Analytic Characterization of the Hessian in Shallow ReLU Models: A Tale of Symmetry

Yossi Arjevani
NYU
yossi.arjevani@gmail.com

Michael Field
UCSB
mikefield@gmail.com

Abstract

We consider the optimization problem associated with fitting two-layers ReLU networks with respect to the squared loss, where labels are generated by a target network. We leverage the rich symmetry structure to analytically characterize the Hessian at various families of spurious minima in the natural regime where the number of inputs d and the number of hidden neurons k is finite. In particular, we prove that for $d \geq k$ standard Gaussian inputs: (a) of the dk eigenvalues of the Hessian, $dk - O(d)$ concentrate near zero, (b) $\Omega(d)$ of the eigenvalues grow linearly with k. Although this phenomenon of extremely skewed spectrum has been observed many times before, to our knowledge, this is the first time it has been established rigorously. Our analytic approach uses techniques, new to the field, from symmetry breaking and representation theory, and carries important implications for our ability to argue about statistical generalization through local curvature.

1 Introduction

Much of the current effort in understanding the empirical success of artificial neural networks is concerned with the geometry of the associated nonconvex optimization landscapes. Of particular importance is the Hessian spectrum which characterizes the local curvature of the loss at different points in the space. This, in turn, allows one to closely examine the dynamics of stochastic first order methods [1, 2], design potentially better optimization methods [3, 4], and argue about various challenging aspects of the network generalization capabilities [5, 6, 7]. Unfortunately, the excessively high cost involved in an exact computation of the Hessian spectrum renders this task prohibitive already for moderate-sized problems.

Existing approaches for addressing this computational barrier use numerical methods for approximating the Hessian spectrum [2, 8], study the limiting spectral density of shallow models w.r.t. *randomly* drawn weights [9, 10, 11], or employ various simplified indirect curvature metrics [12, 5, 7, 13, 14]. Notably, none of these techniques is able to yield an analytic characterization of the Hessian at critical points in high-dimensional spaces.

In this paper, we develop a novel approach for studying the Hessian in a class of student-teacher (ST) models. Concretely, we focus on the squared loss of fitting the ReLU network $\boldsymbol{x} \mapsto \mathbf{1}_k^\top \phi(\boldsymbol{W}\boldsymbol{x})$,

$$\mathcal{L}(\boldsymbol{W}) \doteq \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},I_d)}\left[(\mathbf{1}_k^\top \phi(\boldsymbol{W}\boldsymbol{x}) - \mathbf{1}_k^\top \phi(\boldsymbol{V}\boldsymbol{x}))^2\right], \quad \boldsymbol{W} \in M(k,d), \tag{1.1}$$

where $\phi(z) \doteq \max\{0, z\}$ is the ReLU activation acting coordinate-wise, $\mathbf{1}_k$ is the k-dimensional vector of all ones, $M(k,d)$ denotes the space of all $k \times d$ matrices, and $\boldsymbol{V} \in M(k,d)$ denotes the weight matrix of the target network. The ST framework offers a clean venue for analyzing optimization- and generalization-related aspects of neural network models, and has consequently enjoyed a surge of interest in recent years, e.g., [15, 16, 17, 18, 19, 20, 21, 22], to name a few.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

Perhaps surprisingly, already for this simple model, the rich and perplexing geometry of the induced nonconvex optimization landscape seems to be out of reach of existing analytic methods.

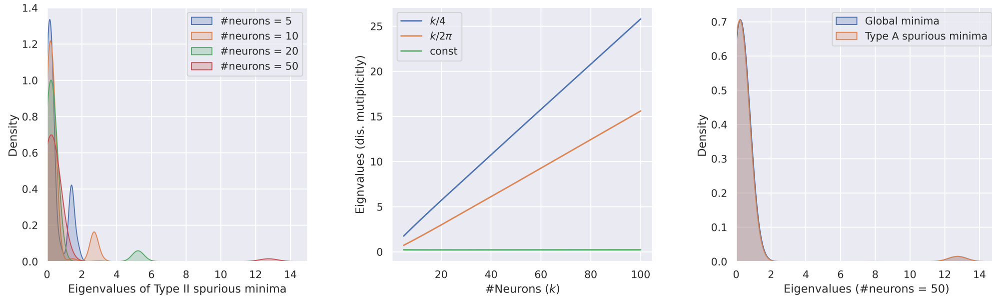

Figure 1: (Left) in congruence with Theorem 2, $1 - \Theta(1/k)$ fraction of the spectral density at type II spurious minima concentrates around $1/4 \pm 1/2\pi$ as the number of neurons k grows simultaneously with the number of inputs. The remaining $\Theta(1/k)$ fraction consists of outliers. (Middle) examining the spectrum of type II minima (disregarding multiplicity) as k grows confirms the existence of $k+1$ outlier eigenvalues, of which k grow at a rate of $k/4$ and one at a rate of $k/2\pi$. (Right) the spectra of global minima and type A spurious minima are almost indistinguishable already for $k = 50$, thus challenging the flat minima conjecture.

The starting point of our approach is the following simple observation: for any permutation matrices $\boldsymbol{P} \in M(k,k)$, $\boldsymbol{Q} \in M(d,d)$, it holds that $\mathcal{L}(\boldsymbol{P}\boldsymbol{W}\boldsymbol{Q}^\top) = \mathcal{L}(\boldsymbol{W})$, for all $\boldsymbol{W} \in M(k,d)$ [23, Section 4.1]. It is natural to ask how the critical points of $\mathcal{L}$ reflect this symmetry. This question was answered in [23] where it was shown that critical points detected by stochastic gradient descent (SGD) remain unchanged under transformations of the form $\boldsymbol{W} \mapsto \boldsymbol{P}\boldsymbol{W}\boldsymbol{Q}^\top$ for large groups of pairs of permutation matrices $(\boldsymbol{P}, \boldsymbol{Q})$. Using these invariance properties, families of critical points of $\mathcal{L}$ were expressed as power series in $1/\sqrt{k}$ leading to, for example, a precise formula for the decay rate of $\mathcal{L}$ [24]. Building on this, we show in this paper how the rich symmetry structure can be used to derive an analytic description of the Hessian spectral density of $\mathcal{L}$, for arbitrarily large, yet finite, values of k. Having this access to precise high-dimensional spectral densities, we revisit a number of hypotheses in the machine learning literature pertaining to curvature, optimization and generalization, and establish or refute them *rigorously* for the first time.

The paper is organized as follows. In Section 2 we state our main results and provide discussions aimed at interpreting our findings in the light of existing literature. Section 3 and Section 4 are devoted to describing our representation theory-based approach; all proofs are deferred to the appendix. Lastly, detailed empirical corroborations of our analysis are given in Section F.

2 Main results and related work

A formal discussion of our main results requires some familiarity with group and representation theory. Here, we provide a high-level description of our contributions, and defer more detailed statements to later sections after the relevant notions have been introduced.

Symmetry-based analysis framework. Utilizing the rich symmetry exhibited by neural network models, we develop a novel framework for analytically characterizing the second-order information of shallow ST ReLU models. In its general form, our main result can be stated as follows.

Theorem 1 (Informal)**.** *Assuming a $k \times k$ orthogonal target matrix $\boldsymbol{V}$ ($d = k$), the spectrum of local minima of $\mathcal{L}$ consists of a fixed number of distinct eigenvalues (ranging between 6 and 22 for high symmetry minima)—independent of the number of neurons k. Moreover, the spectral distribution is massively concentrated in a small number of eigenvalues (ranging between 2 and 4 for high symmetry minima) which accounts for $k^2 - O(k)$ of the spectrum. Similar results hold if $d > k$.*

The theorem is a consequence of the unique *isotypic* decomposition of the Hessian that derives from the invariance properties of $\mathcal{L}$ (see Theorem 4). Using stability arguments, it follows that upon convergence, the spectral density is expected to accumulate in clusters whose number does not depend on k. This is confirmed by empirical results which we provide in section F.1.

Next, we instantiate our framework to the global minima and three families of spurious local minima introduced in [24], referred to as types A, I and II (type II corresponds to the spurious minima described for $6 \leq k \leq 20$ in [22]). A complete description of the minima is provided in Lemma 5 (type II) and in Section D.2 (type A and I).

Theorem 2. *Assuming a $k \times k$ orthogonal target matrix $\boldsymbol{V}$, and $k \geq 6$,*

1. *$\nabla^2\mathcal{L}$ at $\boldsymbol{W} = \boldsymbol{V}$ has 6 distinct strictly positive eigenvalues:*
   (a) $\frac{1}{4} - \frac{1}{2\pi}$ *of multiplicity* $\frac{k(k-1)}{2}$.
   (b) $\frac{1}{4} + \frac{1}{2\pi}$ *of multiplicity* $\frac{k(k-3)}{2}$.
   (c) $\frac{k+1}{4} + O(k^{-1})$ *and* $\frac{1}{4} + O(k^{-1})$ *of multiplicity* $k - 1$.
   (d) $\approx -0.3471 + \frac{k}{2\pi} + O(k^{-1})$ *and* $\approx 0.8471 + \frac{k}{4} + O(k^{-1})$ *of multiplicity one.*
   (e) *The objective value is* 0.
2. *$\nabla^2\mathcal{L}$ at type A spurious local minima has 7 distinct strictly positive eigenvalues:*
   (a) $\frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$ *of multiplicity* $\frac{(k-1)(k-2)}{2}$.
   (b) $\frac{1}{4} + \frac{1}{2\pi} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$ *of multiplicity* $\frac{k(k-3)}{2}$.
   (c) *3 eigenvalues,* $\frac{k+1}{4} + O(k^{-1/2})$, $\frac{1}{4} + O(k^{-1/2})$ *and* $\frac{1}{4} - \frac{1}{2\pi} + O(k^{-1/2})$ *of multiplicity* $k - 1$.
   (d) *2 eigenvalues:* $c_1 + \frac{k}{4} + O(k^{-1/2})$ *and* $c_2 + \frac{k}{2\pi} + O(k^{-1/2})$ *of multiplicity one,* $c_1, c_2 > 0$.
   (e) *The objective value is* $(\frac{1}{2} - \frac{1}{\pi}) + O(k^{-1/2})$ *[24].*
3. *$\nabla^2\mathcal{L}$ at type II spurious local minima has 12 distinct strictly positive eigenvalues:*
   (a) $\frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-3/2})$ *of multiplicity* $\frac{(k-2)(k-3)}{2}$.
   (b) $\frac{1}{4} + \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-3/2})$ *of multiplicity* $\frac{(k-1)(k-4)}{2}$.
   (c) *5 Eigenvalues of multiplicity $k - 2$, of which one grows at a rate of $\frac{k+1}{4} + O(k^{-1})$, and the rest converge to small constants.*
   (d) *5 Eigenvalues of multiplicity 1, of which 2 grow at a rate of $c_3 + \frac{k}{4} + O(k^{-1})$, one grows at a rate of $c_4 + \frac{k}{2\pi} + O(k^{-1})$, $c_3, c_4 > 0$, and the rest converge to small constants.*
   (e) *The objective value is* $(\frac{1}{2} - \frac{2}{\pi^2})k^{-1} + O(k^{-3/2})$ *[24].*

If $d > k$, there will be 2 (resp. 3) additional strictly positive eigenvalues for type A (resp. I or II) minima with total multiplicity $(d - k)k$. The full description, together with that for type I eigenvalues, is given in Section A.

We note that methods for establishing the existence of spurious local minima for $\mathcal{L}$ are computer-aided and applicable only for small-scale problems [22]. Our method establishes the existence of spurious local minimum analytically and for arbitrarily large k and d (assuming $k \leq d$). An additional consequence of Theorem 2 is that *not all local minima are alike*. Below, we discuss the implications of the similarities and the differences between families of minima of $\mathcal{L}$.

Positively-skewed Hessian spectral density. Although first reported nearly 30 years ago [25], to the best of our knowledge, this is the first time that this phenomenon of extremely skewed spectral density has been established *rigorously* for high-dimensional problems (see Figure 1). Early empirical studies of the Hessian spectrum [25] revealed that local minima tend to be extremely ill-conditioned. This intriguing observation was corroborated and further refined in a series of works [26, 27, 28] which studied how the spectrum evolves along the training process. It was noticed that, upon convergence, the spectral density decomposes into two parts: a bulk of eigenvalues concentrated around zero, and a small set of positive outliers located away from zero.

Due to the high computational cost of an exact computation of the Hessian spectrum ($O(k^3d^3)$ for a $k \times d$ weight matrix), this phenomenon of extremely skewed spectral densities has only been confirmed for small-scale networks. Other methods for extracting second-order information in large-scale problems roughly fall into two general categories. The first class of methods approximate the Hessian spectral density by employing various numerical estimation techniques, most notably stochastic Lanczos method (e.g., [2, 8]). These methods have provided various numerical evidences that indicate that a similar skewed spectrum phenomenon also occurs in full-scale modern neural networks. The second class of techniques builds on tools from random matrix theory. This approach yields an exact computation of the limiting spectral distribution (i.e., the number of neurons is taken to infinity), assuming the inputs, as well as the model weights are drawn at random [9, 10, 11]. In contrast, our method gives an exact description of the spectral density for essentially any (finite) number of neurons, and at critical points rather than randomly drawn weight matrices.

The flat minima conjecture and implicit bias. It has long been debated whether some notion of local curvature can be used to explain the remarkable generalization capabilities of modern neural networks [3, 5, 6, 29, 30, 4, 7]. One intriguing hypothesis suggests that minima with wider basins of attraction tend to generalize better. An intuitive possible explanation is that flat minima promote statistical and numerical stability; together with low empirical loss, these ingredients are widely-used to achieve good generalization, cf. [31].

Perhaps surprisingly, our analysis shows that the spectra of global minima and the spurious minima considered in Theorem 2 agree on $k^2 - O(k)$ out of k^2 eigenvalues to within $O(k^{-1/2})$-accuracy ($d = k$). Thus, only the remaining $O(k)$ can potentially account for any essential difference in the local curvature. However, for type A spurious minima, even the remaining $O(k)$ eigenvalues are $O(k^{-1/2})$-far from the spectrum of the global minima. Consequently, in our settings, local second-order curvature *cannot* be used to separate global minima from spurious minima, thus ruling out notions of 'flatness' which rely exclusively on the Hessian spectrum. Of course, other metrics of a 'wideness of basins' may well apply.

Despite being a striking counter-example for a spectral-based notion of flatness, we note that, empirically, under Xavier initialization [32], type A spurious minima are rarely detected by SGD [24]. This stands in sharp contrast to type II minima to which SGD converges with a substantial empirical probability. Thus, for reasons which are yet to be understood, the bias induced by Xavier initialization seems to favor the class of global and type II minima at which the objective value decays with k to zero, rather than type A and type I minima whose objective value converges to strictly positive constants, cf., [33, 34]. We leave further study of this phenomenon, as well as other families of spurious minima, to future work.

Proof technique. Conceptually, the derivation of the eigenvalue estimate in Theorem 2 is based on ideas originating in symmetry-breaking, equivariant bifurcation theory and representation theory. Group invariance properties of the loss function (1.1) imply that the Hessian at symmetric points (under a proper notion of symmetry) must exhibit a certain block structure, and this makes possible an explicit computation of the Hessian spectrum. Empirically, and somewhat miraculously, spurious minima of (1.1) tend to be highly symmetric. As a consequence, their Hessian can be simplified using the same symmetry-based methods. The reminder of the paper is devoted to a formal and more detailed exposition of this approach.

3 The method: a symmetry-based analysis of the Hessian

In order to avoid a long preliminaries section, key ideas and concepts are introduced and organized so as to illuminate our strategy for analyzing the Hessian. We illustrate with reference to the case of global minima where $d = k$ and the target weight matrix $\boldsymbol{V}$ is the identity $\boldsymbol{I}_k$.

3.1 Studying invariance properties via group action

We first review background material on group actions and fix notations (see [35, Chapters 1, 2] for a more complete account). Elementary concepts from group theory are assumed known. We start with two examples that are used later.

Examples 1. (1) The *symmetric group* S_d, $d \in \mathbb{N}$, is the group of permutations of $[d] \doteq \{1, \ldots, d\}$.
(2) Let $\mathrm{GL}(d, \mathbb{R})$ denote the space of invertible linear maps on $\mathbb{R}^d$. Under composition, $\mathrm{GL}(d, \mathbb{R})$ has the structure of a group. The *orthogonal group* $\mathrm{O}(d)$ is the subgroup of $\mathrm{GL}(d, \mathbb{R})$ defined by

$\mathrm{O}(d) = \{A \in \mathrm{GL}(d, \mathbb{R}) \mid \|Ax\| = \|x\|, \text{ for all } x \in \mathbb{R}^d\}$. Both $\mathrm{GL}(d, \mathbb{R})$ and $\mathrm{O}(d)$ can be viewed as groups of invertible $d \times d$ matrices.

Characteristically, these groups consist of *transformations* of a set and so we are led to the notion of a *G-space* X where we have an *action* of a group G on a set X. Formally, this is a group homomorphism from G to the group of bijections of X. For example, S_d naturally acts on $[d]$ as permutations and both $\mathrm{GL}(d, \mathbb{R})$ and $\mathrm{O}(d)$ act on $\mathbb{R}^d$ as linear transformations (or matrix multiplication).

