A. Additional Proofs

A.1. Lower bound on $\lambda_{\min}$

We wish to verify that the Hessian of a point returned by the gradient descent algorithm is positive definite, as well as provide a lower bound for its smallest eigenvalue, avoiding the possibility of errors due to floating-point computations.

Since the Hessians we encounter have relatively small entries and are well-conditioned, it turns out that computing the spectral decomposition in floating-point arithmetic provides a very good approximation of the true spectrum of the matrix. Therefore, instead of performing spectral decomposition symbolically from scratch, our algorithmic approach is to use the floating-point decomposition, and merely bound its error, using simple quantities which are easy to compute symbolically. Specifically, given the (floating-point, possibly approximate) decomposition $UDU^\top$ of a matrix A, we bound the error using the distance of $UDU^\top$ from A, as well as the distance of U from its projection on the subspace of orthogonal matrices given by $\bar{U} := U\left(U^\top U\right)^{-0.5}$. Formally, we use the following algorithm (where numerical computations refer to operations in floating-point arithmetic):

Input: Square matrix $A \in \mathbb{R}^{d\times d}$.
Output: A lower bound on the smallest eigenvalue of A if it is positive-definite and -1 otherwise.
- Numerically compute A', a double precision estimate of A.
- Symbolically compute $\epsilon_1 = ||A - A'||_F$.
- Numerically compute $U, D \in \mathbb{R}^{d\times d}$ s.t. $A' \approx UDU^\top$, D is diagonal.
- Symbolically compute $E = \mathbf{I} - U^\top U$, $A'' = UDU^\top$, $\epsilon_2 = ||A' - A''||_F$.
- Symbolically compute an upper bound $B = 1 + ||U - \mathbf{I}||_F$ on $||U||_{\text{sp}}$.
- Symbolically compute an upper bound $C = ||E||_F$ on $||E||_{\text{sp}}$.
- Let $\lambda_{\min}, \lambda_{\max}$ denote the smallest and largest diagonal entries of D respectively, then symbolically compute $\epsilon_3 = B^2\left(2\lambda_{\max}\left(\frac{1}{\sqrt{1-C}} - 1\right) + \left(\frac{1}{\sqrt{1-C}} - 1\right)^2\right)$.
- Return $\lambda_{\min} - \epsilon_1 - \epsilon_2 - \epsilon_3$ if it is larger than 0 and -1 otherwise.

Algorithm analysis: For the purpose of analyzing the algorithm, the following two lemmas will be used. We will also make use of the following version of Weyl's inequality, stated below for completeness.

Theorem 2 (Weyl's inequality)**.** *Suppose $A, B, P \in \mathbb{R}^{d\times d}$ are real symmetric matrices such that $A - B = P$. Assume that A, B have eigenvalues $\alpha_1 \geq \ldots \geq \alpha_d, \beta_1 \geq \ldots \geq \beta_d$ respectively, and that $||P||_{sp} \leq \epsilon$. Then*

$$|\alpha_i - \beta_i| \leq \epsilon \quad \forall i \in [d]\,.$$

Lemma 4. *For any natural $n \geq 0$ we have*

$$4^{-n}\sum_{k=0}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} = 1.$$

Proof. Clearly, for any $|x| < 1$ we have

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k. \tag{11}$$

Using the generalized binomial theorem, we have for any $|x| < 1$

$$\begin{aligned}\frac{1}{\sqrt{1-x}} &= \sum_{k=0}^{\infty} \binom{k-0.5}{k} x^k = \sum_{k=0}^{\infty} \frac{\prod_{i=0}^{k-1}(k-i-0.5)}{k!} x^k \\ &= \sum_{k=0}^{\infty} \frac{\prod_{i=0}^{k-1}(2k-2i-1)}{2^k k!} x^k = \sum_{k=0}^{\infty} \frac{2^k k! \prod_{i=0}^{k-1}(2k-2i-1)}{4^k (k!)^2} x^k \\ &= \sum_{k=0}^{\infty} \frac{\prod_{i=0}^{k-1}(2k-2i) \prod_{i=0}^{k-1}(2k-2i-1)}{4^k (k!)^2} x^k = \sum_{k=0}^{\infty} \frac{(2k)!}{4^k (k!)^2} x^k \\ &= \sum_{k=0}^{\infty} \binom{2k}{k} 4^{-k} x^k. \end{aligned} \tag{12}$$

Consider the k-th coefficient in the expansion of the square of Eq. (12), which is well defined as the sum converges absolutely for any $|x| < 1$. From Eq. (11), these coefficients are all 1. However, these are also given by the expansion of the square of Eq. (12). Specifically, the k-th coefficient in the square is given as the sum of all x^k coefficients in the expansion of the root, that is, it is a convolution of the coefficients in Eq. (12) with index $\leq k$, thus we have

$$4^{-n} \sum_{k=0}^{n} \binom{2k}{k} \binom{2n-2k}{n-k} = 1.$$

□

Lemma 5. *Let $U^\top U$ be a diagonally dominant matrix, let $E = \mathbf{I} - U^\top U$ satisfying $||E||_{sp} \leq C < 1$. Then $\left(U^\top U\right)^{-0.5} = \sum_{n=0}^{\infty} \binom{2n}{n} 4^{-n} E^n$. Moreover, $E' \coloneqq \sum_{n=1}^{\infty} \binom{2n}{n} 4^{-n} E^n$ satisfies $||E'||_{sp} \leq \left(\frac{1}{\sqrt{1-C}} - 1\right)$.*

Proof. Consider the series given by the partial sums

$$S_n = \sum_{k=0}^{n} \binom{2k}{k} 4^{-k} E^k,$$

and observe that

$$\begin{aligned} U^\top U S_n^2 &= (\mathbf{I} - E) \left(\sum_{k=0}^{n} \binom{2k}{k} 4^{-k} E^k \right)^2 \\ &= (\mathbf{I} - E) \left(\sum_{k=0}^{n} E^k + \sum_{k=n+1}^{2n} \beta_k E^k \right) \\ &= \mathbf{I} - E^{n+1} + (\mathbf{I} - E) E^{n+1} \sum_{k=0}^{n-1} \beta_{n+k+1} E^k, \end{aligned} \tag{13}$$

where the second equality is due to Lemma 4, and holds for some $\beta_k \in (0,1)$, $k \in \{n+1, \dots, 2n\}$. Now, since

$$\begin{aligned}
&\lim_{n\to\infty} \left|\left| (\mathbf{I}-E)\, E^{n+1} \sum_{k=0}^{n-1} \beta_{n+k+1} E^k \right|\right|_{\text{sp}} \\
&\leq \lim_{n\to\infty} ||\mathbf{I}-E||_{\text{sp}}\, ||E||_{\text{sp}}^{n+1} \left|\left| \sum_{k=0}^{n-1} \beta_{n+k+1} E^k \right|\right|_{\text{sp}} \\
&\leq \lim_{n\to\infty} ||\mathbf{I}-E||_{\text{sp}}\, ||E||_{\text{sp}}^{n+1} \left(\sum_{k=0}^{n-1} \beta_{n+k+1} ||E||_{\text{sp}}^k \right) \\
&\leq \lim_{n\to\infty} ||\mathbf{I}-E||_{\text{sp}}\, ||E||_{\text{sp}}^{n+1} \left(\sum_{k=0}^{n-1} C^k \right) \\
&\leq \lim_{n\to\infty} ||\mathbf{I}-E||_{\text{sp}}\, ||E||_{\text{sp}}^{n+1} (1-C)^{-1} \\
&=0,
\end{aligned}$$

we have that Eq. (13) reduces to $\mathbf{I}$ as $n \to \infty$, concluding the proof of the lemma. □

