Chapter 2

THE PROXIMAL POINT METHOD AND OPERATOR SPLITTING METHODS

This chapter reviews related basic concepts as well as the proximal point method, the Euler method, and the splitting method for the fundamental problem $(\mathcal{P}_0)$. The main known convergence results will also be given.

2.1. Monotonicity of Operators and Some Basic Concepts

Suppose T is an operator on $\mathcal{H}$. If the cardinality of $T(w)$ is at most 1 for all $w \in \mathcal{H}$, we call T be single-valued.

The *domain* of an operator T is the set

$$\operatorname{dom}(T) := \{w \in \mathcal{H} \,|\, T(w) \neq \emptyset\}. \tag{2.1.1}$$

If $\operatorname{dom}(T) = \mathcal{H}$, we say that T has *full domain*. The *range* or *image* of T is similarly the set

$$\operatorname{rge}(T) = \{y \in \mathcal{H} \,|\, \exists w \in \mathcal{H} \quad \text{with} \quad y \in T(w)\}. \tag{2.1.2}$$

The domain and range of T are both subsets of $\mathcal{H}$.

The *inverse* $T^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ of an operator T on $\mathcal{H}$ is defined as

$$T^{-1}(y) = \{x \in \mathcal{H} \,|\, y \in T(x)\}, \qquad \forall \quad y \in \mathcal{H}. \tag{2.1.3}$$

Let T_1, T_2 be two operators on $\mathcal{H}$ and λ be a real number in $\mathbb{R}$. Then

$$\begin{aligned} \lambda T_1(x) =& \{\lambda y \in \mathcal{H} \,|\, y \in T_1(x)\}, \qquad \forall \quad x \in \mathcal{H} \\ (T_1 + T_2)(x) =& T_1(x) + T_2(x) \\ =& \{y + z \,|\, y \in T_1(x), z \in T_2(x)\}, \quad \forall \quad x \in \mathcal{H} \end{aligned} \tag{2.1.4}$$

Subset S of $\mathcal{H}$ is *closed* if the cluster points of any sequence $\{w_k\} \subset S$ are still in S. S is a *convex* set if $tx + (1-t)y \in S$ for any $x \in S$, $y \in S$ and $0 < t < 1$.

Definition 1.1 in Chapter 1 gives (maximal) monotonicity of operators. The following includes some well known facts. For more details please refer to Eckstein (1989).

Proposition 2.1.1. *Let T be an operator on $\mathcal{H}$. Then*

(a) *T is (maximal) monotone if and only if T^{-1} is (maximal) monotone.*

(b) *Let d be any member of $\mathcal{H}$. Then T is (maximal) monotone if and only if $T+d$ is (maximal) monotone.*

(c) *If T is maximal monotone, then T is both closed-valued and convex-valued. That is, $T(x)$ is a closed convex set for any x.*

(d) *If T is maximal monotone, the graph $\operatorname{gph} T$ of T is a closed set. Hence, if $x^k \to x$ and $y^k \to y$ as $k \to \infty$ with $y^k \in T(x^k)$, then $y \in T(x)$.*

(e) *If T is monotone, then T is maximal if and only if $\operatorname{rge}(I+T) = \mathcal{H}$, where I is the identity operator on $\mathcal{H}$.*

(f) *If T is monotone, then $J_{\lambda,T} = (I+\lambda T)^{-1}$ is single-valued for any $\lambda > 0$. If T is also maximal, then $J_{\lambda,T}$ has full domain.*

In the following examples of monotone operators, for this dissertation, we define $\overline{\mathbb{R}} = [-\infty, \infty]$:

The *indicator function* $\delta_C : \mathcal{H} \to \overline{\mathbb{R}}$ of a set $C \subset \mathcal{H}$ is given by

$$\delta_C(x) = \begin{cases} 0 & \text{for } x \in C, \\ \infty & \text{otherwise} \end{cases} \tag{2.1.5}$$

The *normal cone* operator $\mathcal{N}_C$ of a set $C \subset \mathcal{H}$ is given by

$$\mathcal{N}_C(x) = \begin{cases} \{y \in \mathcal{H} \,|\, \langle x' - x, y\rangle \le 0, \forall x' \in C\} & \text{for all } x \in C, \\ \emptyset & \text{otherwise} \end{cases} \tag{2.1.6}$$

Now consider an extended-real-valued function $f : \mathcal{H} \to \overline{\mathbb{R}}$.

(a) The (effective) *domain* of f is

$$\operatorname{dom} f = \{x \in \mathcal{H} \,|\, f(x) < \infty\}. \tag{2.1.7}$$

(b) f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathcal{H}$.

(c) The *epigraph* of f is

$$\operatorname{epi} f = \{(x, y) \in \mathcal{H} \times \mathbb{R} \,|\, y \ge f(x)\}. \tag{2.1.8}$$

(d) f is said to be *convex* if $\operatorname{epi} f$ is a convex set in $\mathcal{H} \times \mathbb{R}$.

(e) f is said to be *closed* if $\operatorname{epi} f$ is a closed set in $\mathcal{H} \times \mathbb{R}$.

(f) f is *lower semicontinuous* if for any sequence $\{x^k\} \subset \mathcal{H}$ converging to some $x \in \mathcal{H}$, $f(x) \leq \liminf_{k \to \infty} f(x^k)$.

Recall that the *subgradient operator* $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ of a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is that

$$\partial f(x) = \{y \in \mathcal{H} \,|\, f(x+d) \geq f(x) + \langle d, y \rangle, \ \forall\ d \in \mathcal{H}\}, \quad \forall \ \ x \in \mathcal{H}. \tag{2.1.9}$$

The above definitions directly follow those in Convex Analysis (Rockafellar, [290]). The first three examples in the following are well known facts. Please refer to the papers of Rockafellar.

Example 2.1.1. *$\mathcal{N}_C = \partial(\delta_C)$ for any set $C \subset \mathcal{H}$. If C is convex, then $\mathcal{N}_C$ is monotone. If C is nonempty, closed, and convex, $\mathcal{N}_C$ is maximal monotone.*

Example 2.1.2. *If $f : \mathcal{H} \to \overline{\mathbb{R}}$ is convex, then $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ is a monotone operator. If f is also closed (or lower semicontinuous) and proper, ∂f is maximal monotone on $\mathcal{H}$.*

Example 2.1.3. *Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ and $\mathcal{H}_1$, $\mathcal{H}_2$ are Hilbert spaces. Let $L : \mathcal{H} \to \overline{\mathbb{R}}$ be such that $L(x, y)$ is convex in $x \in \mathcal{H}_1$ and concave in $y \in \mathcal{H}_2$, and*

$$T_L = \begin{pmatrix} \partial_x L \\ -\partial_y L \end{pmatrix} \tag{2.1.10}$$

Then T_L is maximal monotone if L is closed and proper.

Example 2.1.4. *Any constant operator on $\mathcal{H}$ is maximal monotone.*

Proof. Let d be an arbitrary element in $\mathcal{H}$ and $T(x) = \{d\}$ for all $x \in \mathcal{H}$. If $y \neq 0$, one has $\langle (y + x) - x, y \rangle > 0$ for $y + x = x' \in \mathcal{H}$. Hence

$$\mathcal{N}_{\mathcal{H}}(x) = \{y \in \mathcal{H} \,|\, \langle x' - x, y \rangle \leq 0, \forall x' \in \mathcal{H}\} = \{0\} \ \text{for all} \ x \in \mathcal{H}.$$

Therefore

$$T = \mathcal{N}_{\mathcal{H}} + d$$

Example 2.1.1 and the part (b) of Proposition 2.1.1 imply that T is maximal monotone. ∎

Definition 2.1.1. *An operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called strongly monotone if there exists some $\alpha > 0$ such that*

$$\langle x' - x, y' - y\rangle \geq \alpha \|x' - x\|^2, \qquad \forall \quad y \in T(x), \quad y' \in T(x'). \tag{2.1.11}$$

Here α is called modulus.