An example, which we use extensively in studying the invariance properties of $\mathcal{L}$, is given by the action of the group $S_k \times S_d \subset S_{k \times d}$, $k, d \in \mathbb{N}$, on $[k] \times [d]$ defined by

$$(\pi, \rho)(i, j) = (\pi^{-1}(i), \rho^{-1}(j)), \ \pi \in S_k, \rho \in S_d, \ (i, j) \in [k] \times [d]. \tag{3.2}$$

This action induces an action on the space $M(k, d)$ of $k \times d$-matrices $A = [A_{ij}]$ by $(\pi, \rho)[A_{ij}] = [A_{\pi^{-1}(i), \rho^{-1}(j)}]$. The action can be defined in terms of permutation matrices but is easier to describe in terms of rows and columns: $(\pi, \rho)A$ permutes rows (resp. columns) of A according to π (resp. ρ). As mentioned in the introduction, for our choice of $\boldsymbol{V} = \boldsymbol{I}_k$, $\mathcal{L}$ is $S_k \times S_d$-invariant. If $d = k$, define the *diagonal subgroup* ΔS_k of $S_k \times S_k$ by $\Delta S_k = \{(g, g) \mid g \in S_k\}$. Note that $\Delta S_k \approx S_k$. When we restrict the $S_k \times S_k$-action on $M(k, k)$ to ΔS_k, we refer to the diagonal S_k-action, or just the S_k-action on $M(k, k)$. This action of S_k on $M(k, k)$ maps diagonal matrices to diagonal matrices and should not be confused with the actions of S_k on $M(k, k)$ defined by either permuting rows or columns.

Example 2. Take $p, q \in \mathbb{N}$, $p + q = k$, and consider the diagonal action of $S_p \times S_q \subset S_k$ on $M(k, k)$. Write $A \in M(k, k)$ in block matrix form as $A = \begin{bmatrix} A_{p,p} & A_{p,q} \\ A_{q,p} & A_{q,q} \end{bmatrix}$. If $(g, h) \in S_p \times S_q \subset S_k$, then $(g, h)A = \begin{bmatrix} gA_{p,p} & (g, h)A_{p,q} \\ (h, g)A_{q,p} & hA_{q,q} \end{bmatrix}$ where $gA_{p,p}$ (resp. $hA_{q,q}$) are defined via the diagonal action of S_p (resp. S_q) on $A_{p,p}$ (resp. $A_{q,q}$), and $(g, h)A_{p,q}$ and $(h, g)A_{q,p}$ are defined through the natural action of $S_p \times S_q$ on rows and columns. Thus, for $(g, h)A_{p,q}$ (resp. $(h, g)A_{q,p}$) we permute rows (resp. columns) according to g and columns (resp. rows) according to h. In the case when $p = k - 1$, $q = 1$, S_{k-1} will act diagonally on $A_{k-1,k-1}$, fix a_{kk}, and act by permuting the first $(k - 1)$ entries of the last row and column.

Given $\boldsymbol{W} \in M(k, k)$, the largest subgroup of $S_k \times S_k$ fixing $\boldsymbol{W}$ is called the *isotropy* subgroup of $\boldsymbol{W}$ and is used as means of measuring the symmetry of $\boldsymbol{W}$. The isotropy subgroup of $\mathbf{V} \in M(k, k)$ is the diagonal subgroup ΔS_k. Our focus will be on critical points $\boldsymbol{W}$ whose isotropy groups are subgroups of the target matrix $\boldsymbol{V} = \boldsymbol{I}_k$, that is, ΔS_k and ΔS_{k-1} (see Figure 2—we use the notation ΔS_k as the isotropy is a *subgroup* of $S_k \times S_k$). Other choices of target matrices yield different symmetry-breaking of the isotropy of the global minima (see [23] for more details). In the next section, we show how the symmetry of local minima greatly simplifies the analysis of their Hessian.

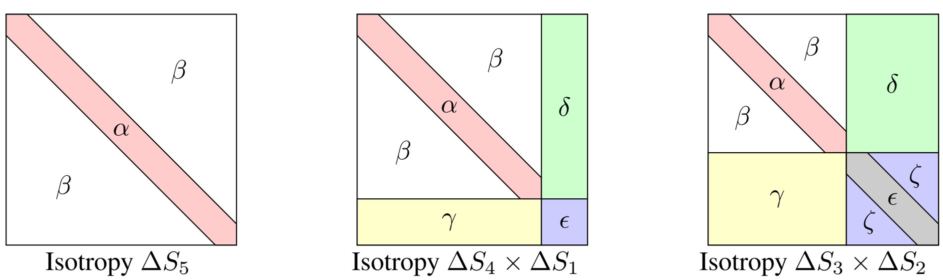

Figure 2: A schematic description of 5×5 matrices with isotropy $\Delta S_5, \Delta S_4 \times \Delta S_1$ and $\Delta S_3 \times S_2$, from left to right (borrowed from [23]). $\alpha, \beta, \gamma, \delta, \epsilon$ and ζ are assumed to be 'sufficiently' different.

3.2 The spectrum of equivariant linear isomorphisms

If G is a subgroup of $\mathrm{O}(d)$, the action on $\mathbb{R}^d$ is called an *orthogonal* representation of G (we often drop the qualifier orthogonal). Denote by $(\mathbb{R}^d, G)$ as necessary. The *degree* of a representation (V, G) is the dimension of V (V will always be a linear subspace of some $\mathbb{R}^n$ with the induced Euclidean inner product). The action of $S_k \times S_d \subset S_{k \times d}$ on $M(k, d)$ is orthogonal with respect to the standard

Euclidean inner product on $M(k,d) \approx \mathbb{R}^{k\times d}$ since the action permutes the coordinates of $\mathbb{R}^{k\times d}$ (equivalently, components of $k \times d$ matrices).

Given two representations (V,G) and (W,G), a map $A : V \to W$ is called G-equivariant if $A(gv) = gA(v)$, for all $g \in G, v \in V$. If A is linear and equivariant, we say A is a *G-map*. Invariant functions naturally provide examples of equivariant maps. Thus the gradient $\nabla\mathcal{F}$ is a $S_k \times S_d$-equivariant self map of $M(k,d)$ and if $\boldsymbol{W}$ is a critical point of $\nabla\mathcal{F}$ with isotropy $G \subset S_k \times S_d$, then $\nabla^2\mathcal{F}(\boldsymbol{W}) : M(k,d) \to M(k,d)$ is a G-map (see [35, 24]). The equivariance of the Hessian is the key ingredient that allows us to study the spectral density at *symmetric* local minima.

A representation $(\mathbb{R}^n, G)$ is *irreducible* if the only linear subspaces of $\mathbb{R}^n$ that are preserved (invariant) by the G-action are $\mathbb{R}^n$ and $\{0\}$. Two orthogonal representations (V,G), (W,G) are *isomorphic* (and have the same *isomorphism class*) if there exists a G-map $A : V \to W$ which is a linear isomorphism. If (V,G), (W,G) are irreducible but not isomorphic then every G-map $A : V \to W$ is zero (as the kernel and the image of a G-map are G-invariant). If (V,G) is irreducible, then the space $\mathrm{Hom}_G(V,V)$ of G-maps (endomorphisms) of V is a real associative division algebra and is isomorphic by a theorem of Frobenius to either $\mathbb{R}, \mathbb{C}$ or $\mathbb{H}$ (the quaternions). The *only* case that will concern us here is when $\mathrm{Hom}_G(V,V) \approx \mathbb{R}$ when we say the representation is *real*.

Example 3. Let $n > 1$. Take the natural (orthogonal) action of S_n on $\mathbb{R}^n$ defined by permuting coordinates. The representation is not irreducible since the subspace $T = \{(x,x,\cdots,x) \in \mathbb{R}^n \mid x \in \mathbb{R}\}$ is invariant by the action of S_n, as is the hyperplane $H_{n-1} = T^\perp = \{(x_1,\cdots,x_n) \mid \sum_{i\in[n]} x_i = 0\}$. It is easy to check that (T, S_n), also called the *trivial* representation of S_n, and (H_{n-1}, S_n), the *standard* representation, are irreducible, real, and not isomorphic.

Every representation $(\mathbb{R}^n, G)$ can be written uniquely, up to order, as an orthogonal direct sum $\oplus_{i\in[m]} V_i$, where each (V_i, G) is an orthogonal direct sum of isomorphic irreducible representations (V_{ij}, G), $j \in [p_i]$, and (V_{ij}, G) is isomorphic to $(V_{i'j'}, G)$ if and only if $i' = i$. The subspaces V_{ij} are *not* uniquely determined if $p_i > 1$. If there are m distinct isomorphism classes $\mathfrak{v}_1, \cdots, \mathfrak{v}_m$ of irreducible representations, then $(\mathbb{R}^n, G)$ may be represented by the sum $p_1\mathfrak{v}_1 + \cdots + p_m\mathfrak{v}_m$, where $p_i \geq 1$ counts the number of representations with isomorphism class $\mathfrak{v}_i$. Up to order, this sum (that is, the $\mathfrak{v}_i$ and their multiplicities) is uniquely determined by $(\mathbb{R}^n, G)$. This is the *isotypic decomposition* of $(\mathbb{R}^n, G)$ (see [36] and Section B). The isotypic decomposition is a powerful tool for extracting information about the spectrum of G-maps.

If $G = S_k$, then every irreducible representation of S_k is real [37, Thm. 4.3]. Suppose, as above, that $(\mathbb{R}^n, S_k) = \oplus_{i\in[m]} V_i$ and $A : \mathbb{R}^n \to \mathbb{R}^n$ is an S_k-map. Since the induced maps $A_{ii'} : V_i \to V_{i'}$ must be zero if $i \neq i'$, A is uniquely determined by the S_k-maps $A_{ii} : V_i \to V_i$, $i \in [m]$. Fix i and choose an S_k-representation (W, S_k) in the isomorphism class $\mathfrak{v}_i$. Choose S_k-isomorphisms $W \to V_{ij}$, $j \in [p_i]$. Then A_{ii} induces $\overline{A}_{ii} : W^{p_i} \to W^{p_i}$ and so determines a (real) matrix $M_i \in M(p_i, p_i)$ since $\mathrm{Hom}_{S_k}(W,W) \approx \mathbb{R}$. Different choices of V_{ij}, or isomorphism $W \to V_{ij}$, yield a matrix similar to M_i. Each eigenvalue of M_i of multiplicity r gives an eigenvalue of A_{ii}, and so of A, of multiplicity $r\,\mathrm{degree}(\mathfrak{v}_i)$.

Fact 1. (Notations and assumptions as above.) If A is the Hessian, all eigenvalues are real and each eigenvalue of M_i of multiplicity r will be an eigenvalue of A with multiplicity $r\,\mathrm{degree}(\mathfrak{v}_i)$. In particular, A has most $\sum_{i\in[m]} p_i$ distinct real eigenvalues—regardless of the dimension of the underlying space.

Our strategy can be now summarized as follows. Given a local minima $\boldsymbol{W}$, we compute the isotropy group $G \subset S_k \times S_d$ of $\boldsymbol{W}$. Since the Hessian of $\mathcal{F}$ at $\boldsymbol{W}$ is a G-map, may use the isotypic decomposition of the action of G on $M(k,d)$ to extract the spectral properties of the Hessian. In our setting, local minima have large isotropy groups, typically, as large as $\Delta(S_p \times S_{k-p})$, $0 \leq p < k/2$. Studying the Hessian at these minima requires the isotopic decomposition corresponding to $\Delta(S_p \times S_{k-p})$, $0 \leq p < k/2$, which we detail in Theorem 4 below.

3.3 The isotypic decomposition of $(M(k,k), S_k)$ and the spectrum at $\mathbf{W} = \mathbf{V}$

Regard $M(k,k)$ as an S_k-space (diagonal action). The trivial representation, denoted by $\mathfrak{t}_k$, and the standard representation, denoted by $\mathfrak{s}_k$, introduced in Example 3 are examples of the many irreducible representations of S_k. In the general theory, each irreducible representation of S_k is associated to a partition of the set $[k]$. The description of the isotypic decomposition of $(M(k,k), S_k)$ is relatively simple and uses just 4 irreducible representations of S_k for $k \geq 4$.

- The trivial representation $\mathfrak{t}_k$ of degree 1.
- The standard representation $\mathfrak{s}_k$ of S_k of degree $k-1$.
- The exterior square representation $\mathfrak{x}_k = \wedge^2 \mathfrak{s}_k$ of degree $\frac{(k-1)(k-2)}{2}$.
- A representation $\mathfrak{y}_k$ of degree $\frac{k(k-3)}{2}$. We describe $\mathfrak{y}_k$ explicitly later in terms of symmetric matrices (formally, it is the representation associated to the partition $(k-2,2)$).

We omit the subscript k when clear from the context. Assume that $k \geq 4$. We begin with a well-known result about the representation $\mathfrak{s} \otimes \mathfrak{s}$ (see, e.g., [37]). If $\mathfrak{s} \odot \mathfrak{s}$ denotes the symmetric tensor product of $\mathfrak{s}$, then

$$\mathfrak{s} \otimes \mathfrak{s} = \mathfrak{s} \odot \mathfrak{s} + \mathfrak{x} = \mathfrak{t} + \mathfrak{s} + \mathfrak{y} + \mathfrak{x}. \tag{3.3}$$

Since all the irreducible S_k-representations are real, they are isomorphic to their dual representations and so we have the isotypic decomposition

$$M(k,k) \quad \approx \quad \mathbb{R}^k \otimes \mathbb{R}^k \approx (\mathfrak{s} + \mathfrak{t}) \otimes (\mathfrak{s} + \mathfrak{t}) = 2\mathfrak{t} + 3\mathfrak{s} + \mathfrak{x} + \mathfrak{y}, \tag{3.4}$$

since $\mathfrak{t} \otimes \mathfrak{s} = \mathfrak{s}$ and $\mathfrak{t} \otimes \mathfrak{t} = \mathfrak{t}$.

Using Fact 1, information can immediately be deduced from Equation (3.4). For example, if $\boldsymbol{W}$ is a critical point of isotropy ΔS_k (a fixed point of the S_k-action on $M(k,k)$), then the spectrum of the Hessian contains at most $2+3+1+1=7$ distinct eigenvalues which distribute as follows: $\mathfrak{t}$ contributes 2 eigenvalues of multiplicity 1, $\mathfrak{s}$ contributes 2 eigenvalues of multiplicity $k-1$, $\mathfrak{x}$ contributes one eigenvalue of multiplicity $\frac{(k-1)(k-2)}{2}$, and $\mathfrak{y}$ contributes one eigenvalue of multiplicity $\frac{k(k-3)}{2}$. This applies to the global minimum $\boldsymbol{W} = \mathbf{V}$ and the spurious minimum of type A.

Next, we would like to compute the actual eigenvalues. We demonstrate the method for the single $\mathfrak{x}$-eigenvalue. Pick a non-zero vector from the $\mathfrak{x}$-representation. For example,

$$\mathfrak{X}^k = \begin{bmatrix} 0 & 1 & \dots & 1 & -(k-2) \\ -1 & 0 & \dots & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & \dots & 0 & 1 \\ (k-2) & -1 & \dots & -1 & 0 \end{bmatrix},$$

where rows and columns sum to zero and the only non-zero entries are in rows and columns 1 and k. Let $\overline{\mathfrak{X}^k} \in \mathbb{R}^{k \times k}$ be defined by concatenating the rows of $\mathfrak{X}^k$. Since $\mathfrak{x}$ only occurs once in the isotopic decomposition and $\nabla^2 \mathcal{L}(\boldsymbol{V})$ is S_k-equivariant, $\overline{\mathfrak{X}^k}$ must be an eigenvector. In particular, $(\nabla^2 \mathcal{L}(\boldsymbol{V}) \overline{\mathfrak{X}^k})_i = \lambda_{\mathfrak{x}} \overline{\mathfrak{X}^k_i}$, all $i \in [k^2]$. Choose i so that $\overline{\mathfrak{X}^k_i} \neq 0$. For example, $\overline{\mathfrak{X}^k_2} = 1$. Matrix multiplication, yields $\lambda_{\mathfrak{x}} = 1/4 - 1/2\pi$ (see Section C for expressions for the Hessian entries).

A similar analysis holds for the eigenvalue associated to $\mathfrak{y}$. The multiple factors $2\mathfrak{t}$ and $3\mathfrak{s}$ are handled by making judicious choices of orthogonal invariant subspaces and representative vectors in $M(k,k)$. A complete derivation of all the eigenvalues, including a detailed list of the representative vectors and expressions for the Hessian of $\mathcal{L}$ at $\boldsymbol{V}$, are provided in the appendix.

4 The Hessian spectrum at spurious minima

Having described the general strategy for analyzing the Hessian spectrum for global minima, we now examine the spectrum at various types of spurious minima. We need two additional ingredients: a specification of the entries of a given family of spurious minima and the respective isotypic decomposition; we begin with the latter.

As discussed in the introduction, the symmetry-based analysis of the Hessian relies on the fact that isotropy groups of spurious minima tend to be (and some provably are) maximal subgroups of the target matrix isotropy. For $\boldsymbol{V} = \boldsymbol{I}$, the relevant maximal isotropy groups are of the form $\Delta(S_p \times S_q)$, $p+q=k$. Below, we provide the corresponding isotypic decompositions. Assume $d = k$ and regard $M(k,k)$ as an $S_p \times S_q$-space, where $S_p \times S_q \subset S_k$ and the (diagonal) action of S_k is restricted to the subgroup $S_p \times S_q$.

