Chapter 6

Conclusions and Future Work

In this final chapter, the main contributions of this thesis are summarized, and possible directions for future work are discussed.

6.1 Summary

The central theme of using optimization in machine learning is to design and analyse new algorithms that efficiently and speedily solve various optimization frameworks that are utilized to solve learning problems. Our work primary focused on leveraging concepts of pure mathematics for designing such machine learning algorithms. We proposed two new iterative schemes for accelerated proximal gradient algorithms. In addition to the proximal gradient algorithms, two new schemes of operator splitting algorithm are proposed, which is also an important contribution of this thesis.

In the infinite-dimensional real space, proximal gradient methods can be interpreted as the forward-backward splitting algorithms. Also, the accelerated gradient algorithms can be interpreted as inertial-based forward-backward splitting methods. In Chapter 3, we proposed new definitions of proximal gradient methods based on the non-traditional fixed point iterative schemes. We applied all the newly proposed algorithms to the task of regression. We also proposed a new inertial-based forward-backward splitting technique and corresponding accelerated gradient algorithm. The convergence and stability of the proposed algorithm are analyzed in this chapter. To test the practical performance of the algorithms, we applied the algorithms to the tasks of the high-dimensional regression/classification problem and experiments are performed on several publicly available real datasets. We also applied the algorithm for the task of unified sparse representation learning, and present the experiments and result analysis with two real cross-modal datasets. In the end, we applied the algorithm to solve two extended-lasso frameworks, which are overlapped group lasso and fused lasso, and showed the performance with three real datasets.

In Chapter 4, a new inertial-based forward-backward algorithm, which used the concept of viscosity-approximation based fixed-point iterative scheme, is proposed. As far as the practical performance of the algorithm concerns, the proposed algorithm demonstrates better empirical convergence rate than the traditional accelerated gradient algorithm. We apply the recent viscosity approximation based forward-backward algorithm [175] to solve the problem of multitask regression and proposed a novel viscosity-approximation-based accelerated gradient algorithm (VAGA) for the problem of multitask regression. The most important fact of the proposed algorithm is that it is the first accelerated gradient algorithm that converges to a solution point strongly in the generalized infinite dimensional Hilbert space. The algorithm is applied to the problem of regularized multitask regression with sparsity-inducing regularizers. Experimental results are presented with three benchmark real datasets. VAGA is also applied to the popular bio-informatics problem of joint splice-site recognition and showed the performance with six different genomes.

Based on the concept of extragradient fixed-point schemes, a new operator splitting method is proposed in Chapter 5, along with its application to the microarray gene analysis. The proposed algorithm is considered to be parallel to the traditional Douglas-Rachford splitting, Peaceman-Rachford splitting, and the forward-backward splitting technique. We propose an extragradient-based operator splitting algorithm (EOSA) for the problem of minimizing the sum of two non-smooth closed convex functions. The linear convergence of the algorithm is shown under few specific assumptions. We have also proposed an accelerated variant of the algorithm and analyzed the convergence of the proposed algorithm. To demonstrate the performance of the proposed algorithms on the real-world problems, we performed extensive experiments with four high-dimensional microarray gene datasets and compared different standard and latest related algorithms.

6.2 Future Work

The contributions in this thesis lead to several future directions as follows:

- A recent research direction in the field of machine learning optimization algorithms is to propose new methods for the learning frameworks under nonconvex settings [117, 85, 89, 146, 220, 39, 147]. Under this setting, either the loss function, or the regularization function, or both the functions are non-convex. Removing the constraint of convexity extends the areas of applications. Investigating the nature of the proposed algorithms under the non-convex frameworks is an interesting line of work.

- The practical problems we currently handled in this thesis concerns the high-dimensional datasets, where the number of instances is significantly lesser than the number of attributes in each instance or example. Another extension may involve handling the case when the number of instances is also very large. Stochastic gradient algorithms work commendably in such cases. It will be an interesting problem to develop novel stochastic formulations for the proposed iterative schemes.

- Deep learning is a current topic of interest in the field of machine learning that gives promising results in many of the application areas such as image processing and speech recognition to mention a few. An interesting direction of research is to apply the proposed methods to the deep learning for efficient optimization. According to the current scenario, stochastic gradient algorithms perform well for deep learning frameworks, applying the stochastic formulations of the proposed work is next idea of research.

- An important analysis for the class of accelerated gradient techniques include the convergence rate analysis. In this thesis, we have provided the empirical rate of convergence of the proposed algorithms using many real benchmark datasets. However, it is required to analyze the rates of convergence of the proposed algorithms.