Report on the paper

Retraction-based first-order feasible methods for difference-of-convex programs with smooth inequality and simple geometric constraints

by *Yongle Zhang, Guoyin Li & Ting Kei Pong*

In this manuscript, the authors propose a first-order feasible method for DC programs constrained by smooth inequalities and an easy-to-project convex feasible set. The convergence analysis is made through the assumption that the smooth constraints are convex, and later extended to the case of one nonconvex constraint. The new scheme is illustrated by some numerical experiments on group LASSO and compressed sensing.

The paper is very well written and it is not hard to follow . The results seem to be correct assuming the validity of those in [43], since most of the arguments follow from that manuscript. Fortunately, [43] is already published on *SIAM J. Optim.*.

In what follows I include a list of comments the author should tackle for the revised version.

1. P3, L42: The closed ball with respect to which norm?
2. P4, L48-49: Please, give a reference for this assertion.
3. P5, Algorithm 1: It is not clear to me what is the role of parameter $\underline{\beta}$. Is it only used as a lower bound for the initial β_k^0? If so, how can you assert that $\inf_k \beta \overline{k} \geq \underline{\beta}$ in Remark 3.1? The same applies to Algorithm 2 and Remark 4.1.
4. P7, L23: I cannot see how [38, Th 2.6] is applied (same doubt in Theorem 4.1, P16 L9). Are you assuming full domain of functions P_1 and P_2? If so, make it more explicit.
5. P8, L43: Why is $\{\xi^k\}$ bounded? Maybe this it related with the full domain asked in item 4. The same for $\{\partial P_1(u^{k_j})\}$.
6. P10, L10: How does one get the term $+\frac{1}{\beta_k}\|u^k - x^k\|^2$? It is supposed that the authors are using the convexity of P_1, but it is not strongly convex.
7. P16, L29: Which is the closed form formula for $\widetilde{\tau}$?
8. P23, L44: How is such τ found?
9. Numerical results: As far as I see, Algorithm 3 does not require to find a completely feasible initial point ($x_0 \in F$) but just in C ($x_0 \in C$). I did not find any statement about how the initial point is chosen for $\mathrm{ESQM}_{\mathrm{ls}}$. In principle, $SPGL1$ and Slater point are not needed. Therefore, it is not fair to say that FPA is faster since the sum of the iteration time plus the required initialization exceeds the CPU time of $\mathrm{ESQM}_{\mathrm{ls}}$. Would it be any other cheaper way to compute x_0?
10. P26: In table 2, the time of SPGL1 for $i = 4$ is greater than for $i = 6$. Does it make any sense?
11. References: It seems that there exists a more recent version (2016) of [38]. Update [43] and all the referenced results from there to its published version (*SIAM J. Optim., 31(3), 2024–2054.*).