Theorem 4. *The isotypic decomposition of $(M(k,k), S_p \times S_q)$ is given by:*

1. *If $p = k-1$, $q = 1$, and $k \geq 5$,*
$$M(k,k) = 5\mathfrak{t} + 5\mathfrak{s}_{k-1} + \mathfrak{x}_{k-1} + \mathfrak{y}_{k-1}.$$

2. *If $q \geq 2$, $k-1 > p > p/2$ and $k \geq 4+q$, then*
$$M(k,k) = 6\mathfrak{t} + 6\mathfrak{s}_p + a\mathfrak{s}_q + \mathfrak{x}_p + \mathfrak{y}_p + b\mathfrak{x}_q + c\mathfrak{y}_q + 2\mathfrak{s}_p \boxtimes \mathfrak{s}_q,$$
where if $q = 2$, then $a = 4, b = c = 0$; if $q = 3$, then $a = 5, b = 1, c = 0$; and if $q \geq 4$, then $a = 6, b = c = 1$.

Theorem 4 implies that the Hessian spectrum of local minima (or critical points) with isotropy $\Delta(S_p \times S_q)$ has at most 12 distinct eigenvalues if (1) applies, and if (2) holds, at most 19 distinct eigenvalues if $q = 2$, at most 21 distinct eigenvalues if $q = 3$, and at most 22 distinct eigenvalues if $q \geq 4$. Moreover, $k^2 - O(k)$ of the k^2 eigenvalues (counting multiplicity) are the $\mathfrak{x}$- and $\mathfrak{y}$-eigenvalues. We omit some less interesting cases when k is small.

Following the same lines of argument described in Section 3.3, our goal is to pick a set of non-zero vectors for each irreducible representation that will allow us to compute the spectrum. While this is simple, estimating the Hessian is not trivial. For this, we need good estimates on the critical points determining the spurious local minima.

In a recent work [24], three infinite families of critical points were described: type A of isotropy ΔS_k, and types I and II of isotropy ΔS_{k-1}. These relatively large isotropy groups made it possible to derive power series in $1/\sqrt{k}$ for the critical points and compute the initial terms. Estimates resulting from these series allow us get sharp estimates on the Hessian which in turn lead to sharp estimates on eigenvalues. The derivation is lengthy and quite technical and is therefore deferred to the appendix. As an illustration of the method, we sketch the derivation of the $\mathfrak{x}$-eigenvalue estimate for the family of type II local minima (case 1 in Theorem 4).

Briefly, if $(\mathfrak{c}_k)_{k \geq 3}$ denotes the sequence of type II critical points of $\mathcal{F}$, then we may represent $\mathfrak{c}_k$ as a point in $M(k,k)^{S_{k-1}} = \{\boldsymbol{W} \mid g\boldsymbol{W} = \boldsymbol{W}, g \in S_{k-1}\}$—the 5-dimensional fixed point space of the (diagonal) action of S_{k-1} on $M(k,k)$. If $\mathfrak{c}_k = (\xi_1^k, \xi_2^k, \xi_3^k, \xi_4^k, \xi_5^k) \in M(k,k)^{S_{k-1}}$, then $\mathfrak{c}_k$ corresponds to $\boldsymbol{W} = [w_{ij}] \in M(k,k)$ where

$$w_{ii} = \begin{cases} \xi_1^k, \ i < k \\ \xi_5^k, \ i = k \end{cases}, \quad w_{ij} = \begin{cases} \xi_2^k, \ i,j < k, \ i \neq j \\ \xi_4^k, \ i < k = j \\ \xi_3^k, \ j < k = i \end{cases}.$$

Lemma 5 ([24, Section 8])**.** *(Notation and assumptions as above.) For large enough k, $\mathfrak{c}_k$ may be written as a convergent power series in $k^{-\frac{1}{2}}$:*

$$\xi_1^k = 1 + \sum_{\ell=4}^{\infty} c_\ell k^{-\ell/2}, \quad \xi_2^k = \sum_{\ell=4}^{\infty} e_\ell k^{-\ell/2}, \quad \xi_3^k = \sum_{\ell=2}^{\infty} f_\ell k^{-\ell/2},$$
$$\xi_4^k = \sum_{\ell=2}^{\infty} g_\ell k^{-\ell/2}, \quad \xi_5^k = -1 + \sum_{\ell=2}^{\infty} d_\ell k^{-\ell/2},$$

where

$$c_4 = \tfrac{8}{\pi}, \qquad d_2 = 2 + \tfrac{8\pi+8}{\pi^2}, \ e_4 = -\tfrac{4}{\pi}, \ f_2 = 2, \ g_2 = -e_4,$$
$$c_5 = -\tfrac{320\pi}{3\pi^4(\pi-2)}, \ d_3 = \tfrac{64\pi-768}{3\pi^4(\pi-2)}, \ e_5 = -\tfrac{32}{\pi^3}, \ f_3 = 0, \ g_3 = -e_5.$$

Proceeding with the lines of argument described in Section 3.3, we use these power series for $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5$ to derive estimate for the Hessian entries (see Table 1), which in turn give:

$$\begin{aligned}\left(\nabla^2 \mathcal{L}(\mathfrak{c}_k)\overline{\mathfrak{X}^{k-1}}\right)_2 &= (H_{22}^{11} - H_{23}^{11} - H_{33}^{12} + 2H_{13}^{12} - H_{12}^{12})\overline{\mathfrak{X}^{k-1}} \\ &= \left(\frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2})\right)\overline{\mathfrak{X}^{k-1}}_2,\end{aligned}$$

Hessian Entry	Estimate	Hessian Entry	Estimate
H^{11}_{22}	$\frac{1}{2} - \frac{1}{\pi k} + O(k^{-2})$	H^{11}_{23}	$O(k^{-2})$
H^{12}_{33}	$\frac{1}{4} + O(k^{-2})$	H^{12}_{13}	$O(k^{-2})$
H^{12}_{12}	$\frac{1}{2\pi} + O(k^{-2})$		

Table 1: Estimates of the Hessian entries for type II critical points based on the formula provided in Section C and Lemma 5 below. H^{pq} denotes the (p, q)'th $k \times k$ block of the $k^2 \times k^2$ matrix $\nabla^2 \mathcal{L}(\mathfrak{c}_k)$.

showing that $\frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2})$ is an eigenvalue of $\nabla^2 \mathcal{L}(\mathfrak{c}_k)$ of multiplicity $\frac{(k-2)(k-3)}{2}$ (note that the computation implicitly relies on the symmetry of the entries of $\nabla^2 \mathcal{L}(\mathfrak{c}_k)$). The complete derivation of the eigenvalue estimates stated in Theorem 2 is provided in Sections A-E.

5 Conclusion

We exploit the presence of rich symmetry in ST two-layers ReLU models to derive an analytic characterization of the Hessian spectrum in the natural regime where the number of inputs and hidden neurons is finite. This allow us, for the first time, to rigorously confirm (and refute) various hypotheses regarding the mysterious generalization abilities of neural networks. The methods described in the paper apply more broadly [23], and yield different spectral properties for the Hessian that vary by the choice of the underlying distributions, activation functions and architectures. The approach we wish to put forward follow in the tradition of mathematics and physics in that we start with a symmetric model, for which we can prove detailed analytic results, and subsequently break symmetry to get insight into the general theory (since critical points are non-degenerate, the results we obtain are robust under symmetry breaking perturbations of $\mathbf{V}$ [35, 9.2]; see also [22, Cor. 1]).

Some of the results derived in this work seem to challenge several research directions. Although much effort has been invested in establishing conditions under which no spurious minima exist [38, 39, 40], we prove the existence of *infinite* families of spurious minima for a simple shallow ReLU model. The hope for nonconvex optimization landscapes with no spurious minima requires therefore further refinement, at least for certain parameter regimes. Secondly, as demonstrated by type A and type II minima, not all local minima are alike. In particular, the hidden mechanism under which such spurious minima are alleviated may be somewhat different. Lastly, it is the authors' belief that a deep understanding of basic models, such as ST models, is a prerequisite for any general theory aimed at explaining the success of deep learning.

Acknowledgements

Part of this work was completed while YA was visiting the Simons Institute for the Foundations of Deep Learning program. We thank Amir Ofer, Itai Safran, Ohad Shamir, Michal Shavit and Daniel Soudry for valuable discussions. Thanks also to Bob Howlett, University of Sydney, for help with the representation theory of S_n.

Broader Impact

To the best of our knowledge, there are no ethical aspects or future societal consequences directly involved in our work.

References

[1] Zhanxing Zhu, Jingfeng Wu, Bing Yu, Lei Wu, and Jinwen Ma. The anisotropic noise in stochastic gradient descent: Its behavior of escaping from sharp minima and regularization effects. In *Proc. Int. Conf. Mach. Learn.*, pages 7654–7663, 2019.

[2] Behrooz Ghorbani, Shankar Krishnan, and Ying Xiao. An investigation into neural net optimization via hessian eigenvalue density. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 2232–2241. PMLR, 2019.

[3] Sepp Hochreiter and Jürgen Schmidhuber. Flat minima. *Neural Computation*, 9(1):1–42, 1997.

[4] Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, and Riccardo Zecchina. Entropy-sgd: Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12): 124018, 2019.

[5] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016.

[6] Stanisław Jastrzkebski, Zachary Kenton, Devansh Arpit, Nicolas Ballas, Asja Fischer, Yoshua Bengio, and Amos Storkey. Three factors influencing minima in sgd. *arXiv preprint arXiv:1711.04623*, 2017.

[7] Laurent Dinh, Razvan Pascanu, Samy Bengio, and Yoshua Bengio. Sharp minima can generalize for deep nets. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1019–1028. JMLR. org, 2017.

[8] Vardan Papyan. The full spectrum of deepnet hessians at scale: Dynamics with sgd training and sample size. *arXiv preprint arXiv:1811.07062*, 2018.

[9] Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems*, pages 2637–2646, 2017.

[10] Jeffrey Pennington and Pratik Worah. The spectrum of the fisher information matrix of a single-hidden-layer neural network. In *Advances in Neural Information Processing Systems*, pages 5410–5419, 2018.

[11] Cosme Louart, Zhenyu Liao, Romain Couillet, et al. A random matrix approach to neural networks. *The Annals of Applied Probability*, 28(2):1190–1248, 2018.

[12] Ian J. Goodfellow and Oriol Vinyals. Qualitatively characterizing neural network optimization problems. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[13] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. In *Advances in Neural Information Processing Systems*, pages 6389–6399, 2018.

[14] Felix Draxler, Kambis Veschgini, Manfred Salmhofer, and Fred A. Hamprecht. Essentially no barriers in neural network energy landscape. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 1308–1317. PMLR, 2018.

[15] Alon Brutzkus and Amir Globerson. Globally optimal gradient descent for a convnet with gaussian inputs. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 605–614. JMLR. org, 2017.

[16] Simon S. Du, Jason D. Lee, Yuandong Tian, Aarti Singh, and Barnabás Póczos. Gradient descent learns one-hidden-layer CNN: don't be afraid of spurious local minima. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 1338–1347, 2018.

[17] Yuanzhi Li and Yang Yuan. Convergence analysis of two-layer neural networks with relu activation. In *Advances in Neural Information Processing Systems*, pages 597–607, 2017.

[18] Soheil Feizi, Hamid Javadi, Jesse Zhang, and David Tse. Porcupine neural networks:(almost) all local optima are global. *arXiv preprint arXiv:1710.02196*, 2017.

[19] Qiuyi Zhang, Rina Panigrahy, Sushant Sachdeva, and Ali Rahimi. Electron-proton dynamics in deep learning. *arXiv preprint arXiv:1702.00458*, pages 1–31, 2017.

[20] Rong Ge, Jason D. Lee, and Tengyu Ma. Learning one-hidden-layer neural networks with landscape design. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[21] Yuandong Tian. An analytical formula of population gradient for two-layered relu network and its applications in convergence and critical point analysis. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3404–3413. JMLR. org, 2017.

[22] Itay Safran and Ohad Shamir. Spurious local minima are common in two-layer relu neural networks. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 4430–4438, 2018.

[23] Yossi Arjevani and Michael Field. On the principle of least symmetry breaking in shallow relu models. *arXiv preprint arXiv:1912.11939*, 2019.

[24] Yossi Arjevani and Michael Field. Symmetry & critical points for a model shallow neural network. *CoRR*, abs/2003.10576, 2020.

[25] Léon Bottou. Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nımes*, 91 (8):12, 1991.

[26] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.

[27] Levent Sagun, Leon Bottou, and Yann LeCun. Eigenvalues of the hessian in deep learning: Singularity and beyond. *arXiv preprint arXiv:1611.07476*, 2016.

[28] Levent Sagun, Utku Evci, V Ugur Guney, Yann Dauphin, and Leon Bottou. Empirical analysis of the hessian of over-parametrized neural networks. *arXiv preprint arXiv:1706.04454*, 2017.

[29] Lei Wu, Zhanxing Zhu, et al. Towards understanding generalization of deep learning: Perspective of loss landscapes. *arXiv preprint arXiv:1706.10239*, 2017.

[30] Zhewei Yao, Amir Gholami, Qi Lei, Kurt Keutzer, and Michael W Mahoney. Hessian-based analysis of large batch training and robustness to adversaries. In *Advances in Neural Information Processing Systems*, pages 4949–4959, 2018.

[31] Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. Learnability, stability and uniform convergence. *Journal of Machine Learning Research*, 11(Oct):2635–2670, 2010.

[32] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[33] Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *The Journal of Machine Learning Research*, 19(1):2822–2878, 2018.

[34] Suriya Gunasekar, Jason D Lee, Daniel Soudry, and Nati Srebro. Implicit bias of gradient descent on linear convolutional networks. In *Advances in Neural Information Processing Systems*, pages 9461–9471, 2018.

[35] Michael J. Field. *Dynamics and symmetry*, volume 3 of *ICP Advanced Texts in Mathematics*. Imperial College Press, London, 2007. ISBN 978-1-86094-828-2; 1-86094-828-6. doi: 10.1142/9781860948541.

[36] Charles Benedict Thomas. *Representations of finite and Lie groups*. World Scientific, 2004.

[37] William Fulton and Joe Harris. Representation theory, volume 129 of. *Graduate Texts in Mathematics*, 1991.

[38] Abbas Kazemipour, Brett Larsen, and Shaul Druckmann. No spurious local minima in deep quadratic networks. *CoRR*, abs/2001.00098, 2020.

[39] Daniel Soudry and Yair Carmon. No bad local minima: Data independent training error guarantees for multilayer neural networks. *CoRR*, abs/1605.08361, 2016.

[40] Rong Ge, Jason D. Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 2973–2981, 2016.

[41] Charles Benedict Thomas. *Representations of finite and Lie groups*. World Scientific, 2004.

[42] GD James. The representation theory of the symmetric groups. Springer, 1978.

A hitchhiker's guide to the appendix. The appendix is organized as follows. In Section A, we provide a description of the Hessian spectrum of type I spurious minima, as well as the additional eigenvalues which correspond to the $d > k$ case. This completes the statement of Theorem 2 given in the main paper. Next, we devote Section B to representation-theoretic preliminaries for the group action under consideration. Concretely, we compute the relevant isotypic decompositions and list our choice of representative vectors. In Section C, we use the symmetry of the Hessian (w.r.t. the group action) to simplify and specialize the generic expressions of the Hessian entries to the families of spurious minima considered in this paper. Once the ΔS_k-case is completed (see Section C.3), we show how to fully analyze the Hessian spectrum of global minima in a relatively simple way using symmetry (see Example 22). In Section D, the long groundwork laid in previous sections is put to use for deriving the Hessian spectrum of types A, I and II minima for $k = d$. The derivation of the additional $d > k$ case eigenvalues is presented in Section E. In Section F, we demonstrate the eigenvalue bulks phenomenon for perturbed minima, as discussed in the follow-up discussion of Theorem 1. We conclude with numerical estimates for the Hessian spectrum which we obtain through LinAlg, a linear algebra package of Python. The numerical results confirm our analytic characterization of the Hessian spectra.

A Type I Hessian spectrum and proof of Theorem 2

Below, we provide a description of the Hessian spectrum for type I spurious minima. This completes the statement of Theorem 2 given in the main paper.

Theorem 2 (Cont.). Assuming a $k \times k$ orthogonal target matrix $\boldsymbol{V}$, and $k \geq 6$, $\nabla^2 \mathcal{L}$ at type I spurious local minima has 12 distinct strictly positive eigenvalues:

1. $\frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$ of multiplicity $\frac{(k-2)(k-3)}{2}$.
2. $\frac{1}{4} + \frac{1}{2\pi} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$ of multiplicity $\frac{(k-1)(k-4)}{2}$.
3. 5 Eigenvalues of multiplicity $k-2$, of which one grows at a rate of $\frac{k+1}{4} + O(k^{-1})$, and the rest converge to small constants.
4. 5 Eigenvalues of multiplicity 1, of which 2 grow at a rate of $c_3 + \frac{k}{4} + o(1)$, one grows at a rate of $c_4 + \frac{k}{2\pi} + o(1)$, $c_3, c_4 > 0$, and the rest converge to small constants.

We extend Theorem 2 to allow for $d > k$. Recall that if $d > k$, we append $d - k$ zeros to the end of each row of $\mathbf{V}$ to define $\mathbf{V} \in M(k, d)$. We denote the resulting objective function by $\mathcal{F}_n$, where $n = d - k$ and so $\mathcal{F}_0 = \mathcal{F}$, the objective function of Theorem 2.

