

1 C Further Experiments and Experimental Details

2 C.3 Experiments with Locked Representation (fixed labeling of the figure in appendix C.3 of 3 the main paper)

4 We also note that we did one more comparison, which is not reported in the main paper. Namely:

- 5 • *Trainable Temperature and Relative Bias with Explicit Adapter.* We initialize at $t = 10 =$
6 $e^{t'}, b = 0, \delta = \frac{e^x}{1+e^x}$ with $x = 1/2$ and run Adam on $\{V_i\}_{i=1}^N, t', b_{\text{rel}}, x$ for the loss
7 $\mathcal{L}^{\text{RB-Sig}}(\{A_{\text{locked}}^\delta(U_i)\}_{i=1}^N, \{A_{\text{trainable}}^\delta(V_i)\}_{i=1}^N; e^{t'}, b_{\text{rel}})$ and initial learning rate 0.01. Since
8 the adapter is an invertible transformation on the representations, we reported the inner
9 products both with the adapter and without it (that is, we invert by removing the last
10 coordinate and dividing by δ .)

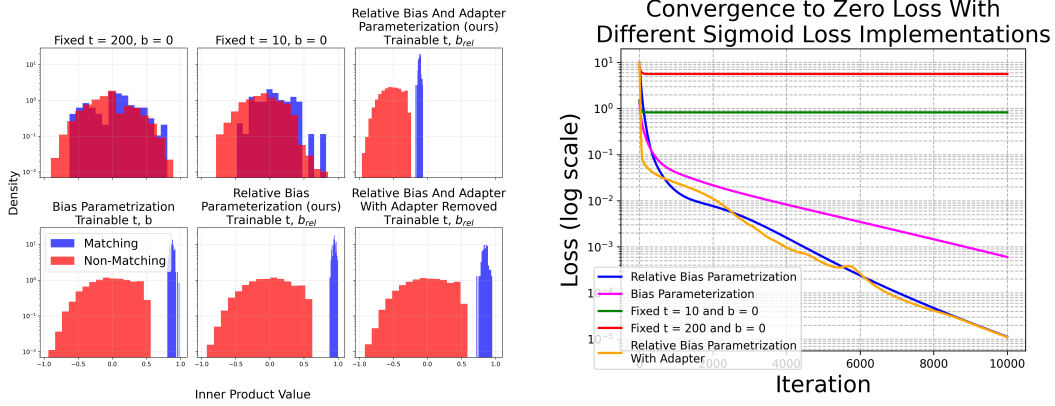


Figure 1: Inner-product separation and loss convergence under six sigmoid-loss parameterizations. *Left:* Log-density histograms of inner-product scores for matching (blue) versus non-matching (red) pairs, evaluated under fixed inverse temperature $t = 200, b = 0$, fixed $t = 10, b = 0$, trainable bias b , our relative-bias parameterization (trainable b_{rel}), and the same two schemes with the adapter removed; only the trainable-bias models show clear separation. *Right:* Sigmoid-loss trajectories (log scale) over 10,000 iterations for the same six settings; only those variants that learn both bias and inverse temperature reach zero loss, and our relative-bias parameterization (with and without adapter) converges most rapidly.

11 We can overall see that the performance of $\mathcal{L}^{\text{RB-Sig}}$ algorithm with an adapter and without is rather
12 comparable and the inner product separations are similar. One difference to note is that the training
13 with adapter seems less stable. Thus, we believe that in practice not using the adapter might be the
14 better approach.