Turning back to the algorithm analysis, we wish to numerically compute the eigenvalues of A and bound their deviation due to roundoff errors. Other than the inaccuracy in computing $A'' \approx A'$, another obstacle is that U is not exactly orthogonal, however it is very close to orthogonal in the sense that $E = \mathbf{I} - U^\top U$ has a small norm. Let $\bar{U} = U \left(U^\top U\right)^{-0.5}$ be the projection of U onto the space of orthogonal matrices in $\mathbb{R}^{d\times d}$. Clearly, $\left(U^\top U\right)^{-0.5}$ is well defined if $U^\top U$ is diagonally-dominant, hence positive-definite, which can be easily verified. Also,

$$\begin{aligned}
\bar{U}^\top \bar{U} &= U \left(U^\top U\right)^{-0.5} \left(U \left(U^\top U\right)^{-0.5} \right)^\top \\
&= U \left(U^\top U\right)^{-0.5} \left(U^\top U\right)^{-0.5} U^\top \\
&= U \left(U^\top U\right)^{-1} U^\top \\
&= U U^{-1} \left(U^\top\right)^{-1} U^\top \\
&= \mathbf{I}.
\end{aligned}$$

We now upper bound $\left|\left|A'' - \bar{A}\right|\right|_{\text{sp}}$, where $\bar{A} = \bar{U} D \bar{U}^\top$ and therefore its spectrum is given to us explicitly as the diagonal entries of D, $\text{diag}\,(D)$. Compute

$$\begin{aligned}
\left|\left|A'' - \bar{A}\right|\right|_{\text{sp}} &= \left|\left| U D U^\top - \bar{U} D \bar{U}^\top \right|\right|_{\text{sp}} \\
&= \left|\left| U D U^\top - U \left(U^\top U\right)^{-0.5} D \left(U \left(U^\top U\right)^{-0.5} \right)^\top \right|\right|_{\text{sp}} \\
&= \left|\left| U \left(D - \left(U^\top U\right)^{-0.5} D \left(U^\top U\right)^{-0.5} \right) U^\top \right|\right|_{\text{sp}} \\
&= \left|\left| U \left(D - (\mathbf{I} + E')\, D\, (\mathbf{I} + E') \right) U^\top \right|\right|_{\text{sp}} \\
&= \left|\left| U \left(E'D + DE' + E'^2 \right) U^\top \right|\right|_{\text{sp}} \\
&\leq ||U||_{\text{sp}}^2 \left|\left| E'D + DE' + E'^2 \right|\right|_{\text{sp}} \\
&\leq ||U||_{\text{sp}}^2 \left(2\, ||D||_{\text{sp}}\, ||E'||_{\text{sp}} + ||E'||_{\text{sp}}^2 \right) \\
&\leq B^2 \left(2\lambda_{\max} \left(\frac{1}{\sqrt{1-C}} - 1 \right) + \left(\frac{1}{\sqrt{1-C}} - 1 \right)^2 \right) \\
&= \epsilon_3.
\end{aligned}$$

Estimating the spectrum diag (D) of A using the spectrum of $\bar{A}$ yields an approximation error of

$$\begin{aligned}\left|\left|A-\bar{A}\right|\right|_{\text{sp}} &= \left|\left|A-A'+A'-A''+A''-\bar{A}\right|\right|_{\text{sp}} \\ &\le \left|\left|A-A'\right|\right|_{\text{sp}} + \left|\left|A'-A''\right|\right|_{\text{sp}} + \left|\left|A''-\bar{A}\right|\right|_{\text{sp}} \\ &\le \epsilon_1+\epsilon_2+\epsilon_3,\end{aligned}$$

where in the last inequality we used the fact that the Frobenius norm upper bounds the spectral norm, which also proves that C is an upper bound on $||E'||_{\text{sp}}$. Verifying the upper bound given by B, we compute

$$||U||_{\text{sp}} = ||U-\mathbf{I}+\mathbf{I}||_{\text{sp}} \le ||U-\mathbf{I}||_{\text{sp}} + ||\mathbf{I}||_{\text{sp}} \le 1+||U-\mathbf{I}||_F \,.$$

Whenever U is close to unity, this provides a sharper upper bound than taking $C=||U||_F$.

Finally, applying Weyl's inequality (Thm. 2) to A and $\bar{A}$, we have that the spectra of the two cannot deviate by more than $\epsilon_1+\epsilon_2+\epsilon_3$, concluding the proof of the algorithm.

A.2. Upper Bound on Remainder Term $R_{\mathbf{w}_1^n,\mathbf{u}}$

In a nutshell, to derive an upper bound L on the third order term in Eq. (3), we show that the second order term in any direction is L-Lipschitz. Recalling that the purpose of this upper bound is to provide the radius of the ball enclosing a minimum in the vicinity of $\mathbf{w}_1^n$ (see Lemma 1), we observe, however, that Lemma 9 suggests L depends on the norm each neuron attains inside the ball, and therefore also on the radius of the ball enclosing the minimum. To circumvent this circular dependence between the radius and the third order bound, we first fixed the radius around $\mathbf{w}_1^n$ where we bound the third order term[5], and then checked whether the resulting radius enclosing the ball is smaller than the one used for the bound, thus validating the result.

In what follows, the ball where the third order bound is derived on is referred to as some compact subset of the weight space $A\subseteq\mathbb{R}^{kn}$. We now define some notation that will be used throughout the rest of this section. Given A, define

$$\mathbf{w}_{\min} = \min_{\mathbf{w}_1^n\in A}\min_{i\in[n]} ||\mathbf{w}_i||_2\,,$$

$$\mathbf{w}_{\max} = \max_{\mathbf{w}_1^n\in A}\max_{i\in[n]} ||\mathbf{w}_i||_2\,.$$

That is, $\mathbf{w}_{\min}$ and $\mathbf{w}_{\max}$ are the neurons with minimal and maximal norm among all possible network weights in the set A, respectively. Similarly, defining $\mathbf{v}_{\max}$ to be the target parameter vector with maximal 2-norm, the necessary bound is now given by the following theorem:

Theorem 3. *Suppose $\nabla^2F(\cdot)$ is differentiable on $A\subseteq\mathbb{R}^{kn}$. Then*

$$\sup_{\substack{\mathbf{w}_1^n\in A\\ \mathbf{u}:||\mathbf{u}||_2=1}}\sum_{i_1,i_2,i_3}\frac{\partial^3}{\partial w_{i_1}\partial w_{i_2}\partial w_{i_3}}F\left(\mathbf{w}_1^n\right)u_{i_1}u_{i_2}u_{i_3}\le L_A,$$

where

$$L_A \coloneqq \frac{n}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2}\left(\sqrt{2}\left(n-1\right)\left(||\mathbf{w}_{\max}||+||\mathbf{w}_{\min}||\right)+k\,||\mathbf{v}_{\max}||\right).$$

To prove the theorem, we will first need the following two lemmas.