Theorem 6. *Assume the conditions of Theorem 2 and let $d > k \geq 6$. Set $n = d - k$. The sequence of spurious minima described in Lemma 5 uniquely determines a sequence $(\mathfrak{c}_k^n)$ of critical points defining spurious minima for $\mathcal{F}_n$ which have isotropy $\Delta S_{k-1} \times S_n$ (S_n permutes columns). In particular, $\mathcal{F}_n$ is real analytic at $\mathfrak{c}_k^n$, $k \geq 6$, and the spectrum of the Hessian of $\mathcal{F}_n$ will be the union of the spectrum of the Hessian of $\mathcal{F}_0$, together with 3 strictly positive eigenvalues $\lambda_1, \lambda_2, \lambda_3$ satisfying*

1. *λ_1 has multiplicity $n(k-2)$ and $\lambda_1 = \frac{1}{4} + O(k^{-1})$.*
2. *λ_2 has multiplicity n and $\lambda_2 = \frac{1}{4} + O(k^{-\frac{1}{2}})$.*
3. *λ_3 has multiplicity n and $\lambda_3 = \frac{k+1}{4} + O(k^{-1})$.*

For the spurious minima of Lemma 5, $\mathcal{F}_n(\mathfrak{c}_k^n) = (\frac{1}{2} - \frac{2}{\pi^2})k^{-1} + O(k^{-\frac{3}{2}})$.

The proof is given in Section E.

B The isotypic decomposition of $(M(k,k), G)$

In this section our aim is give, with minimal prerequisites, the results needed from the representation theory of the symmetric group. A little background in character theory would be helpful for checking a few statements (for example, showing specific representations of S_k are irreducible or real)—for this the introductory text [41] would suffice. The first three or four lectures in [37] give a good, but terse, introduction to the representation theory of S_k. There are many texts covering the general theory, for example [42], but a lot of work is often required to extract the information needed here. Moreover, the representation theory of S_k is special because the ground field can be taken to be $\mathbb{R}$ (or the rationals). Many introductory texts on representation theory work over the complex field: the proofs are often much easier but it is often awkward to translate to results over the real field.

B.1 The isotypic decomposition

We begin with a precise version of the orthogonal decomposition described in Section 3.2. Suppose $V \subset \mathbb{R}^m$ is a linear subspace, with Euclidean inner product induced from $\mathbb{R}^m$, and (V, G) is an orthogonal G-representation.

Lemma 7. *The representation (V, G) may be written as an orthogonal direct sum $\bigoplus_{i=1}^m (\oplus_{j=1}^{p_i} V_{ij})$ where $V_{ij} \subset V$, (V_{ij}, G) is irreducible, and (V_{ij}, G) is isomorphic to $(V_{\ell k}, G)$ iff $i = \ell$, and $j, k \in [p_i]$. The subspaces $\oplus_{j=1}^{p_i} V_{ij}$ are unique, $i \in [m]$.*

Proof Induction on $n = \dim(V)$. Trivial for $n = 1$. Assume proved for all representations of degree less than n. If (V, G) is of degree n either it is irreducible, and there is nothing to prove, or not. If not, there exists a proper G-invariant linear subspace V_1 of V. By the orthogonality of the action, $V_2 = V_1^\perp$ is G-invariant and so (V, G) is the orthogonal direct sum of representations (V_1, G) and (V_2, G). Apply the inductive hypothesis to (V_1, G) and (V_2, G). The proof of uniqueness is straightforward and we omit the details. $\square$

If $p_i = 1$, for all $i \in m$, the orthogonal decomposition given by the lemma is unique, up to order; otherwise the decomposition is not unique. For this reason, Theorem 4 was formulated in terms of isomorphism *classes* rather than in terms of specific subspaces.

In spite of the lack of uniqueness of Lemma 7, in some cases there may be *natural* choices of invariant subspace for the irreducible components. This is exactly the situation for the isotypic decomposition of $(M(k, k), G)$, $G = S_p \times S_{k-p}$, given in Theorem 4. This naturality allows us to give natural constructions of the matrices M_i, $i \in [m]$, used for determining the spectrum of G-maps $A : M(k, k) \to M(k, k)$.

Example 8. The isotypic decomposition for $(M(k, k), S_k)$ is $2\mathfrak{t} + 3\mathfrak{s} + \mathfrak{x} + \mathfrak{y}$, $k \geq 4$. The subspace of $M(k, k)$ determined by $2\mathfrak{t}$ is the set of all $k \times k$ matrices $\mathcal{T} = \{T_{a,b} \mid a, b \in \mathbb{R}\}$ where the diagonal entries of $T_{a,b}$ all equal a and the off-diagonal entries all equal b. There are many ways to write $\mathcal{T}$ as an orthogonal direct sum. For example, $\mathcal{T} = T_{1,1}\mathbb{R} \oplus T_{\frac{2}{k}, -\frac{1}{k(k-1)}}\mathbb{R}$. However, there is only one natural way: $\mathcal{T} = T_{1,0}\mathbb{R} \oplus T_{0,1}\mathbb{R}$. Define $\mathfrak{D}_1^k = T_{1,0}$, $\mathfrak{D}_2^k = T_{0,1}$. If we take the *standard* realization of $(\mathfrak{t}, S_k)$ to be $(\mathbb{R}, S_k)$, where S_k acts trivially on $\mathbb{R}$, then we have natural S_k-maps $\alpha_1, \alpha_2 : \mathbb{R} \to M(k, k)$ defined by $\alpha_i(t) = t\mathfrak{D}_i^k$, $i = 1, 2$. If $A : M(k, k) \to M(k, k)$ is an S_k-map, then A restricts to the S_k-map $A_\mathfrak{t} : \mathcal{T} \to \mathcal{T}$ and $A_\mathfrak{t}$ uniquely determines a 2×2-matrix $[a_{ij}]$ by $A_\mathfrak{t}(\mathfrak{D}_i^k) = a_{i1}\mathfrak{D}_1^k + a_{i2}\mathfrak{D}_2^k$, $i = 1, 2$. The eigenvalues (and multiplicities in this case) of $A_\mathfrak{t} : \mathcal{T} \to \mathcal{T}$ are the same as the eigenvalues of $[a_{ij}]$. If we choose a different orthogonal decomposition of $\mathcal{T}$, we get a different 2×2-matrix that is similar to $[a_{ij}]$ and so has the same eigenvalues.

In the isotypic decompositions of $M(k, k)$ we consider in detail here, only $\mathfrak{t}$ and $\mathfrak{s}$ occur with multiplicity greater than 1 (later we address the exterior tensor product representation $2\mathfrak{s}_p \boxtimes \mathfrak{s}_q$—but methods are the same). Before describing how we handle the factors $\mathfrak{s}$, we need a more explicit description of the representation $(M(k, k), S_k)$.

B.2 Decomposition of $(M(k, k), S_k)$ into spaces of matrices.

Assume $k \geq 4$ in what follows (results are easily obtained if $k \leq 3$ but are not interesting for our applications).

Let $\mathbb{D}_k$ denote the space of diagonal $k \times k$-matrices, $\mathbb{A}_k$ the space of skew-symmetric $k \times k$-matrices, and $\mathbb{S}_k$ the space of symmetric $k \times k$-matrices with diagonal entries zero. We have the orthogonal direct sum decomposition

$$M(k, k) = \mathbb{D}_k \oplus \mathbb{A}_k \oplus \mathbb{S}_k$$

Since S_k acts diagonally on $M(k, k)$, this direct sum is S_k-invariant.

Recall that $H_{k-1} \subset \mathbb{R}^k$ is the hyperplane $\sum_{i\in[k]} x_i = 0$. In Example 3, we defined (H_{k-1}, S_k) and (T, S_k) to be the standard and trivial representations of S_k. We write here $(\mathbb{R}, S_k)$, rather than (T, S_k), but caution that there is always at least one non-trivial representation of S_k on $\mathbb{R}$. However, these representations do not not occur here. View (H_{k-1}, S_k) and $(\mathbb{R}, S_k)$ as standard *models* or *realizations* of the isomorphism classes $\mathfrak{s}_k$ and $\mathfrak{t}$.

Lemma 9. *$\mathbb{D}_k$ is the orthogonal S_k-invariant direct sum $\mathbb{D}_{k,1} \oplus \mathbb{D}_{k,2}$, where*

1. *$\mathbb{D}_{k,1}$ is the space of diagonal matrices with all entries equal and is naturally isomorphic to (T, S_k).*
2. *$\mathbb{D}_{k,2}$ is the $(k-1)$-dimensional space of diagonal matrices with diagonal entries summing to zero and is naturally isomorphic to (H_{k-1}, S_k).*

In particular, the isotypic decomposition of $(\mathbb{D}, S_k)$ is $\mathfrak{t} + \mathfrak{s}_k$.

Proof For (1), define the S_k map $\mathbb{R}\to\mathbb{D}_{k,1}$ by $t \mapsto t\mathfrak{D}_1^k$ and for (2), map $(x_1, \cdots, x_k) \in H_{k-1}$ to the diagonal matrix D with entries $d_{ii} = x_i$, $i \in [k]$. □

The lemma gives a simple instance of natural choices of subspace in the isotopic decomposition as well as a natural choice of matrix $\mathfrak{D}_1^k \in \mathbb{D}_{k,1}$ corresponding to $1 \in T$ (we give a choice of matrix for $\mathbb{D}_{k,2}$ shortly).

Next we extend the previous lemma to $\mathbb{A}_k$ and $\mathbb{S}_k$ and give and define *explicit* matrices in the isotypic components.

Lemma 10. *$\mathbb{A}_k$ is the orthogonal S_k-invariant direct sum $\mathbb{A}_{k,1} \oplus \mathbb{A}_{k,2}$, where*

1. *$\mathbb{A}_{k,1}$ is the $(k-1)$-dimensional space of matrices $[a_{ij}]$ for which there exists $(x_1, \cdots, x_k) \in H_{k-1}$ such that for all $i, j \in [k]$, $a_{ij} = x_i - x_j$,*
2. *$\mathbb{A}_{k,2}$ consists of all skew-symmetric matrices with row sums zero.*

As representations, $(\mathbb{A}_{k,1}, S_k)$ is isomorphic to (H_{k-1}, S_k) and $(\mathbb{A}_{k,2}, S_k)$ is isomorphic to $(\wedge^2 H_{k-1}, S_k)$. In particular, the isotypic decomposition of $(\mathbb{A}_k, S_k)$ is $\mathfrak{s}_k + \mathfrak{x}_k$.

Proof The isotypic decomposition of $(\mathbb{A}_k, S_k)$ and irreducibility of the exterior square representation may be found in [42, 37]. Alternatively, use the explicit description and character theory to verify irreducibility. □

Lemma 11. *$\mathbb{S}_k$ is the orthogonal S_k-invariant direct sum $\mathbb{S}_{1,k} \oplus \mathbb{S}_{2,k} \oplus \mathbb{S}_{3,k}$, where*

1. *$\mathbb{S}_{1,k}$ is the 1-dimensional space of symmetric matrices with diagonal entries zero and all off diagonal entries equal.*
2. *$\mathbb{S}_{2,k}$ is the $(k-1)$-dimensional space of matrices $[a_{ij}] \in \mathbb{S}_k$ for which there exists $(x_1, \cdots, x_k) \in H_{k-1}$ such that for all $i, j \in [k]$, $i \neq j$, $a_{ij} = x_i + x_j$.*
3. *$\mathbb{S}_{3,k}$ consists of all symmetric matrices in $\mathbb{S}_k$ with all row (equivalently, column) sums zero.*
4. *dim$(\mathbb{S}_{3,k}) = \frac{k(k-3)}{2}$.*

The representations $(\mathbb{S}_{k,i}, S_k)$ are irreducible, $i \in [3]$: $(\mathbb{S}_{k,1}, S_k)$ is isomorphic to the trivial representation, $(\mathbb{S}_{k,2})$ is isomorphic to the standard representation and $(\mathbb{S}_{k,3}, S_k)$ is isomorphic to the S_k-representation associated to the partition $(k-2, 2)$ (isomorphism type $\mathfrak{y}_k$).

Proof It is straightforward to check the orthogonality, (1–4) and the S_k-invariance of the decomposition. The isotypic decomposition of $\mathbb{S}_{1,k} \oplus \mathbb{S}_{2,k}$ is $\mathfrak{t} + \mathfrak{s}_k$. It is known that the isotypic decomposition of $(\mathbb{S}_k, S_k)$ is $\mathfrak{t} + \mathfrak{s}_k + \mathfrak{y}_k$ [42, 37]. Since we have already identified the factors $\mathfrak{t}, \mathfrak{s}_k$, $(\mathbb{S}_{k,3}, S_k)$ has isomorphism type $\mathfrak{y}_k$. Alternatively, use the explicit description of $(\mathbb{S}_{k,3}, S_k)$ and character theory to verify irreducibility—which is all we need. □

B.3 The general method

We have now identified three sub-representations in $(M(k,k), S_k)$ that are isomorphic to the standard representation (H_{k-1}, S_k). Moreover lemmas 9, 10, and 11 give explicit parametrizations of the representations in terms of the standard representation (H_{k-1}, S_k). Choose a non-zero vector in H_{k-1}, for example $(1, -1, 0, \cdots, 0)$. Denote the corresponding elements in $\mathbb{D}_k$, $\mathbb{A}_k$ and $\mathbb{S}_k$ by $\mathfrak{S}_1^k$, $\mathfrak{S}_2^k$ and $\mathfrak{S}_3^k$ respectively. Then

$$\mathfrak{S}_1^k = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix},$$

$$\mathfrak{S}_2^k = \begin{bmatrix} 0 & 2 & 1 & \dots & 1 & 1 \\ -2 & 0 & -1 & \dots & -1 & -1 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad \mathfrak{S}_3^k = \begin{bmatrix} 0 & 0 & 1 & \dots & 1 & 1 \\ 0 & 0 & -1 & \dots & -1 & -1 \\ 1 & -1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & 0 & \dots & 0 & 0 \\ 1 & -1 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Suppose $A : M(k,k) \to M(k,k)$ is an S_k-map. Set $\mathbb{V} = \mathbb{D}_{k,2} \oplus \mathbb{A}_{k,1} \oplus \mathbb{S}_{k,2}$ so that $\mathbb{V}$ has isotypic decomposition $3\mathfrak{s}_k$. Setting $A_\mathfrak{s} = A|\mathbb{V}$, we have $A_\mathfrak{s} : \mathbb{V} \to \mathbb{V}$. Since $\mathfrak{s}_k$ is a real representation,

$$A(\mathfrak{S}_i^k) = \sum_{j \in [3]} a_{ij} \mathfrak{S}_j^k, \; i \in [3],$$

where $[a_{ij}]$ is a real 3×3-matrix. The eigenvalues of the matrix $[a_{ij}]$ give the eigenvalues of $A_\mathfrak{s} : \mathbb{V} \to \mathbb{V}$ (with multiplicities multiplied by $(k-1)$).

We have shown how to deal with multiple factors of $\mathfrak{t}$ and $\mathfrak{s}$. For the representations $\mathfrak{x}_k$ and $\mathfrak{y}_k$, we have $A|\mathfrak{x}_k = \lambda_\mathfrak{x} I$, $A|\mathfrak{y}_k = \lambda_\mathfrak{y} I$. It is enough to compute $A(M)_i$ where M is a non-zero matrix in $\mathbb{A}_{k,2}$ (resp. $\mathbb{S}_{k,3}$) with $M_i \neq 0$ (M in vectorized form so $i \in [k^2]$). To simplify computations, we choose matrices with many zeros and take

$$\mathfrak{X}^k = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 & -(k-2) \\ -1 & 0 & 0 & \dots & 0 & 1 \\ -1 & 0 & 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 0 & 0 & \dots & 0 & 1 \\ (k-2) & -1 & -1 & \dots & -1 & 0 \end{bmatrix} \in \mathbb{A}_{k,2}$$

and

$$\mathfrak{Y}^k = \begin{bmatrix} 0 & k-3 & 3-k & \dots & 0 & 0 & 0 \\ k-3 & 0 & 0 & \dots & -1 & -1 & -1 \\ 3-k & 0 & 0 & \dots & 1 & 1 & 1 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \end{bmatrix} \in \mathbb{S}_{k,3}$$

B.4 Isotypic decomposition of $(M(k,k), S_p \times S_q)$

Assume $p + q = k$, regard $S_p \times S_q$ as a subgroup of S_k and restrict the diagonal action of S_k on $M(k,k)$ to $S_p \times S_q$ to define $M(k,k)$ as an $S_p \times S_q$-space. We assume $k > p > k/2$ so that $S_p \times S_q$ will be a maximal intransitive subgroup of S_k [23, 24]. Clearly, $M(k,k)$ decomposes as an orthogonal $S_p \times S_q$-invariant direct sum

$$M(k,k) = M(p,p) \oplus M(p,q) \oplus M(q,p) \oplus M(q,q),$$

where $M(p,p)$ is an S_p-space and $M(q,q)$ is an S_q space (diagonal actions). We regard $M(p,q)$ and $M(q,p)$ as $S_p \times S_q$-spaces. Thus, S_p acts on $M(p,q)$ (resp. $M(q,p)$) by permuting rows (resp. columns) and S_q acts on $M(p,q)$ (resp. $M(q,p)$) by permuting columns (resp. rows). At first sight this convention may seem confusing but observe that the map $M(p,q) \to M(q,p); A \mapsto A^T$, is a linear isomorphism and an $S_p \times S_q$-map. Hence the representations $(M(p,q), S_p \times S_q)$ and $(M(q,p), S_p \times S_q)$ are isomorphic.

If $A \in M(k,k)$, write A in block form as $A = \begin{bmatrix} A_{p,p} & A_{p,q} \\ A_{q,p} & A_{q,q} \end{bmatrix}$, where $A_{r,s} \in M(r,s)$, $(r,s) \in \{p,q\}$. Certain special block matrices will be needed for the analysis of the eigenvalue structure. We make use of the matrices defined in the previous section.

