \forall k \geq 0,$$

$$\alpha_k = \lambda \|\mathbf{M}\| \mu_k \quad \forall k \geq 0,$$

$$\beta_0 = \|\mathbf{p}^0 - J_{\lambda B}(\mathbf{p}^0 + \lambda \mathbf{w}^0)\|,$$

$$\beta_k = \lambda \nu_k \quad \forall k \geq 1,$$

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
 \mathbf{0} &\in \partial_x [f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \tfrac{1}{2}\lambda \|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2]_{\mathbf{x}=\bar{\mathbf{x}}^k} \\
 &\Rightarrow \mathbf{0} \in \partial f(\bar{\mathbf{x}}^k) + \mathbf{M}^\top \mathbf{p}^k + \lambda \mathbf{M}^\top (\mathbf{M}\bar{\mathbf{x}}^k - \mathbf{w}^k) \\
 &\Rightarrow \mathbf{0} \in \partial f(\bar{\mathbf{x}}^k) + \mathbf{M}^\top \tilde{\mathbf{p}}^k \\
 &\Rightarrow -\mathbf{M}^\top \tilde{\mathbf{p}}^k \in \partial f(\bar{\mathbf{x}}^k) \\
 &\Rightarrow \bar{\mathbf{x}}^k \in \partial f^*(-\mathbf{M}^\top \tilde{\mathbf{p}}^k) \\
 &\Rightarrow -\mathbf{M}\bar{\mathbf{x}}^k \in \partial [f^* \circ (-\mathbf{M}^\top)](\tilde{\mathbf{p}}^k) = A\tilde{\mathbf{p}}^k.
 \end{aligned}$$

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

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

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

we obtain

$$\|\mathbf{x}^{k+1} - \bar{\mathbf{x}}^k\| \leq \mu_k, \quad \|\mathbf{q}^k - \tilde{\mathbf{p}}^k\| \leq \lambda \|\mathbf{M}\| \mu_k,$$

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

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

and also

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

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

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

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

As

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

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

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

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

Thus, (Y3) holds for k , and (Y1) holds for $k+1$ by $\|\mathbf{p}^{k+1} - \tilde{\mathbf{s}}^k\| \leq \lambda \nu_k$. 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 $\mathbf{p}^k \rightarrow \mathbf{p}^*$ and $\mathbf{w}^k \rightarrow \mathbf{w}^*$, where $(\mathbf{p}^*, \mathbf{w}^*) \in B$ and $\mathbf{p}^* + \lambda \mathbf{w}^* = \mathbf{z}^*$. By rearranging the multiplier update formula, we have

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

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 $-\mathbf{M}^\top \tilde{\mathbf{p}}^k \in \partial f(\bar{\mathbf{x}}^k)$, or, equivalently, $(-\mathbf{M}^\top \tilde{\mathbf{p}}^k, \bar{\mathbf{x}}^k) \in \partial f$, for all k . Using

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

we have by taking limits that $\tilde{\mathbf{p}}^k \rightarrow \mathbf{p}^*$, and since $\|\mathbf{x}^k - \bar{\mathbf{x}}^k\| \leq \mu_k \rightarrow 0$, we also have $\bar{\mathbf{x}}^k \rightarrow \mathbf{x}^*$. Therefore, $(-\mathbf{M}^\top \mathbf{p}^*, \mathbf{x}^*) \in \partial f$ by the limit property for maximal monotone operators (e.g., [4]). 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. \square

The convergence of the alternating direction method of multipliers with either approximate minimization or relaxation factors was previously unknown, and, due to the complexities of the convergence proofs, would have been difficult to derive from first principles. Thus, Theorem 7 demonstrates the power of the monotone operator theoretical framework.

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

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

has been satisfied. For more implementable stopping criteria, which, under appropriate assumptions, imply these kinds of conditions, we refer to Rockafellar [48]. Essentially, if f is strongly convex, such a condition is implied by a certain bound on the smallest-magnitude subgradient of the minimand at \mathbf{x}^{k+1} . Thus, for any \mathbf{x} such that $\partial_{\mathbf{x}}[f(\mathbf{x}) + \langle \mathbf{p}^k, \mathbf{M}\mathbf{x} \rangle + \frac{1}{2}\lambda \|\mathbf{M}\mathbf{x} - \mathbf{w}^k\|^2]$ contains a member of sufficiently small norm, one may halt the minimization and set $\mathbf{x}^{k+1} = \mathbf{x}$. This idea is adapted from a stopping rule for the method of multipliers due to Kort and Bertsekas [27] (see also [2, p. 329]). A similar discussion applies to the computation of \mathbf{w}^{k+1} .