Lemma 6. *Suppose $\nabla^2F(\cdot)$ is differentiable on $A\subseteq\mathbb{R}^{kn}$. Then*

- *$h_1(\mathbf{w},\mathbf{v})$ is $\frac{||\mathbf{v}_{\max}||}{\pi||\mathbf{w}_{\min}||^2}$ Lipschitz in $\mathbf{w}$ on A.*
- *$h_1(\mathbf{w}_1,\mathbf{w}_2)$ is $\frac{\sqrt{2}||\mathbf{w}_{\max}||}{\pi||\mathbf{w}_{\min}||^2}$ Lipschitz in $(\mathbf{w}_1,\mathbf{w}_2)$ on A.*

[5] specifically, the radius was chosen to be a 10^{-3} fraction of $\max_{i\in[n]}||\mathbf{w}_i||_2$. Testing this value, we observed that restricting the radius further only slightly improved the bound

- *$h_2(\mathbf{w}_1, \mathbf{w}_2)$ is $\frac{\sqrt{2}}{\pi||\mathbf{w}_{\min}||}$ Lipschitz in $(\mathbf{w}_1, \mathbf{w}_2)$ on A.*

Proof. We begin with computing some useful derivatives:

$$\frac{\partial}{\partial \mathbf{w}} \cos(\theta_{\mathbf{w},\mathbf{v}}) = \frac{\partial}{\partial \mathbf{w}} \frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||} = \frac{\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{\mathbf{w}}{||\mathbf{w}||^2} \frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||} = \frac{\mathbf{n}_{\mathbf{v},\mathbf{w}}}{||\mathbf{w}||}.$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}} \sin(\theta_{\mathbf{w},\mathbf{v}}) &= \frac{\partial}{\partial \mathbf{w}} \sqrt{1 - \left(\frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)^2} = \left(-\frac{\frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}}{\sqrt{1 - \left(\frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\right) \frac{\mathbf{n}_{\mathbf{v},\mathbf{w}}}{||\mathbf{w}||} \\ &= -\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}|| \sin(\theta_{\mathbf{w},\mathbf{v}})} \mathbf{n}_{\mathbf{v},\mathbf{w}} = -\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}||} \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}.\end{aligned}$$

$$\frac{\partial}{\partial \mathbf{w}} \theta_{\mathbf{w},\mathbf{v}} = \frac{\partial}{\partial \mathbf{w}} \arccos\left(\frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right) = -\frac{1}{\sqrt{1 - \left(\frac{\mathbf{w}^\top \mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}} \frac{\mathbf{n}_{\mathbf{v},\mathbf{w}}}{||\mathbf{w}||} = -\frac{\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}}{||\mathbf{w}||}.$$

Now, differentiating the spectral norms of h_1 and h_2 using Lemma 9 yields

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}} ||h_1(\mathbf{w}, \mathbf{v})||_{\text{sp}} &= \frac{\partial}{\partial \mathbf{w}} \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{\pi\,||\mathbf{w}||} \\ &= -\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}} + \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \bar{\mathbf{w}} \\ &= \frac{||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right),\end{aligned}$$

therefore

$$\begin{aligned}&\left|\left|\frac{\partial}{\partial \mathbf{w}} ||h_1(\mathbf{w}, \mathbf{v})||_{\text{sp}}\right|\right|_2 \\ =&\left|\left|\frac{||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\right|\right|_2 \\ =&\frac{||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \sqrt{\left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)^\top \left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)} \\ =&\frac{||\mathbf{v}||}{\pi\,||\mathbf{w}||^2} \sqrt{\sin^2(\theta_{\mathbf{w},\mathbf{v}})\,||\bar{\mathbf{w}}||^2 + \cos^2(\theta_{\mathbf{w},\mathbf{v}})\,||\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}||^2} \\ =&\frac{||\mathbf{v}||}{\pi\,||\mathbf{w}||^2}.\end{aligned}$$

Next, differentiating with respect to $\mathbf{v}$ gives

$$\begin{aligned}\frac{\partial}{\partial \mathbf{v}} ||h_1(\mathbf{w}, \mathbf{v})||_{\text{sp}} &= \frac{\partial}{\partial \mathbf{v}} \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{\pi\,||\mathbf{w}||} \\ &= -\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{\pi\,||\mathbf{w}||\,||\mathbf{v}||} \bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}} + \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})}{\pi\,||\mathbf{w}||} \bar{\mathbf{v}} \\ &= \frac{1}{\pi\,||\mathbf{w}||} \left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\right),\end{aligned}$$

so

$$\left|\left|\frac{\partial}{\partial \mathbf{v}}\left|\left|h_1\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}\right|\right|_2=\frac{1}{\pi\left|\left|\mathbf{w}\right|\right|}\left(\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{v}}-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\right)=\frac{1}{\pi\left|\left|\mathbf{w}\right|\right|}.$$

Concluding the derivation for the spectral norm of the gradient of h_1 we get

$$\left|\left|\frac{\partial}{\partial\left(\mathbf{w},\mathbf{v}\right)}\left|\left|h_1\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}\right|\right|_2=\sqrt{\left(\frac{1}{\pi\left|\left|\mathbf{w}\right|\right|}\right)^2+\left(\frac{\left|\left|\mathbf{v}\right|\right|}{\pi\left|\left|\mathbf{w}\right|\right|^2}\right)^2}=\frac{1}{\pi\left|\left|\mathbf{w}\right|\right|^2}\sqrt{\left|\left|\mathbf{w}\right|\right|^2+\left|\left|\mathbf{v}\right|\right|^2}. \tag{14}$$

Similarly, for h_2 we have

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}}\left|\left|h_2\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}&=\frac{\partial}{\partial \mathbf{w}}\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}+\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)\\&=\frac{1}{2\pi}\left(\frac{\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}}{\left|\left|\mathbf{w}\right|\right|}-\frac{\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\left|\left|\mathbf{w}\right|\right|}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\&=\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{w}\right|\right|}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}},\end{aligned}$$

thus

$$\left|\left|\frac{\partial}{\partial \mathbf{w}}\left|\left|h_2\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}\right|\right|_2=\left|\left|\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{w}\right|\right|}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right|\right|_2=\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{w}\right|\right|}\leq\frac{1}{\pi\left|\left|\mathbf{w}\right|\right|}.$$