Block matrix decompositions related to $\mathfrak{x}$

Define

$$\mathfrak{X}^{p,p} = \begin{bmatrix} \mathfrak{X}^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \quad \mathfrak{X}^{q,q} = \begin{bmatrix} \mathbf{0}_{q,q} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{X}^q \end{bmatrix},$$

where for the definition of $\mathfrak{X}^{q,q}$ it is assumed that $q \geq 3$.

Block matrix decompositions related to $\mathfrak{y}$

Define

$$\mathfrak{y}^{p,p} = \begin{bmatrix} \mathfrak{y}^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \quad \mathfrak{y}^{q,q} = \begin{bmatrix} \mathbf{0}_{q,q} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{y}^q \end{bmatrix},$$

where for the definition of $\mathfrak{Y}^{q,q}$ it is assumed that $q > 3$.

Block matrix decompositions related to $\mathfrak{t}$

Let $\mathcal{I}_{r,s}$ denote the $r \times s$-matrix with all entries equal to 1. Define

$$\mathfrak{D}_1^{p,p} = \begin{bmatrix} \mathfrak{D}_1^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \qquad \mathfrak{D}_2^{p,p} = \begin{bmatrix} \mathfrak{D}_2^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix},$$

$$\mathfrak{D}_1^{q,q} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{D}_1^q \end{bmatrix}, \qquad \mathfrak{D}_2^{q,q} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{D}_2^q \end{bmatrix},$$

$$\mathfrak{D}_3^{p,q} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathcal{I}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \qquad \mathfrak{D}_3^{q,p} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathcal{I}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix},$$

where for the definition of $\mathfrak{D}_2^{q,q}$, it is assumed that $q \geq 2$.

Block matrix decompositions related to $\mathfrak{s}$

Define $\mathfrak{S}_r^{p,q}, \mathfrak{S}_c^{p,q} \in M(p,q)$ by

$$\mathfrak{S}_r^{p,q} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & 0 & \dots & 0 \end{bmatrix}, \mathfrak{S}_c^{p,q} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ -1 & -1 & -1 & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

and $\mathfrak{S}_r^{q,p}, \mathfrak{S}_c^{q,p} \in M(q,p)$ by $\mathfrak{S}_r^{q,p} = (\mathfrak{S}_c^{p,q})^T$, $\mathfrak{S}_c^{q,p} = (\mathfrak{S}_r^{p,q})^T$. Note that $\mathfrak{S}_r^{p,q}$ and $\mathfrak{S}_c^{q,p}$ are only defined if $q \geq 2$. Set

$$\begin{aligned}
\mathfrak{S}_1^{p,p} &= \begin{bmatrix} \mathfrak{S}_1^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, & \mathfrak{S}_1^{q,q} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{S}_1^q \end{bmatrix}, \\
\mathfrak{S}_2^{p,p} &= \begin{bmatrix} \mathfrak{S}_2^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, & \mathfrak{S}_2^{q,q} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{S}_2^q \end{bmatrix}, \\
\mathfrak{S}_3^{p,p} &= \begin{bmatrix} \mathfrak{S}_3^p & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, & \mathfrak{S}_3^{q,q} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & \mathfrak{S}_3^q \end{bmatrix}, \\
\mathfrak{S}_4^{p,q} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathfrak{S}_c^{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, & \mathfrak{S}_4^{q,p} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathfrak{S}_r^{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \\
\mathfrak{S}_5^{p,q} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathfrak{S}_r^{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, & \mathfrak{S}_5^{q,p} &= \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathfrak{S}_c^{q,p} & \mathbf{0}_{q,q} \end{bmatrix}.
\end{aligned}$$

The first column defines representative elements in each of the factors comprising $5\mathfrak{s}_p$ that lie in the image of of $(1,-1,0,\cdots,0) \in H_{p-1}$ by the natural map of (H_{p-1}, S_p) onto that factor; similarly for the second column (with p replaced by q).

Block matrix decompositions related to $\mathfrak{s}_p \boxtimes \mathfrak{s}_q$

Recall that $\mathfrak{s}_p \boxtimes \mathfrak{s}_q$ is the exterior tensor product of the S_p-representation $\mathfrak{s}_p$ and the S_q-representation $\mathfrak{s}_q$. The degree of $\mathfrak{s}_p \boxtimes \mathfrak{s}_q$ is $(p-1)(q-1)$. Since $p > k/2$, $p \neq q$ and so $\mathfrak{s}_p \boxtimes \mathfrak{s}_q$ is irreducible. Just as we view $(M(p,q), S_p \times S_q)$ and $(M(q,p), S_p \times S_q)$ as isomorphic representations, we regard $\mathfrak{s}_q \boxtimes \mathfrak{s}_p$ as the isomorphism class of an $S_p \times S_q$ representation and then $\mathfrak{s}_p \boxtimes \mathfrak{s}_q = \mathfrak{s}_q \boxtimes \mathfrak{s}_p$.

Assume $q \geq 2$. Define

$$\mathfrak{H}_u^{p,q} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in M(p,q), \quad \mathfrak{H}_l^{q,p} = (\mathfrak{H}_u^{p,q})^T \in M(q,p)$$

Define

$$\mathfrak{H}^{p,q} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathfrak{H}_u^{p,q} \\ \mathbf{0}_{q,p} & \mathbf{0}_{q,q} \end{bmatrix}, \quad \mathfrak{H}^{q,p} = \begin{bmatrix} \mathbf{0}_{p,p} & \mathbf{0}_{p,q} \\ \mathfrak{H}_l^{q,p} & \mathbf{0}_{q,q} \end{bmatrix},$$

Note that $\mathfrak{H}_1^{p,q} \in M(p,q)$, $\mathfrak{H}_l^{q,p} \in M(q,p)$. The isotopic decomposition of $(M(p,q), S_p \times S_q)$ (equivalently, $(M(q,p), S_p \times S_q)$ is

$$(\mathfrak{s}_p + \mathfrak{t}) \boxtimes (\mathfrak{s}_q + \mathfrak{t})) = \mathfrak{s}_p \boxtimes \mathfrak{s}_q + \mathfrak{s}_p + \mathfrak{s}_q + \mathfrak{t}$$

Hence $M(p,q) \oplus M(q,p)$ contributes $2\mathfrak{s}_p \boxtimes \mathfrak{s}_q + 2\mathfrak{s}_p + 2\mathfrak{s}_q + 2\mathfrak{t}$ to the isotypic decomposition of $(M(k,k), S_p \times S_q)$.

If $A : M(p,q) \to M(q,p)$ is an $S_p \times S_q$-map, then $A(\mathfrak{H}_u^{p,q}) = c\mathfrak{H}_l^{q,p}$, for some $c \in \mathbb{R}$. as an $S_p \times S_q$ representation (we switch the order of the action on the target). In particular, the linear isomorphism $H : M(p,q) \to M(q,p)$, $A \mapsto A^T$, is an $S_p \times S_q$-map and $H(\mathfrak{H}_1^{p,q}) = \mathfrak{H}_1^{q,p}$. In order to compute spectrum associated to $2\mathfrak{s}_p \boxtimes \mathfrak{s}_q$, we use the representative matrices $\mathfrak{H}^{p,q}, \mathfrak{H}^{q,p}$. Trivial factors add the representative matrices $\mathfrak{D}_3^{p,q}, \mathfrak{D}_3 q, p$ and $\mathfrak{s}_p, \mathfrak{s}_q$ add the 4 representative matrices $\mathfrak{S}_i^{r,s}$, where $i = 4, 5$ and $r, s \in \{p, q\}$, $r \neq s$.

All the algebra is now in place for computing the spectra of $S_p \times S_q$-maps of $M(k,k)$, where p, q, k satisfy the conditions of Theorem 4.

C Computation of the Hessian of $\mathcal{F}$.

We assume that $d = k$ (the case $d > k$ is done is Section E). We make use of the computations given in [22, 4.3.1] where the parameters are viewed as column vectors rather than as row vectors, the natural choice for the matrix formalism. The result, however, is independent of whatever viewpoint is adopted. Here we represent as columns (labelled by superscripts) to keep compatibility with notation in [22]. The result we give applies to the case $\boldsymbol{W} = \mathbf{V}$ although the Hessian formula [22, 4.1.1] is not well-defined when $\boldsymbol{W} = \mathbf{V}$ (division by zero). We remark that $\mathcal{F}(\boldsymbol{W})$ is C^2 at $\boldsymbol{W} = \mathbf{V}$, but not real analytic, or even smooth [24].

It follows from our analysis that if none of the parameter vectors $\boldsymbol{w}^j$ in $\boldsymbol{W}$ has isotropy ΔS_k, then the Hessian depends only on (a) the angles between between parameter vectors and (b) angles between parameter vectors and the the target parameters $\mathbf{v}^j$, $j \in [k]$ determining $\mathbf{V}$. In particular, *there is no dependence on the norms* $\|\boldsymbol{w}^j\|$. If $\boldsymbol{W}$ has isotropy ΔS_{k-1}, and $\boldsymbol{w}^k \neq \mathbf{0}$, then a similar result holds but now with mild dependence on norms of parameters. Isotropy groups which are not diagonal often lead to parallel parameter vectors and loss of differentiability of $\mathcal{F}$ (see [24, Ex. 4.9]).

Henceforth, we always assume that no rows of $\boldsymbol{W} = [\boldsymbol{w}^1, \cdots, \boldsymbol{w}^k]$ are parallel. In particular, that $\boldsymbol{w}^j \neq \mathbf{0}$ and $|\langle \boldsymbol{w}^i, \boldsymbol{w}^j \rangle| \neq \|\boldsymbol{w}^i\| \|\boldsymbol{w}^j\|$, $i, j \in [k]$, $i \neq j$.

C.1 Formula for the Hessian $\mathcal{F}$ at $\mathbf{W} = [\mathbf{w}^1, \cdots, \mathbf{w}^k]$

We recall some results and notation from [22, 4.4.1]. Specifically, for non-parallel $\boldsymbol{w}, \mathbf{v} \in \mathbb{R}^k$, let $\theta_{\boldsymbol{w},\mathbf{v}} \in (0, \pi)$ denote the angle between $\boldsymbol{w}, \mathbf{v}$ and define

1. $\mathbf{n}_{\boldsymbol{w},\mathbf{v}} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} - \cos(\theta_{\boldsymbol{w},\mathbf{v}}) \frac{\mathbf{v}}{\|\mathbf{v}\|}$.

2. $\bar{\mathbf{n}}_{\boldsymbol{w},\mathbf{v}} = \frac{\mathbf{n}_{\boldsymbol{w},\mathbf{v}}}{\|\mathbf{n}_{\boldsymbol{w},\mathbf{v}}\|}$.

Note that

$$\|\mathbf{n}_{\boldsymbol{w},\mathbf{v}}\| = \sin(\theta_{\boldsymbol{w},\mathbf{v}}). \tag{C.5}$$

If $\mathbf{v}, \boldsymbol{w}$ are parallel but not zero, then $\mathbf{n}_{\boldsymbol{w},\mathbf{v}} = \mathbf{0}$ and we *define* $\bar{\mathbf{n}}_{\boldsymbol{w},\mathbf{v}} = \mathbf{0}$—this choice gives the correct value for the Hessian H of $\mathcal{F}$ if $\boldsymbol{W} = \mathbf{V}$.

We write the Hessian H of $\mathcal{F}$ as a $k \times k$-matrix of $k \times k$-blocks: $H = [H^{pq}]$. Since H is symmetric, $H^{pq} = (H^{qp})^T$, $p, q \in [k]$, and H^{pp} is symmetric. Each block H^{pq} corresponds to derivatives with respect to $\boldsymbol{w}^p, \boldsymbol{w}^q$.

Let $\mathbf{I} = \mathbf{I}_k \in M(k,k)$ denote the identity matrix. Given non-parallel $\boldsymbol{w}, \mathbf{v} \in \mathbb{R}^k$, define $h_1, h_2 \in M(k,k)$ by

$$h_1(\boldsymbol{w}, \mathbf{v}) = \frac{\sin(\theta_{\boldsymbol{w},\mathbf{v}})\|\mathbf{v}\|}{2\pi\|\boldsymbol{w}\|}\left(\mathbf{I} - \frac{\boldsymbol{w}\boldsymbol{w}^T}{\|\boldsymbol{w}\|^2} + \bar{\mathbf{n}}_{\mathbf{v},\boldsymbol{w}}\bar{\mathbf{n}}_{\mathbf{v},\boldsymbol{w}}^T\right), \tag{C.6}$$

$$h_2(\boldsymbol{w}, \mathbf{v}) = \frac{1}{2\pi}\left((\pi - \theta_{\boldsymbol{w},\mathbf{v}})\mathbf{I} + \frac{\bar{\mathbf{n}}_{\boldsymbol{w},\mathbf{v}}\mathbf{v}^T}{\|\mathbf{v}\|} + \frac{\bar{\mathbf{n}}_{\mathbf{v},\boldsymbol{w}}\boldsymbol{w}^T}{\|\boldsymbol{w}\|}\right). \tag{C.7}$$

Lemma 12 ([22, Theorem 5])**.** *The Hessian $H = [H^{pq}]$ of $\mathcal{F}$ at the critical point $\boldsymbol{W} = [\boldsymbol{w}^1, \cdots, \boldsymbol{w}^k]$ is given by*

$$\begin{aligned} H^{pp} &= \frac{1}{2}\mathbf{I} + \sum_{q\in\mathbf{k}} \left(h_1(\boldsymbol{w}^p, \boldsymbol{w}^q) - h_1(\boldsymbol{w}^p, \mathbf{v}^q)\right),\ p \in [k], \\ H^{pq} &= h_2(\boldsymbol{w}^p, \boldsymbol{w}^q),\ p, q \in [k],\ p \neq q. \end{aligned}$$

C.2 Expressions for h_1, h_2.

We work towards obtaining more geometric expressions for the blocks H^{pq}. This will involve a careful analysis of the terms h_1, h_2 in the preceeding lemma. The term $h_1(\boldsymbol{w}^p, \mathbf{v}^p)$, used only in the description of the diagonal blocks, is particularly tricky as $\boldsymbol{w}^p$ is often close to being parallel to $\mathbf{v}^p$ in our applications.

Let $\langle\ ,\ \rangle$ denote the standard Euclidean inner product on $\mathbb{R}^k$ and $\veebar$ denote the *exclusive* or.

Lemma 13. *If $q \in [k]$, $\boldsymbol{w} = [w_1, \cdots, w_k]^T \in \mathbb{R}^k$ and $\boldsymbol{w}, \mathbf{v}^q$ are not parallel, then*

$$\begin{aligned} h_1(\boldsymbol{w}, \mathbf{v}^q)_{ij} &= \frac{\sin(\alpha_{\boldsymbol{w}q})}{2\pi\|\boldsymbol{w}\|}\left(\delta_{ij} - \frac{w_i w_j}{\|\boldsymbol{w}\|^2} + K_{ij}^{\boldsymbol{w}q}\frac{w_i w_j}{\|\boldsymbol{w}\|^2}\right),\ (i,j) \neq (q,q) \\ h_1(\boldsymbol{w}, \mathbf{v}^q)_{qq} &= \frac{\sin^3(\alpha_{\boldsymbol{w}q})}{\pi\|\boldsymbol{w}\|}, \end{aligned}$$

where $\alpha_{\boldsymbol{w}q} = \cos^{-1}\left(\frac{\langle\boldsymbol{w},\mathbf{v}^q\rangle}{\|\boldsymbol{w}\|}\right)$ and

$$K_{ij}^{\boldsymbol{w}q} = \begin{cases} -1,\ i \veebar j = q, \\ \frac{w_q^2}{\sum_{\ell\neq q} w_\ell^2} = \cot^2(\alpha_{\boldsymbol{w}q}),\ i, j \neq q. \end{cases}$$

Proof The proof is a straightforward computation using (C.6) and we only give details for $h_1(\boldsymbol{w}, \mathbf{v}^q)_{qq}$. By (C.6) and (C.5), we have

$$h_1(\boldsymbol{w}, \mathbf{v}^q)_{qq} = \frac{\sin(\alpha_{\boldsymbol{w}q})}{2\pi\|\boldsymbol{w}\|}\left[1 - \frac{w_q^2}{\|\boldsymbol{w}\|^2} + \left(1 - \cos(\alpha_{\boldsymbol{w}q})\frac{w_q}{\|\boldsymbol{w}\|}\right)^2 / \sin^2(\alpha_{\boldsymbol{w}q})\right].$$

Since $\cos(\alpha_{\boldsymbol{w}q}) = \frac{w_q}{\|\boldsymbol{w}\|}$, $1 - \cos(\alpha_{\boldsymbol{w}q})\frac{w_q}{\|\boldsymbol{w}\|} = 1 - \cos^2(\alpha_{\boldsymbol{w}q}) = \sin^2(\alpha_{\boldsymbol{w}q})$. Hence $h_1(\boldsymbol{w}, \mathbf{v}^q)_{qq} = \frac{\sin(\alpha_{\boldsymbol{w}q})}{2\pi\|\boldsymbol{w}\|}\left(1 - \cos^2(\alpha_{\boldsymbol{w}q}) + \sin^2(\alpha_{\boldsymbol{w}q})\right)$ giving the result since $1 - \cos^2(\alpha_{\boldsymbol{w}q}) + \sin^2(\alpha_{\boldsymbol{w}q}) = 2\sin^2(\alpha_{\boldsymbol{w}q})$. $\square$

Remark 14. $K_{ij}^{\boldsymbol{w}q}$ is well-defined if $i,j\neq q$—since $\boldsymbol{w}$ and $\mathbf{v}^q$ are not parallel. Moreover, even though $K_{ij}^{\boldsymbol{w}q}$ may be large, because of the division by $\sum_{p\neq q} w_p^2$, $|K_{ij}^{\boldsymbol{w}q} w_i w_j| \leq \|\boldsymbol{w}\|/2$ by the Cauchy-Schwartz inequality. This allows us to show the formula we derive below for the Hessian applies when $\boldsymbol{W}=\mathbf{V}$, even though $\boldsymbol{w}^i$ is parallel to $\mathbf{v}^i$, $i\in[k]$, and that $\mathcal{F}$ is C^2 at $\boldsymbol{W}=\mathbf{V}$.