Thus, T^{-1} is strongly monotone if there exists a $\beta > 0$ such that

$$\langle x' - x, y' - y\rangle \geq \beta \|y' - y\|^2, \qquad \forall \quad y \in T(x), \quad y' \in T(x'). \tag{2.1.12}$$

The following is elementary:

Example 2.1.5. *Any positive semidefinite matrix $A : \mathbb{R}^n \to \mathbb{R}^n$ is maximal monotone. Moreover, if A is actually positive definite, then A is strongly monotone.*

Now let us consider some properties of the strong monotonicity of T.

Proposition 2.1.2. *If T^{-1} is strongly monotone, then T is single-valued.*

Proof. For any $x \in \mathcal{H}$ and $y, y' \in T(x)$, there is a $\beta > 0$ such that

$$0 = \langle x - x, y' - y\rangle \geq \beta \|y' - y\|^2.$$

This implies $y' = y$. □

Proposition 2.1.3. *If T is strongly monotone with modulus α, then*

$$\|y_1 - y_2\| \geq \alpha \|x_1 - x_2\|, \quad \textit{for any} \quad y_i \in T(x_i), i = 1, 2. \tag{2.1.13}$$

Proof. Since

$$\alpha \|x_1 - x_2\|^2 \leq \langle x_1 - x_2, y_1 - y_2\rangle \leq \|x_1 - x_2\| \|y_1 - y_2\|$$

Thus, the conclusion is true. □

Let T be an operator over $\mathcal{H}$. Recall that T is *Lipschitz continuous* with constant M if for any $x_i \in \mathcal{H}$ and $y_i \in T(x_i)$ for $i = 1, 2$, we have

$$\|y_1 - y_2\| \leq M \|x_1 - x_2\|. \tag{2.1.14}$$

Proposition 2.1.4. *If T is Lipschitz continuous, then T is single-valued.*

Proof. Taking $x_1 = x_2$ in (2.1.14) implies $y_1 = y_2$. □

Proposition 2.1.5. *If T^{-1} is strongly monotone with β, then T is Lipschitz continuous with constant β^{-1}.*

Proof. If T^{-1} is strongly monotone with β, then

$$\beta\|y_1 - y_2\|^2 \leq \langle x_1 - x_2, y_1 - y_2\rangle \leq \|x_1 - x_2\|\|y_1 - y_2\|$$

for any $y_1 \in T(x_1)$ and $y_2 \in T(x_2)$. Hence

$$\|y_1 - y_2\| \leq \beta^{-1}\|x_1 - x_2\|. \qquad \square$$

2.2. Support Functions, Lagrangian, Primal and Dual Problems

Let W be a nonempty subset in $\mathcal{H}$ and $Q : \mathcal{H} \to \mathcal{H}$ be a symmetric ($Q^* = Q$) bounded linear operator.

(a) W is a convex *polyhedron* if it can be expressed as the intersection of some finite collection of closed half-spaces.

(b) The *support function* of a convex set W is defined as

$$\sigma_W(z) = \sup_{w\in W} \langle w, z\rangle, \forall z \in \mathcal{H} \tag{2.2.1}$$

(c) The *monitoring function* $\rho_{W,Q}$ of W and Q with convex polyhedral W and monotone Q is defined at $z \in \mathcal{H}$ as

$$\rho_{W,Q}(z) = \sup_{w\in W} \{\langle w, z\rangle - \tfrac{1}{2}\langle w, Qw\rangle, \tag{2.2.2}$$

It is obvious that $\rho_{W,Q} = \sigma_W$ for $Q = 0$.

Proposition 2.2.1 (Rockafellar, 1987). *The function $\rho_{W,Q}$ is lower semicontinuous, convex, and piecewise linear-quadratic: its effective domain is a nonempty convex polyhedron that can be decomposed into finitely many polyhedral convex sets, on each of which $\rho_{W,Q}$ is quadratic or linear.*

As examples, we give the following for $\mathcal{H} = \mathbb{R}^n$.

Example 2.2.1. Let $W = \mathbb{R}^n_+$, $Q = 0$. Then

$$\rho_{W,Q}(z) = \sigma_{\mathbb{R}^n_+}(z) = \begin{cases} 0 & \text{if} \quad z \leq 0, \\ -\infty & \text{otherwise} \end{cases} \tag{2.2.3}$$

So when $z = q - Du$, we have that $\rho_{W,Q}(q - Du) = 0$ if $Du \geq q$.

Example 2.2.2. Let Q be positive definite and $W = \mathbb{R}^n$. Then

$$\rho_{\mathbb{R}^n,Q}(z) = \tfrac{1}{2} z \cdot Q^{-1} z \tag{2.2.4}$$

Example 2.2.3. Let $Q = 0$ and

$$W = \{ z \in \mathbb{R}^n_+ \mid \|z\|_1 = 1 \} \tag{2.2.5}$$

Then

$$\rho_{W,Q}(z) = \sigma_W(z) = \max_{i=1,\dots,n} z_i = \|z\|_\infty \quad \text{for} \quad z = (z_1, \dots, z_n) \tag{2.2.6}$$

Now let us consider concepts of primal problems, dual problems, and their Lagrangian functions.

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces and $L : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R}$ be such that $L(u, v)$ is convex in $u \in \mathcal{H}_1$ and concave in $v \in \mathcal{H}_2$.

Consider the following saddle point problem:

$$\operatorname*{minimax}_{u \in U,\, v \in V} L(u, v) \tag{2.2.7}$$

where *minimax* means that one minimizes L in u and maximizes in v and $U \subset \mathcal{H}_1$ and $V \subset \mathcal{H}_2$ are nonempty closed convex sets.

Associated with L, U, and V, the *primal* and *dual* problems are respectively

$$\text{minimize} \quad f(u) \quad \text{for all} \quad u \in U, \quad \text{with} \quad f(u) := \sup_{v \in V} L(u, v), \tag{2.2.8}$$

$$\text{maximize} \quad g(v) \quad \text{for all} \quad v \in V, \quad \text{with} \quad g(v) := \inf_{u \in U} L(u, v). \tag{2.2.9}$$

We specify L as a *Lagrangian* function.

A pair of elements $\bar{u} \in U$ and $\bar{v} \in V$ is said to give a *saddle point* of L on $U \times V$ when the following inequalities hold:

$$L(u, \bar{v}) \geq L(\bar{u}, \bar{v}) \geq L(\bar{u}, v) \qquad \text{for all} \quad u \in U, \quad v \in V.$$

A case of importance is that of the extended linear-quadratic programming. For this case, $L(u, v)$ becomes the following:

$$L(u, v) = p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - v{\cdot}Du, \tag{2.2.10}$$

where $\mathcal{H}_1 = \mathbb{R}^k$, $\mathcal{H}_2 = \mathbb{R}^l$, U and V are polyhedral and $P \in \mathbb{R}^{k\times k}$ and $Q \in \mathbb{R}^{l\times l}$ are symmetric and positive semidefinite, $D \in \mathbb{R}^{l\times k}$ for the extended linear-quadratic case here.

When $L(u, v)$ is given by (2.2.10), the (2.2.8) and (2.2.9) have the following expressions:

$$\min f(u) = p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + \rho_{V,Q}(q - Du) \quad \text{over} \quad u \in U, \tag{2.2.11}$$

$$\max g(v) = q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - \rho_{U,P}(D^* v - p) \quad \text{over} \quad v \in V. \tag{2.2.12}$$

Examples for the primal and dual problems (2.2.11) and (2.2.12) can be find in Rockafellar (1987) etc.

