Report on

MONOTONE OPERATORS AND THE PROXIMAL POINT ALGORITHM*(PPA)

and

AUGMENTED LAGRANGIANS AND APPLICATIONS OF THE PROXIMAL POINT ALGORITHM IN CONVEX PROGRAMMING*(AL)

by Rockafellar

Wang Yufan 11910902

This report mainly shows my understanding of essay PPA and AL. PPA mainly discussed the convergence of proximal algorithm which is applied to maximal monotone operators while AL introduces three algorithms to solve convex program by applying proximal point algorithm for maximal monotone operators.

The proximal point algorithm generates by

$$z^{k+1} \approx P_k(z^k) \qquad P_k = (I + c^k T)^{-1}$$

where c_k is a sequence of positive real number. The proximate will be reduced due to the variation of the property of T. The strongly monotone of T may make it easier to deal with the convergence while it may excludes some of the most important applications. We now introduce two general criteria for the approximate calculation:

$$||z^{k+1} - P_k(z^k)|| \leq \varepsilon_k \qquad \sum \varepsilon_k \leq \infty \quad (A)$$

$$||z^{k+1} - P_k(z^k)|| \leq \delta_k ||z^{k+1} - z^k|| \qquad \sum \delta_k \leq \infty \quad (B)$$

These two are very important in the argument of the convergence of algorithm.

To deal with the convergence of the algorithm, we shall make use of the mapping

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

since

$$0 \in T(z) \iff Q_k(z) = 0$$

by using the properties of Q_k we can show the result of convergence in certain conditions.

Theorem 1. $\{z_k\}$ are any sequence generated by the proximal point algorithm under criterion (A) with $\{c_k\}$ bounded away from zero. Suppose $\{z_k\}$ is bounded; this holds under the preceding assumption if and only if there exists at least one solution to $0 \in T(z)$. Then $\{z_k\}$ converges in the weak topology to a point z^∞ satisfying $0 \in T(z^\infty)$, and

$$0 = \lim_{k \to \infty} ||Q_k(z^k)|| = \lim_{k \to \infty} ||z^{k+1} - z^k||$$

That is, $\{z_k\}$ converges to z^∞ by applying the properties of Q_k as an intermediary under these conditions. Moreover, we can estimate the rate of convergence in some ways.

Theorem 2. $\{z_k\}$ are any sequence generated by the proximal point algorithm using criterion (B) with $\{c_k\}$ nondecreasing, assume that $\{z_k\}$ is bounded and that T^{-1} is Lipschitz continuous at 0 with modulus a; let

$$\mu_k = a/(a^2 + {c_k}^2)^{1/2} < 1$$

Then $\{z_k\}$ converges strongly to $\bar{z}$, the unique solution to 0 T(z). Moreover, there is an index k such that

$$||z^{k+1} - \bar{z}|| \leq \theta_k ||z^k - \bar{z}||$$

for all $k \geq \bar{k}$, where

$$1 > \theta_k \equiv (\mu_k + \delta_k)/(1 - \delta_k)$$

for all $k \geq \bar{k}$, and

$$\theta_k \to \mu_\infty$$

Now we find a coefficient θ_k to estimate the rate of convergence. Since we have already finished the preperation work about the convergence, we can finally move on to introduce the three algorithms.

We consider the convex program

$$\min f_0(x) \quad s.t. \quad f_1(x) \leq 0, ..., f_m(x) \leq 0 \qquad (P)$$

we now introduce the three algorithms.

Primal application: proximal minimization algorithm. The proximal minimization algorithm of (P) is

$$x^{k+1} \approx \arg \min_{x \in \mathbf{R}^n} f(x) + (1/2c_k)|x - x^k|^2$$

by applying Theorem 1 we can get the convergence of proximal minimization algorithm under certain conditions.

Theorem 3. Let the proximal minimization algorithm be executed with stopping criterion (A) applied to the minimization function. If the sequence $\{x_k\}$ is bounded, then $\{x_k\}$ converges to some optimal solution x^∞ of (P). Moreover

$$0 \leq f(x^{k+1}) - \min P \leq {c_k}^{-1} |x^{k+1} - x^\infty| (\epsilon_k + |x^{k+1} - x^k|) \to 0$$

The boundedness of x^k under (A) is in fact equivalent to the existence of an optimal solution to (P). By applying the general rate-of-convergence theorem for the proximal point algorithm(Theorem 2), we can moreover get the rate of convergence of it as $|x^{k+1} - \bar{x}| \leq \theta_k |x^k - \bar{x}| < |x^k - \bar{x}|$ under stopping criterion (B).