Lemma 15. *(Notation and assumptions as above.) For $i,j,p,q\in[k]$, $p\neq q$,*

$$\begin{aligned} h_1(\boldsymbol{w}^p,\boldsymbol{w}^q)_{ij} \quad &= \quad \frac{\sin(\Theta_{pq})\|\boldsymbol{w}^q\|}{2\pi\|\boldsymbol{w}_p\|}\left(\delta_{ij}-\frac{w_i^p w_j^p}{\|\boldsymbol{w}^p\|^2}\right)+ \\ &\frac{\|\boldsymbol{w}^q\|}{2\pi\|\boldsymbol{w}_p\|\sin(\Theta_{pq})}\left(\frac{w_i^q w_j^q}{\|\boldsymbol{w}^q\|^2}-\cos(\Theta_{pq})\frac{w_i^p w_j^q+w_i^q w_j^p}{\|\boldsymbol{w}^p\|\|\boldsymbol{w}^q\|}+\cos^2(\Theta_{pq})\frac{w_i^p w_j^p}{\|\boldsymbol{w}^p\|^2}\right), \end{aligned}$$

where $\Theta_{pq}=\cos^{-1}\left(\frac{\langle\boldsymbol{w}^p,\boldsymbol{w}^q\rangle}{\|\boldsymbol{w}^p\|\|\boldsymbol{w}^q\|}\right)$.

Proof A straightforward computation using (C.6). □

Lemma 16. *(Notation and assumptions as above.) For $i,j,p,q\in[k]$, $p\neq q$,*

$$h_2(\boldsymbol{w}^p,\boldsymbol{w}^q)_{ij}=\frac{(\pi-\Theta_{pq})\delta_{ij}}{2\pi}+\frac{1}{2\pi\sin(\Theta_{pq})}L_{ij}^{pq},$$

where

$$L_{ij}^{pq}=\frac{w_i^p w_j^q+w_i^q w_j^p}{\|\boldsymbol{w}^p\|\|\boldsymbol{w}^q\|}-\cos(\Theta_{pq})\left(\frac{w_i^p w_j^p}{\|\boldsymbol{w}^p\|^2}+\frac{w_i^q w_j^q}{\|\boldsymbol{w}^q\|^2}\right).$$

Proof A straightforward computation using (C.7). □

Remarks 17. (1) Since no columns of $\boldsymbol{W}$ are parallel, division by $\sin(\Theta_{pq})$ is safe in both lemmas. In our applications, $\sin(\Theta_{pq})$ will typically be close to 1 for large k.
(2) If we let α_{pq} denote the angle between $\boldsymbol{w}^p$ and $\mathbf{v}^q$, $p,q\in[k]$, then all of the terms in second lemma can be written in terms of the angles α_{iq} and Θ_{pq} with no norm terms appearing ($w_q^p/\|\boldsymbol{w}^p\|=\cos(\alpha_{pq})$). This is true in the first lemma if all $\boldsymbol{w}^j$ have the same norm.

C.3 The Hessian of critical points with isotropy ΔS_k.

We give a formula for the Hessian at critical points $\boldsymbol{W}=[\boldsymbol{w}^1,\cdots,\boldsymbol{w}^k]$ with isotropy ΔS_k. We continue to assume $k=d$. Since isotropy is S_k, columns can never be parallel: if two columns are parallel, then since $\boldsymbol{W}$ is fixed by ΔS_k, all columns must be equal and so the isotropy of $\boldsymbol{W}$ is strictly bigger than ΔS_k.

Since $\boldsymbol{W}$ has isotropy ΔS_k, $\|\boldsymbol{w}^i\|$ is independent of $i\in[k]$ and we set $\|\boldsymbol{w}^i\|=\tau$, $i\in[k]$. Set $w_i^i=R$, $w_i^j=S$, $i,j\in[k]$, $j\neq i$, so that the diagonal entries of $\boldsymbol{W}$ are all equal to R, the off-diagonal entries all equal to S. Since the isotropy of $\boldsymbol{W}$ is ΔS_k, $S\neq R$. Define the angles

1. $\Theta=\cos^{-1}\left(\frac{\langle\boldsymbol{w}^i,\boldsymbol{w}^j\rangle}{\tau^2}\right)$, $i,j\in[k], i\neq j$.
2. $\alpha=\cos^{-1}\left(\frac{\langle\boldsymbol{w}^i,\mathbf{v}^j\rangle}{\tau}\right)$, $i,j\in[k], i\neq j$.
3. $\beta=\cos^{-1}\left(\frac{\langle\boldsymbol{w}^i,\mathbf{v}^i\rangle}{\tau}\right)$, $i\in[k]$.

and note that Θ, α and β are independent of $i,j\in[k]$.

For $i,j,p\in[k]$, we tabulate the possible values of $A_{ij}^p\doteq w_i^p w_j^p/\tau^2$.

$$\begin{aligned} A_{ij}^p &= R^2/\tau^2=\cos^2(\beta), \quad i=j=p \\ A_{ij}^p &= RS/\tau^2=\cos(\alpha)\cos(\beta), \quad i\veebar j=p \\ A_{ij}^p &= S^2/\tau^2=\cos^2(\alpha), \quad i,j\neq p, \end{aligned}$$

and define

$$
\begin{aligned}
A^{pq}_{ij} &= \cos^2(\alpha) + \cos^2(\beta), & B^{pq}_{ij} &= 2\cos(\alpha)\cos(\beta), \quad i,j \in \{p,q\}, i = j \\
&= 2\cos(\alpha)\cos(\beta), & &= \cos^2(\alpha) + \cos^2(\beta), \quad i,j \in \{p,q\}, i \neq j \\
&= \cos(\alpha)\cos(\beta) + \cos^2(\alpha), & &= \cos(\alpha)\cos(\beta) + \cos^2(\alpha), \quad i \veebar j \in \{p,q\} \\
&= 2\cos^2(\alpha) & &= 2\cos^2(\alpha), \quad i,j \notin \{p,q\}.
\end{aligned}
$$

Note that A^{pq}_{ij}, B^{pq}_{ij} are symmetric in p,q and i,j.

Lemma 18 (Off diagonal blocks)**.** *(Notation and assumptions as above.) If $p,q,i,j \in [k]$, with $p \neq q$, then*

$$
H^{pq}_{ij} = \frac{(\pi - \Theta)\delta_{ij}}{2\pi} + \frac{B^{pq}_{ij} - \cos(\Theta)A^{pq}_{ij}}{2\pi\sin(\Theta)}.
$$

In particular $H^{pq} = H^{qp}$ and are symmetric matrices.

Proof Immediate from Lemma 16, the definitions of A^{p}_{ij}, A^{pq}_{ij} and B^{pq}_{ij} and the symmetry of A^{pq}_{ij}, B^{pq}_{ij} n p,q and i,j. □

We need a preliminary result before we give a precise description of $[H^{ij}]$.

Lemma 19. *(Notation and assumptions as above.) If $p,q,i,j \in [k]$, $p \neq q$, then*

$$
\begin{aligned}
h_1(\boldsymbol{w}^p, \boldsymbol{w}^q)_{ij} &= \frac{\sin(\Theta)}{2\pi}\left(\delta_{ij} - \frac{w^p_i w^p_j}{\tau^2}\right) + \\
&\quad \frac{1}{2\pi\tau^2\sin(\Theta)}\left(w^q_i w^q_j - \cos(\Theta)(w^p_i w^q_j + w^q_i w^p_j) + \cos^2(\Theta) w^p_i w^p_j\right) \\
&= \frac{\sin(\Theta)}{2\pi}\left(\delta_{ij} - A^p_{ij}\right) + \frac{1}{2\pi\sin(\Theta)}\left(A^q_{ij} - \cos(\Theta)B^{pq}_{ij} + \cos^2(\Theta)A^p_{ij}\right)
\end{aligned}
$$

Let $p,q,i,j \in [k]$. If $K^{pq}_{ij} \doteq K^{\boldsymbol{w}^p q}_{ij}$, then

1. $p \neq q$

$$
\begin{aligned}
h_1(\boldsymbol{w}^p, \mathbf{v}^q)_{ij} &= \frac{\sin(\alpha)}{2\pi\tau}\left(\delta_{ij} - A^p_{ij} + K^{pq}_{ij}A^p_{ij}\right), \ (i,j) \neq (q,q) \\
&= \frac{\sin^3(\alpha)}{\pi\tau}, \ i = j = q,
\end{aligned}
$$

where if $i,j \neq q$,

$$
K^{pq}_{ij}A^p_{ij} = \begin{cases}
\cos^2(\beta)\cot^2(\alpha), \ i = j = p \\
-\cos(\alpha)\cos(\beta), \ i,j \in \{p,q\}, \ i \neq j \\
\cos(\alpha)\cos(\beta)\cot^2(\alpha), \ i \veebar j = p, \ i,j \neq q \\
-\cos^2(\alpha), \ i \veebar j = q, \ i,j \neq p \\
\cos^2(\alpha)\cot^2(\alpha), \ i,j \notin \{p,q\}.
\end{cases}
$$

2. $p = q$

$$
\begin{aligned}
h_1(\boldsymbol{w}^q, \mathbf{v}^q)_{ij} &= \frac{\sin(\beta)}{2\pi\tau}\left(\delta_{ij} - A^q_{ij} + K^{qq}_{ij}A^q_{ij}\right), \ (i,j) \neq (q,q) \\
&= \frac{\sin^3(\beta)}{\pi\tau}, \ i = j = q,
\end{aligned}
$$

where if $i,j \neq q$,

$$
K^{qq}_{ij}A^q_{ij} = \begin{cases}
\frac{\cos^2(\beta)}{k-1}, \ i,j \neq q. \\
-\cos(\alpha)\cos(\beta), \ i = q \neq j, \ j = q \neq i
\end{cases}
$$

Proof We verify the statements concerning the terms $K_{ij}^{qq} A_{ij}^{q}$. First note that both K_{ij}^{qq} and A_{ij}^{q} are symmetric in i, j. If $i, j \neq q$, then $K_{ij}^{qq} = \frac{(w_q^q)^2}{\sum_{\ell \neq q} (w_\ell^q)^2} = \frac{R^2}{(k-1)S^2}$, and $A_{ij}^{q} = \cos^2(\alpha) = \frac{S^2}{\tau^2}$. Hence $K_{ij}^{qq} A_{ij}^{q} = \frac{R^2}{\tau^2(k-1)} = \frac{\cos^2(\beta)}{k-1}$. On the other hand if $i = q, j \neq q$, $K_{qj}^{qq} = -1$ and $A_{qj}^{q} = \cos(\alpha)\cos(\beta)$. Hence $K_{qj}^{qq} A_{qj} = -\cos(\alpha)\cos(\beta)$. □

Proposition 20 (Diagonal blocks)**.** *(Notation and assumptions as above.) Let $i, j, p \in [k]$.*

1.

$$\begin{aligned} H_{pp}^{pp} &= \frac{1}{2} + \frac{(k-1)\sin^2(\beta)}{2\pi}\left(\sin(\Theta) - \frac{\sin(\alpha)}{\tau}\right) - \frac{\sin^3(\beta)}{\pi\tau} + \\ &\frac{(k-1)}{2\pi}\left(\frac{1}{\sin(\Theta)}\big(\cos(\alpha) - \cos(\Theta)\cos(\beta)\big)^2 - \frac{\cot(\alpha)\cos(\alpha)\cos^2(\beta)}{\tau}\right). \end{aligned}$$

2. If $i \neq p$, then

$$\begin{aligned} H_{ii}^{pp} &= \frac{1}{2} + \frac{(k-2)\sin^2(\alpha)}{2\pi}\left(\sin(\Theta) - \frac{\sin(\alpha)}{\tau}\right) + \\ &\frac{\sin^2(\alpha)}{\pi}\left(\frac{\sin(\Theta)}{2} - \frac{\sin(\alpha)}{\tau}\right) + \frac{(k-2)}{2\pi}\left(\frac{\cos^2(\alpha)}{\sin(\Theta)}(1-\cos(\Theta))^2 - \frac{\cot(\alpha)\cos^3(\alpha)}{\tau}\right) + \\ &\frac{1}{2\pi\sin(\Theta)}(\cos(\beta) - \cos(\Theta)\cos(\alpha))^2 - \frac{\sin(\beta)}{2\pi\tau}\left(\sin^2(\alpha) + \frac{\cos^2(\beta)}{k-1}\right) \end{aligned}$$

1. If $i \veebar j = p$, then

$$\begin{aligned} H_{1j}^{pp} &= -\frac{(k-1)\cos(\alpha)\cos(\beta)}{2\pi}\left(\sin(\Theta) - \frac{\sin(\alpha)}{\tau}\right) + \\ &\frac{(k-2)\cos(\alpha)}{2\pi}\left(\frac{\cos(\alpha) - \cos(\Theta)(\cos(\alpha)+\cos(\beta)) + \cos^2(\Theta)\cos(\beta)}{\sin(\Theta)}\right) - \\ &\frac{(k-2)}{2\pi\tau}\sin(\alpha)\cos(\alpha)\cos(\beta)\cot^2(\alpha) + \frac{\sin(\alpha)\cos(\alpha)\cos(\beta)}{2\pi\tau} + \\ &\frac{1}{2\pi\sin(\Theta)}\big(\cos(\alpha)\cos(\beta) - \cos(\Theta)(\cos^2(\alpha)+\cos^2(\beta)) + \cos^2(\Theta)\cos(\alpha)\cos(\beta)\big) + \\ &\frac{\cos(\alpha)\sin(\beta)\cos(\beta)}{\pi\tau} \end{aligned}$$

2. If $i, j \neq p$, then

$$\begin{aligned} H_{ij}^{pp} &= -\frac{(k-1)\cos^2(\alpha)}{2\pi}\left(\sin(\Theta) - \frac{\sin(\alpha)}{\tau}\right) + \\ &\frac{(k-3)\cos^2(\alpha)}{2\pi}\left(\frac{(1-\cos(\Theta))^2}{\sin(\Theta)} - \frac{\sin(\alpha)\cot^2(\alpha)}{\tau}\right) + \\ &\frac{\cos(\alpha)}{\pi}\left(\frac{\cos(\beta) - \cos(\Theta)(\cos(\beta)+\cos(\alpha)) + \cos^2(\Theta)\cos(\alpha)}{\sin(\Theta)} + \frac{\sin(\alpha)\cos(\alpha)}{\tau}\right) + \\ &\frac{\sin(\beta)}{2\pi\tau}\left(\cos^2(\alpha) - \frac{\cos^2(\beta)}{k-1}\right) \end{aligned}$$

Proposition 21 (Off diagonal blocks)**.** *(Notation and assumptions as above.) Given $i, j, p, q \in [k]$, $p \neq q$.*

1. $i \in \{p, q\}$,

$$H_{ii}^{pq} = \frac{\pi - \Theta}{2\pi} + \frac{1}{2\pi\sin(\Theta)}\big(2\cos(\alpha)\cos(\beta) - \cos(\Theta)(\cos^2(\alpha)+\cos^2(\beta))\big)$$

2. $i \notin \{p,q\}$,

$$H^{pq}_{ii} = \frac{\pi-\Theta}{2\pi} + \frac{\cos^2(\alpha)}{\pi\sin(\Theta)}\big(1-\cos(\Theta)\big)$$

3. $i,j \in \{p,q\}$, $i \neq j$,

$$H^{pq}_{ij} = \frac{1}{2\pi\sin(\Theta)}\big(\cos^2(\alpha)+\cos^2(\beta)-2\cos(\Theta)\cos(\alpha)\cos(\beta)\big)$$

4. $i \veebar j \in \{p,q\}$,

$$H^{pq}_{ij} = \frac{\cos(\alpha)(\cos(\alpha)+\cos(\beta))}{2\pi\sin(\Theta)}\big(1-\cos(\Theta)\big)$$

5. $i,j \notin \{p,q\}$,

$$H^{pq}_{ij} = \frac{\cos^2(\alpha)}{\pi\sin(\Theta)}\big(1-\cos(\Theta)\big)$$

Proof Both results follow straightforwardly from the definitions of A^p_{ij}, B^{pq}_{ij} and Lemmas 18, 19. $\square$

Example 22 (Spectrum of the Hessian at $\boldsymbol{W} = \mathbf{V}$)**.** For $i,j \in [k]$, let δ_{ij} be equal to 1 if $i=j$, and 0 otherwise, and $\boldsymbol{\delta}_{ij} \in M(k,k)$ be the matrix with i,j entry equal to 1 and all other entries zero. If $\boldsymbol{W} = \mathbf{V}$, take $\alpha = \Theta = \pi/2$ and $\beta = 0$ in Propositions 20, 21. The Hessian $H = [H^{pq}]$ of $\mathcal{F}$ at $\mathbf{V}$ is then given by

$$\begin{aligned} H^{pp} &= \frac{1}{2}\mathbf{I},\ p \in [k], \\ H^{pq} &= \frac{1}{4}\left(\mathbf{I} + \frac{2}{\pi}(\boldsymbol{\delta}_{pq}+\boldsymbol{\delta}_{qp})\right),\ p \neq q. \end{aligned}$$

In particular

$$\begin{aligned} H^{pp}_{ij} &= \begin{cases} \frac{1}{2},\ i=j \\ 0,\ i\neq j \end{cases}, \qquad p \in [k], \\ H^{pq}_{ij} &= \begin{cases} \frac{1}{4},\ i=j \\ \frac{1}{2\pi},\ i,j \in \{p,q\},\ i \neq j, \qquad p,q \in [k], p \neq q \\ 0,\ \text{otherwise} \end{cases} \end{aligned}$$

Let $\mathbf{r}_i$ denote row i of H, $i \in [3]$.