For the gradient with respect to $\mathbf{v}$ we have

$$\begin{aligned}\frac{\partial}{\partial \mathbf{v}}\left|\left|h_2\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}&=\frac{\partial}{\partial \mathbf{v}}\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}+\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)\\&=\frac{1}{2\pi}\left(\frac{\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}}{\left|\left|\mathbf{v}\right|\right|}-\frac{\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\left|\left|\mathbf{v}\right|\right|}\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\right)\\&=\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{v}\right|\right|}\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}},\end{aligned}$$

which implies

$$\left|\left|\frac{\partial}{\partial \mathbf{v}}\left|\left|h_2\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}\right|\right|_2=\left|\left|\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{v}\right|\right|}\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\right|\right|_2=\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{v}\right|\right|}.$$

Concluding the derivation for the spectral norm of the gradient of h_2 we get

$$\begin{aligned}\left|\left|\frac{\partial}{\partial\left(\mathbf{w},\mathbf{v}\right)}\left|\left|h_2\left(\mathbf{w},\mathbf{v}\right)\right|\right|_{\text{sp}}\right|\right|_2&=\sqrt{\left(\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{w}\right|\right|}\right)^2+\left(\frac{1-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{2\pi\left|\left|\mathbf{v}\right|\right|}\right)^2}\\&\leq\frac{1}{\pi}\sqrt{\frac{1}{\left|\left|\mathbf{w}\right|\right|^2}+\frac{1}{\left|\left|\mathbf{v}\right|\right|^2}}.\end{aligned} \tag{15}$$

Finally, since a differentiable function is L-Lipschitz if and only if its gradient's 2-norm is bounded by L, the lemma follows from substituting $\mathbf{w}_{\min}, \mathbf{w}_{\max}, \mathbf{v}_{\max}$ in Eq. (14) and Eq. (15). ☐

Lemma 7. *Suppose $\nabla^2 F\left(\cdot\right)$ is differentiable on $A\subseteq\mathbb{R}^{kn}$. Then $\left|\left|\nabla^2 F\left(\cdot\right)\right|\right|_{sp}$ is L_A-Lipschitz in $\mathbf{w}_1^n$ on A.*

Proof. Since Lemma 6 implies the Lipschitzness of the spectral norms of $\tilde{h}_1, \tilde{h}_2$ in $\mathbf{w}_1^n\in\mathbb{R}^{kn}$, we let $\mathbf{w}_1^n=\left(\mathbf{w}_1,\dots,\mathbf{w}_n\right)$,

$\mathbf{w}_1'^n = (\mathbf{w}_1', \dots, \mathbf{w}_n') \in A$, then compute

$$\begin{aligned}
&\left|\left|\nabla^2 F\left(\mathbf{w}_1^n\right) - \nabla^2 F\left(\mathbf{w}_1'^n\right)\right|\right|_{\text{sp}} \\
&= \left|\left| \frac{1}{2}\mathbf{I} + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \tilde{h}_1\left(\mathbf{w}_i, \mathbf{w}_j\right) - \sum_{\substack{i=1,\dots,n \\ j=1,\dots,k}} \tilde{h}_1\left(\mathbf{w}_i, \mathbf{v}_j\right) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \tilde{h}_2\left(\mathbf{w}_i, \mathbf{w}_j\right) \right.\right. \\
&\qquad \left.\left. - \left(\frac{1}{2}\mathbf{I} + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \tilde{h}_1\left(\mathbf{w}_i', \mathbf{w}_j'\right) - \sum_{\substack{i=1,\dots,n \\ j=1,\dots,k}} \tilde{h}_1\left(\mathbf{w}_i', \mathbf{v}_j\right) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \tilde{h}_2\left(\mathbf{w}_i', \mathbf{w}_j'\right) \right) \right|\right|_{\text{sp}} \\
&= \left|\left| \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \left(\tilde{h}_1\left(\mathbf{w}_i, \mathbf{w}_j\right) - \tilde{h}_1\left(\mathbf{w}_i', \mathbf{w}_j'\right)\right) + \sum_{\substack{i=1,\dots,n \\ j=1,\dots,k}} \left(\tilde{h}_1\left(\mathbf{w}_i, \mathbf{v}_j\right) - \tilde{h}_1\left(\mathbf{w}_i', \mathbf{v}_j\right)\right) \right.\right. \\
&\qquad \left.\left. + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \left(\tilde{h}_2\left(\mathbf{w}_i, \mathbf{w}_j\right) - \tilde{h}_2\left(\mathbf{w}_i', \mathbf{w}_j'\right)\right) \right|\right|_{\text{sp}} \\
&\leq \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \left|\left|\tilde{h}_1\left(\mathbf{w}_i, \mathbf{w}_j\right) - \tilde{h}_1\left(\mathbf{w}_i', \mathbf{w}_j'\right)\right|\right|_{\text{sp}} + \sum_{\substack{i=1,\dots,n \\ j=1,\dots,k}} \left|\left|\tilde{h}_1\left(\mathbf{w}_i, \mathbf{v}_j\right) - \tilde{h}_1\left(\mathbf{w}_i', \mathbf{v}_j\right)\right|\right|_{\text{sp}} \\
&\qquad + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \left|\left|\tilde{h}_2\left(\mathbf{w}_i, \mathbf{w}_j\right) - \tilde{h}_2\left(\mathbf{w}_i', \mathbf{w}_j'\right)\right|\right|_{\text{sp}} \\
&\leq \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \frac{\sqrt{2}\left|\left|\mathbf{w}_{\max}\right|\right|}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2} \left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 + \sum_{\substack{i=1,\dots,n \\ j=1,\dots,k}} \frac{\left|\left|\mathbf{v}_{\max}\right|\right|}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2} \left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \frac{\sqrt{2}}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|} \left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 \\
&= \left(n\left(n-1\right) \frac{\sqrt{2}\left|\left|\mathbf{w}_{\max}\right|\right|}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2} + nk \frac{\left|\left|\mathbf{v}_{\max}\right|\right|}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2} + n\left(n-1\right) \frac{\sqrt{2}}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|} \right) \left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 \\
&= \frac{n}{\pi\left|\left|\mathbf{w}_{\min}\right|\right|^2} \left(\sqrt{2}\left(n-1\right)\left(\left|\left|\mathbf{w}_{\max}\right|\right| + \left|\left|\mathbf{w}_{\min}\right|\right|\right) + k\left|\left|\mathbf{v}_{\max}\right|\right|\right) \left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 .
\end{aligned}$$

