Inria international program

Associate Team acronym: BIGFOKS2

Principal investigator (Inria): *Francis Bach - SIERRA project-team*

Principal investigator (partner): *Chiranjib Bhattacharyya - Professor at Indian Institute of Science (IISc), Bangalore*

Other participants:
(a) Raman Sankaran (PhD Student, CSA, IISc)
(b) Bibaswan Chatterjee (M.E. student, CSA, IISc)
(c) Rakesh Shivanna (Google Inc, alumnus of CSA, IISc)
(d) Sesh Kumar (INRIA, Paris)
(e) Achintya Kundu (PhD student, CSA, IISc)

1. Abstract of the scientific program

In this project we identify two central problems in Machine Learning in the Big-Data setting. The goal is to design efficient first order methods for these two problems. In the following we describe in brief the problems and relevant research directions.

1.) Kernel methods for learning

Many real world problems involve observations which may not have an explicit feature representation. Instead, often pairwise relationship between examples maybe known, e.g. E-commerce companies selling books may have information about the similarity between a pair of books. Protein structure prediction is another example where no explicit feature representation is known but pairwise relationship can be easily obtained using some alignment. Kernel functions are mathematical constructs which naturally incorporate similarities between. These functions are dot-products in a Hilbert space and implicitly define a feature representation. Learning Kernel based prediction function in the Big-Data setting is an ill-addressed problem. There are two specific challenges that needs to be addressed. Firstly, Design of kernel functions for diverse data-types such as graphs remain a challenge. Secondly, efficient learning with these kernel functions often require solving complicated convex programs which are not suited for Big-Data setting.

2.) Using Prox-functions for solving large-scale Convex relaxations of Submodular minimization problems arising in Machine Learning

Many important problems, in a wide variety of domains, can be posed as minimization of submodular functions (e.g., image segmentation). It has been long noted that submodular functions on discrete problems play the same role as convex functions on continuous functions. Recent results show that one can design specific regularizers based on submodular functions, which exhibit favorable behaviors. However existing algorithms for such optimization problems still depend on Interior point (IP) methods (or methods of similar complexity, such as the minimum-norm-point algorithm). As is well known that IP methods are not useful in the Big- Data setting due to its O(N3) time complexity. The cubic complexity arises due to the inherent Newton step (or linear systems) in the algorithm. The goal of this research will be to address this issue and design first-order methods which avoid these Newton steps. State-of-the-art algorithms in convex optimization require the notion of proximal minimization which in turn requires the notion of prox-functions. Prox-functions of Lovasz-extension-based regularized are not known. As part of this research we would like to design suitable Prox-functions which are not expensive and yet can yield an accurate minimization of the resulting problem at hand.

2. Scientific progress

We consider problems arising in the context of regularized learning. It is well known in the machine learning literature that regularizers encode the prior information regarding the structure of the models (their sparsity pattern), which often lead to better solutions. The major questions which arise in the context of regularization are

(a) What is the generalization error and does the solution lead to consistent solutions ?
(b) How do we solve the problem computationally ?
(c) How efficient is the solver ?

We first consider the problem of graph transduction, and learning to predict the labels on the vertices. We show that regularization with the spectral norm leads to better generalization, and propose an efficient algorithm based on inexact proximal operator, to solve the optimization problem. In the second work, we consider a particular family of subquadratic norms, which are derived as ℓ_2-relaxations of cardinality based submodular functions. We show that these family of norms have the ability of simultaneously identifying the groups of correlated variables, and learn the model parameters. We also propose an efficient solver based on the Pool-Adjacent-Violators algorithm to solve the optimization problem. In the third work, we consider the general class of subquadratic norms, which are increasingly used in the field of structured sparsity. We derive general purpose algorithms for the instances where the proximal operator for the norm is not easily computable. In the fourth work, we consider convex optimization problems constrained over intersection of many simple sets, where we propose first-order algorithms with improved convergence rates.