The eigenvalue associated to $\mathfrak{x}_k$.

Assume $k > 3$ and take $\mathfrak{X}^k \in \mathbb{A}_{k,2} \subset M(k,k)$ with vectorization $\overline{\mathfrak{X}^k}$ (by rows) as defined previously. Computing $\langle \mathbf{r}_2, \overline{\mathfrak{X}^k}\rangle$, we find that

$$H(\mathfrak{X}^k)_{12} = \langle \mathbf{r}_2, \overline{\mathfrak{X}^k}\rangle = \frac{1}{4} - \frac{1}{2\pi}.$$

Since $\mathfrak{X}_{12} = 1$, the eigenvalue $\lambda_{\mathfrak{x}}$ associated to $\mathfrak{x}_k$ is $\frac{1}{4} - \frac{1}{2\pi}$, and has multiplicity $(k-1)(k-2)/2$.

The eigenvalue associated to $\mathfrak{y}_k$.

Using the same method as above, $\lambda_{\mathfrak{y}} = \frac{1}{4} + \frac{1}{2\pi}$, for all $k \geq 5$.

The eigenvalues associated to $\mathfrak{t}$.

In this case, we compute $\langle \mathbf{r}_i, \mathfrak{D}^k_j\rangle$, for $i,j \in \mathbf{2}$ (see Example8 for $\mathfrak{D}^k_j$), to find the matrix $\begin{bmatrix} \frac{1}{2} + \frac{k-1}{2\pi} & \frac{k-1}{4} \\ \frac{1}{4} & \frac{1}{2\pi} + \frac{k}{4} \end{bmatrix}$ giving the eigenvalues $\lambda^1_{\mathfrak{t}}, \lambda^2_{\mathfrak{t}}$ associated to the factor $2\mathfrak{t}$. For $k = 6$, we find $\lambda^1_{\mathfrak{t}} = 0.8896627389$, $\lambda^2_{\mathfrak{t}} = 2.0652669197$. As functions of k, $\lambda^i_{\mathfrak{t}}$ monotonically increase like $c_i k$, where $c_1 \approx 0.16$ and $c_2 = 0.25$.

The eigenvalues associated to $\mathfrak{s}_k$.

Denote the matrix associated to the factor $3\mathfrak{s}_k$ by $B = [\beta_{ij}] \in M(3,3)$. That is,

$$H(\mathfrak{S}_i) = \beta_{i1}\mathfrak{S}_1 + \beta_{i2}\mathfrak{S}_2 + \beta_{i3}\mathfrak{S}_3,\ i \in \mathbf{3}.$$

Since $H(\mathfrak{S}_i)_j = \langle \mathbf{r}_j, \mathfrak{S}_i\rangle$, it follows that for $i \in \mathbf{3}$,

$$\langle \mathbf{r}_1, \mathfrak{S}_i\rangle = \beta_{i1}, \quad \langle \mathbf{r}_2, \mathfrak{S}_i\rangle = 2\beta_{i2}, \quad \langle \mathbf{r}_3, \mathfrak{S}_i\rangle = \beta_{i2} + \beta_{i3}$$

where the factor 2 in the second equation occurs since the 12-component of $\mathfrak{S}_2$ is 2. Setting $h_{ij} = \langle \mathbf{r}_j, \mathfrak{S}_i\rangle$,

$$[h_{ij}] = \begin{bmatrix} \frac{1}{2} - \frac{1}{2\pi} & -\frac{1}{4} & 0 \\ -\frac{k}{4} & \frac{k+2}{4} - \frac{1}{\pi} & \frac{1}{4} - \frac{1}{2\pi} \\ \frac{k-2}{4} & -\frac{k-2}{4} & \frac{1}{4} + \frac{1}{2\pi} \end{bmatrix},$$

and so

$$B = \begin{bmatrix} \frac{1}{2} - \frac{1}{2\pi} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{k}{4} & \frac{k+2}{8} - \frac{1}{2\pi} & -\frac{k}{8} \\ \frac{k-2}{4} & -\frac{k-2}{8} & \frac{k}{8} + \frac{1}{2\pi} \end{bmatrix}.$$

Since this equation has real roots for all $k \geq 5$, we can solve in terms of trigonometric functions using the formula of Franccois Viéte. From this we find that for $k = 6$, the eigenvalues are

$$\lambda_{\mathfrak{s}}^1 = 1.712918525755,\ \lambda_{\mathfrak{s}}^2 = 0.287081474245,\ \lambda_{\mathfrak{s}}^3 = 0.090845056908.$$

Numerical examination of eigenvalues for different values of k reveals that the last eigenvalue is constant and equal to $\frac{1}{4} - \frac{1}{2\pi}$—the same as $\lambda_{\mathfrak{x}}$. It may be shown that the characteristic equation of B has the factorization

$$\left(\lambda - \frac{1}{4} + \frac{1}{2\pi}\right)\left(\lambda^2 - (\frac{k}{4} + \frac{1}{2})\lambda + \frac{1}{16}(k - 4\pi^2 + 4\pi + 1)\right).$$

Analysis of the roots of the quadratic term reveal that $\lambda_{\mathfrak{s}}^1 = \frac{k+1}{4} + O(k^{-1})$, and $\lambda_{\mathfrak{s}}^2 = \frac{1}{4} + O(k^{-1})$ is monotone decreasing with limit 0.25. In particular, the eigenvalues of the Hessian are uniformly bounded above zero.

C.4 The Hessian at critical points with isotropy S_{k-1}.

We assume $\boldsymbol{W} = [\boldsymbol{w}^1, \cdots, \boldsymbol{w}^k]$ is a critical point of $\mathcal{F}$ with isotropy S_{k-1} and that $\boldsymbol{w}_k \neq \mathbf{0}$. The isotropy S_{k-1} then guarantees that no two columns of $\boldsymbol{W}$ are parallel. We give an angle representation of the Hessian at $\boldsymbol{W}$. In this case, we need 7 angles which we describe below. Since $\boldsymbol{W}$ has isotropy S_{k-1}, $\|\boldsymbol{w}^i\|$ is independent of $i \in [k-1]$. Set $\|\boldsymbol{w}^i\| = \tau$, $i < k$, and $\|\boldsymbol{w}^k\| = \tau_k$.

Define the angles

1. $\Theta = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^i, \boldsymbol{w}^j\rangle}{\tau^2}\right)$, $i, j < k$, $i \neq j$.
2. $\Lambda = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^i, \boldsymbol{w}^k\rangle}{\tau\tau_k}\right)$, $i < k$.
3. $\alpha_{ii} = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^i, \mathbf{v}^i\rangle}{\tau}\right) = \cos^{-1}\left(\frac{\boldsymbol{w}_i^i}{\tau}\right)$, $i < k$.
4. $\alpha_{ij} = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^i, \mathbf{v}^j\rangle}{\tau}\right) = \cos^{-1}\left(\frac{\boldsymbol{w}_j^i}{\tau}\right)$, $i, j < k$, $i \neq j$.
5. $\alpha_{ik} = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^i, \mathbf{v}^k\rangle}{\tau}\right) = \cos^{-1}\left(\frac{\boldsymbol{w}_k^i}{\tau}\right)$, $i < k$.
6. $\alpha_{kk} = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^k, \mathbf{v}^k\rangle}{\tau_k}\right) = \cos^{-1}\left(\frac{\boldsymbol{w}_k^k}{\tau_k}\right)$.
7. $\alpha_{kj} = \cos^{-1}\left(\frac{\langle \boldsymbol{w}^k, \mathbf{v}^j\rangle}{\tau_k}\right) = \cos^{-1}\left(\frac{\boldsymbol{w}_j^k}{\tau_k}\right)$, $j < k$.

So as to simplify and shorten some of the expressions involved in the description of the Hessian, we set

1. $\cos(\alpha_{ij}) = \mathfrak{cos}_{ij}$, $\cos(\alpha_{ik}) = \mathfrak{cos}_{ik}$, and $\cos(\alpha_{ii}) = \mathfrak{cos}_{ii}$, $i,j < k$.
2. $\cos(\alpha_{kj}) = \mathfrak{cos}_{kj}$ and $\cos(\alpha_{kk}) = \mathfrak{cos}_{kk}$.
3. $\cot(\alpha_{ij}) = \mathfrak{cot}_{ij}$, $\cot(\alpha_{ij}) = \mathfrak{cot}_{ik}$, and $\cot(\alpha_{kj}) = \mathfrak{cot}_{kj}$
4. $\sin(\alpha_{ij}) = \mathfrak{sin}_{ij}$, $\sin(\alpha_{ik}) = \mathfrak{sin}_{ik}$ and $\sin(\alpha_{ii}) = \mathfrak{sin}_{ii}$
5. $\sin(\alpha_{kj}) = \mathfrak{sin}_{kj}$ and $\sin(\alpha_{kk}) = \mathfrak{sin}_{kk}$.

Along similar lines to the previous section, we define and tabulate the values of A^p_{ij}, A^{pq}_{ij} and B^{pq}_{ij}, $i,j,p \in [k]$. For $i,j < k$, define $\rho = w^i_{ii}$, $\varepsilon = w^i_{ij}$, $\zeta = w^i_{ik}$, $\eta = w^i_{kj}$ and $\nu = w^k_{kk}$. Note that by S_{k-1} symmetry, $\rho, \varepsilon, \zeta, \eta$ do not depend on the choice of $i,j \in [k-1]$.

1. $p < k$.

$$\begin{aligned}
A^p_{ij} &= \rho^2/\tau^2 = \mathfrak{cos}^2_{ii}, \quad i = j = p \\
A^p_{ij} &= \rho\varepsilon/\tau^2 = \mathfrak{cos}_{ii}\mathfrak{cos}_{ij}, \quad i,j < k,\ i \veebar j = p \\
A^p_{ij} &= \rho\zeta/\tau^2 = \mathfrak{cos}_{ii}\mathfrak{cos}_{ik}, \quad i \veebar j = k,\ i \veebar j = p \\
A^p_{ij} &= \varepsilon^2/\tau^2 = \mathfrak{cos}^2_{ij}, \quad i,j < k,\ i,j \neq p \\
A^p_{ij} &= \varepsilon\zeta/\tau^2 = \mathfrak{cos}_{ij}\mathfrak{cos}_{ik}, \quad i \veebar j = k,\ i,j \neq p \\
A^p_{ij} &= \zeta^2/\tau^2 = \mathfrak{cos}^2_{ik}, \quad i = j = k
\end{aligned}$$

2. $p = k$.

$$\begin{aligned}
A^k_{ij} &= \nu^2/\tau^2_k = \mathfrak{cos}^2_{kk}, \quad i = j = k \\
A^k_{ij} &= \nu\eta/\tau^2_k = \mathfrak{cos}_{kk}\mathfrak{cos}_{kj}, \quad i \veebar j = k \\
A^k_{ij} &= \eta^2/\tau^2_k = \mathfrak{cos}^2_{kj}, \quad i,j < k
\end{aligned}$$

1. If $p,q,i,j < k$, $p \neq q$, define

$$\begin{aligned}
A^{pq}_{ij} &= \mathfrak{cos}^2_{ii} + \mathfrak{cos}^2_{ij}, & B^{pq}_{ij} &= 2\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}, \quad i = j \in \{p,q\} \\
&= 2\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}, & &= \mathfrak{cos}^2_{ii} + \mathfrak{cos}^2_{ij}, \quad i,j \in \{p,q\}, i \neq j \\
&= \mathfrak{cos}_{ij}\mathfrak{cos}_{ii} + \mathfrak{cos}^2_{ij}, & &= \mathfrak{cos}_{ij}\mathfrak{cos}_{ii} + \mathfrak{cos}^2_{ij}, \quad i \veebar j \in \{p,q\} \\
&= 2\mathfrak{cos}^2_{ij} & &= 2\mathfrak{cos}^2_{ij}, \quad i,j \notin \{p,q\}
\end{aligned}$$

2. If $p,q < k$, $p \neq q$, $i \veebar j = k$, define

$$\begin{aligned}
A^{pq}_{ij} &= \mathfrak{cos}_{ik}(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) & B^{pq}_{ij} &= \mathfrak{cos}_{ik}(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}), \quad i \veebar j \in \{p,q\} \\
&= 2\mathfrak{cos}_{ik}\mathfrak{cos}_{ij} & &= 2\mathfrak{cos}_{ik}\mathfrak{cos}_{ij}, \quad i,j \notin \{p,q\}
\end{aligned}$$

3. $p,q < k$, $p \neq q$, $i = j = k$, define

$$A^{pq}_{kk} = 2\mathfrak{cos}^2_{ik} \qquad B^{pq}_{kk} = 2\mathfrak{cos}^2_{ik}$$

4. If $p \veebar q = k$, and $i,j < k$ define

$$\begin{aligned}
A^{pq}_{ij} &= \mathfrak{cos}^2_{kj} + \mathfrak{cos}^2_{ii}, & B^{pq}_{ij} &= 2\mathfrak{cos}_{kj}\mathfrak{cos}_{ii}, \quad i = j \in \{p,q\} \\
&= \mathfrak{cos}^2_{kj} + \mathfrak{cos}_{ij}\mathfrak{cos}_{ii} & &= \mathfrak{cos}_{kj}(\mathfrak{cos}_{ij} + \mathfrak{cos}_{ii}), \quad i \veebar j \in \{p,q\} \\
&= \mathfrak{cos}^2_{kj} + \mathfrak{cos}^2_{ij}, & &= 2\mathfrak{cos}_{kj}\mathfrak{cos}_{ij}, \quad i,j \notin \{p,q\}
\end{aligned}$$

5. If $p \veebar q = k$, and $i \veebar j = k$ define

$$\begin{aligned}
A^{pq}_{ij} &= \mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \mathfrak{cos}_{ik}\mathfrak{cos}_{ii}, & B^{pq}_{ij} &= \mathfrak{cos}_{kk}\mathfrak{cos}_{ii} + \mathfrak{cos}_{kj}\mathfrak{cos}_{ik},\ i,j \in \{p,q\} \\
&= \mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \mathfrak{cos}_{ik}\mathfrak{cos}_{ij} & &= \mathfrak{cos}_{kk}\mathfrak{cos}_{ij} + \mathfrak{cos}_{kj}\mathfrak{cos}_{ik},\ i \veebar j \in \{p,q\}
\end{aligned}$$

6. If $p \veebar q = k$, and $i = j = k$ define

$$A^{pq}_{kk} = \mathfrak{cos}^2_{kk} + \mathfrak{cos}^2_{ik}, \qquad B^{pq}_{kk} = 2\mathfrak{cos}_{kk}\mathfrak{cos}_{ik}$$

Note that A^{pq}_{ij}, B^{pq}_{ij} are symmetric in p, q and i, j.

Proposition 23 (Off diagonal blocks)**.** *(Notation and assumptions as above.) If $p, q, i, j \in [k]$, with $p \neq q$. then*

1. *If $p, q < k$,*
$$H^{pq}_{ij} = \frac{(\pi - \Theta)\delta_{ij}}{2\pi} + \frac{B^{pq}_{ij} - \cos(\Theta)A^{pq}_{ij}}{2\pi\sin(\Theta)}.$$

2. *If $p \veebar q = k$,*
$$H^{pq}_{ij} = \frac{(\pi - \Lambda)\delta_{ij}}{2\pi} + \frac{B^{pq}_{ij} - \cos(\Lambda)A^{pq}_{ij}}{2\pi\sin(\Lambda)}.$$

In particular, $H^{pq} = H^{qp}$ and are symmetric matrices.

Proof Along exactly the same lines as that of Lemma 18. $\square$

Before giving the main lemma for computation of the terms $h_1(\boldsymbol{w}^p, \boldsymbol{w}^q)$, and $h_1(\boldsymbol{w}^p, \mathbf{v}^q)$, we need to extend the definition of K^{pq}_{ij} to allow for S_{k-1} symmetry.