□

Proof of Thm. 3. Let $\mathbf{w}_1^n, \mathbf{w}_1'^n \in A$. For any $\mathbf{u} \in \mathbb{R}^{kn}$ with $\left|\left|\mathbf{u}\right|\right|_2 = 1$ we have using Lemma 7

$$\begin{aligned}
&\left|\mathbf{u}^\top \nabla^2 F\left(\mathbf{w}_1^n\right)\mathbf{u} - \mathbf{u}^\top \nabla^2 F\left(\mathbf{w}_1'^n\right)\mathbf{u}\right| \\
=&\left|\mathbf{u}^\top \left(\nabla^2 F\left(\mathbf{w}_1^n\right) - \nabla^2 F\left(\mathbf{w}_1'^n\right)\right)\mathbf{u}\right| \\
\leq&\left|\left|\nabla^2 F\left(\mathbf{w}_1^n\right) - \nabla^2 F\left(\mathbf{w}_1'^n\right)\right|\right|_{\text{sp}} \\
\leq&L_A\left|\left|\mathbf{w}_1^n - \mathbf{w}_1'^n\right|\right|_2 ,
\end{aligned}$$

therefore the differentiable on A, $\mathbb{R}^{kn} \to \mathbb{R}$ function $t \mapsto \mathbf{u}^\top \nabla^2 F\left(\mathbf{w}_1^n + t\mathbf{u}\right)\mathbf{u}$ is L_A-Lipschitz for any $\mathbf{u} \in \mathbb{R}^{kn}$, $\left|\left|\mathbf{u}\right|\right| = 1$, hence its derivative on A is upper bounded by L_A. Namely, we have that

$$\sup_{\substack{\mathbf{w}_1^n \in A \\ \mathbf{u}:\left|\left|\mathbf{u}\right|\right|_2 = 1}} \sum_{i_1,i_2,i_3} \frac{\partial^3}{\partial w_{i_1}\partial w_{i_2}\partial w_{i_3}} F\left(\mathbf{w}_1^n\right) u_{i_1}u_{i_2}u_{i_3} \leq L_A.$$

□

A.3. Lipschitzness of $F\left(\mathbf{w}_1^n\right)$ and Proof of Lemma 2

In this subsection, we turn to proving a Lipschitz bound on the objective in Eq. (6), implying Lemma 2 and showing that the local minimum identified in Eq. (1) is necessarily non-global. A straightforward approach would be to globally upper bound $||\nabla F\left(\mathbf{w}_1^n\right)||$ (excluding the neighborhood of some singular points). However, this approach is quite loose, since it does not take advantage of the fact that the gradients $\nabla F\left(\mathbf{w}_1^n\right)$ close to our points of interest are very small. Instead, we first derive a Lipschitz bound on $\nabla^2 F\left(\mathbf{w}_1^n\right)$, implying that $\nabla F\left(\mathbf{w}_1^n\right)$ does not vary too greatly, and therefore remains small for any $\mathbf{w}_1'^n$ in the ball enclosing $\mathbf{w}_1^n$, providing a stronger bound than the more naive approach.

Theorem 4. *Suppose F is thrice-differentiable on $A \subseteq \mathbb{R}^{kn}$. Then for any $\mathbf{w}_1'^n \in A$,*

$$\left|F\left(\mathbf{w}_1'^n\right) - F\left(\mathbf{w}_1^n\right)\right| \leq ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 \left(LH\,||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2\right),$$

where

$$LH := \frac{1}{2} + n\left(n-1\right)\left(\frac{||\mathbf{w}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||} + \frac{1}{2}\right) + \frac{nk\,||\mathbf{v}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||}.$$

To prove the theorem, we will need the following lemma:

Lemma 8. *Suppose F is thrice-differentiable on $A \subseteq \mathbb{R}^{kn}$. Then*

$$\sup_{\mathbf{w}_1^n \in A} \left|\left|\nabla^2 F\left(\mathbf{w}_1^n\right)\right|\right|_{sp} \leq \frac{1}{2} + n\left(n-1\right)\left(\frac{||\mathbf{w}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||} + \frac{1}{2}\right) + \frac{nk\,||\mathbf{v}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||}.$$

Proof. Recall the Hessian of the objective as defined in Eq. (10). Using Lemma 9, the fact that the spectral norms of h_1, h_2 and $\tilde{h}_1, \tilde{h}_2$ are identical, and the fact that $\sin\left(x\right) \leq x$ for any $x > 0$, we have for any $W \in A$

$$\begin{aligned}
\left|\left|\nabla^2 F\left(\mathbf{w}_1^n\right)\right|\right|_{\text{sp}} &= \left|\left|\frac{1}{2}\mathbf{I} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \tilde{h}_1\left(\mathbf{w}_i, \mathbf{w}_j\right) - \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,k}} \tilde{h}_1\left(\mathbf{w}_i, \mathbf{v}_j\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \tilde{h}_2\left(\mathbf{w}_i, \mathbf{w}_j\right)\right|\right|_{\text{sp}} \\
&\leq \frac{1}{2} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \left|\left|\tilde{h}_1\left(\mathbf{w}_i, \mathbf{w}_j\right)\right|\right|_{\text{sp}} + \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,k}} \left|\left|\tilde{h}_1\left(\mathbf{w}_i, \mathbf{v}_j\right)\right|\right|_{\text{sp}} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \left|\left|\tilde{h}_2\left(\mathbf{w}_i, \mathbf{w}_j\right)\right|\right|_{\text{sp}} \\
&\leq \frac{1}{2} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \frac{\sin\left(\theta_{\mathbf{w}_i,\mathbf{w}_j}\right)||\mathbf{w}_j||}{2\pi\,||\mathbf{w}_i||} + \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,k}} \frac{\sin\left(\theta_{\mathbf{w}_i,\mathbf{v}_j}\right)||\mathbf{v}_j||}{2\pi\,||\mathbf{w}_i||} \\
&\quad + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \frac{1}{2\pi}\left(\pi - \theta_{\mathbf{w}_i,\mathbf{w}_j} + \sin\left(\theta_{\mathbf{w}_i,\mathbf{w}_j}\right)\right) \\
&\leq \frac{1}{2} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \frac{||\mathbf{w}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||} + \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,k}} \frac{||\mathbf{v}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \frac{1}{2\pi}\pi \\
&\leq \frac{1}{2} + n\left(n-1\right)\left(\frac{||\mathbf{w}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||} + \frac{1}{2}\right) + \frac{nk\,||\mathbf{v}_{\max}||}{2\pi\,||\mathbf{w}_{\min}||}.
\end{aligned}$$

□

Proof of Thm. 4. For some $\mathbf{u} \in \mathbb{R}^{kn}$, consider the function $g_{\mathbf{u}}\left(t\right) = \mathbf{u}^\top \nabla F\left(\mathbf{w}_1^n + t\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right)\right)$

Since F is thrice-differentiable, we have from the mean value theorem that there exists some $t_{\mathbf{u}}$ such that

$$\begin{aligned}
\mathbf{u}^\top\left(\nabla F\left(\mathbf{w}_1'^n\right) - \nabla F\left(\mathbf{w}_1^n\right)\right) &= \frac{g_{\mathbf{u}}\left(1\right) - g_{\mathbf{u}}\left(0\right)}{1-0} \\
&= g_{\mathbf{u}}'\left(t_{\mathbf{u}}\right) \\
&= \mathbf{u}^\top \nabla^2 F\left(\mathbf{w}_1^n + t_{\mathbf{u}}\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right)\right)\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right)
\end{aligned}$$