The basic facts about the relationship between (2.2.11) and (2.2.12) are given by the following:

Theorem 2.2.2 (Rockafellar and Wets 1986). *If either (2.2.11) or (2.2.12) has finite optimal value, or if both problems have feasible solutions, then both optimal values are finite and equal, and both problems have optimal solutions. In this case a pair $(\bar{u}, \bar{v})$ is a saddlepoint of $L(u, v)$ in (2.2.10) relative to $u \in U$ and $v \in V$ if and only if $\bar{u}$ is an optimal solution to (2.2.11) and $\bar{v}$ is an optimal solution to (2.2.12).*

2.3. Some Examples for the Fundamental Problem

The fundamental problem $(\mathcal{P}_0)$ in Chapter 1 covers many models from optimization, optimal control, engineering, business, industrial management, numerical differential equations, etc.. In the following we shall give some examples in these fields to show that application.

Example 2.3.1. (*Minimization problems of convex functions.*)

Consider a convex unconstrained problem:

$$\text{minimize} \quad f(x), \tag{2.3.1}$$

where $f : \mathcal{H} \to \mathbb{R}$ is convex and closed. Example 2.1.2 tells that $T = \partial f$ is maximal monotone. However, Rockafellar (1966) proved that the global minimizers of f are the points of x such that $0 \in \partial f(x)$. That is, solving (2.3.1) is equivalent to solving the fundamental problem $(\mathcal{P}_0)$ with $T = \partial f$.

Example 2.3.2. (*Box constrained quadratic problems.*)

Consider the following box constrained QP problem:

$$\begin{aligned} \text{minimize} \quad & c{\cdot}x + \tfrac{1}{2}x{\cdot}Px \\ \text{subject to} \quad & x \in U. \end{aligned} \tag{2.3.2}$$

Where $U \subset \mathbb{R}^n$ is a nonempty *box*, i.e., products of closed intervals, and $P \in \mathbb{R}^{n\times n}$ is a semi-positive definite matrix.

Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be

$$T(x) = Px + c + \mathcal{N}_U(x). \tag{2.3.3}$$

Then T is maximal monotone from Example 2.1.2 and solving (2.3.2) is the same as solving the fundamental problem $(\mathcal{P}_0)$ with T given by (2.3.3) from the optimality conditions of the problem (2.3.2).

Example 2.3.3. (*Saddle point problems of convex-concave functions.*)

Consider the saddle point problem (2.2.7). When L is also closed, then T defined by

$$T = \begin{pmatrix} \partial_u L + \mathcal{N}_U \\ -\partial_v L + \mathcal{N}_V \end{pmatrix} \tag{2.3.4}$$

is maximal monotone from Example 2.1.3, and Rockafellar [287] proved that the global saddle points of L on $U \times V$ are the elements of $z = (u, v)$ such that $0 \in T(z)$ with T given by (2.3.4).

Example 2.3.4. (*Linear programming problems.*)

Consider the following standard LP problem:

$$\begin{aligned} &\text{maximize} \quad c{\cdot}u \\ &\text{subject to} \quad Au \leq b, \qquad u \in \mathbb{R}^n_+. \end{aligned} \tag{2.3.5}$$

Where $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$.

From the Lagrangian of (2.3.5), Example 2.2.1 and the results of Rockafellar [287], we know that solving the above LP problem (2.3.5), and its dual, is equivalent to solving the fundamental problem $(\mathcal{P}_0)$ with T as follows:

$$T(u, v) = \begin{pmatrix} c + A^* v + \mathcal{N}_U(u) \\ b - Au + \mathcal{N}_V(v) \end{pmatrix}, \quad \text{for} \quad z = \begin{pmatrix} u \\ v \end{pmatrix}, \tag{2.3.6}$$

where $\mathcal{H} = \mathbb{R}^n \times \mathbb{R}^m$, $U = \mathbb{R}^n_+$, and $V = \mathbb{R}^m_+$. It is elementary that T in (2.3.6) is maximal monotone.

Example 2.3.5. (*Quadratic programming problems.*)

Consider the following general QP problem:

$$\begin{aligned} &\text{minimize} \quad c{\cdot}u + \tfrac{1}{2} u{\cdot}Qu \\ &\text{subject to} \quad Au \leq b, \\ &\qquad\qquad\quad u \in U. \end{aligned} \tag{2.3.7}$$

Where $U \subset \mathbb{R}^n$ is a polyhedron and $Q \in \mathbb{R}^{n \times n}$ is a semi-positive definite matrix.

Similarly, from the Lagrangian of (2.3.7), Example 2.2.1 and the results in Section 2.2, solving the QP problem (2.3.7) and its dual is equivalent to solving the fundamental problem with T as follows:

$$T(u, v) = \begin{pmatrix} Qu + c + A^* v + \mathcal{N}_U(u) \\ b - Au + \mathcal{N}_V(v) \end{pmatrix}, \tag{2.3.8}$$

where $\mathcal{H}$ and V are the same as those in Example 2.3.4.

Example 2.3.6. (*Basic piecewise linear programming.*)

Consider the following piecewise LP problem:

$$\begin{aligned} &\text{minimize} \quad p{\cdot}u + \|q - Du\|_\infty \\ &\text{subject to} \quad u \in U, \end{aligned} \tag{2.3.9}$$

where $U \subset \mathbb{R}^n$ is a polyhedron and $D \in \mathbb{R}^{m\times n}$.

Similarly, solving the problem (2.3.9), and its dual problem, is the same as solving the fundamental problem $(\mathcal{P}_0)$ with T as

$$T(u,v) = \begin{pmatrix} p - D^* v + \mathcal{N}_U(u) \\ Du - q + \mathcal{N}_V(v) \end{pmatrix}. \tag{2.3.10}$$

Here $\mathcal{H} = \mathbb{R}^n \times \mathbb{R}^m$ and

$$V = \{v \in \mathbb{R}^m_+ \,|\, \|v\|_1\} \tag{2.3.11}$$

which is the *unit simplex* in $\mathbb{R}^m$. The maximal monotonicity of T can be obtained from Example 2.2.3 and Example 2.1.3.

Example 2.3.7. (*Continuous-time model of extended linear-quadratic programming.*)

Consider the following continuous-time problem from Rockafellar (1987). The time interval $[t_0, t_1]$ is fixed here.

The model problem is: over the dynamical system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + b(t) \quad \text{a.e.,} \quad x(t_0) = B_e u_e + b_e \tag{2.3.12}$$

with control space

$$\mathcal{U} = \{(u, u_e) \,|\, u \in \mathcal{L}^2, u(t) \in U(t) \quad \text{a.e.,} \quad u_e \in U_e\} \tag{2.3.13}$$

minimize the functional

$$\begin{aligned} \mathcal{F}(u,u_e) := &\int_{t_0}^{t_1} [p(t){\cdot}u(t) + \tfrac{1}{2}u(t){\cdot}P(t)u(t) - c(t){\cdot}x(t)]\,dt + [p_e{\cdot}u_e \\ &+ \tfrac{1}{2}u_e{\cdot}P_e u_e - c_e{\cdot}x(t_1)] + \rho_{V_e,Q_e}(q_e - C_e x(t_1) - D_e u_e) \\ &+ \int_{t_0}^{t_1} \rho_{V(t),Q(t)}(q(t) - C(t)x(t) - D(t)u(t))\,dt. \end{aligned} \tag{2.3.14}$$

where the sets $U(t) \subset \mathbb{R}^k$, $U_e \subset \mathbb{R}^{k_e}$, $V(t) \subset \mathbb{R}^l$, and $V_e \subset \mathbb{R}^{l_e}$ are polyhedral convex and the matrices $P(t) \in \mathbb{R}^{k\times k}$, $P_e \in \mathbb{R}^{k_e\times k_e}$, $Q(t) \in \mathbb{R}^{l\times l}$ and $Q_e \in \mathbb{R}^{l_e\times l_e}$ are positive semidefinite. All the data elements in the above problem are assumed to depend continuously on t. The ρ terms are defined by the (2.2.2). In general, ρ is piecewise linear-quadratic convex functions that may take on the value ∞.

