

1 Response to reviewers concerning the manuscript

2 #10658: Finer Metagenomic Reconstruction via Biodiversity Optimization.

3 Thank you to the reviewers for their thorough evaluation of this submission. One reviewer stood out with a "reject"
4 decision which does not seem justified to us, given that the work clearly seems to have sparked the reviewer's interest.
5 The work is innovative (as acknowledged by all reviewers) and deserves fast publication in a venue with an appreciative
6 audience. NeurIPS seemed to us to be this venue, and the quality of the reviews confirmed our impression. As the
7 reviewer puts it, "The main reason for the low score is that the paper is not self-contained, and tries to cover too much
8 ground, losing focus". This may be a fair objection, but to us it also contradicts the recommendation to also discuss
9 "other notions of diversity", "related sparse formulations", and "other sparse solvers". We will try to apply "significant
10 reorganization and compression", so long as the 8-page limit allows it.

11 Besides some restructuring (which was also advised by another reviewer) and typo fixing, another concern that was
12 raised more than once was the performance of the method on real-life metagenomes when sequencing noise is present.
13 This is actually some work in progress, with modifications (with a "group-lasso like formulation") required to the
14 optimization algorithm in order to preserve accuracy and speed. Producing "a software implementation [which] would
15 be highly beneficial to biologists" is indeed our next goal. An initial implementation is indeed provided in the GitHub
16 repository linked to in Section 6, line 247 (redacted per NeurIPS requirements). But all of this goes beyond the scope of
17 this paper, which is more theoretical and focuses on presenting in the simplest setting the novel idea of minimizing
18 biodiversity for metagenomic purposes, proving results about it, and numerically demonstrating the superiority of this
19 approach.

20 Finally, here are some responses to other selected comments:

- 21 • the broader impact statement will be expanded;
- 22 • the reason why plain ℓ_1 -minimization is excluded is that (in the noiseless setting) its success has actually
23 nothing to do with the objective function, only with the constraints due to the reconstruction of probability vectors;
- 24 • "figure 1, why random similarity performs better than the identity matrix? And why does quikr (which assumes
25 identity similarity) performs better than identity?": We also found the superior performance of a random
26 similarity matrix over an identity similarity matrix surprising. We posit that an identity matrix is so biologically
27 unrealistic (given the similarity of organism 16s rRNA sequences) that even considering a random similarity
28 matrix will lead to better results. For the second part of the question, the Quikr algorithm only corresponds to
29 (MinDiv) when $q = 1$, $\mathbf{Z} = \mathbf{I}$, and $\lambda \rightarrow \infty$ (lines 241-242). In Figure 1, $\lambda = 10,000$ and $q = 0.01$, hence the
30 performance difference. We neglected to mention q was set to 0.01 in this section (but did in all others) so will
31 clarify this oversight in revision;
- 32 • "Are there other applications beyond metagenomics?" we do not have another example at the moment, but we
33 share the reviewer's opinion that "This sounds like a very general formulation which should have applications
34 to diverse fields."
- 35 • "You discuss concavity of $D_{z,q}$ in some length. Is it central to the discussion – how is it used?": As mentioned
36 in lines 153-159, the purpose of discussing concavity is to demonstrate/prove that this is a challenging,
37 NP-hard optimization problem in general, convergence to global minimizers of traditional approaches are not
38 guaranteed (lines 167-172), and hence emphasize the surprising result that in certain situations the problem
39 reduces to a simple linear program (line 209). In addition, concavity is used to demonstrate that (MinDiv)
40 always has a sparse minimizer (Proposition 3), as well as to provide a lower bound in Proposition 6.
- 41 • "Can the author(s) comment on the bound of sparsity in line 86 and proposition 3. The bound seems to be
42 large and can be even larger than the dimension of \mathbf{x} . Does it really imply sparse solutions?": The bound
43 mentioned in line 86 (Proposition 6) does not comment on the sparsity of minimizers to (MinDiv) but rather
44 on the value of the objective function in a certain case. However, Proposition 3 does indeed guarantee the
45 existence of an m -sparse minimizer where m is the number of rows in the matrix \mathbf{A} . Hence why in the start of
46 Section 2, we frame the whole discussion where $m \ll N$.