Supplemental Material: Geometry of Neural Network Loss Surfaces via Random Matrix Theory

1. Computing the normalized index

One way to obtain an expression for the normalized index is to rewrite eqn. (18) as $f(G) = z$ (where $f(G) = \mathcal{R}_H(G) + 1/G$), so that $G = f^{-1}(z)$. Integrating the inverse of a function requires only integration of the function itself (Laisant, 1905),

$$\int f^{-1}(z)dz = zf^{-1}(z) - F \circ f^{-1}(z) + C\,, \tag{S1}$$

where F is the antiderivative of f. This relation gives,

$$\alpha(\epsilon,\phi) = 1 - \frac{1}{\pi}\mathrm{Im}\Big[\epsilon G_H(0)^2 + \log G_H(0) - \log\big(1 - \phi G_H(0)\big)/\phi\Big]\,. \tag{S2}$$

An explicit representation of $G_H(0)$ and thus $\alpha(\epsilon,\phi)$ is possible by solving the cubic equation in eqn. (18). The full result is very long and unenlightening, but we find that for small α,

$$\alpha(\epsilon,\phi) \approx \alpha_0(\phi)\left|\frac{\epsilon-\epsilon_c}{\epsilon_c}\right|^{3/2}\,, \quad \epsilon_c = \frac{1}{16}(1 - 20\phi - 8\phi^2 + (1+8\phi)^{3/2})\,, \tag{S3}$$

where ϵ_c is the critical value of ϵ below which all critical points are minimizers.

2. On the assumption that the weights are I.I.D. random normal variables

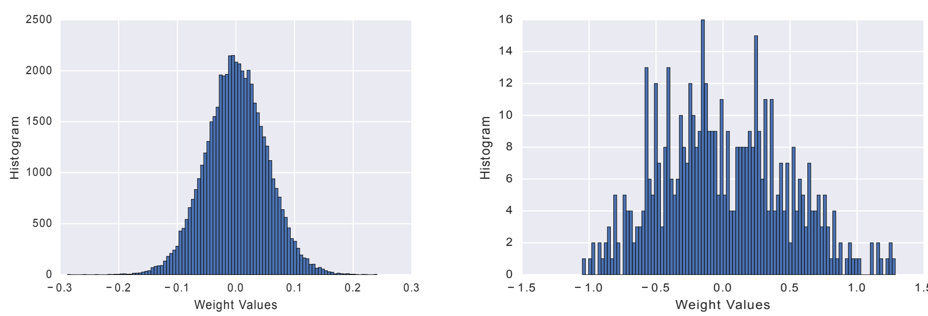

Figure S1. Histogram of weight matrix entries ($W^{(1)}$ left, $W^{(2)}$ right) after training a 50 hidden-unit single-layer ReLU network on a subset of 500 grayscaled CIFAR-10 images. Left, right histograms have a total of 51,200 and 500 entries.

3. Spectral density of H_1 for single-hidden-layer ReLU networks

From (Dupic & Castillo, 2014) and referring to eqn. (27), the density can be written as

$$\rho_{H_1}(\lambda) = \big(1 - \min\big(1, \frac{\alpha}{2}\big)\big)\delta(\lambda) + \frac{\alpha^2|\lambda|}{2\epsilon}\,\rho_c\Big(\frac{\alpha^2\lambda^2}{2\epsilon}, \frac{\alpha}{2}\Big)\,, \tag{S4}$$

where $1/\alpha = \phi/2 = n/m$,

$$\rho_c(x,\alpha) \quad = \quad \frac{\sqrt{3}}{6\pi x\sqrt[3]{2}}\big(r_+ - r_-\big)\,\mathbf{1}_{x\in[\theta(1-\alpha)x_-, x_+]}\,, \tag{S5}$$

and,

$$r_\pm \quad = \quad \sqrt[3]{9(2+\alpha)(x-\xi_0) \pm 6\sqrt{3(x-x_-)x(x_+-x)}}\,, \tag{S6}$$

$$x_\pm \quad = \quad \frac{8 + 20\alpha - \alpha^2 \pm \sqrt{\alpha}(8+\alpha)^{3/2}}{8}\,, \tag{S7}$$

$$\xi_0 \quad = \quad -\frac{2(-1+\alpha)^3}{9(2+\alpha)}\,. \tag{S8}$$

4. Free independence and the evolution of eigenvalues over training

We plot the eigenvalues of the Hessian $H_0 + H_1$ and the transformed Hessian $H_0 + QH_1Q^T$ as the parameters evolve over training. The training set is CIFAR-10 downsampled to 4x4 images, grayscaled and whitened. We train a 16-20-16 ReLU autoencoding network without biases on the first 150 images of the dataset using full-batch gradient descent with learning rate 0.05. The parameters are initialized as zero-mean Gaussians with variance 2 over the number of incoming units.