Rockafellar ([287], [273], [275]) has proved that solving the above primal problem and its dual problem is equivalent to solving the fundamental problem $(\mathcal{P}_0)$ of finding a $(u, v) \in \mathcal{U} \times \mathcal{V}$ such that $0 \in T(u, v)$ for $t \in [t_0, t_1]$ a.e., where $u = [u(t), u_e]$, $v = [v(t), v_e]$,

$$\begin{aligned} T(u,v) = &\begin{pmatrix} \mathcal{N}_{\mathcal{U}}(u) \\ \mathcal{N}_{\mathcal{V}}(v) \end{pmatrix} \\ &+ \begin{pmatrix} p(t) + P(t)u(t) - D(t)^* v(t) - B(t)^* y(t) \\ p_e + P_e u_e - D_e^* v_e - B_e^* y(t_0) \\ C(t)x(t) + Q(t)v(t) + D(t)u(t) - q(t) \\ C_e x(t_1) + Q_e v_e + D_e u_e - q_e \end{pmatrix}, \end{aligned} \tag{2.3.15}$$

$$\mathcal{V} = \{(v, v_e) \,|\, v \in \mathcal{L}^2, v(t) \in V(t) \quad \text{a.e.,} \quad v_e \in V_e\} \tag{2.3.16}$$

and

$$\mathcal{H} = \mathcal{L}^2([t_0, t_1], \mathbb{R}^k) \times \mathbb{R}^{k_e} \times \mathcal{L}^2([t_0, t_1], \mathbb{R}^l) \times \mathbb{R}^{l_e}. \tag{2.3.17}$$

2.4. The Euler Method

This section reviews the Euler method for Problem $(\mathcal{P}_0)$ and its convergent properties by Gol'shtein, etc..

Consider minimizing a convex function f. A popular way to solve it is the *gradient method* when f is smooth. Suppose $\mathcal{H}$ is $\mathbb{R}^n$. Then for any starting point z^0, the gradient method scheme is

$$z^{k+1} = z^k - \lambda_k \nabla f(z^k), \tag{2.4.1}$$

here $\{\lambda_k\}_{k=0}^{\infty} \subset (0, \infty)$ is the *step size* sequence. Later in this dissertation, the $\{\lambda_k\}$ we will use has the same definition. We shall not describe it again.

In this case, the method (2.4.1) solves the problem: finding a point $z^* \in \mathbb{R}^n$ such that $\nabla f(z^*) = 0$.

Naturally, we can generalize this method to solve our fundamental problem by replacing the gradient with a general monotone operator T. For given initial point z^0 and step sizes λ_k, the modified method has the following scheme:

$$z^{k+1} \in z^k - \lambda_k T(z^k) = (I - \lambda_k T)(z^k). \tag{2.4.2}$$

We call the above generalized gradient method the *Euler algorithm* or *forward step method.*

The convergence for the Euler method requires some restrictive conditions. Even for the special case of the gradient method (2.4.1), convergence proofs generally require the step sizes λ_k to follow rules such as the exact line search, Armijo's rule, or Goldstein's rule, etc.. For more details, please refer to the Ph.D dissertation of J. Eckstein.

There are some convergence results for (2.4.1) without sampling of the objective function under restrictive assumptions (Levitin and Polyak 1966, Polyak 1987). Here we just give some known convergence results for a maximal monotone T.

Proposition 2.4.1. *Suppose $\mathcal{H} = \mathbb{R}^n$ and T is maximal monotone. If $0 < \inf\{\lambda_k\}$ and $\{z^k\}$ generated by the Euler scheme (2.4.2) converges to a point $z \in \mathbb{R}^n$, then z is a zero point of T.*

Proof. From (2.4.2), we have

$$w^k := \frac{z^k - z^{k+1}}{\lambda_k} \in T(z^k). \tag{2.4.3}$$

The given conditions imply that $w^k \to 0$ as $k \to \infty$. Therefore, Part (d) of Proposition 2.1.1 gives that $0 \in T(z)$. $\square$

Theorem 2.4.2 (Gol'shtein 1975). *Let $\mathcal{H} = \mathbb{R}^n$ and T be maximal monotone on $\mathbb{R}^n$. Suppose T^{-1} is strongly monotone with modulus β and $\{z^k\}$, $\{w^k\}$ are generated by*

$$z^{k+1} = z^k - \lambda_k w^k, \tag{2.4.4}$$

$$\|w^k - T(z^k)\| \le \varepsilon_k, \tag{2.4.5}$$

$$0 < \inf_{k\ge 0} \lambda_k \le \sup_{k\ge 0} \lambda_k < 2\beta, \tag{2.4.6}$$

$$\sum\nolimits_{k=0}^{\infty} \varepsilon_k < \infty. \tag{2.4.7}$$

Then the sequence $\{z^k\}$ converges to a zero point of T if $T^{-1}(0) \neq \emptyset$.

2.5. The Proximal Point Algorithm (Backward Step Method)

In 1976 Rockafellar developed a method to solve the fundamental problem $(\mathcal{P}_0)$ and termed it the *proximal point algorithm.* The method generates for any starting point z^0 a sequence $\{z^k\}$ in $\mathcal{H}$ by the scheme

$$z^{k+1} = (I + \lambda_k T)^{-1}(z^k) = P_k(z^k) = J_{\lambda_k, T}(z^k). \tag{PPA}$$

Proposition 2.5.1. *Given any maximal monotone operator T, real number $\lambda > 0$ and $x \in \mathcal{H}$, then $0 \in T(x)$ if and only if $(I + \lambda T)^{-1}(x) = \{x\}$.*

Proof. By direct calculation,

$$y \in T(x) \text{ iff } \; x + \lambda y \in (I + \lambda T)(x) \text{ iff } \; \{x\} = (I + \lambda T)^{-1}(x + \lambda y)$$

Hence,

$$0 \in T(x) \text{ if and only if } \; \{x\} = (I + \lambda T)^{-1}(x).$$

$\square$

Proposition 2.5.2 (Rockafellar, [287]). *If $T = T_L$ of (2.1.10), then $\bar{z} = (I + \lambda T)^{-1}(z^k)$ is equivalent to*

$$\bar{z} = (\bar{x}, \bar{y}) = \operatorname{argminimax} \Lambda_k(x, y) \tag{2.5.1}$$

where

$$\Lambda_k(x, y) = L(x, y) + \frac{1}{2\lambda}\|x - x^k\|^2 - \frac{1}{2\lambda}\|y - y^k\|^2 \tag{2.5.2}$$

Proposition 2.5.3 (Rockafellar [278]). *Under the conditions of Example 2.1.2 and* $T = \partial f$, $x^{k+1} = P_k(x^k)$ *reduces to*

$$\begin{aligned} x^{k+1} &= \operatorname{argmin} \Phi_k(x) \\ \Phi_k(x) &= f(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2 \end{aligned} \tag{2.5.3}$$

Minty (1962) proved the *proximal mapping* $P_k = (I + \lambda_k T)^{-1}$ (the name was from Moreau (1965) for the case of $T = \partial f$) is single-valued and nonexpansive, and $P_k(z) = z$ if and only if $0 \in T(z)$.

The early convergence results for the proximal point algorithm have been given by Martinet (1970, 1972) for specific cases. He showed with $\lambda_k \equiv \lambda$ that

(a) if $T(z) = T_0(z) + \mathcal{N}_D(z)$ for $z \in D$ otherwise $T(z) = \emptyset$ with D nonempty closed convex bounded and T_0 single-valued, monotone, and hemicontinuous(i.e. continuous along each line segment in H with respect to the weak topology), then $\{z^k\}$ from (PPA) converges weakly to a zero point z^* of T.

(b) if $T = \partial f$ with a lower semicontinuous convex proper function f and the level sets $\{z \in \mathcal{H} \,|\, f(z) \leq \alpha\}$, $\alpha \in \mathbb{R}$ are all weakly compact, then $\{f(z^k)\}$ converges to $f(z^*) = \min f$.