Problem 1. *Spectral Norm Regularization of Orthonormal Representations for Graph Transduction*

Authors. *R. Shivanna (Google Inc), B. Chatterjee (M.E. student, CSA, IISc), R. Sankaran (PhD student, CSA, IISc), C. Bhattacharyya (Professor, CSA, IISc), F. Bach (INRIA-SIERRA project team, ENS, Paris).*

(Published in NIPS, 2015)

For learning labels on vertices of a graph, it is known that embedding a graph on a unit sphere leads to better generalization. However, the choice of optimal embedding and an efficient algorithm to compute the same remains an open issue. In this paper, we show that orthonormal representations, a class of unit-sphere graph embeddings are PAC learnable. The existing analysis does not directly apply, and we propose an alternative PAC-based bounds which do not depend on the VC dimension of the underlying function class but are related to the famous Lovasz Theta function.

Results presented

The main contribution of the paper is SEER, a spectral regularized orthonormal embedding for graph transduction, derived from the PAC bound. SEER is posed as a non-smooth convex function over an elliptope. These problems are usually solved as semidefinite programs (SDPs) with time complexity O(n^6). Projecting on elliptopes is a demanding problem and remains the main computational bottleneck. We tackle this challenge by presenting Infeasible Inexact proximal (IIP) method: an Inexact proximal method which performs subgradient procedure on an approximate projection, not necessarily feasible. We show that IIP converges to the optimal with the rate O(1/sqrt(T)). IIP is generally applicable and we use it to compute SEER where the approximate projection step is computed by an accelerated gradient descent procedure, such as FISTA. We show that if number of iterations in the FISTA step are at least sqrt(T), the IIP method yields the same O(1/sqrt(T)) convergence. The proposed algorithm easily scales to 1000's vertices, while the standard SDP computation does not scale beyond few hundred vertices. As an interesting aside, we can easily compute Lovasz Theta, and show similar speedups. Furthermore, the analysis presented here easily extends to the multiple graphs setting.

Problem 2. *Identifying groups of strongly correlated variables through Smoothed Ordered Weighted L_1-norms*

Authors. *R. Sankaran (PhD student, CSA, IISc), C. Bhattacharyya (Professor, CSA, IISc), F. Bach (INRIA-SIERRA project team, ENS, Paris).*

(Published in AISTATS 2017)

The failure of LASSO to identify groups of correlated predictors in linear regression has sparked significant research interest. Recently, various norms were proposed, which can be best described as instances of ordered weighted ℓ_1 norms (OWL), as an alternative to ℓ_1 regularization used in LASSO. OWL can identify groups of correlated variables but it forces the model to be constant within a group. This artifact induces unnecessary bias in the model estimation. In this paper we take a submodular perspective and show that OWL can be posed as the Lovasz extension of a suitably defined submodular function. The submodular perspective not only explains the group-wise constant behavior of OWL, but also suggests alternatives.

Results Presented

The main contribution of this paper is smoothed OWL (SOWL), a new family of norms, which not only identifies the groups but also allows the model to be flexible inside a group. We establish several algorithmic and theoretical properties of SOWL including group identification and model consistency. We also provide algorithmic tools to compute the SOWL norm and its proximal operator, whose computational complexity (O(d log d)) is significantly better than that of general purpose solvers (O(d^2 log d)). In our experiments, SOWL compares favorably with respect to OWL in the regimes of interest.

Problem 3. *Convex optimization over intersection of simple sets: improved convergence rate guarantees via an exact penalty approach*

Authors. *A. Kundu (PhD student, CSA, IISc), C. Bhattacharyya (Professor, CSA, IISc), F. Bach (INRIA-SIERRA project team, ENS, Paris).*

(Accepted for publication in AISTATS 2018)

We consider the problem of minimizing a convex function over the intersection of finitely many simple sets which are easy to project onto. This is an important problem arising in various domains such as machine learning. The main difficulty lies in finding the projection of a point in the intersection of many sets. Existing approaches yield an infeasible point with an iteration-complexity of $O(1/\varepsilon^2)$ for nonsmooth problems with no guarantees on the in-feasibility.