Taking $\mathbf{u} = \nabla F\left(\mathbf{w}_1'^n\right) + \nabla F\left(\mathbf{w}_1^n\right)$ and recalling that from Lemma 8 we have that $\sup_{\mathbf{w}_1'^n \in A} \left|\left|\nabla^2 F\left(\mathbf{w}_1'^n\right)\right|\right|_{\text{sp}}$ is bounded by LH, we get

$$
\begin{aligned}
& ||\nabla F\left(\mathbf{w}_1'^n\right)||_2^2 - ||\nabla F\left(\mathbf{w}_1^n\right)||_2^2 \\
= & \left(\nabla F\left(\mathbf{w}_1'^n\right) + \nabla F\left(\mathbf{w}_1^n\right)\right)^\top \left(\nabla F\left(\mathbf{w}_1'^n\right) - \nabla F\left(\mathbf{w}_1^n\right)\right) \\
= & \left(\nabla F\left(\mathbf{w}_1'^n\right) + \nabla F\left(\mathbf{w}_1^n\right)\right)^\top \nabla^2 F\left(\mathbf{w}_1^n + t_{\mathbf{u}}\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right)\right)\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right) \\
\leq & ||\nabla F\left(\mathbf{w}_1'^n\right) + \nabla F\left(\mathbf{w}_1^n\right)||_2 \left|\left|\nabla^2 F\left(\mathbf{w}_1^n + t_{\mathbf{u}}\left(\mathbf{w}_1'^n - \mathbf{w}_1^n\right)\right)\right|\right|_{\text{sp}} ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 \\
\leq & \left(||\nabla F\left(\mathbf{w}_1'^n\right)||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2\right) LH \, ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 \, .
\end{aligned}
$$

Dividing by $||\nabla F\left(\mathbf{w}_1'^n\right)||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2$ and rearranging yields

$$
||\nabla F\left(\mathbf{w}_1'^n\right)||_2 \leq LH \, ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2 \, .
$$

That is, the target function F is $(LH \, ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2)$-Lipschitz on A, thus

$$
|F\left(\mathbf{w}_1'^n\right) - F\left(\mathbf{w}_1^n\right)| \leq ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 \left(LH \, ||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 + ||\nabla F\left(\mathbf{w}_1^n\right)||_2\right) .
$$

□

Proof of Lemma 2. For A which is a ball of radius r centered at $\mathbf{w}_1^n = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$, we have that $||\mathbf{w}_{\max}|| = \max_i ||\mathbf{w}_i|| + r$ as well as $||\mathbf{w}_{\min}|| = \min_i ||\mathbf{w}_i|| - r$. Plugging this in Thm. 4 and substituting $||\mathbf{w}_1'^n - \mathbf{w}_1^n||_2 \leq r$ completes the proof of the lemma. □

A.4. Technical Proofs

A.4.1. Derivation of $\nabla^2 F\left(\mathbf{w}_1^n\right)$

Theorem 5. *The Hessian of F at point $\mathbf{w}_1^n = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ with respect to target values $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ is given on the main diagonals*

$$
\frac{\partial^2 F}{\partial \mathbf{w}_i^2} = \frac{1}{2}\mathbf{I} + \sum_{\substack{j=1 \\ j \neq i}}^{n} h_1\left(\mathbf{w}_i, \mathbf{w}_j\right) - \sum_{j=1}^{k} h_1\left(\mathbf{w}_i, \mathbf{v}_j\right),
$$

and on the off-diagonals by

$$
\frac{\partial^2 F}{\partial \mathbf{w}_i \partial \mathbf{w}_j} = h_2\left(\mathbf{w}_i, \mathbf{w}_j\right),
$$

where

$$
h_1\left(\mathbf{w}, \mathbf{v}\right) = \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right) ||\mathbf{v}||}{2\pi \, ||\mathbf{w}||} \left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}^\top\right),
$$

and

$$
h_2\left(\mathbf{w}, \mathbf{v}\right) = \frac{1}{2\pi}\left(\left(\pi - \theta_{\mathbf{w},\mathbf{v}}\right)\mathbf{I} + \bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^\top\right).
$$

Proof. By a straightforward calculation, we have

$$
\begin{aligned}
\frac{\partial^2 f\left(\mathbf{w}, \mathbf{v}\right)}{\partial w_i^2} = & \frac{1}{2\pi}\left(||\mathbf{v}|| \left(\left(\frac{1}{||\mathbf{w}||} - \frac{w_i^2}{||\mathbf{w}||^3}\right)\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right) - \frac{w_i}{||\mathbf{w}||}\frac{\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}}{\sqrt{1 - \left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{v_i}{||\mathbf{w}|| \, ||\mathbf{v}||} - \frac{w_i}{||\mathbf{w}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}|| \, ||\mathbf{v}||}\right)\right)\right. \\
& \left. + \frac{v_i}{\sqrt{1 - \left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{v_i}{||\mathbf{w}|| \, ||\mathbf{v}||} - \frac{w_i}{||\mathbf{w}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}|| \, ||\mathbf{v}||}\right)\right) \\
= & \frac{1}{2\pi}\left(\frac{||\mathbf{v}||}{||\mathbf{w}||}\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right) + w_i^2\frac{||\mathbf{v}|| \cos\left(2\theta_{\mathbf{w},\mathbf{v}}\right)}{||\mathbf{w}||^3 \sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)} - 2w_i v_i \frac{\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{||\mathbf{w}||^2 \sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)} + v_i^2 \frac{1}{||\mathbf{w}|| \, ||\mathbf{v}|| \sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}\right)
\end{aligned}
$$

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial w_i w_j} = \frac{1}{2\pi}\left(||\mathbf{v}||\left(-\frac{w_i w_j}{||\mathbf{w}||^3}\sin(\theta_{\mathbf{w},\mathbf{v}}) - \frac{w_i}{||\mathbf{w}||}\frac{\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{v_j}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{w_j}{||\mathbf{w}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)\right.$$
$$\left. + \frac{v_i}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{v_j}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{w_j}{||\mathbf{w}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)$$
$$= \frac{1}{2\pi}\left(w_i w_j \frac{||\mathbf{v}||\cos(2\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}||^3\sin(\theta_{\mathbf{w},\mathbf{v}})} - (w_i v_j + w_j v_i)\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})} + v_i v_j \frac{1}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})}\right)$$

