

1 We thank the anonymous reviewers for their feedback, which we will incorporate into future drafts. We are grateful for
2 the many typographical corrections and suggested edits. As to the main concerns raised by the reviewers, in the interest
3 of brevity we have paraphrased them below in bold, and provided itemized responses.

- 4 • **A similar figure to Figure 1 should be in the introduction, to better explain identifiability (Reviewer #1)**

5 We agree and will add this in the final version.

- 6 • **The algorithm(s) should be broken out into a clear algorithm box for ease of parsing / implementation.
7 (Reviewer #1, #3)**

8 We agree and will add this in the final version.

- 9 • **It would be useful to provide a Python implementation. (Reviewer #1, #2)**

10 For reproducibility, implementation of all models compared in this paper is included with our submission. We
11 are also planning to publish a pip-installable package for box embeddings which will include GumbelBox,
12 SmoothBox, and the original Box Embeddings.

- 13 • **Do we use comparable amounts of parameters for vector vs. box-based models? (Reviewer #1, #2)**

14 For fair comparison we have used the same number of parameters for all models in experiments 3 and 4, (Table
15 2 & 3). Note: Gumbel box only has two parameters per entity per dimension, not three - the scale parameter is
16 a global hyperparameter. The experimental code included with the submission contains all hyperparameter
17 settings used.

- 18 • **GaussianBox section is confusing to include prominently (Reviewer #1)**

19 We thank the reviewer for suggesting these required re-structures to improve clarity. We agree with all of it,
20 and will definitely incorporate them on the final version of the paper.

- 21 • **What properties does GumbelBox have compared to SmoothBox from Li et al.? (Reviewer #3)**

22 Gumbel box (with the softplus approximation) can actually be interpreted as a generalization of SmoothBox
23 with a nonzero temperature on the intersection operation. This is mentioned in the appendix, but we will make
24 this more clear in the camera ready.

25 Adding this nonzero temperature improves learning over SmoothBox in a number of settings (ie. the cases
26 mentioned in the synthetic experiments) by improving local identifiability. Even in scenarios without identifi-
27 ability issues, GumbelBox has denser gradients, since all parameters of the boxes being intersected contribute
28 to the intersection. This is always beneficial, but particularly so for cases where multiple boxes are intersected,
29 a point we can highlight more in the final version.

- 30 • **How does wall-clock time compare to the smoothed box model and other embeddings? (Reviewer #3)**

31 Since we compute the expected intersection / volume in closed form rather than by sampling, the overall
32 training time is not appreciably slower. For example, in case of WordNet (Table-2), 150 epochs take around 35
33 minutes for SmoothBox and 41 minutes for GumbelBox with the same hyperparameters (batch size of 256,
34 embedding dimension of 5, and number of negative samples 16) on GeForce RTX 2080 Ti (averaged over 20
35 runs).

- 36 • **Include a conclusion section (Reviewer #4)**

37 We will include a conclusion and future work section in the final version.

- 38 • **Approximations on lines 179-185 are discussed too vaguely, please elaborate on possible failure modes.
39 (Reviewer #1)**

40 We agree that, even though the empirical results are strong, these approximations could use a bit more
41 discussion. We will add a discussion and visualize graphically the impact of these approximations and edge
42 cases where they could be inaccurate.

- 43 • **Make the statements regarding local identifiability clearer and more rigorous. (Reviewer #3)**

44 While the local identifiability improvements are discussed in the paper, Figures 4 and 5 in the appendix give
45 more intuition to why the GumbelBox model has improved local identifiability compared to SoftBox. We will
46 move these figures to the main body of the work, which may help clarify the argument.

47 The paper also identifies certain specific classes of parameters for which GumbelBox is locally identifiable
48 and SmoothBox is not, which we demonstrate in the initial synthetic experiments, however we agree that more
49 formal statements can be made. For the camera ready we can include a rigorous characterization of these
50 classes and proof of this fact.

51 We can also prove that in 1-dimension, or when the per-dimension ratios are fixed, the GumbelBox model is
52 entirely locally identifiable. (Note: all local identifiability is modulo a global translation of the parameters.)