For the general strong convergence of the proximal point algorithm, Rockafellar [291] obtained the following three convergence theorems:

Theorem 2.5.4. *If* T *is strongly monotone and* $0 < \inf\{\lambda_k\}$, *then* $\{z^k\}$ *converges strongly to the unique zero point of* T *at least as fast as linear rate. In addition, if* $\lambda_k \to \infty$, *the convergence is superlinear.*

For a looser relation between z^{k+1} and $P_k(z^k)$, he considered two criteria for the approximate calculation of $P_k(z^k)$ there:

$$\|z^{k+1} - P_k(z^k)\| \leq \varepsilon_k, \qquad \sum_{k=0}^{\infty} \varepsilon_k < \infty, \tag{2.5.4}$$

$$\|z^{k+1} - P_k(z^k)\| \leq \delta_k \|z^{k+1} - z^k\|, \qquad \sum_{k=0}^{\infty} \delta_k < \infty. \tag{2.5.5}$$

Theorem 2.5.5. *Let $\{z^k\}$ be any sequence generated by the proximal point algorithm under criterion (2.5.4) with $0 < \inf\{\lambda_k\}$. Suppose $\mathrm{zer}(T) \neq \emptyset$. Then $\{z^k\}$ converges weakly to a zero point of T.*

Definition 2.5.1. *Let T be a maximal monotone operator. We say that T^{-1} is Lipschitz continuous at 0 (with modulus $a \geq 0$) if $T^{-1}(0) = \{\bar{z}\}$, for some $\theta > 0$ we have*

$$\|z - \bar{z}\| \leq a\|w\| \quad \textit{whenever} \quad z \in T^{-1}(w) \quad \textit{and} \quad \|w\| \leq \theta. \tag{2.5.6}$$

Theorem 2.5.6. *Let $\{z^k\}$ be any sequence generated by the proximal point algorithm under criterion (2.5.5) with $\{\lambda_k\}$ nondecreasing. Assume that $\{z^k\}$ is bounded and T^{-1} is Lipschitz continuous at 0. Then $\{z^k\}$ converges strongly to the unique zero point of T linearly. Moreover, if $\lambda_\infty = \infty$, the convergence is superlinear.*

There are also other convergence theorems. Here we shall give an additional two without proofs. Eckstein (1989) described a theorem by referring to that of Brezis and Lions (1978). That is,

Theorem 2.5.7. *Let $\mathcal{H} = \mathbb{R}^n$ and $\{z^k\}$ be any sequence generated by (PPA) with $T^{-1}(0) \neq \emptyset$. Suppose $\sum_{k=0}^{\infty} \lambda_k^2 = \infty$. Then $\{z^k\}$ converges to a zero point of T.*

From the work of Gol'shtein (1975) and Gol'shtein and Tret'yakov (1979), Eckstein introduced another kind of relaxation for (PPA).

Theorem 2.5.8. *Let $\lambda > 0$ and $\varepsilon_k \geq 0$ for $k \geq 0$. Assume that $T^{-1}(0) \neq \emptyset$ and*

$$\|w^k - (I + \lambda T)^{-1}(z^k)\| \leq \varepsilon_k, \qquad \sum_{k=0}^{\infty} \varepsilon_k < \infty, \tag{2.5.7}$$

$$0 < \inf_{k \geq 0} \sigma_k \leq \sup_{k \geq 0} \sigma_k < 2, \qquad z^{k+1} = (1 - \sigma_k) z^k + \sigma_k w^k. \tag{2.5.8}$$

Then $\{z^k\}$ converges to a zero point of T.

There are many extensions, or generalizations, or convergence discussions, and/or applications for the proximal point algorithms. Interested person may refer to Bertsekas and Tseng (1994), Censor and Zenios (1992), Eckstein (1993), Ferris (1991), Goldstein (1987), Güler (1991, 1992), Ha (1990), Kassay (1985), Luque (1984, 1986), Qian (1992), Spingarn (1982), Teboulle (1992), Tossings (1994), Wright (1990), etc.

2.6. Summary of Operator Splitting Methods and Double Backward Splitting Algorithms

The next five sections shall review the methods and their main properties of operator splitting for Problem $(\mathcal{P}_0)$. One may refer to Eckstein (1989) for more details.

Generally, suppose that T_1, T_2 and T are any maximal monotone operators such that $T = T_1 + T_2$. An *operator splitting method* for T is one that attempts to solve $(\mathcal{P}_0)$ repeated application of the combinations of operators of the form I, T_1, T_2, $(I + \lambda T_1)^{-1}$ and/or $(I + \lambda T_2)^{-1}$, where λ is some positive real number.

Several kinds of splitting algorithms have been developed in published literature. The most prominent are

(BFBS) basic forward-backward splitting method:

$$z^{k+1} \in (I + \lambda_k T_2)^{-1}(I - \lambda_k T_1)(z^k),$$

(DBS) double backward splitting method:

$$z^{k+1} = (I + \lambda_k T_2)^{-1}(I + \lambda_k T_1)^{-1}(z^k),$$

(PRS) Peaceman-Rachford splitting method:

$$z^{k+1} \in (I + \lambda_k T_2)^{-1}(I - \lambda_k T_1)(I + \lambda_k T_1)^{-1}(I - \lambda_k T_2)(z^k),$$

(DRS) Douglas-Rachford splitting method:

$$z^{k+1} \in (I + \lambda_k T_1)^{-1}[(I + \lambda_k T_2)^{-1}(I - \lambda_k T_1) + \lambda_k T_1](z^k),$$

(SPI) Spingarn's partial inverse method:

$$z^{k+1} = \tfrac{1}{2}[(I + \lambda T_1)^{-1}(z^k - w^k) + (I + \lambda T_2)^{-1}(z^k + w^k)],$$

$$w^{k+1} = w^k + \tfrac{1}{2}[(I + \lambda T_1)^{-1}(z^k - w^k) - (I + \lambda T_2)^{-1}(z^k + w^k)].$$

Definition 2.6.1. *Let T, T_1, and T_2 be maximal monotone operators over $\mathcal{H}$ such that $T = T_1 + T_2$. For a sequence of positive scalars $\{\lambda_k\}$, a sequence $\{z^k\} \subset \mathcal{H}$ is said to be generated by the double backward splitting algorithm (DBSA) for any start point z^0 if it satisfies the scheme (DBS) for all $k \geq 0$.*

For the general case, Lions (1978) proved that the DBSA is *unconditionally stable*, i.e., z^k remains bounded independently of k for any $\lambda_k \equiv \lambda$, but it does not converge to a zero point of T for any $\lambda_k \equiv \lambda$, except with some special modification. This is the main advantage and disadvantage of the DBSA.

Passty (1979) proved the weak ergodic convergence of the DBSA for the general maximal monotone T_1 and T_2 under the conditions

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \qquad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \tag{2.6.1}$$

Theorem 2.6.1 (Passty 1979). *Let $\{z^k\}$ be a sequence generated by the DBSA. If $\{\lambda_k\}$ satisfies (2.6.1) and*

$$w_m = \sum_{k=0}^{m} \lambda_k z^k \Big/ \sum_{k=0}^{m} \lambda_k, \tag{2.6.2}$$

then either (i) $\|w_m\| \to \infty$ for $T^{-1}(0) = \emptyset$; or (ii) $\{w_m\}$ converges weakly to a zero of T.

The ergodic convergence in the above does not mean the convergence of $\{z^k\}$. So far stronger results do not exist for the general maximal monotone case. Under some stronger conditions, Passty proved the strong convergence of $\{z^k\}$.

Theorem 2.6.2 (Passty 1979). *Let $\{z^k\}$ be a sequence generated by the DBSA with $\{\lambda_k\}$ satisfying (2.6.1). If either (i) T_2 is strongly monotone; or (ii) $T^{-1}(0)$ has nonempty interior, then $\{z^k\}$ converges strongly to a zero of T.*

The difficulty for the convergence of $\{z^k\}$ mainly come from the disadvantage proven by Lions. Therefore, $\lambda_k \to 0$ is required to guarantee that $\{z^k\}$ converges to a zero of T. The applications of the DBSA appear limited.