6. Concerning finite termination

The device of the splitting operator allows many results related to the proximal point algorithm to be carried over to Douglas–Rachford splitting and its special cases. In this section we briefly give a negative result that suggests that one aspect of proximal point theory, that of finite termination, will be difficult or impossible to carry over. We concentrate on a certain “staircase” property of monotone operators.

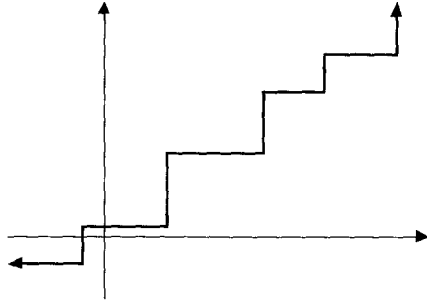
A monotone operator T on a Hilbert space \mathcal{H} is said to be *staircase* if for all $y \in \text{im } T$, there exists some $\delta(y) > 0$ such that

$$w \in Tx, \|w - y\| < \delta(y) \Rightarrow y \in Tx.$$

T is called *locally staircase at zero* if $0 \in \text{im } T$ and such a condition holds for the single case $y = 0$, that is, there exists $\delta > 0$ such that

$$w \in Tx, \|w\| < \delta(y) \Rightarrow 0 \in Tx.$$

We use the term “staircase” because an operator on \mathbb{R}^1 with the staircase property has a graph that resembles a flight of stairs (see Figure 3). The idea of a staircase operator is closely related to the so-called “diff-max” property of convex functions [8, 9, 29]. In brief, a convex function h is diff-max if and only if $(\partial h)^{-1}$ is staircase. In general, if the closed convex function h is polyhedral on \mathbb{R}^n , both ∂h and $(\partial h)^{-1} = \partial h^*$ are staircase [8].

Fig. 3. A staircase operator on \mathbb{R}^1 .

Luque [32], building on earlier observations by Rockafellar [48], proved that the exact proximal point algorithm, when each iterate is computed exactly, converges finitely when applied to any operator T which is locally staircase at zero. The basic proof is very simple: suppose we have $z^{k+1} = (I + \lambda T)^{-1} z^k$ for all $k \geq 0$. Then $(z^{k-1} - z^k)/\lambda \in Tz^k$ for all $k \geq 1$. For large enough k , we have $\|(z^{k-1} - z^k)/\lambda\| < \delta$, implying $0 \in Tz^k$ and $z^{k+1} = z^k$. This basic line of analysis dates back to the finite-termination results of [1] for the method of multipliers.

We will now show, however, that $S_{\lambda, A, B}$ need not be staircase, even if both A and B are staircase.

Theorem 9. *There exist maximal monotone operators A and B on \mathbb{R}^n , both staircase, such that $S_{\lambda, A, B}$ is not staircase for some choice of $\lambda > 0$, and the exact proximal point algorithm, without over- or under-relaxation, does not converge finitely when applied to $S_{\lambda, A, B}$.*

Proof. We need to consider only the case of \mathbb{R}^2 , and operators of the form $N_V = V \times V^\perp$, where V is a linear subspace, which were seen to be maximal monotone in the previous section. All operators of this form are staircase (in fact, for any $y \in V^\perp$, $\delta(y)$ may be taken arbitrarily large). Define the following linear subspaces of \mathbb{R}^2 :

$$W = \{(x_1, x_2) \mid x_2 = 0\} = \{(x_1, 0) \mid x_1 \in \mathbb{R}\},$$

$$U = \{(x_1, x_2) \mid x_2 = x_1\} = \{(x, x) \mid x \in \mathbb{R}\}.$$

Then

$$W^\perp = \{(x_1, x_2) \mid x_1 = 0\} = \{(0, x_2) \mid x_2 \in \mathbb{R}\},$$

$$U^\perp = \{(x_1, x_2) \mid x_2 = -x_1\} = \{(-z, z) \mid z \in \mathbb{R}\}.$$

Following the discussion of partial inverses in the previous sections,

$$\begin{aligned} S_{1, N_W, N_U} &= \{(x_W + y_{W^\perp}, y_W + x_{W^\perp}) \mid x \in U, y \in U^\perp\} \\ &= \{((x, z), (-z, x)) \mid a, b \in \mathbb{R}\}. \end{aligned}$$

Now, $S_{1,N_W,N_U}((x_1, x_2)) \ni (0, 0)$ if and only if $x_1 = x_2 = 0$. Thus S_{1,N_W,N_U} is not locally staircase at zero, and cannot be staircase.