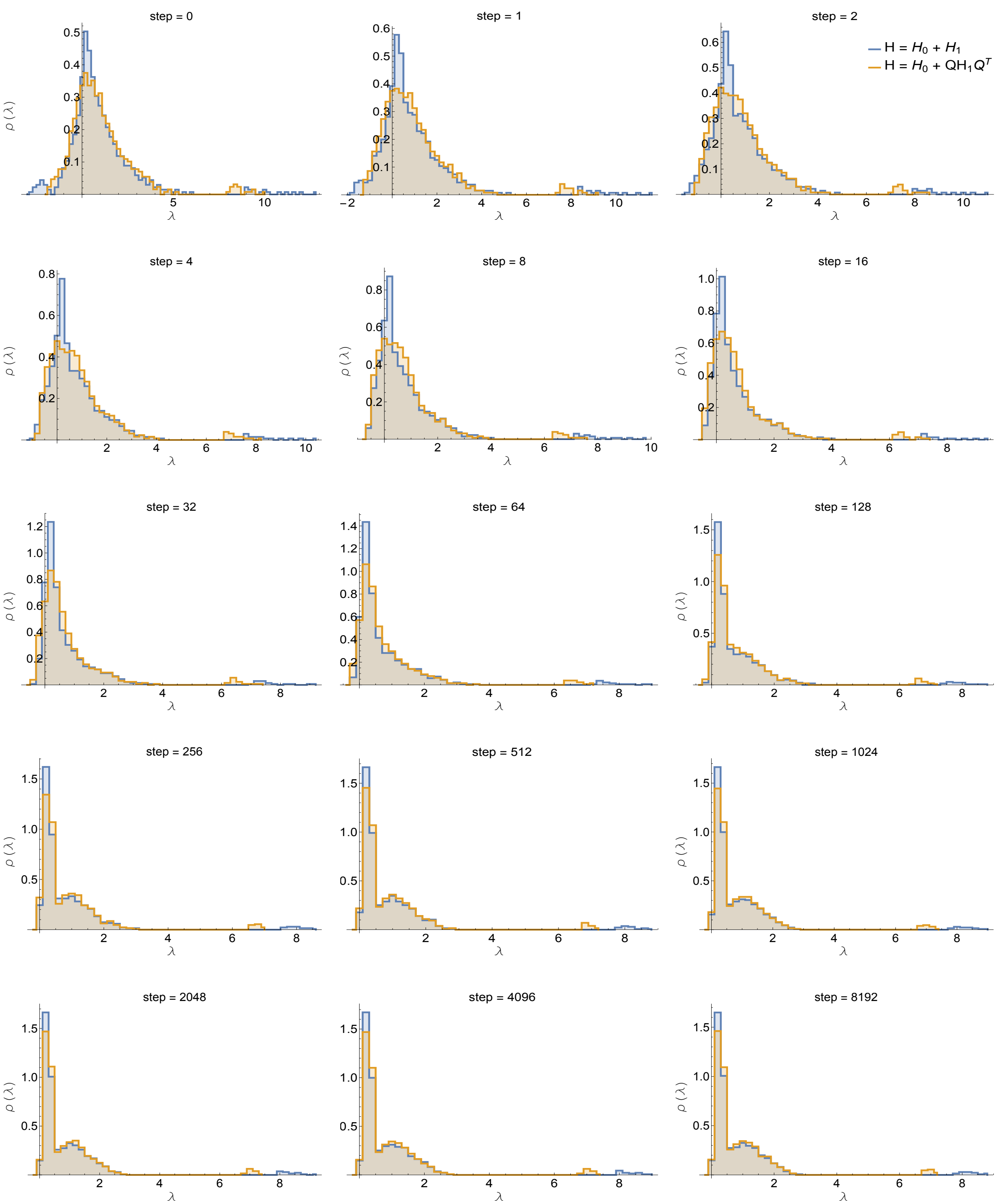

Figure S2. Evolution of the eigenvalues of the Hessian over training.