**Reviewer #1:**

We did our best to explain the lower bound construction in the most pedagogically principled way. In fact we also tried to create a 2-dimensional picture to illustrate our construction, however decided against it as the phenomenon at play seems to be inherently high-dimensional and low-dimensional pictures are misleading (in particular it is not possible to represent the "valley" to the optimum in low dimensions). If you have other suggestions to clarify the construction we would be happy to hear them!

We also would like to point out that, while you are certainly correct that there are only few papers on lower bounds for parallel non-smooth optimization, it might be misguided to characterize the problem as "not very well-studied". There are thousands of papers on variants of this topic, and arguably the reason for the paper shortage on the most canonical variant (the one studied in this paper) is that progress was hard to come by after Nemirovski's seminal contributions from the 90s. We believe that our new ideas will be of value to both the machine learning community and to the broader optimization community, despite the fact that the current algorithm is not practical.

**Reviewer #2:**

Thank you for the kind words!

**Reviewer #3:**

Thank you for your positive feedback! Here is a detailed answer to your questions:

1. Thank you for the suggestion, we will consider re-organizing the material (it is a challenge to make this fit in 8 pages!).
2. This is in fact a great open problem! As far as we know, for the tiny epsilon regime (say $1/\mathrm{poly}(d)$), the best lower bound used to be Nemirovski's $d^{1/3}$, which is now $d^{1/2}$ thanks to the wall function argument of our paper. We will clarify this point in the paper.
3. Agreed, thanks!
4. The intuition is actually quite simple: line 139 on page 4 shows that the lower bound construction would imply the $d^{1/3}/\epsilon^{2/3}$ if one could take delta (the radius outside of which the "walling" starts) to be epsilon. This is technically difficult because for such some delta, our proposed wall function will "leak information" between dimensions. However it seems believable that there could be a more complicated wall function that could work at such tiny walling radii. We will add a comment along those lines in the paper.
5. You are correct that a Gaussian kernel in lieu of $\chi$ could possibly work, however as far as we can see this would make the proof longer. Our current proof crucially relies on the fact that the tail of $\chi$ is zero, which wouldn't be true with a Gaussian kernel.