2.7. Peaceman-Rachford Splitting Algorithms

Definition 2.7 1. *Let T, T_1, and T_2 be maximal monotone operators over $\mathcal{H}$ such that $T = T_1 + T_2$, and $\{\lambda_k\} \subset (0, \infty)$. We call a sequence $\{z^k\}$ generated by the Peaceman-Rachford splitting algorithm (PRSA) for any starting point $z^0 \in \mathrm{dom}(T_2)$ if it satisfies the scheme (PRS) in Section 2.6 for all $k \geq 0$.*

The PRSA was formed by Peaceman and Rachford to solve parabolic and elliptic differential equations for the special linear case of T_1 and T_2. For the general maximal monotone operators, Lions (1978), Lions and Mercier (1979) analyzed the convergence and stability. They concluded that the PRSA is unconditionally stable, i.e., z^k remains bounded independently of k for any λ_k.

Let $J_i^\lambda = (I + \lambda T_i)^{-1}$ for $i = 1, 2$. When Lions and Mercier analyzed the convergence of the PRSA, they considered special sequences generated from the PRSA with $\lambda_k \equiv \lambda > 0$:

(i) Choose any $z^0 \in \mathrm{dom}(T_1 + T_2)$ and $y^0 \in T_2(z^0)$

(ii) Let $u^0 = z^0 + \lambda y^0$

(iii) Find the unique $u^{k+1} = (2J_1^\lambda - I)(2J_2^\lambda - I)(u^k)$ for $k = 0, 1, 2, \ldots$

(iv) $z^{k+1} = J_2^\lambda(u^{k+1})$.

Eckstein (1989) defined a *canonical Peaceman-Rachford* method, i.e., a sequence $\{(z^k, y^k)\}$ with $y^k \in T_2(z^k)$ may be obtained by the following procedure:

(i') Choose any $z^0 \in \mathrm{dom}(T_1 + T_2)$ and $y^0 \in T_2(z^0)$

(ii') Find the unique $x^k \in T_1(v^k)$ such that $v^k + \lambda_k x^k = z^k - \lambda^k y^k$ for $k = 0, 1, \ldots$

(iii') Find the unique $y^{k+1} \in T_2(z^{k+1})$ such that $z^{k+1} + \lambda_k y^{k+1} = v^k - \lambda_k x^k$.

Proposition 2.7.1 (Eckstein 1989). *For given z^0 and y^0, the Lions-Mercier method (i)-(iv) and the canonical Peaceman-Rachford method generate the same $\{z^k\}$ with $\lambda_k \equiv \lambda > 0$. Furthermore, $u^k = z^k + \lambda y^k$.*

Proposition 2.7.2 (Lions and Mercier 1979). *Suppose*

$$\text{there exists} \quad \bar{z} \in \mathcal{H}, \bar{x} \in T_1(\bar{z}), \bar{y} \in T_2(\bar{z}) \quad \text{such that} \quad \bar{x} + \bar{y} = 0. \tag{2.7.1}$$

Let $\{u^k\}$ and $\{z^k\}$ are generated from the Lions-Mercier method, and

$$w^k = 2z^k - u^k, y^k = \frac{u^k - z^k}{\lambda}, x^k = \frac{w^k - u^{k+1}}{2\lambda}.$$

Then, the sequences z^k, u^k, w^k, x^k, y^k remain bounded and

$$\lim_{k\to\infty} (y^k - \bar{y}, z^k - \bar{z}) = 0; \tag{2.7.2}$$

$$\lim_{k\to\infty} (x^k - \bar{x}, \frac{u^{k+1} + w^k}{2} - \bar{z}) = 0. \tag{2.7.3}$$

□

We call an operator T to have *the L-M property* if, for all $x^k, x \in \text{dom}(T)$ such that $T(x^k)$ is bounded, $x^k \overset{w}{\to} \bar{x}$ ($x^k \to \bar{x}$ weakly), and

$$\langle T(x^k) - T(x), x^k - x\rangle \to 0 \quad \text{as} \quad k \to \infty,$$

then one has $x = \bar{x}$.

Theorem 2.7.3 (Lions and Mercier 1979). *Suppose (2.7.1) is true and $\{z^k\}$ is generated from the Lions-Mercier method (i)-(iv).*

(a) *If T_2 is single valued and satisfies the L-M property, then z^k converges weakly to the unique zero point $\bar{z}$ of $T_1 + T_2$.*

(b) *If T_2 is strongly monotone, then z^k converges strongly to the unique $\bar{z}$.*

(c) *If T_2^{-1} is strongly monotone and univoque, then T_2 satisfies the conditions of Part (a).*

(d) *If T_2 is strictly monotone and weakly closed, then z^k converges weakly to the unique $\bar{z}$.*

(e) *Let* u^k *be the sequence from (iii) and* $w^k = 2z^k - u^k$. *If* T_1 *is single valued and satisfies the L-M property, then* $(u^{k+1} + w^k)/2$ *converges weakly to the unique* $\bar{z}$.

2.8. Douglas-Reachford Splitting Algorithms

Definition 2.8.1. *Let* T, T_1, *and* T_2 *be maximal monotone operators over* $\mathcal{H}$ *such that* $T = T_1 + T_2$, *and* $\{\lambda_k\} \subset (0, \infty)$. *We call a sequence* $\{z^k\}$ *generated by the Douglas-Rachford splitting algorithm (DRSA) for any starting point* $z^0 \in \mathrm{dom}(T_2)$ *if it satisfies the scheme (DRS) in Section 2.6 for all* $k \geq 0$.

The DRSA was originally formed by Douglas and Rachford to apply to get the numerical solution of the heat conduction problem in 2 and 3 space variables in 1956 for the special linear case of T_1 and T_2. For general maximal monotone operators, Lions and Mercier (1979) analyzed the convergence and stability. They concluded that the DRSA is also unconditionally stable, i.e., z^k remains bounded independently of k for any λ_k.

Similarly, let $J_i^\lambda = (I + \lambda T_i)^{-1}$ for $i = 1, 2$. When Lions and Mercier analyzed the convergence of the PRSA, they considered special sequences generated from the PRSA with $\lambda_k \equiv \lambda > 0$:

(a) Choose any $z^0 \in \mathrm{dom}(T_1 + T_2)$ and $x^0 \in T_1(z^0)$

(b) Let $u^0 = z^0 + \lambda x^0$

(c) Find the unique $u^{k+1} = J_2^\lambda(2J_1^\lambda - I)(u^k) + (I - J_1^\lambda)(u^k)$ for $k = 0, 1, 2, \ldots$

(d) $z^{k+1} = J_1^\lambda(u^{k+1})$.

Eckstein (1989) defined a *tight Douglas-Rachford* method, i.e., a sequence $\{(z^k, x^k)\}$ with $x^k \in T_1(z^k)$ that may be obtained by the following procedure:

(a') Choose any $z^0 \in \mathrm{dom}(T_1 + T_2)$ and $x^0 \in T_1(z^0)$

(b') Find the unique $y^k \in T_2(v^k)$ such that $v^k + \lambda_k y^k = z^k - \lambda^k x^k$ for $k = 0, 1, \ldots$

(c') Find the unique $x^{k+1} \in T_1(z^{k+1})$ such that $z^{k+1} + \lambda_k x^{k+1} = v^k + \lambda_k x^k$.

Proposition 2.8.1 (Eckstein 1989). *For given z^0 and x^0, the Lions-Mercier method (a)-(d) and the tight Douglas-Rachford method generate the same $\{z^k\}$ with $\lambda_k \equiv \lambda > 0$. Furthermore, $u^k = z^k + \lambda x^k$.*

The convergence conditions for the Lions-Mercier method (a)-(d) is weaker than those for the double backward algorithm and the PRSA.

