A Reflected Forward-Backward Splitting Method with an Inertial Term for Monotone Inclusions Involving Lipschitzian Operators

Author : Peerapong Tawong , Student id: 650510533
Advisor : Asst. Prof. Dr. Supalak Phothi

FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY

INTRODUCTION

Solving operator inclusion problems in Hilbert spaces plays an important role in the fields of functional analysis and convex optimization, particularly problems of the form involving the sum of two operators

$$0 \in A(x) + B(x)$$

The Reflected Forward–Backward Splitting Method is an important algorithm used to solve problems of this form under the assumptions that A is a maximal monotone operator and B is a Lipschitz continuous operator. This algorithm can be written as an iterative process involving the resolvent operator, which allows the convergence properties to be analyzed using tools from functional analysis.

The motivation for this research arises from the observation that an inertial term can be added into the method, potentially leading to faster convergence. Therefore, the inertial term was added to the reflected forward–backward splitting method previously studied in the seminar course (206390). This leads to the research question of how such properties can be connected to applied problems in Machine Learning.

In particular, the LASSO problem, which is an optimization problem involving the minimization of the sum of convex functions, can be reformulated as a problem involving the sum of monotone operators.

PRELIMINARIES

Definition Let A be any operator. The set of zeros of A is defined by

$$\mathrm{zer}(A) = \{x \in H : 0 \in Ax\}$$

Definition Let $A : H \to 2^H$ be a multi-valued operator. The graph of A is defined by

$$\mathrm{gra}(A) \subset H \times H = \{(x, y) : y \in Ax\}$$

Definition Let $A : H \to 2^H$. The operator A is called a *monotone operator* if

$$\langle x - y,\, u - v\rangle \geq 0, \quad \forall (x, u), (y, v) \in \mathrm{gra}(A)$$

If $B : H \to H$ is a single-valued operator, then B is monotone if

$$\langle x - y,\, Bx - By\rangle \geq 0, \quad \forall (x, y) \in H \times H$$

Definition Let $A : H \to 2^H$ be a monotone operator. The operator A is called a *maximal monotone operator* if there does not exist an operator C such that

$$\mathrm{gra}(A) \subset \mathrm{gra}(C)$$

Definition An operator B is called a *Lipschitz operator* if there exists a constant $\mu > 0$ such that

$$\|Bx - By\| \leq \mu\|x - y\|, \quad \forall x, y \in H$$

Proposition Let A be a maximal monotone operator. If $x_n \rightharpoonup x$ and $y_n \to y$, and $(x_n, y_n) \in \mathrm{gra}(A)$ for all n, then $(x, y) \in \mathrm{gra}(A)$

Theorem Let $A : H \to 2^H$ be a maximal monotone operator and $B : H \to H$ be a monotone and Lipschitz continuous operator. Then $A + B$ is a maximal monotone operator

Proposition Let C be a nonempty subset of H and let (x_k) be a sequence in H such that $\|x_k - x\|$ converges for every $x \in C$. If every weak cluster point of the sequence (x_k) belongs to C, that is,

$$\{x \in H : x_{k_j} \rightharpoonup x \text{ for some subsequence } (x_{k_j}) \text{ of } (x_k)\} \subset C,$$

then the sequence (x_k) converges weakly to a point in C

MAIN RESULTS

RESULTS FROM THEOREM

From the problem of finding a zero of the sum of operators

$$0 \in A(x) + B(x)$$

where A is a maximal monotone operator and B is a Lipschitz continuous operator

From the iterative scheme of the *Reflected Forward–Backward Splitting Method with an inertial term*,

$$y_n = 2x_n - x_{n-1}$$
$$x_{n+1} = (Id + \gamma A)^{-1}(x_n + \theta(x_n - x_{n-1}) - \gamma B y_n)$$

the theorem implies that the sequence (x_n) weakly converges to x, which is the zero of the sum of the operators, when $\theta \in \left[0, \frac{1}{8}\right)$ and $\gamma \in \left(0, \frac{(\sqrt{2}-1)(1-8\theta)}{\mu}\right)$

NEURAL NETWORK CLASSIFICATION

1. The Iris dataset consisting of 150 samples, each with four features and one class label, is divided into a training set and a test set. The training set contains 105 samples, with 35 samples from each class, while the test set contains 45 samples, with 15 samples from each class.
2. (I) The neural network is constructed with 33 hidden nodes.
   (II) Let $W \in \mathbb{R}^{4\times 33}$ be the weight matrix of the hidden layer.
   (III) Let $b \in \mathbb{R}^{1\times 33}$ be the bias vector of the hidden layer.
   (IV) Let "ReLU" be the Activation Function.
3. Let $X_{\text{test}} \in \mathbb{R}^{45\times 4}$ denote the test data matrix containing all 45 test samples, where each sample consists of four features.

   ***Thus, the training process reduces to solving a LASSO problem to determine the weight matrix of the output layer $\beta \in \mathbb{R}^{33\times 3}$.
4. The predicted outputs for the training data are computed as $Y_{\text{test}} = \big(\mathrm{ReLU}(X_{\text{test}}W + b)\big)\beta$ and the predicted class is chosen as the class with the largest value in each row.
5. Compute Accuracy of test set by using confusion matrix

LASSO PROBLEM

Let $H_{\text{train}} = \big(\mathrm{ReLU}(X_{\text{train}}W + b)\big)\beta$ be the feature matrix obtained from the hidden layer, and let Y_{train} be the one-hot encoding matrix of the true classes.

Consider the problem of finding β that minimizes the following function

$$\min_{\beta} \lambda\|\beta\|_1 + \frac{1}{2}\|H_{\text{train}}\beta - Y_{\text{train}}\|_2^2$$

where $\lambda > 0$ is a parameter controlling sparsity.

which is equivalent to

$$0 \in \partial\lambda\|\beta\|_1 + \nabla\frac{1}{2}\|H_{\text{train}}\beta - Y_{\text{train}}\|_2^2$$

Therefore, the problem can be written in the form

$$0 \in A(\beta) + B(\beta)$$

Let (β_n) is a sequence generated by Reflected Forward–Backward Splitting Method with an inertial term, the theorem implies that the sequence (β_n) weakly converges to β, which is the solution of LASSO problem.

A table showing the comparison of the overall accuracy values of the test data for each method according to the number of iterations when $\gamma = \dfrac{0.32}{\mu}$

Iteration	Accuracy values of the test data (%)	
	RFBS Method with $\theta = 0.0025$	RFBS Method
500	79.2593	77.7778
1000	83.7037	83.7037
1500	88.1481	88.1481
2000	89.6296	89.6296
2500	95.5556	95.5556
3000	95.5556	95.5556
3500	97.0370	95.5556
4000	97.0370	97.0370
4250	98.5185	97.0370
>4500	98.5185	98.5185