Hence

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial \mathbf{w}^2} = \frac{1}{2\pi}\left(\frac{||\mathbf{v}||}{||\mathbf{w}||}\sin(\theta_{\mathbf{w},\mathbf{v}})\,\mathbf{I} + \frac{||\mathbf{v}||\cos(2\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}||^3\sin(\theta_{\mathbf{w},\mathbf{v}})}\mathbf{w}\mathbf{w}^\top\right. \tag{16}$$
$$\left. - \frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})}\left(\mathbf{w}\mathbf{v}^\top + \mathbf{v}\mathbf{w}^\top\right) + \frac{1}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})}\mathbf{v}\mathbf{v}^\top\right)$$
$$= \frac{||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\sin(\theta_{\mathbf{w},\mathbf{v}})\,\mathbf{I} + \frac{\cos(2\theta_{\mathbf{w},\mathbf{v}})}{\sin(\theta_{\mathbf{w},\mathbf{v}})}\bar{\mathbf{w}}\bar{\mathbf{w}}^\top - \frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{\sin(\theta_{\mathbf{w},\mathbf{v}})}\left(\bar{\mathbf{w}}\bar{\mathbf{v}}^\top + \bar{\mathbf{v}}\bar{\mathbf{w}}^\top\right) + \frac{1}{\sin(\theta_{\mathbf{w},\mathbf{v}})}\bar{\mathbf{v}}\bar{\mathbf{v}}^\top\right)$$
$$= \frac{||\mathbf{v}||}{2\pi\sin(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{w}||}\left(\sin^2(\theta_{\mathbf{w},\mathbf{v}})\left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top\right) + \left(\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}}\right)\left(\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}}\right)^\top\right). \tag{17}$$

Recall the definition of $\mathbf{n}$ in Eq. (9), we have that

$$\begin{aligned}||\mathbf{n}||^2 &= \left(\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}}\right)^\top\left(\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}}\right)\\ &= \bar{\mathbf{v}}^\top\bar{\mathbf{v}} - 2\cos(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{v}}^\top\bar{\mathbf{w}} + \cos^2(\theta_{\mathbf{w},\mathbf{v}})\,\bar{\mathbf{w}}^\top\bar{\mathbf{w}}\\ &= 1 - \cos^2(\theta_{\mathbf{w},\mathbf{v}})\\ &= \sin^2(\theta_{\mathbf{w},\mathbf{v}}).\end{aligned}$$

Therefore Eq. (16) can be written as

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial \mathbf{w}^2} = \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})\,||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}^\top\right).$$

Differentiating with respect to different individual parameter vectors, we have

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial w_i \partial v_i} = \frac{1}{2\pi}\left(\frac{w_i}{||\mathbf{w}||}\left(\frac{v_i}{||\mathbf{v}||}\sin(\theta_{\mathbf{w},\mathbf{v}}) - ||\mathbf{v}||\frac{\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{w_i}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{v_i}{||\mathbf{v}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)\right.$$

$$\left. + (\pi - \theta_{\mathbf{w},\mathbf{v}}) + \frac{v_i}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{w_i}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{v_i}{||\mathbf{v}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)$$

$$= \frac{1}{2\pi}\left(-w_i^2\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}^2||\sin(\theta_{\mathbf{w},\mathbf{v}})} + w_i v_i \frac{1}{||\mathbf{w}||\,||\mathbf{v}||}\left(\sin(\theta_{\mathbf{w},\mathbf{v}}) + \frac{1}{\sin(\theta_{\mathbf{w},\mathbf{v}})} + \frac{\cos^2(\theta_{\mathbf{w},\mathbf{v}})}{\sin(\theta_{\mathbf{w},\mathbf{v}})}\right)\right.$$

$$\left. - v_i^2\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{v}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})} + (\pi - \theta_{\mathbf{w},\mathbf{v}})\right)$$

$$= \frac{1}{2\pi}\left(-w_i^2\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}^2||\sin(\theta_{\mathbf{w},\mathbf{v}})} + w_i v_i\frac{2}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})} - v_i^2\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{v}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})} + (\pi - \theta_{\mathbf{w},\mathbf{v}})\right)$$

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial w_i \partial v_j} = \frac{1}{2\pi}\left(\frac{w_i}{||\mathbf{w}||}\left(\frac{v_j}{||\mathbf{v}||}\sin(\theta_{\mathbf{w},\mathbf{v}}) - ||\mathbf{v}||\frac{\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{w_j}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{v_j}{||\mathbf{v}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)\right.$$

$$\left. + \frac{v_i}{\sqrt{1-\left(\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||||\mathbf{v}||}\right)^2}}\left(\frac{w_j}{||\mathbf{w}||\,||\mathbf{v}||} - \frac{v_j}{||\mathbf{v}||^2}\frac{\mathbf{w}^\top\mathbf{v}}{||\mathbf{w}||\,||\mathbf{v}||}\right)\right)$$

$$= \frac{1}{2\pi}\left(-w_i w_j\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}^2||\sin(\theta_{\mathbf{w},\mathbf{v}})} + w_i v_j\frac{1}{||\mathbf{w}||\,||\mathbf{v}||}\left(\sin(\theta_{\mathbf{w},\mathbf{v}}) + \frac{\cos^2(\theta_{\mathbf{w},\mathbf{v}})}{\sin(\theta_{\mathbf{w},\mathbf{v}})}\right)\right.$$

$$\left. + w_j v_i\frac{1}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})} - v_i v_j\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{v}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})}\right)$$

$$= -\frac{1}{2\pi}\left(w_i w_j\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{w}^2||\sin(\theta_{\mathbf{w},\mathbf{v}})} - w_i v_j\frac{1}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})}\right.$$

$$\left. - w_j v_i\frac{1}{||\mathbf{w}||\,||\mathbf{v}||\sin(\theta_{\mathbf{w},\mathbf{v}})} + v_i v_j\frac{\cos(\theta_{\mathbf{w},\mathbf{v}})}{||\mathbf{v}||^2\sin(\theta_{\mathbf{w},\mathbf{v}})}\right).$$

Hence

$$\frac{\partial^2 f(\mathbf{w},\mathbf{v})}{\partial \mathbf{w}\partial \mathbf{v}} = \left(\frac{\pi - \theta_{\mathbf{w},\mathbf{v}}}{2\pi}\right)\mathbf{I} + \frac{1}{2\pi\sin(\theta_{\mathbf{w},\mathbf{v}})}\left((\bar{\mathbf{w}}+\bar{\mathbf{v}})(\bar{\mathbf{w}}+\bar{\mathbf{v}})^\top - (1+\cos(\theta_{\mathbf{w},\mathbf{v}}))\left(\bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{v}}\bar{\mathbf{v}}^\top\right)\right)$$

$$= \left(\frac{\pi - \theta_{\mathbf{w},\mathbf{v}}}{2\pi}\right)\mathbf{I} + \frac{1}{2\pi\sin(\theta_{\mathbf{w},\mathbf{v}})}\left(\bar{\mathbf{w}}\bar{\mathbf{v}}^\top + \bar{\mathbf{v}}\bar{\mathbf{w}}^\top - \cos(\theta_{\mathbf{w},\mathbf{v}})\bar{\mathbf{w}}\bar{\mathbf{w}}^\top - \cos(\theta_{\mathbf{w},\mathbf{v}})\bar{\mathbf{v}}\bar{\mathbf{v}}^\top\right)$$

$$= \left(\frac{\pi - \theta_{\mathbf{w},\mathbf{v}}}{2\pi}\right)\mathbf{I} + \frac{1}{2\pi\sin(\theta_{\mathbf{w},\mathbf{v}})}\left((\bar{\mathbf{w}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\bar{\mathbf{v}})\bar{\mathbf{v}}^\top + (\bar{\mathbf{v}} - \cos(\theta_{\mathbf{w},\mathbf{v}})\bar{\mathbf{w}})\bar{\mathbf{w}}^\top\right)$$

$$= \frac{1}{2\pi}\left((\pi - \theta_{\mathbf{w},\mathbf{v}})\mathbf{I} + \bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^\top\right).$$