Theorem 2.8.2 (Lions and Mercier 1979). *Suppose the condition (2.6.1) is true. Let $\{z^k\}$ and $\{u^k\}$ be generated from the Lions-Mercier method (a)-(d). Then u^k converges weakly to $\hat{u} \in \mathcal{H}$, as $k \to \infty$, such that $\hat{z} = J_1^\lambda(\hat{u})$ is a zero point of $T = T_1 + T_2$. Furthermore,*

(1) if T_1 is linear, z^k converges weakly to a zero point of T;

(2) if J_1^λ is weakly closed, z^k converges weakly to a zero point of T;

(3) if J_2^λ is compact, z^k converges strongly to a zero point of T;

(4) if J_1^λ is compact and also weakly closed, z^k converges strongly to a zero point of T;

(5) if T_1 and T_2 are odd, z^k converges strongly to a solution of T;

(6) if T_1 or T_2 is strongly monotone, z^k converges strongly to the unique solution of T;

(7) if T is maximal monotone, any subsequence of $\{z^k\}$ converges weakly to a solution of T.

Lions and Mercier (1979) also gave an evaluation of the speed of convergence in the same analysis as follows:

Theorem 2.8.3. *Suppose the condition (2.6.1) is true. Let $\{z^k\}$ and $\{u^k\}$ be generated from the L-M method (a)-(d). If T_1 is both strongly monotone with modulus α and Lipschitz with constant M, then there exists a constant C_1 such that*

$$\|u^k - \bar{u}\| \le C_1\gamma^k; \|z^k - \bar{z}\| \le C_1\gamma^k,$$

where $\bar{z} = J_1^\lambda(\bar{u})$ is the unique solution of T and

$$\gamma = \left[1 - \frac{2\lambda\alpha}{(1+\lambda M)^2}\right]^{1/2}.$$

Eckstein has also analyzed some properties of the L-M method (a)-(d) in the dissertation for his Ph.D. Some his results follow:

An operator S_λ is called as a *splitting operator* of T_1 and T_2 with respect to $\lambda > 0$ if

$$S_\lambda = (G_\lambda)^{-1} - I \quad \text{with} \quad G_\lambda = J_2^\lambda(2J_1^\lambda - I) + (I - J_1^\lambda). \tag{2.8.1}$$

Theorem 2.8.4 (Eckstein 1989). *S_λ is a (maximal) monotone operator provided T_1 and T_2 are (maximal) monotone. Furthermore, the iteration (c) of the L-M method is equivalent to applying the proximal point algorithm (PPA) to the operator S_λ with the proximal point step size λ_k fixed at 1. Hence, if $T^{-1}(0) = \emptyset$, then $\{u^k\}$ from the L-M method (a)-(c) is an unbounded sequence.*

The following useful generalization, also by Eckstein, is applicable for the relaxation and approximation of the L-M method (a)-(d):

Theorem 2.8.5. *Let $\mathcal{H} = \mathbb{R}^n$, and suppose $T^{-1}(0) \neq \emptyset$. Given some $u^0 \in \mathbb{R}^n$, $\lambda > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, and $\eta_k \in \mathbb{R}$ such that sequences $\{u^k\}$, $\{a^k\}$ and $\{b^k\}$ satisfy the following conditions:*

$$\|a^k - J_1^\lambda(u^k)\| \leq \beta_k \forall k \geq 0$$

$$\|b^k - J_2^\lambda(2a^k - u^k)\| \leq \alpha_k \forall k \geq 0$$

$$\|u^{k+1} = u^k + \eta_k(b^k - a^k) \forall k \geq 0$$

$$\sum\nolimits_{k=0}^{\infty} \alpha_k < \infty$$

$$\sum\nolimits_{k=0}^{\infty} \beta_k < \infty$$

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

Then $\{u^k\}$ converges to some $\bar{u}$ such that $\bar{z} = J_1^\lambda(\bar{u}) \in T^{-1}(0)$.

2.9. Spingarn's Partial Inverse Algorithms

Definition 2.9.1 (Spingarn 1983). *Let E and W be mutually complementary subspaces of the Hilbert space $\mathcal{H}$, and let T be a monotone operator on $\mathcal{H}$. Then the partial inverse T_E of T, with respect to E, is the operator with the following graph:*

$$\operatorname{gph} T_E = \{ (x_E + y_W, y_E + x_W) \in \mathcal{H} \times \mathcal{H} : y \in T(x) \} \tag{2.9.1}$$

where $x = x_E + x_W$ with $x_E \in E$, $x_W \in W$ and similar to y.

Spingarn proved that for any subspace E, T_E is (maximal) monotone if and only if T is (maximal) monotone and $0 \in T_E(z)$ iff $z_W \in T(z_E)$. Hence, he suggested applying the proximal point algorithm to partial inverse operator T_E to solve the following problem:

$$\text{to find} \quad x \in E \quad \text{and} \quad y \in W \quad \text{such that} \quad y \in T(x). \tag{2.9.2}$$

He proposed first obtaining a z such that $0 \in T_E(z)$ by the proximal point algorithm and then projecting z onto E and W to get x and y respectively. Spingarn (1983) termed this the *method of partial inverse.*

Since the method of partial inverses is basically a proximal point algorithm, the convergence results by Spingarn are similar to those of the proximal point algorithm by Rockafellar (1976).

Spingarn also applied the method of partial inverse to the following problem of related sums of monotone operators: Given maximal monotone mappings $T_1, \ldots, T_m$ on $\mathcal{H}$ and positive (nonzero) real numbers $\eta_1, \ldots, \eta_m$, find $y_1, \ldots, y_m$ and $x \in \mathcal{H}$ such that

$$\begin{aligned} &(i) \quad y_i \in T_i(x), i = 1, \ldots, m \\ &(ii) \quad \eta_1 y_1 + \ldots + \eta_m y_m = 0. \end{aligned} \tag{2.9.3}$$

Here we are interested that $m = 2$, and $\eta_1 = \eta_2 = \lambda$. In this case, Chen and Rockafellar (1992) deduced the (SPI) scheme in Section 2.6 according to the method of partial inverses.

Eckstein (1989) proved that $T_E = S_\lambda$ with $\lambda = 1$, and T_1, T_2 in (2.8.1) as T, $\mathcal{N}_E$ respectively. Hence, the method of partial inverses with proximal point step sizes as 1 is a special case of the Lions-Mercier method (a)-(d) for E as a linear subspace of $\mathcal{H}$, where $\mathcal{N}_E$ is the normal cone.

2.10. Basic Forward-Backward Splitting Methods

Let $\mathcal{H}$ be a real Hilbert space, and T_1, T_2 be maximal monotone operators on $\mathcal{H}$ such that $T = T_1 + T_2$ is maximal. For a sequence $\{\lambda_k\}_{k=0}^{\infty}$ of positive scalars, a sequence $\{z^k\}_{k=0}^{\infty} \subset \mathcal{H}$ is said to be generated by the *basic forward-backward splitting* (BFBS) method due to Passty (1979), if

$$z^{k+1} \in (I + \lambda_k T_2)^{-1}(I - \lambda_k T_1)(z^k), \qquad \forall \quad k \geq 0. \tag{2.10.1}$$

This method generalizes the *gradient projection* method, which was originally proposed by Goldstein (1964) and, independently, by Levitin and Polyak (1966) for convex optimization. The lesser known *subgradient projection* method, e.g., Polyak (1978, 1987), is also an example of the BFBS method. Further special cases of the (2.10.1) have been examined by Bruck (1975, 1977). Passty (1979), Gabay (1983), Tseng (1990, 1991), and Patriksson (1993) have also considered the methods of this class more generally.

Proposition 2.10.1. *Given $\lambda > 0$ and two maximal monotone operators T_1 and T_2 on a Hilbert space $\mathcal{H}$, then*

$$z \in (I + \lambda T_2)^{-1}(I - \lambda T_1)(z) \quad \textit{iff} \quad 0 \in T(z) = (T_1 + T_2)(z) \tag{2.10.2}$$

If $\{z^k\}$ conforms to the forward-backward recursion and, $z^k = z^{k+1}$, then $0 \in (T_1 + T_2)(z^k)$.

