

... - 8 September

Time is going fast and the last month of the second year of my PhD is already going. It is sad, that M2 defences are exactly in this moment and we have so few time to discuss my results... But I remember, that the last year of my PhD is my work by myself so let me start a cycle made of 4 last reports.

**Once upon ...** Frankly speaking, a lot of things happened since I wrote my last report. I've visited a couple of conferences and met some people, discussions with whom I found very interesting. But I ain't going to write it here because it is not so interesting to re-read after.

In addition, there were few meetings with Massih about multi-class classification algorithm but in the end of a day we decided to postpone it. The main idea was the following. Considering an algorithm from [here](#) we want to make a distributed version of it. For now on it is not clear (at least for me) what exactly we want to distribute and how. The trivial distributing of some *for loop* seems useless in terms of math for me and I don't want to be a part of it. Anyway, for now this topic is away from my priority list.

Finally, all the other projects that I have/had are better to be omitted.

**Experiments on SPY-DR without reconditioning.** First, we were thinking about submitting just experimental article on AAAI conference with experiments made on different datasets with SPY-DR Algorithm as the main contribution. In the end of a day I have a lot of experiments, also some, that diverge (so I ain't as stupid idiot as I though before, who couldn't prove the algorithm). At the same time, this experiments don't give any intuition about the way of probability selection. Of course there is a trade-off between *time to identify* and *small size of communications*. In addition, sometimes full algorithm converges slower in terms of iterations even. The only explanation, that I have in mind is in the delays, that bring some additional uncertainty to our algorithm. (May be some additional experiments have to be done to check this theory and may be add it to the article.)

**Meeting 28.08.** During the meeting we discussed all the projects that I have and decided to focus on the main one about SPY-DR. The main conclusion, that Jerome has already written in our SLACK channel is the following.

- Paper (aim to AISTATS)
  - ✓ plug intro in draft + update refs,
  - ✗ write proof of convergence,
  - ✗ plug the proof of improvement result in the draft (as appendix),
  - ✗ add in appendix a short conclusion on numerical exps with SPY-DR showing divergence, check out the convex case: same conv with different rate ( $\mu=0$ ) + new improvement results;
- Experiments
  - ✗ relevant exps are comparaison DAve vs ASPY-DR (vs ASPY) + if possible large-scale data sets (Massih),
  - ✓ write the fucking code and then test it on a small example.

**Discussion about results.** Let me first talk about experimental part. I spent all the end of the week, weekend and Monday as well on the writing part of the task. And, to be honest, I finished writing part of *fucking code* but as far as it is **fucking** it does not work well. I decided to postpone the experimental part for now and concentrate on the theoretical and writing part of article to clarify for myself at least all the bottlenecks in it.

No I am ready to present my results in the writing part.

First, I moved *Introduction, Related works and SPY Chapters* in the new L<sup>A</sup>T<sub>E</sub>X document. In the same time I adapted the proof of “convergence” for SPY for the case of different probabilities to show the problem that comes up with this approach<sup>1</sup>. In addition I’ve added some additional paragraphs in *Introduction* but this work is not finished yet. Finally, I’ve added all the theory about ASPY-DR without proofs to have a clear view on the shape of the article but I suddenly stopped at the moment, when *Identification result* should be presented. I realized, that I haven’t got a clear understanding of proof and decided to concentrate on the supplementary and more precisely on proofs.

The proof of SPY was not so hard to rewrite and there was no underwater rocks there, but in contrast I had a lot of problems with an identification part and Catalyst part itself as far as we’ve decided with Franck to present it in general way from the beginning<sup>2</sup>. Skipped the part with Catalyst itself I started proving the almost surely convergence of algorithm and realized also, that it is not so trivial. Frankly speaking it is trivial, but in the beginning I was completely lost. In the end of the day the proof was added to the appendix.tex file under the assumption that we have a proof of Catalyst.

Now let me write about the problem that I have with a proof of Catalyst. I realized that in fact, the Theorem III.3 from [here](#) does not suit us fully. In the same time as the proof of Catalyst as well. More precisely, our version of Catalyst is FGM with different functions  $f_k$ <sup>3</sup> that all have the same minimizer  $x^*$  with inexact gradient<sup>4</sup>. The basic Catalyst proof works with inexactness and Theorem III.3 works with different  $f_k$ . So now my sub-problem that I’m solving is in combining this two proofs to have a proof of Catalyst.

**Some other ideas.** Let us recall, why we need Catalyst. We decided to use it not because of acceleration, that it brings, but only because of reconditioning that comes together with acceleration. But this is not the only way to recondition the problem. Some other possibilities are:

- use GD with Moreau-Yosida envelope (could be added as another example as far as it is much easier to analyze than FGM);
- use  $\ell_2$  regularization to recondition the problem (the big drawback is in maximal possible accuracy that becomes bigger if hyper parameter is bigger)<sup>5</sup>

In addition, I plan to test Catalyst with some small and stupid function inside to see whether o have a mistake in catalyst or in it’s asynchronous implementation.

---

<sup>1</sup>I think that there is no reason to copy-paste all the proofs, formulas and so on here, as far as they have been already written in overleaf document.

<sup>2</sup>with different  $\kappa$ ’s

<sup>3</sup>thanks to different  $\kappa$ ’s

<sup>4</sup>thanks to the inexact Moreau-Yosida envelope in Catalyst

<sup>5</sup>may be I’m totally wrong with it, sorry then(