Results presented. By reformulating the problem through exact penalty functions, we derive first-order algorithms which not only guarantees that the distance to the intersection is small but also improve the complexity to $O(1/\sqrt{\varepsilon})$ and $O(1/ \varepsilon)$ for smooth functions. For composite and smooth problems, this is achieved through a saddle-point reformulation where the proximal operators required by the primal-dual algorithms can be computed in closed form. We illustrate the benefits of our approach on a graph transduction problem and on graph matching.

Problem 4. *Treating Structured Sparsity With Subquadratic norms : A Primal-Dual Approach.*

Authors*. R. Sankaran (PhD student, CSA, IISc), C. Bhattacharyya (Professor, CSA, IISc), F. Bach (INRIA-SIERRA project team, ENS, Paris).*

(Technical Report)

Reweighted least-squares formulations appear naturally in machine learning and signal processing. The particular instance we tackle in this paper is the regularization by subquadratic norms, i.e., norms that are concave functions of the squares of the components of its arguments. These norms include the ℓ_1-norm, all combinations of ℓ_2-norms of subvectors, H-norms, or norms that emerge as convex relaxations of combinatorial penalties: they may all be formulated as the infimum of squared weighted ℓ_2-norms. The traditional algorithms associated with these formulations require to solve ℓ_2 -regularized problems at each iteration, which can be costly, even with quadratic data-fitting terms.

Results presented.

In this paper, we derive primal-dual formulations that extend the classical algorithms based on proximal operators of the norm, to the instances of subquadratic norms, which typically do not have efficient algorithms for their proximal operators. Particularly, we derive algorithms with convergence rate of O(1/T), which can be accelerated to a rate of O(1/T^2) in favourable cases. For norms where the proximal operator for the norm is available, we show that the running-time is similar, while for general cases, we improve significantly over subgradient descent (the only generic available first-order method). In the last year, we conducted numerical simulations comparing the existing algorithms with the proposed ones in the context of tree structured sparsity, where we show that the proposed solutions clearly outperform the existing benchmarks. We are planning to submit this work in upcoming conferences.

3. Next year's work program

We plan to pursue our collaboration. Through our most recent interactions, an avenue of research has emerged as particularly promising, to deal with the two main parts of the project, namely **Large-scale optimization with infinitely many constraints.** That is, we will consider a formalization of the constrained convex optimization problem by decomposing it into infinitely many simple sets by using randomization.

We plan to have further exchanges between the two teams: one student from IISc to Paris, and Francis Bach to Bangalore.

4. Record of activities

We list below the visits undertaken as part of this project (this includes activities of the earlier joint team financed by IFCAM)

	Visiting institution	Starting date	Ending date
Raman Sankaran	INRIA, Paris	May 15, 2014	July 31, 2014
Raman Sankaran	INRIA, Paris	Jan 5, 2015	Feb 5, 2015
Chiranjib Bhattacharyya	INRIA, Paris	May 1, 2015	May 22, 2015
Sesh Kumar	IISc, Bangalore	Sept 14, 2015	Oct 14, 2015
Francis Bach	IISc, Bangalore	July 1, 2016	July 7, 2016
Chiranjib Bhattacharyya	INRIA, Paris	Nov 1, 2016	Nov 11, 2016
Chiranjib Bhattacharyya	INRIA, Paris	June 30, 2017	July 10, 2017
Achintya Kundu	INRIA, Paris	Sept 6, 2017	Dec 4, 2017

5. Production

Publications and Preprints

1. Spectral Norm Regularization of Orthonormal Representations for Graph Transduction - NIPS, 2015.
2. Identifying groups of strongly correlated variables through Smoothed Ordered Weighted ℓ_1-norms - AISTATS 2017.
3. Convex optimization over intersection of simple sets: improved convergence rate guarantees via an exact penalty approach - AISTATS 2018.
4. Primal-dual Algorithms for Reweighted Least-Squares Formulations - Technical Report.