Summary of the thesis: Adaptive first-order methods for structured optimization

PhD Student: Flavia Chorobura

Supervisor: Ion Necoara

In this thesis we consider structured problems, such as composite optimization problems (i.e., minimization of a sum of two terms, one smooth and one nonsmooth) or composite inclusions (i.e., find a zero of a sum of three operators). These type of problems have applications in many fields such as signal processing, topic modelling collaborative filtering for systems recommedation, semi-supervised learning, triangulation in computer vision, among others. For such problems we design first order methods, such as coordinate descent and forward-backard-forward algorithms that use novel adaptive stepsize strategies. In more details, in most of the works from the literature on coordinate descent algorithms, a composite optimization problem is considered, where the first term is assumed to have block coordinate-wise Lipschitz gradient and the second term is separable. In our work, we considered the nonseparable case for the second term.

More explicit, in Chapters 3 and 4, the second term is assumed twice differentiable. More specifically, in Chapter 3 two algorithms were designed for solving such problems, a coordinate proximal gradient algorithm and a coordinate gradient descent algorithm. For the second algorithm, novel adaptive stepsize strategies were proposed under additional assumptions on the second term. On other hand, in Chapter 4 the first term is assumed smooth along a family of subspaces and a sthocastic proximal gradient method is designed, such that the prox of the second term is computed along a subspace. For all the algorithms proposed in Chapter 3 and 4 we derive sublinear, linear or superlinear convergences rates in the convex and nonconvex cases.

Further, in Chapter 5, the minimization of a general convex nonsmooth function is considered and a smoothing framework for the original function is introduced. We apply random (accelerated) coordinate descent algorithms for solving the corresponding smooth aproximation. Finally, a relative randomized coordinate descent algorithm for solving nonseparable minimization problems with the objective function relative smooth along coordinates with respect to a (possibly nonseparable) differentiable function is considered. For all these algorithms we derive sublinear and/or linear convergence rates in

the original function.

Moreover, in Chapter 6, we consider a sum of three operators, with nontrivial assumptions for the operators compared with the existing literature, which covers for instance optimization problems with smooth and convex functional constraints (e.g., quadratically constrained quadratic programs) or with pseudo-convex objective functions over a simple closed convex set (e.g., quadratic over linear fractional programs). A forward-backward-forward algorithm is considered, and two adaptive stepsize strategies are proposed, which require finding the root of a certain nonlinear equation. Asymptotic, sublinear and linear convergence rates are derived for this algorithm in the (uniformly) pseudo-monotone case, which also includes the monotone case.

Finally, in Chapter 7, numerical simulations are presented, which proves the efficiency of our algorithms from Chapters 3 to 6 in big data applications ranging from matrix factorization to logistic regression, support vector machine classification and minimum eigenvalue problems. The numerical tests are performed on real data and the numerical results show the superior performance of our algorithms when compared with some state of the art optimization and software.

Keywords: convex and nonconvex optimization, nonsmooth and nonseparable objective function, composite minimization, growth condition, coordinate descent, pseudo-monotone operators, forward-backward-forward splitting, adaptive stepsize, convergence analysis, big data applications