Appendix

A. Experiment

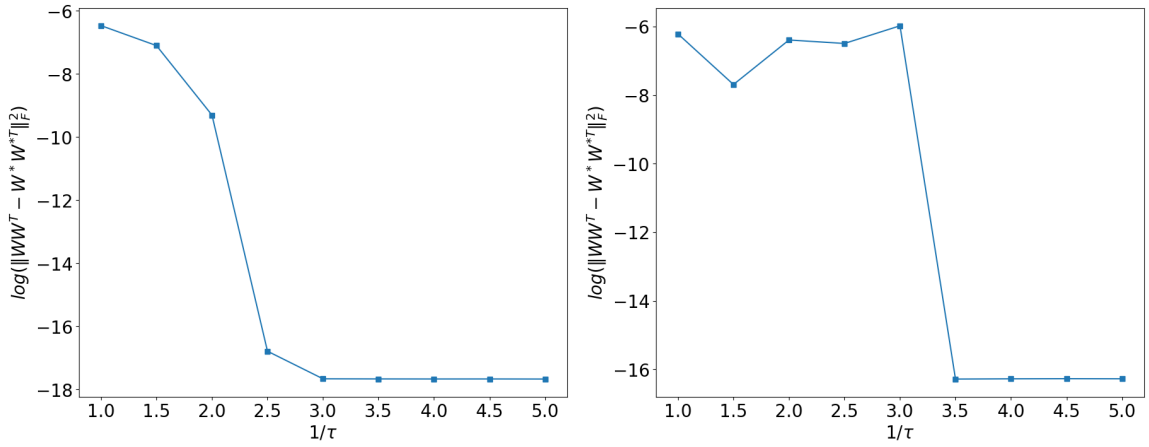
A.1. CE loss Solution Converges to HardMax Solution

We have numerically tested Lemma 3.1 in numerous cases and observed the convergence of the solution. Specifically, the numerical solution of the CE loss with a decreasing temperature τ converges to the solution of the hardmax loss.

Suppose $\boldsymbol{W}$ is the optimal solution for minimizing the CE loss and $\boldsymbol{W}^*$ is the optimal solution for minimizing the HardMax loss. We initially use Procrustes analysis to eliminate potential rotation and permutation ambiguity. We then measure their distance as follows:

$$\|\boldsymbol{W}\boldsymbol{W}^T - \boldsymbol{W}^*\boldsymbol{W}^{*T}\|_F^2 \tag{10}$$

We set the temperature $1/\tau = [1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]$. The solution is extremely close and the distance is minimal, so we use the log scale. The result is depicted in 6.



(a) $d = 5, K = 16$ (b) $d = 6, K = 27$

Figure 6. Optimal solution of CE loss with varying temperature converge to Hardmax solution for $d = 5, K = 16$(left) and $d = 6, K = 27$(right).