

Télécom Paris is looking for a PhD student in optimization

October 21, 2020

Context

Optimising optimisation algorithms is a natural concern of the community. The problem is as follows: for a given function class, what algorithm has the best performance to minimise these functions? Conventionally, we try to bound the number of iterations required to get an approximated solution as a function of the desired accuracy. This provides a guarantee (called upper bound) on the worst case possible among the class. On the other hand, we are also interested in the minimum number of iterations needed to minimize the most recalcitrant functions of the class. This number is called a lower bound. When the upper and lower bounds are equal (or more generally of the same order of magnitude), we say that the algorithm found is optimal. This type of consideration was the basis of the research effort on the accelerated gradient method and on the method of interior points.

In this project, we are particularly interested in the problem of a optimising a convex function under affine constraints. This problem is found in many fields (statistical learning, signal processing, shape optimization, operational research. . .) and corresponds to the prototype of the problem requiring operator splitting. Proximal methods, such as the inexact augmented Lagrangian, the primal-dual hybrid gradient and the ADMM, are commonly used to solve such problems because they solve large problems in a reasonable time.

It has been shown that if we compute the average of the iterates that they produce, these methods reach the optimal convergence rate $O(1 / k)$ on the class of convex functions f . This is remarkable but does not necessarily correlate with the user experience. In practice, we will instead favour the final iterate to the average of iterates and we will observe a linear convergence, ie an exponential decay of the measure of optimality. This gap between theory and practice happens because the user does not try to solve the most difficult problem among the convex problems under affine constraints. To obtain finer results, the class of functions studied must be restricted.

For example, if the Lagrangian is strongly convex-concave, then Chambolle and Pock proposed a primal-dual algorithm with linear convergence. Similarly, if f is quadratic, then the exact augmented Lagrangian method converges linearly.

These hypotheses are restrictive: for example, the Lagrangian of a constrained optimization problem is never strongly convex-concave. Nevertheless, the metric sub-regularity of the generalized gradient of the Lagrangian is enough to have the linear convergence. This hypothesis covers the two previously cited cases. Moreover, in the case where there is no constraint, it is equivalent to the quadratic error bound property.

Main conjecture of the project

Let's start by describing the situation in the unconstrained case $\min_x f(x)$ with f convex, which is better understood. We say that f satisfies the quadratic error bound, if there exists $\mu > 0$ such that $f(x) - \min f \geq \mu \text{dist}(x, \arg \min f)^2$. Let us denote this function class F_μ . The gradient method converges at the speed $O((1 - \mu)^k)$ on F_μ and at the speed $O(1/k)$ on F_0 . As it is the same algorithm whatever the value of μ , the gradient method is adaptive with respect to this parameter.

Moreover, the restarted accelerated gradient method converges at the speed $O((1 - \sqrt{\mu})^k)$, which is optimal. The restart frequency depends on μ and for $\mu = 0$, we obtain the accelerated gradient method in $O(1/k^2)$, which is also optimal. At the cost of a logarithmic term, we can implement a method that automatically adjusts the restart frequency to μ and is at the same time optimal on all classes F .

On the side of the convex optimization under affine constraints, if we have the metric sub-regularity with parameter $m > 0$, which means that the gradient G of the Lagrangian satisfies $\text{dist}(0, G(z)) \geq m \text{dist}(z, G^{-1}(0))$, the hybrid primal-dual gradient (PDHG) converges at the speed $O((1 - m^2)^k)$. If $m = 0$, we have a convergence in $O(1/\sqrt{k})$. Here too, the algorithm is written in the same way regardless of the value of m . As explained previously, in the case $m = 0$, averaging the iterates of PDHG gives an optimal algorithm for $m = 0$ with the $O(1/k)$ complexity, but deteriorates the performances when $m > 0$.

The conjecture we would like to solve is based on an analogy with the case without constraint.

1. There exists an algorithm A_m that solves convex optimization with constraints at the speed $O(1 - m)^k$ under the assumption of metric sub-regularity with parameter $m > 0$ of the generalised gradient of the Lagrangian.
2. This speed is optimal for this class of functions.
3. Moreover, the limiting algorithm A_0 has the speed $O(1/k)$ for convex functions.

The main contribution is to replace m^2 with m in the convergence rate. This corresponds paradigm shift in the speed of resolution of the problem and therefore a major advance in the field.

Applicant's profile

Master's degree in mathematics or data science. Good background in optimization.

Languages: French or English

Laboratory

The work will be done under the supervision of Olivier Fercoq within the Laboratoire Traitement et Communication de l'Information of Télécom Paris.

How to apply?

Send to olivier.fercoq@telecom-paris.fr your CV, cover letter and transcript of grades of Master's studies.