Recall the objective in Eq. (6), we have that its Hessian is comprised of $n \times n$ blocks of size $d \times d$ each. On the main diagonal we therefore have

$$\frac{\partial^2 F}{\partial \mathbf{w}_i^2} = \frac{\partial^2}{\partial \mathbf{w}_i^2}\left(\frac{1}{2}f\left(\mathbf{w}_i,\mathbf{w}_i\right) + \sum_{\substack{j=1\\ j\neq i}}^{n} f\left(\mathbf{w}_i,\mathbf{w}_j\right) + \sum_{j=1}^{k} f\left(\mathbf{w}_i,\mathbf{v}_j\right)\right)$$
$$= \frac{1}{2}\mathbf{I} + \sum_{\substack{j=1\\ j\neq i}}^{n} h_1\left(\mathbf{w}_i,\mathbf{w}_j\right) - \sum_{j=1}^{k} h_1\left(\mathbf{w}_i,\mathbf{v}_j\right),$$

and on the off diagonal we have

$$\frac{\partial^2 F}{\partial \mathbf{w}_i \partial \mathbf{w}_j} = h_2\left(\mathbf{w}_i,\mathbf{w}_j\right).$$

□

A.4.2. THE SPECTRAL NORM OF h_1 AND h_2

Lemma 9. *We have that*

- $||h_1\left(\mathbf{w},\mathbf{v}\right)||_{sp} = \frac{\sin(\theta_{\mathbf{w},\mathbf{v}})||\mathbf{v}||}{\pi||\mathbf{w}||}$.
- $||h_2\left(\mathbf{w},\mathbf{v}\right)||_{sp} = \frac{1}{2\pi}\left(\pi - \theta_{\mathbf{w},\mathbf{v}} + \sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)$.

Proof. To find the spectral norm, we compute the spectra of h_1, h_2.

- Clearly, for any $\mathbf{u} \in \mathbb{R}^d$ orthogonal to both $\bar{\mathbf{w}}, \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}$ we have

  $$h_1\left(\mathbf{w},\mathbf{v}\right)\mathbf{u} = \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}^\top\right)\mathbf{u} = \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\mathbf{u}.$$

  Thus $\frac{\sin(\theta_{\mathbf{w},\mathbf{v}})||\mathbf{v}||}{2\pi||\mathbf{w}||}$ is an eigenvalue of h_1 with multiplicity at least $d-2$. Since $\bar{\mathbf{w}}, \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}$ are orthogonal, their corresponding eigenvalues comprise the rest of the spectrum of h_1. Compute

  $$h_1\left(\mathbf{w},\mathbf{v}\right)\bar{\mathbf{w}} = \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}^\top\right)\bar{\mathbf{w}}$$
  $$= \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\bar{\mathbf{w}} - \bar{\mathbf{w}}\,||\bar{\mathbf{w}}||^2\right)$$
  $$= 0.$$

  Hence 0 is the eigenvalue of $\bar{\mathbf{w}}$. Also,

  $$h_1\left(\mathbf{w},\mathbf{v}\right)\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}} = \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\mathbf{I} - \bar{\mathbf{w}}\bar{\mathbf{w}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}^\top\right)\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}$$
  $$= \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{2\pi\,||\mathbf{w}||}\left(\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}} + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\,||\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}||^2\right)$$
  $$= \frac{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)||\mathbf{v}||}{\pi\,||\mathbf{w}||}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}.$$

  Therefore $\frac{\sin(\theta_{\mathbf{w},\mathbf{v}})||\mathbf{v}||}{\pi||\mathbf{w}||}$ is the largest eigenvalue of h_1.

- Once again, for any $\mathbf{u} \in \mathbb{R}^d$ orthogonal to both $\bar{\mathbf{v}}, \bar{\mathbf{w}}$ we have

  $$h_2\left(\mathbf{w},\mathbf{v}\right)\mathbf{u} = \frac{1}{2\pi}\left(\left(\pi - \theta_{\mathbf{w},\mathbf{v}}\right)\mathbf{I} + \bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^\top + \bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^\top\right)\mathbf{u} = \frac{1}{2\pi}\left(\pi - \theta_{\mathbf{w},\mathbf{v}}\right)\mathbf{u}.$$

Thus $\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)$ is an eigenvalue of h_2 with multiplicity at least $d-2$. We now show the remaining two eigenvalues correspond to the eigenvectors $\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}$ and $\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}$.

$$
\begin{aligned}
&h_2\left(\mathbf{w},\mathbf{v}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\mathbf{I}+\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\frac{\bar{\mathbf{w}}-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{v}}}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}-\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}\frac{\bar{\mathbf{v}}-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{w}}}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\frac{1-\cos^2\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}-\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\frac{1-\cos^2\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)-\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\right)\\
=&\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}-\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}-\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right).
\end{aligned}
$$

Hence $\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}-\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)$ is an eigenvalue of h_2. Similarly, we have

$$
\begin{aligned}
&h_2\left(\mathbf{w},\mathbf{v}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\mathbf{I}+\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\bar{\mathbf{w}}^{\top}\frac{\bar{\mathbf{w}}-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{v}}}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}+\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\bar{\mathbf{v}}^{\top}\frac{\bar{\mathbf{v}}-\cos\left(\theta_{\mathbf{w},\mathbf{v}}\right)\bar{\mathbf{w}}}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\frac{1-\cos^2\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}+\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}\frac{1-\cos^2\left(\theta_{\mathbf{w},\mathbf{v}}\right)}{\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)}\right)\\
=&\frac{1}{2\pi}\left(\left(\pi-\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)+\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right)\right)\\
=&\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}+\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)\left(\bar{\mathbf{n}}_{\mathbf{w},\mathbf{v}}+\bar{\mathbf{n}}_{\mathbf{v},\mathbf{w}}\right).
\end{aligned}
$$

Therefore $\frac{1}{2\pi}\left(\pi-\theta_{\mathbf{w},\mathbf{v}}+\sin\left(\theta_{\mathbf{w},\mathbf{v}}\right)\right)$ is the largest eigenvalue of h_2.

□