Dual application: method of multipliers. We take the dual problem of (P), that is

$$\max \inf_{x \in C} f_0(x) + y_1 f_1(x) + ... + y_m f_m(x) \qquad (D)$$

The augmented Lagrangian is

$$L(x, y, c) = f_0(x) + \sum_{i=1}^{m} (y_i f_i(x) + (c/2) f_i(x)^2)$$

Then we use Newton's Method to update y^k so we can use it to solve the minimization of augmented Lagrangian

$$Y(x^{k+1}, y^k, c_k) = c_k \nabla_y L(x^{k+1}, y^k, c_k)$$

So the method of multipliers can be expressed as

$$x^{k+1} \approx \arg \min_{x \in \mathbf{R}^n} L(x, y^k, c_k)$$

$$y^{k+1} = Y(x^{k+1}, y^k, c_k)$$

by applying Theorem 1 again we can get the convergence of method of multipliers under certain conditions.

Theorem 4. Let the method of multipliers be executed with stopping criterion (A) applied to $L(x, y^k, c_k)$. If the sequence $\{y_k\}$ is bounded, then $\{y_k\}$ converges to some optimal solution y^∞ of (D), and $\{x_k\}$ is asymptotically minimizing for (P). Moreover

$$f_i(x^{k+1}) \leq c_k^{-1} |y^{k+1} - y^k| \to 0 \quad for \quad i = 1, ..., m$$

$$f_0(x^{k+1}) - asym \inf(P) \leq (1/2c_k)[\epsilon_k^2 + |y^k|^2 - |y^{k+1}|^2]$$

The boundedness of $\{y_k\}$ under (A) is actually equivalent to the existence of an optimal solution to (D). By applying the general rate-of-convergence theorem for the proximal point algorithm(Theorem 2) again, we can moreover get the rate of convergence of it as $|y^{k+1} - \bar{y}| \leq \theta_k |y^k - \bar{y}| < |y^k - \bar{y}|$ and $|x^{k+1} - \bar{x}| \leq \theta'_k |y^k - \bar{y}|$ under stopping criterion (B). Now we move on to the last algorithm.

Minimax application: the proximal method of multipliers. It is actually a combination of proximal point algorithm and method of multipliers. Instead of the augmented Lagrangian in method of multipliers, we use the proximal function of augmented Lagrangian. In other words, we approximate the optimal point by augmented Lagrangian but not the original function, that is

$$x^{k+1} \approx \arg \min_{x \in \mathbf{R}^n} L(x, y^k, c_k) + (1/2c_k)|x - x^k|^2$$

$$y^{k+1} = Y(x^{k+1}, y^k, c_k)$$

again by applying Theorem 1 we get its convergence.

Theorem 5. Let the method of multipliers be executed with stopping criterion (A) applied to $L(x, y^k, c_k)$. If the sequence $\{(x_k, y_k)\}$ is bounded, then $\{(x_k, y_k)\}$ converges to some optimal solution $\{(x_\infty, y_\infty)\}$ of (P,D), and min(P)=max(D). Moreover

$$y_i^{k+1} f_i(x^{k+1}) \to 0 \quad for \quad i = 1, ..., m$$

$$f_0(x^{k+1}) - \min(P) \leq {c_k}^{-1} |x^{k+1} - x^\infty| (\epsilon_k + |x^{k+1} - x^k|) - \sum_{i=1}^{m} y_i^{k+1} f_i(x^{k+1}) \to 0$$

$$f_0(x^{k+1}) - \min(P) \leq {c_k}^{-1} |y^\infty| |y^{k+1} - y^k| \to 0$$

the difference of this method comparing to the former two is that it calculates the optimal solution of both original program and dual program.By applying the general rate-of-convergence theorem for the proximal point algorithm(Theorem 2), we can get the rate of convergence a bit differently, which is $|(x^{k+1}, y^{k+1}) - (\bar{x}, \bar{y})| \leq \theta_k |(x^k, y^k) - (\bar{x}, \bar{y})| < |(x^k, y^k) - (\bar{x}, \bar{y})|$ under stopping criterion (B). By fixing x^k of y^k, we can obtain the rate of convergence of single variable respectively.

By comparing the three algorithm, we find something to consider. Actually the augmented Lagrangian is considered to solve the dual problem while the proximal minimization solves the original problem, both can leads to an optimal solution of the original problem if there is no dual gap. The proximal method of multipliers, considering the minimization of augmented Lagrangian as the original problem and applying proximal minimization to it, is actually a practical combination of these two.