Let $S = S_{1,N_W,N_U}$. Then $J_S = (I + S)^{-1} = \{((x - z, x + z), (x, z)) \mid x, z \in \mathbb{R}\}$, or by change of variables,

$$J_S = \{((a, b), \tfrac{1}{2}(a + b, -a + b)) \mid a, b \in \mathbb{R}\}.$$

Thus application of the operator J_S is equivalent to multiplication (on the left) by the matrix

$$J = \tfrac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

To obtain finite convergence of the iteration $z^{k+1} = J_S(z^k)$ from any starting point z^0 other than $(0, 0)$ would require that J be singular, which it is not. \square

Lefebvre and Michelot [29] do present a mild positive result relating to partial inverses (and hence to Douglas–Rachford splitting), but under fairly stringent assumptions. Although Luque’s finite convergence theory may be hard to use in the context of Douglas–Rachford splitting, his convergence *rate* techniques do have application. They have already been used in the context of partial inverses [53, 54, 55], and we will exploit them in other splitting contexts in future papers.

References

- [1] D.P. Bertsekas, “Necessary and sufficient conditions for a penalty method to be exact,” *Mathematical Programming* 9 (1975) 87–99.
- [2] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982).
- [3] D.P. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
- [4] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert* (North-Holland, Amsterdam, 1973).
- [5] H. Brézis and P.-L. Lions, “Produits infinis de resolventes,” *Israel Journal of Mathematics* 29 (1978) 329–345.
- [6] V. Doležal, *Monotone Operators and Applications in Control and Network Theory* (Elsevier, Amsterdam, 1979).
- [7] J. Douglas and H.H. Rachford, “On the numerical solution of heat conduction problems in two and three space variables,” *Transactions of the American Mathematical Society* 82 (1956) 421–439.
- [8] R. Durier, “On locally polyhedral convex functions,” Working paper, Université de Dijon (Dijon, 1986).
- [9] R. Durier and C. Michelot, “Sets of efficient points in a normed space,” *Journal of Mathematical Analysis and its Applications* 117 (1986) 506–528.
- [10] J. Eckstein, “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 (Cambridge, MA, 1988).
- [11] J. Eckstein, “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 (Cambridge, MA, 1989).

- [12] M. Fortin and R. Glowinski, "On decomposition-coordination methods using an augmented Lagrangian," in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (North-Holland, Amsterdam, 1983).
- [13] D. Gabay, "Applications of the method of multipliers to variational inequalities," in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems* (North-Holland, Amsterdam, 1983).
- [14] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Computers and Mathematics with Applications* 2 (1976) 17–40.
- [15] R. Glowinski and P. Le Tallec, "Augmented Lagrangian methods for the solution of variational problems," MRC Technical Summary Report #2965, Mathematics Research Center, University of Wisconsin-Madison (Madison, WI, 1987).
- [16] R. Glowinski and A. Marroco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problemes de Dirichlet non lineaires," *Revue Française d'Automatique, Informatique et Recherche Opérationnelle* 9(R-2) (1975) 41–76.
- [17] E.G. Gol'shtein, "Method for modification of monotone mappings," *Ekonomika i Matemacheskaya Metody* 11 (1975) 1142–1159.
- [18] E.G. Gol'shtein, "Decomposition methods for linear and convex programming problems," *Matekon* 22(4) (1985) 75–101.
- [19] E.G. Gol'shtein, "The block method of convex programming," *Soviet Mathematics Doklady* 33 (1986) 584–587.
- [20] E.G. Gol'shtein, "A general approach to decomposition of optimization systems," *Soviet Journal of Computer and Systems Sciences* 25(3) (1987) 105–114.
- [21] E.G. Gol'shtein and N.V. Tret'yakov, "The gradient method of minimization and algorithms of convex programming based on modified Lagrangian functions," *Ekonomika i Matemacheskaya Metody* 11(4) (1975) 730–742.
- [22] E.G. Gol'shtein and N.V. Tret'yakov, "Modified Lagrangians in convex programming and their generalizations," *Mathematical Programming Study* 10 (1979) 86–97.
- [23] M.R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications* 4 (1969) 303–320.
- [24] M.C. Joshi and R.K. Bose, *Some Topics in Nonlinear Functional Analysis* (Halsted/Wiley, New Delhi, 1985).
- [25] R.I. Kachurovskii, "On monotone operators and convex functionals," *Uspekhi Matemacheskaya Nauk* 15(4) (1960) 213–215.
- [26] R.I. Kachurovskii, "Nonlinear monotone operators in Banach space," *Russian Mathematical Surveys* 23(2) (1968) 117–165.
- [27] B. Kort and D.P. Bertsekas, "Combined primal–dual and penalty methods for convex programming," *SIAM Journal on Control and Optimization* 14 (1976) 268–294.
- [28] J. Lawrence and J.E. Spingarn, "On fixed points of non-expansive piecewise isometric mappings," *Proceedings of the London Mathematical Society* 55 (1987) 605–624.
- [29] O. Lefebvre and C. Michelot, "About the finite convergence of the proximal point algorithm," in: K.-H. Hoffmann et al., eds., *Trends in Mathematical Optimization: 4th French–German Conference on Optimization. International Series of Numerical Mathematics No. 84* (Birkhäuser, Basel, 1988).
- [30] P.-L. Lions, "Une méthode itérative de resolution d'une inequation variationnelle," *Israel Journal of Mathematics* 31 (1978) 204–208.
- [31] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM Journal on Numerical Analysis* 16 (1979) 964–979.
- [32] F.J. Luque, "Asymptotic convergence analysis of the proximal point algorithm," *SIAM Journal on Control and Optimization* 22 (1984) 277–293.
- [33] G.I. Marchuk, *Methods of Numerical Mathematics* (Springer, New York, 1975).
- [34] B. Martinet, "Regularisation d'inequations variationnelles par approximations successives," *Revue Française d'Informatique et de Recherche Operationelle* 4(R-3) (1970) 154–158.
- [35] B. Martinet, "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 (1972) 163–165.
- [36] G.J. Minty, "On the maximal domain of a 'monotone' function," *Michigan Mathematical Journal* 8 (1961) 135–137.
- [37] G.J. Minty, "Monotone (nonlinear) operators in Hilbert space," *Duke Mathematics Journal* 29 (1962) 341–346.