Proof.

$$\begin{aligned} & z \in (I + \lambda T_2)^{-1}(I - \lambda T_1)(z) \\ \Longleftrightarrow & (I + \lambda T_2)(z) \cap (I - \lambda T_1)(z) \neq \emptyset \\ \Longleftrightarrow & T_2(z) \cap (-T_1(z)) \neq \emptyset \\ \Longleftrightarrow & 0 \in (T_1 + T_2)(z) \end{aligned}$$

. $\square$

In general, the forward-backward splitting algorithm will not converge. For complete general maximal monotone operators T_1 and T_2, the basic forward-backward splitting method appears to have first been formulated by Passty (1979). However, the convergent result is very weak.

Passty's Theorem. *Let $\mathcal{H}$ be any real Hilbert space and let T, T_1 and T_2 be maximal monotone operators such that $T = T_1 + T_2$. Further, let $\{z^k\}_{k=0}^{\infty} \subset \mathcal{H}, \{y^k\}_{k=0}^{\infty} \subset \mathcal{H}, \{w^k\}_{k=0}^{\infty} \subset \mathcal{H}$, and $\{\lambda\}_{k=0}^{\infty} \subset (0, \infty)$ be sequences satisfying the conditions for all $k \geq 0$,*

$$\begin{aligned} z^{k+1} &= (I + \lambda_k T_2)^{-1}(z^k - \lambda_k y^k), \\ y^k &\in T_1(z^k), \\ w^k &= \frac{\sum_{j=0}^{k} \lambda_j z^{j+1}}{\sum_{j=0}^{k} \lambda_j}. \end{aligned} \tag{2.10.3}$$

where $\{y^k\}$ is bounded, and

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \tag{2.10.4}$$

Then, if $\mathrm{zer}(T) = \emptyset$, $\|w^k\| \to \infty$, while if $\mathrm{zer}(T) \neq \emptyset$, $\{w^k\}$ converges weakly to some zero of T.

This convergence result is not as useful as convergence of $\{z^k\}$ due to the numerical difficulties of computing w^k for large k. It may be also very difficult in practice to ensure that $\{y^k\}$ is indeed a bounded sequence.

Fortunately many special cases of the method by (2.10.1) are known to converge. These include the gradient projection method due to Goldstein (1964), and Levitin and Polyak (1966), the subgradient projection method in Polyak (1978, 1987), etc., in which $T_2 = \mathcal{N}_C$ (normal cone operator) of some nonempty closed convex set $C \subset \mathcal{H}$, and T_1 is the gradient mapping ∇f or subgradient ∂f associated with a convex function defined throughout C. The result of the gradient projection method is introduced as follows:

G-L-P Theorem. *Let $T_2 = \mathcal{N}_C$, where C is some nonempty, closed, convex subset of a Hilbert space $\mathcal{H}$. Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function, differentiable* *on C, such that ∇f meets the Lipschitz condition*

$$\|\nabla f(x) - \nabla f(x')\| \le L\|x - x'\|, \forall x, x' \in C \tag{2.10.5}$$

for some $L \ge 0$. Then, for $T_1 = \partial f$ and any sequence of step sizes $\{\lambda_k\}_{k=0}{}^{\infty}$, such that

$$0 < \inf_k \lambda_k \le \sup_k \lambda_k < \frac{2}{L} \tag{2.10.6}$$

the BFBS method

$$x^{k+1} \in (I + \lambda_k T_2)^{-1}(x^k - \lambda_k T_1(x^k)) = P_C(x^k - \lambda_k \nabla f(x^k)), \forall k \ge 0 \tag{2.10.7}$$

converges to a zero of $T_1 + T_2$, i.e., a minimizer of f on C.

Now the results of Gabay (1983) and Tseng (1991) are introduced.

Gabay's Theorem. *Suppose T_1 and T_2 are maximal monotone with $T = T_1 + T_2$.*

(a) If T_1 is Lipschitz-continuous (with constant M) and strongly monotone (with modulus α), then the sequence $\{z^k\}$ generated from (2.10.1) converges strongly to $z \in \mathcal{H}$ with $0 \in T(z)$ if

$$0 < \lambda_k \equiv \lambda < 2\alpha M^{-2}. \tag{2.10.8}$$

(b) If $T^{-1}(0) \ne \emptyset$ and T_1^{-1} is strongly monotone (with modulus β), then the sequence $\{z^k\}$ converges weakly to z for all $0 < \lambda_k \equiv \lambda < 2\beta$.

Tseng got the similar result to Part (b) of Gabay's result over finite dimensional space $\mathbb{R}^n$ for $\{\lambda_k\}$ not fixed-valued.

Tseng's Theorem. *If $\mathcal{H} = \mathbb{R}^n$, $\mathrm{zer}(T) \ne \emptyset$, T_1^{-1} is strongly monotone with modulus β and*

$$0 < \inf_{k \ge 0} \lambda_k \le \sup_{k \ge 0} \lambda_k < 2\beta, \tag{2.10.9}$$

then the $\{z^k\}$ from (2.10.1) converges to a zero of T.

When $\mathcal{H} = \mathbb{R}^n$, the G-L-P Theorem is just a corollary of Tseng's Theorem. This is due to a lemma of Gol'shtein and Tret'yakov (1975).

Lemma 2.10.2. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable. For any $M > 0$, the following two conditions are equivalent:*
(i) $\|\nabla h(x) - \nabla h(y)\| \leq M\|x - y\| \qquad \forall x, y \in \mathbb{R}^n$
(ii) $\langle \nabla h(x) - \nabla h(y), x - y\rangle \geq M^{-1}\|\nabla h(x) - \nabla h(y)\|^2 \qquad \forall x, y \in \mathbb{R}^n.$

In 1980, Cohen [51] develops a decomposition method which was declared *auxiliary problem principle* for minimization and saddle point problems. The basic idea of his method is similar to the above BFBS method except replacing I in (2.10.1) by a group of uniformly strongly monotone and uniformly Lipschitz continuous operators $\{\nabla\varphi_k\}$ (refer to corresponding definitions 5.4.1 and 5.4.2 in Chapter 5) when the iterative sequence $\{z^k\}$ is generated. He proved weak and strong convergence for his algorithm. These convergence results are special cases of the theorems in Section 5.4 of this dissertation. Later, he and co-workers extended his algorithm to many other problems ([52] – [59]) although they consider T_2 being a normal cone operator for most cases. Thus, the method becomes an extension of the projection method in many of its applications.

Dafermos (1983), Tseng (1990), Noor (1992), etc., did similar extensions by replacing I in (2.10.1) by a more general operator. However, the T_2 in their algorithms is a normal cone operator. Hence, their methods are extensions of the projection method.

In 1993, Patriksson presented a unified framework of iterative algorithms in his Ph. D dissertation. He consider the following generalized zero point problem: find a $x^* \in X$ such that

$$0 \in (F + U + \mathcal{N}_X)(x^*), \tag{2.10.10}$$

where X is a nonempty, closed and convex set in $\mathbb{R}^n$ and $F : X \to \mathbb{R}^n$, $U : X \rightrightarrows \mathbb{R}^n$ are two operators, such that U is monotone.

In his algorithm for solving the problem (2.10.10), he combined line searches with a generalized scheme of forward-backward splitting:

$$Y(x^k) = [\Phi^k + U + \mathcal{N}_X]^{-1}[\Phi^k - F](x^k), \tag{2.10.11}$$

here Φ^k is a monotone operator. He called his algorithm *cost approximation algorithm.*

Patriksson did not give a convergence analysis of his algorithm for general F and for general monotone operators U and Φ^k. However, he did discuss convergence for his algorithm when he applied the algorithm to problems in nonlinear programs, nondifferentiable optimization, variational inequalities, in which Φ^k is a strongly monotone operator (in some special cases, just monotonicity is necessary) and F, U are some special monotone operators. The convergence results of his algorithm mainly depend on conditions on F, Φ^k, and X, not on U.