Report on "The Complexity of Deterministic Guaranteed Automatic Algorithms: Cones, Not Balls" by Y. Ding et al.

The paper provides results on the information complexity of problems in the worst case setting, where the term "worst case" has to be somehow adapted to the setting of the paper. This comes from the unboundedness of the considered function classes. I think the results are interesting and the considered problem is reasonable, thus the paper is a recommendable contribution to the field of Information-based Complexity. In particular, it is my opinion that the idea behind, i.e. considering cones as classes of input functions instead of balls and using this additional knowledge for the construction of adaptive algorithms, is excellent.

However, due to the huge amount of details and definitions, that are not really needed for the understanding of the (theoretical) results, I suggest a major revision of the first three sections.

To be more precise, the paper considers problems defined by solution operators $S\colon \mathcal{F} \to \mathcal{H}$ for some Banach spaces $\mathcal{F}$ and $\mathcal{H}$, which satisfy $S(cf) = cS(f)$, $c \geq 0$. The corresponding (semi-)norms are given by $|\cdot|_{\mathcal{F}}$ resp. $\|\cdot\|_{\mathcal{H}}$. (Please clearify whether $\mathcal{F}$ is a Banach space *or* $|\cdot|_{\mathcal{F}}$ is a semi-norm!) The goal is to find an algorithm, say A, that uses only finitely many pieces of information of the input, given by linear functionals, and has a preferably small error

$$e(A, f) := \|S(f) - A(f)\|_{\mathcal{H}}.$$

The number of information functionals that are used by the algorithm A for the input $f \in \mathcal{F}$, i.e. $\mathrm{cost}(A, f)$, is called the cost of A for f. Clearly, for non-adaptive algorithms $\mathrm{cost}(A, f)$ is independent of f and it is not possible (except trivial examples) to give bounds on $e(A, f)$ without additional knowledge on f.

The focus of the paper is on adaptive algorithms. In this case it is at least not impossible to find a set $\mathcal{N} \subset \mathcal{F}$ and algorithms A_ε, $\varepsilon > 0$, that satisfy $e(A_\varepsilon, f) \leq \varepsilon$ for all $f \in \mathcal{N}$, but, of course, now the cost of $A_\varepsilon(f)$ depends on f.

Let $\mathcal{G} \supset \mathcal{F}$ be another semi-normed vector space with a weaker semi-norm $|\cdot|_{\mathcal{G}}$ and define, for a given $\tau > 0$, the cone of functions

$$\mathcal{C}_\tau = \{f \in \mathcal{F} : |f|_{\mathcal{F}} \leq \tau |f|_{\mathcal{G}}\}.$$

The main contribution of the paper is that, if one considers only functions from the class $\mathcal{C}_\tau$, there exist such algorithms A_ε such that

$$\sup_{\varepsilon>0} \sup_{f\in\mathcal{C}_\tau} \frac{\mathrm{cost}(A_\varepsilon, f)}{\mathrm{comp}(\varepsilon/|f|_{\mathcal{G}})} < \infty,$$

where $\text{comp}(\varepsilon)$ denotes the minimal cost needed by a non-adaptive algorithm to obtain an error ε for all functions from the unit ball of $\mathcal{F}$. (If I understood the results correctly.) This means that the (ε- and $|f|_{\mathcal{G}}$-dependencies of the) presented algorithms are not much worse than the optimal ones, which seems to be a very nice result.
Finally, the authors prove lower bounds on the complexity of algorithms of "their type", show a kind of (almost) optimality of the presented algorithms and conclude with 2 examples. All this is reasonable and interesting.

I now turn to specific criticisms on the paper. These should be read as suggestions for improvement of the readability, especially for the readership of the Journal of Complexity. Note that the points below concern mostly the Sections 1, 2 and 3. Sections 4–8 are well structured and clearly written.

Major Critism:

1. Please avoid any appearance of cost budget $N_{\max}$, warning flag W and success $\text{succ}(A, \ldots)$. These definitions are only for implementation issues. "Users" should know that an algorithm that is given by an iterative procedure with stopping rule (as the algorithms of the paper), which is not terminated if the cost budget is reached, cannot guarantee an error bound.

2. Concerning the table on page 2:
The paper presents "guaranteed" algorithms, which are necessarily "adaptive" for the considered (unbounded) classes of functions. Also "optimal" is a desirable property. In contrast, "tunable" seems to be superfluous and I don't know what "automatic" means. Why is an algorithm, that needs the width of a cone, "automatic" while algorithms, that use the radius of a ball, are not?

3. The presented results rely on numerous assumptions which are spread over the whole article. In particular, it is assumed that
   - there is a non-adaptive algorithm G_n for approximating $|f|_{\mathcal{G}}$ with explicit one-sided worst case error bounds $h_-(n)$ and $h_+(n)$ for the unit ball in $\mathcal{F}$,
   - there is an algorithm A_n for approximating S with explicit worst case error bounds $h(n)$ resp. $\tilde{h}(n)$ for the unit ball in $\mathcal{F}$ resp. $\mathcal{G}$ and
   - (For Algorithm 2:) there exists a sequence $\mathcal{I} = \{n_1, n_2, \ldots\} \subset \mathbb{N}$ with $n_{i+1} \le r n_i$ for some $r < \infty$ such that A_n and G_n, $n \in \mathcal{I}$, use the same information and A_{n_i} resp. G_{n_i} is "embedded" in $A_{n_{i+1}}$ resp. $G_{n_{i+1}}$ (i.e. the information is *nested*).

   These general assumptions should be stated earlier in the paper.

4. The final formulas for $\mathcal{N}$ in Theorems 1, 2, 5 and 7 are only correct under the assumption that $\widetilde{N}_{\max} \in \mathcal{I}$ resp. $N_{\max}/r \in \mathcal{I}$. Otherwise, the algorithm $A_{\widetilde{N}_{\max}}$ resp. $A_{N_{\max}/r}$ is possibly never calculated in the procedure.

Further Comments:

- There is a wrong equality in the proof of Thm. 1 and in the proof of Thm. 2 you call $h(n)$ etc. wrongly non-decreasing.
- I wouldn't call the presented algorithms "optimal", since they are only assumed to have the correct order of convergence and the correct scaling in $|f|_{\mathcal{G}}$. Maybe consider a change to "almost optimal" or the like.
- Please add the proof of the last statements in Thm. 5 and 7.
- What is $\mathcal{I}$ in Section 5?
- The set $\mathcal{I}$ in Section 6 does not satisfy the property for nested information.
- p. 23, line -1: $f^{(k)}$ instead of f^k
- p. 24, line 1: you should say what $\|\cdot\|_1$ is
- p. 24, line 6: "consider algorithms consider algorithms"
- p. 24, eq. (40): Is the formula correct?
- p. 24, eq. ($41a$): Write $x \in [x_i, x_{i+1}]$. Isn't there a wrong $(n-1)$?
- p. 27, first equation: $\|f_0'\|_1 = 1$
- I do not comment on typos/inaccuracies in Sections 1–3 since there is anyhow a revision needed.