- [38] G.J. Minty, “On the monotonicity of the gradient of a convex function,” *Pacific Journal of Mathematics* 14 (1964) 243–247.
- [39] D. Pascali and S. Sburian, *Nonlinear Mappings of Monotone Type* (Editura Academeie, Bucarest, 1978).
- [40] G.B. Passty, “Ergodic convergence to a zero of the sum of monotone operators in Hilbert space,” *Journal of Mathematical Analysis and Applications* 72 (1979) 383–390.
- [41] M.J.D. Powell, “A method for nonlinear constraints in minimization problems,” in: R. Fletcher, ed., *Optimization* (Academic Press, New York, 1969).
- [42] R.T. Rockafellar, “Characterization of the subdifferentials of convex functions,” *Pacific Journal of Mathematics* 17 (1966) 497–510.
- [43] R.T. Rockafellar, “Local boundedness of nonlinear, monotone operators,” *Michigan Mathematical Journal* 16 (1969) 397–407.
- [44] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
- [45] R.T. Rockafellar, “On the maximal monotonicity of subdifferential mappings,” *Pacific Journal of Mathematics* 33 (1970) 209–216.
- [46] R.T. Rockafellar, “On the maximality of sums of nonlinear monotone operators,” *Transactions of the American Mathematical Society* 149 (1970) 75–88.
- [47] R.T. Rockafellar, “On the virtual convexity of the domain and range of a nonlinear maximal monotone operator,” *Mathematische Annalen* 185 (1970) 81–90.
- [48] R.T. Rockafellar, “Monotone operators and the proximal point algorithm,” *SIAM Journal on Control and Optimization* 14 (1976) 877–898.
- [49] R.T. Rockafellar, “Augmented Lagrangians and applications of the proximal point algorithm in convex programming,” *Mathematics of Operations Research* 1 (1976) 97–116.
- [50] R.T. Rockafellar, “Monotone operators and augmented Lagrangian methods in nonlinear programming,” in: O.L. Mangasarian, R.M. Meyer and S.M. Robinson, eds., *Nonlinear Programming 3* (Academic Press, New York, 1977).
- [51] R.T. Rockafellar and R.J.-B. Wets, “Scenarios and policy aggregation in optimization under uncertainty,” *Mathematics of Operations Research* 16(1) (1991) 119–147.
- [52] J.E. Spingarn, “Partial inverse of a monotone operator,” *Applied Mathematics and Optimization* 10 (1983) 247–265.
- [53] J.E. Spingarn, “A primal–dual projection method for solving systems of linear inequalities,” *Linear Algebra and its Applications* 65 (1985) 45–62.
- [54] J.E. Spingarn, “Application of the method of partial inverses to convex programming: decomposition,” *Mathematical Programming* 32 (1985) 199–233.
- [55] J.E. Spingarn, “A projection method for least-squares solutions to overdetermined systems of linear inequalities,” *Linear Algebra and its Applications* 86 (1987) 211–236.
- [56] P. Tseng, “Applications of a splitting algorithm to decomposition in convex programming and variational inequalities,” *SIAM Journal on Control and Optimization* 29(1) (1991) 119–138.