Appendix

A. Distribution of the eigenvalues of H in trained models

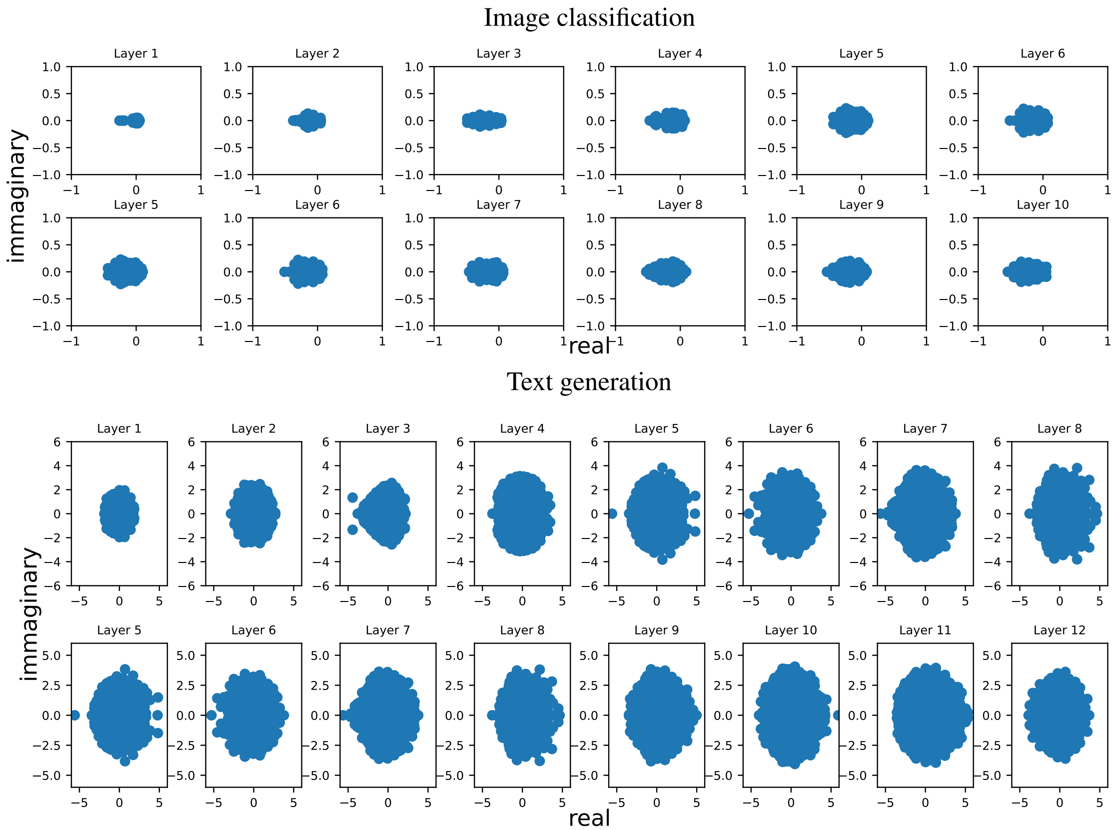

Figure 8. **Distributions of eigenvalues of H** Vision models have distributions skewing to the negatives while language models have symmetrically distributed eigenvalues.

B. Proofs

Proposition 1 ((Meyer & Stewart, 2023))**.** *Given Assumption 3.1, all eigenvalues of* $\mathbf{A}$ *lie within* $(-1, 1]$*. There is one largest eigenvalue that is equal to* 1*, with corresponding unique eigenvector* $\mathbf{1}$*. No eigenvectors of* $\mathbf{A}$ *are equal to* 0*.*

Proof. First, because $\mathbf{A}$ is positive, by the Perron-Frobenius Theorem (Meyer & Stewart, 2023) all eigenvalues of $\mathbf{A}$ are in $\mathbb{R}$ (and so there exist associated eigenvectors that are also in $\mathbb{R}$). Next, recall the definition of an eigenvalue λ and eigenvector $\mathbf{v}$: $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Let us write the equation for any row $i \in \{1, \ldots, n\}$ explicitly:

$$a_{i1}v_1 + \cdots + a_{in}v_n = \lambda v_i.$$

Further let,

$$v_{\max} := \max\{|v_1|, \ldots, |v_n|\} \tag{8}$$

Note that $v_{\max} > 0$, otherwise it is not a valid eigenvector. Further let $k_{\max}$ be the index of $\mathbf{v}$ corresponding to $v_{\max}$. Then