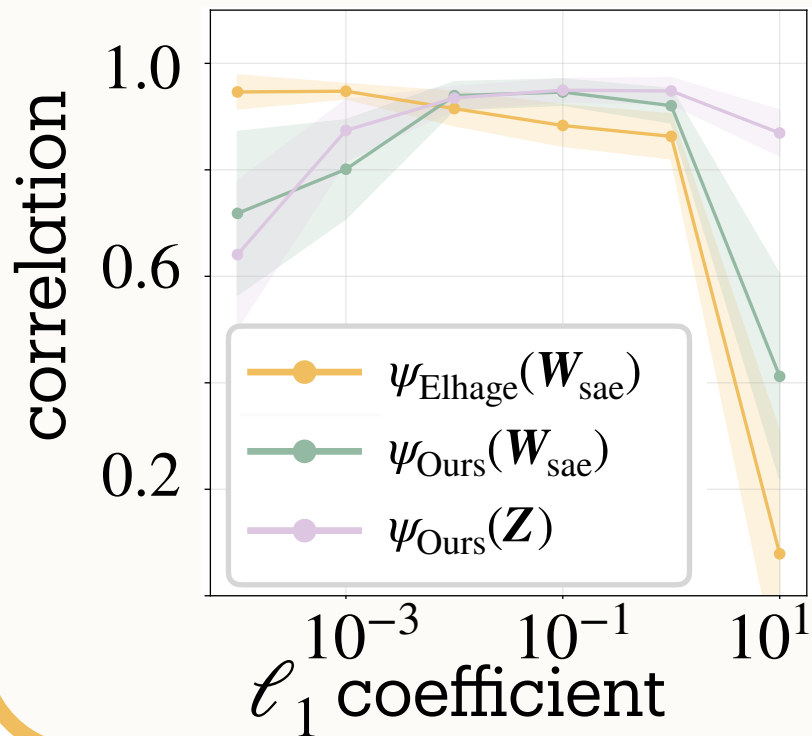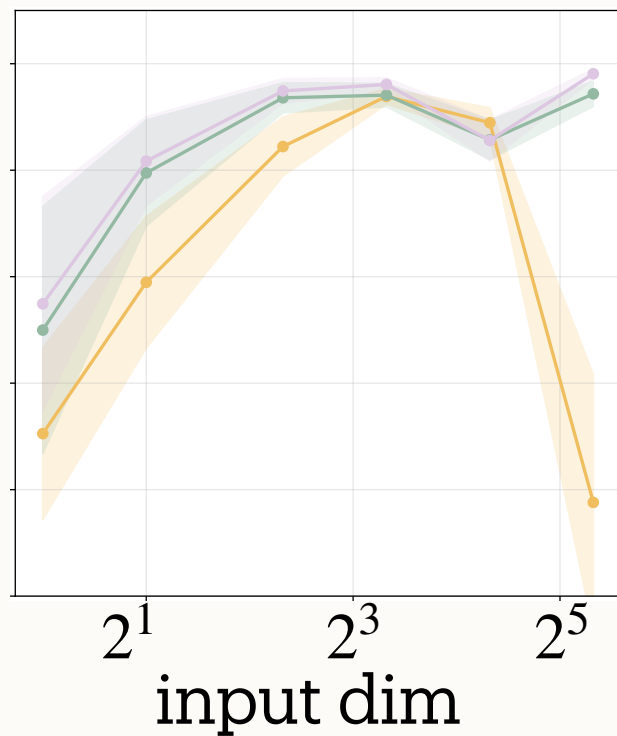ℓ_1 regularization | scaling model | scaling dictionary

Legend:
- $\psi_{\text{Elhage}}(\boldsymbol{W}_{\text{sae}})$
- $\psi_{\text{Ours}}(\boldsymbol{W}_{\text{sae}})$
- $\psi_{\text{Ours}}(\boldsymbol{Z})$

y-axis: correlation

x-axes: ℓ_1 coefficient, input dim, dict size / input dim