Given $i, j, p, q \in [k]$, $p \neq q$, define for $(i, j) \neq (q, q)$,

$$K^{pq}_{ij} = \begin{cases} -1, i \veebar j = q \\ \mathfrak{cot}^2_{ij}, p \neq q,\ p, q < k,\ i, j \neq q \\ \mathfrak{cot}^2_{ik}, q = k,\ p < k,\ i, j \neq q \\ \mathfrak{cot}^2_{kj}, p = k,\ q < k,\ i, j \neq q \end{cases}$$

In case $p = q$, it is more convenient to give the values of $K^{qq}_{ij}A^q_{ij}$ rather than K^{qq}_{ij}, for $(i, j) \neq (q, q)$ ($K^{qq}_{qq}A^q_{qq}$ is not defined). Given $i, j, p, q \in [k]$, with $p = q$, and $(i, j) \neq (q, q)$,

1. If $q \neq k$, then
$$K^{qq}_{ij}A^q_{ij} = \begin{cases} \frac{\mathfrak{cos}^2_{ii}\mathfrak{cos}^2_{ij}}{(k-2)\mathfrak{cos}^2_{ij} + \mathfrak{cos}^2_{ik}},\ i, j \notin \{q, k\} \\ \frac{\mathfrak{cos}^2_{ii}\mathfrak{cos}^2_{ik}}{(k-2)\mathfrak{cos}^2_{ij} + \mathfrak{cos}^2_{ik}},\ i, j = k \\ \frac{\mathfrak{cos}^2_{ii}\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{(k-2)\mathfrak{cos}^2_{ij} + \mathfrak{cos}^2_{ik}},\ i \veebar j = k,\ i, j \neq q \\ -\mathfrak{cos}_{ij}\mathfrak{cos}_{ii},\ i \veebar j = q,\ i, j \neq k \\ -\mathfrak{cos}_{ik}\mathfrak{cos}_{ii},\ i \veebar j = q,\ i \veebar j = k \end{cases}$$

2. $q = k$, then
$$K^{kk}_{ij}A^k_{ij} = \begin{cases} \frac{\mathfrak{cos}^2_{kk}}{k-1},\ i, j \neq k, \\ -\mathfrak{cos}_{kj}\mathfrak{cos}_{kk},\ i \veebar j = k \end{cases}$$

Remark 24. The expressions $K^{qq}_{ij}A^q_{ij}$ are all bounded by 1. For the type II critical points of $\mathcal{F}$, $K^{qq}_{ij}A^q_{ij}$ may be very small. For example, if $q \neq k$ and $i, j \notin \{q, k\}$, then $K^{qq}_{ij}A^q_{ij} = O(k^{-2})$. On the other hand if $i, j = k \neq q$, $K^{qq}_{ij}A^q_{ij} \approx 1$ for large k. ✠

Lemma 25. *(Notation and assumptions as above.) If $p, q, i, j \in [k]$, $p \neq q$, then*

1. *If $p, q < k$*
$$h_1(\boldsymbol{w}^p, \boldsymbol{w}^q)_{ij} = \frac{\sin(\Theta)}{2\pi}\left(\delta_{ij} - A^p_{ij}\right) + \frac{1}{2\pi\sin(\Theta)}\left(A^q_{ij} - \cos(\Theta)B^{pq}_{ij} + \cos^2(\Theta)A^p_{ij}\right)$$

2. *if $p < k, q = k$*
$$h_1(\boldsymbol{w}^p, \boldsymbol{w}^k)_{ij} = \frac{\tau_k\sin(\Lambda)}{2\pi\tau}\left(\delta_{ij} - A^p_{ij}\right) + \frac{\tau_k}{2\pi\tau\sin(\Lambda)}\left(A^k_{ij} - \cos(\Lambda)B^{pk}_{ij} + \cos^2(\Lambda)A^p_{ij}\right)$$

3. *if* $p = k, q < k$.

$$h_1(\boldsymbol{w}^k, \boldsymbol{w}^q)_{ij} \quad = \quad \frac{\tau \sin(\Lambda)}{2\pi\tau_k}\left(\delta_{ij} - A^k_{ij}\right) + \frac{\tau}{2\pi\tau_k \sin(\Lambda)}\left(A^q_{ij} - \cos(\Lambda)B^{kq}_{ij} + \cos^2(\Lambda)A^k_{ij}\right)$$

If $p, q, i, j \in [k]$, *then*

1. $p \neq q$, $p, q < k$.

$$\begin{aligned} h_1(\boldsymbol{w}^p, \mathbf{v}^q)_{ij} &= \frac{\sin(\alpha_{ij})}{2\pi\tau}\left(\delta_{ij} - A^p_{ij} + K^{pq}_{ij}A^p_{ij}\right),\ (i,j) \neq (q,q) \\ &= \frac{\sin^3(\alpha_{ij})}{\pi\tau},\ i = j = q, \end{aligned}$$

2. $p \neq q$, $p < k = q$,

$$\begin{aligned} h_1(\boldsymbol{w}^p, \mathbf{v}^k)_{ij} &= \frac{\sin(\alpha_{ik})}{2\pi\tau}\left(\delta_{ij} - A^p_{ij} + K^{pk}_{ij}A^p_{ij}\right),\ (i,j) \neq (k,k) \\ &= \frac{\sin^3(\alpha_{ik})}{\pi\tau},\ i = j = k, \end{aligned}$$

3. $q < k = p$,

$$\begin{aligned} h_1(\boldsymbol{w}^k, \mathbf{v}^q)_{ij} &= \frac{\sin(\alpha_{kj})}{2\pi\tau_k}\left(\delta_{ij} - A^k_{ij} + K^{kq}_{ij}A^k_{ij}\right),\ (i,j) \neq (q,q) \\ &= \frac{\sin^3(\alpha_{kj})}{\pi\tau_k},\ i = j = q, \end{aligned}$$

4. $p = q < k$,

$$\begin{aligned} h_1(\boldsymbol{w}^p, \mathbf{v}^p)_{ij} &= \frac{\sin(\alpha_{ii})}{2\pi\tau}\left(\delta_{ij} - A^p_{ij} + K^{pp}_{ij}A^p_{ij}\right),\ (i,j) \neq (p,p) \\ &= \frac{\sin^3(\alpha_{ii})}{\pi\tau},\ i = j = p, \end{aligned}$$

5. $p = q = k$,

$$\begin{aligned} h_1(\boldsymbol{w}^k, \mathbf{v}^k)_{ij} &= \frac{\sin(\alpha_{kk})}{2\pi\tau_k}\left(\delta_{ij} - A^k_{ij} + K^{kk}_{ij}A^k_{ij}\right),\ (i,j) \neq (k,k) \\ &= \frac{\sin^3(\alpha_{kk})}{\pi\tau_k},\ i = j = k, \end{aligned}$$

Proposition 26 (Off diagonal blocks)**.** *Assume* $p, q, i, j \in [k]$ *and* $p \neq q$.

(A) *If* $p, q < k$, $i = j$, *then if*

(a) $i \notin \{p, q\}$, $i < k$,

$$H^{pq}_{ii} \quad = \quad \frac{\pi - \Theta}{2\pi} + \frac{\mathfrak{cos}^2_{ij}(1 - \cos(\Theta))}{\pi \sin(\Theta)}$$

(b) $i \in \{p, q\}$,

$$H^{pq}_{ii} \quad = \quad \frac{\pi - \Theta}{2\pi} + \frac{2\mathfrak{cos}_{ii}\mathfrak{cos}_{ij} - (\mathfrak{cos}^2_{ii} + \mathfrak{cos}^2_{ij})\cos(\Theta)}{2\pi \sin(\Theta)}$$

(c) $i = k$,

$$H^{pq}_{kk} \quad = \quad \frac{\pi - \Theta}{2\pi} + \frac{\mathfrak{cos}^2_{ik}(1 - \cos(\Theta))}{\pi \sin(\Theta)}$$

(B) If $p \veebar q = k$*,* $i = j$*, then if*

(a) $i < k$*,* $i = p \veebar q$*,*

$$H_{ii}^{pq} = \frac{\pi - \Lambda}{2\pi} + \frac{2\mathfrak{cos}_{ii}\mathfrak{cos}_{kj} - (\mathfrak{cos}_{ii}^2 + \mathfrak{cos}_{kj}^2)\cos(\Lambda)}{2\pi\sin(\Lambda)}$$

(b) $i < k$*,* $i \notin \{p, q\}$*,*

$$H_{ii}^{pq} = \frac{\pi - \Lambda}{2\pi} + \frac{2\mathfrak{cos}_{ij}\mathfrak{cos}_{kj} - (\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{kj}^2)\cos(\Lambda)}{2\pi\sin(\Lambda)}$$

(c) $i = k$*,*

$$H_{kk}^{pq} = \frac{\pi - \Lambda}{2\pi} + \frac{2\mathfrak{cos}_{ik}\mathfrak{cos}_{kk} - (\mathfrak{cos}_{ik}^2 + \mathfrak{cos}_{kk}^2)\cos(\Lambda)}{2\pi\sin(\Lambda)}$$

(C) If $p, q < k$*,* $i \neq j$*, then if*

(a) $i, j \notin \{p, q\}$*,* $i, j < k$*,*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{ij}^2(1 - \cos(\Theta))}{\pi\sin(\Theta)}$$

(b) $i \veebar j \in \{p, q\}$*,* $i, j < k$*,*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{ij}\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}^2}{2\pi\sin(\Theta)}\big(1 - \cos(\Theta)\big)$$

(c) $i, j \in \{p, q\}$*,* $i, j < k$*,*

$$H_{ij}^{pq} = \frac{1}{2\pi\sin(\Theta)}\big(\mathfrak{cos}_{ii}^2 + \mathfrak{cos}_{ij}^2 - 2\cos(\Theta)\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}\big)$$

(d) $i \veebar j \in \{p, q\}$*,* $i \veebar j = k$*,*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{ik}(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij})}{2\pi\sin(\Theta)}\big(1 - \cos(\Theta)\big)$$

(e) $i, j \notin \{p, q\}$*,* $i \veebar j = k$*,*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{ik}\mathfrak{cos}_{ij}}{\pi\sin(\Theta)}\big(1 - \cos(\Theta)\big)$$

(D) If $p \veebar q = k$*,* $i \neq j$*, then if*

(a) $i, j \notin \{p, q\}$*,* $i, j < k$*,*

$$H_{ij}^{pq} = \frac{2\mathfrak{cos}_{ij}\mathfrak{cos}_{kj} - (\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{kj}^2)\cos(\Lambda)}{2\pi\sin(\Lambda)}$$

(b) $i \veebar j \in \{p, q\}$*,* $i, j < k$*,*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{kj}(\mathfrak{cos}_{ij} + \mathfrak{cos}_{ii}) - \cos(\Lambda)(\mathfrak{cos}_{kj}^2 + \mathfrak{cos}_{ij}\mathfrak{cos}_{ii})}{2\pi\sin(\Lambda)}$$

(c) $i \veebar j = k$*,* $i, j \in \{p, q\}$*.*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{kk}\mathfrak{cos}_{ii} + \mathfrak{cos}_{kj}\mathfrak{cos}_{ik} - \cos(\Lambda)\big(\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \mathfrak{cos}_{ik}\mathfrak{cos}_{ii}\big)}{2\pi\sin(\Lambda)}$$

(d) $i \veebar j = k$*, and* $i \veebar j \notin \{p, q\}$*.*

$$H_{ij}^{pq} = \frac{\mathfrak{cos}_{kk}\mathfrak{cos}_{ij} + \mathfrak{cos}_{ik}\mathfrak{cos}_{kj} - \cos(\Lambda)\big(\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \mathfrak{cos}_{ik}\mathfrak{cos}_{ij}\big)}{2\pi\sin(\Lambda)}$$

In particular, $H^{pq} = H^{qp}$ and the matrices H^{pq} are all symmetric.

Proposition 27 (Diagonal blocks)**.** *Assume $p, i, j \in [k]$.*

(A) If $i = j$, then if

(a) $i \notin \{p, k\}$, $p < k$,

$$\begin{aligned} H_{ii}^{pp} &= \frac{1}{2} + \frac{(k-2)\mathfrak{sin}_{ij}^2}{2\pi}\left(\sin(\Theta) - \frac{\mathfrak{sin}_{ij}}{\tau}\right) + \\ &\frac{(k-3)}{2\pi}\left(\frac{\mathfrak{cos}_{ij}^2(1-\cos(\Theta))^2}{\sin(\Theta)} - \frac{\mathfrak{sin}_{ij}\mathfrak{cot}_{ij}^2\mathfrak{cos}_{ij}^2}{\tau}\right) + \\ &\frac{(\mathfrak{cos}_{ii} - \cos(\Theta)\mathfrak{cos}_{ij})^2}{2\pi\sin(\Theta)} - \frac{\mathfrak{sin}_{ij}^3}{2\pi\tau} + \frac{\mathfrak{sin}_{ij}^2(\tau_k\sin(\Lambda) - \mathfrak{sin}_{ik})}{2\pi\tau} - \frac{\mathfrak{sin}_{ik}\mathfrak{cot}_{ik}^2\mathfrak{cos}_{ij}^2}{2\pi\tau} + \\ &\frac{\tau_k}{2\pi\tau\sin(\Lambda)}\left(\mathfrak{cos}_{kj} - \cos(\Lambda)\mathfrak{cos}_{ij}\right)^2 - \frac{\mathfrak{sin}_{ii}}{2\pi\tau}\left(\mathfrak{sin}_{ij}^2 + \frac{\mathfrak{cos}_{ii}^2\mathfrak{cos}_{ij}^2}{(k-2)\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{ik}^2}\right) \end{aligned}$$

(b) $i = p$, $p < k$,

$$\begin{aligned} H_{pp}^{pp} &= \frac{1}{2} + \frac{(k-2)\mathfrak{sin}_{ii}^2}{2\pi}\left(\sin(\Theta) - \frac{\mathfrak{sin}_{ij}}{\tau}\right) + \\ &\frac{(k-2)}{2\pi}\left(\frac{\left(\mathfrak{cos}_{ij} - \cos(\Theta)\mathfrak{cos}_{ii}\right)^2}{\sin(\Theta)} - \frac{\mathfrak{sin}_{ij}\mathfrak{cos}_{ii}^2\mathfrak{cot}_{ij}^2}{\tau}\right) + \\ &\frac{\mathfrak{sin}_{ii}^2}{2\pi\tau}\left(\tau_k\sin(\Lambda) - \mathfrak{sin}_{ik}\right) - \frac{\mathfrak{sin}_{ii}^3}{\pi\tau} + \\ &\frac{\tau_k}{2\pi\tau\sin(\Lambda)}\left(\mathfrak{cos}_{kj} - \cos(\Lambda)\mathfrak{cos}_{ii}\right)^2 - \frac{\mathfrak{sin}_{ik}\mathfrak{cot}_{ik}^2\mathfrak{cos}_{ii}^2}{2\pi\tau} \end{aligned}$$

(c) if $i = k$, $p < k$,

$$\begin{aligned} H_{kk}^{pp} &= \frac{1}{2} + \frac{(k-2)\mathfrak{sin}_{ik}^2}{2\pi}\left(\sin(\Theta) - \frac{\mathfrak{sin}_{ij}}{\tau}\right) + \\ &\frac{(k-2)\mathfrak{cos}_{ik}^2}{2\pi}\left(\frac{(1-\cos(\Theta))^2}{\sin(\Theta)} - \frac{\mathfrak{sin}_{ij}\mathfrak{cot}_{ij}^2}{\tau}\right) + \\ &\frac{\mathfrak{sin}_{ik}^2}{2\pi\tau}\left(\tau_k\sin(\Lambda) - \mathfrak{sin}_{ii}\right) - \frac{\mathfrak{sin}_{ik}^3}{\pi\tau} + \\ &\frac{\tau_k}{2\pi\tau\sin(\Lambda)}\left(\mathfrak{cos}_{kk} - \cos(\Lambda)\mathfrak{cos}_{ik}\right)^2 - \\ &\frac{\mathfrak{sin}_{ii}}{2\pi\tau}\left(\frac{\mathfrak{cos}_{ii}^2\mathfrak{cos}_{ik}^2}{(k-2)\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{ik}^2}\right) \end{aligned}$$

(d) $i \neq k$, $p = k$,

$$\begin{aligned} H_{ii}^{kk} &= \frac{1}{2} + \frac{(k-1)\mathfrak{sin}_{kj}^2}{2\pi\tau_k}\left(\tau\sin(\Lambda) - \mathfrak{sin}_{kj}\right) + \\ &\frac{(k-2)}{2\pi\tau_k}\left(\frac{\tau(\mathfrak{cos}_{ij} - \cos(\Lambda)\mathfrak{cos}_{kj})^2}{\sin(\Lambda)} - \mathfrak{sin}_{kj}\mathfrak{cos}_{kj}^2\mathfrak{cot}_{kj}^2\right) - \\ &\frac{\mathfrak{sin}_{kj}^3}{2\pi\tau_k} - \frac{\mathfrak{sin}_{kk}}{2\pi\tau_k}\left(\mathfrak{sin}_{kj}^2 + \frac{\mathfrak{cos}_{kk}^2}{k-1}\right) + \\ &\frac{\tau}{2\pi\tau_k\sin(\Lambda)}\left(\mathfrak{cos}_{ii} - \cos(\Lambda)\mathfrak{cos}_{kj}\right)^2 \end{aligned}$$

(e) $i = k$, $p = k$,

$$H_{kk}^{kk} = \frac{1}{2} + \frac{(k-1)\mathfrak{sin}_{kk}^2}{2\pi\tau_k}\big(\tau\sin(\Lambda) - \mathfrak{sin}_{kj}\big) + \frac{(k-1)}{2\pi\tau_k}\left(\frac{\tau(\mathfrak{cos}_{ik} - \cos(\Lambda)\mathfrak{cos}_{kk})^2}{\sin(\Lambda)} - \mathfrak{cot}_{kj}^2\mathfrak{cos}_{kk}^2\mathfrak{sin}_{kj}\right) - \frac{\mathfrak{sin}_{kk}^3}{\pi\tau_k}$$

(B) If $i \neq j$, then if

(a) $i, j \notin \{p, k\}$, $p < k$,

$$\begin{aligned} H_{ij}^{pp} = {} & \frac{\mathfrak{cos}_{ij}^2(k-2)}{2\pi}\left(\frac{\mathfrak{sin}_{ij}}{\tau} - \sin(\Theta)\right) + \frac{(k-4)\mathfrak{cos}_{ij}^2}{2\pi\sin(\Theta)}\big(1-\cos(\Theta)\big)^2 - \\ & \frac{(k-4)\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}^2\mathfrak{cot}_{ij}^2}{2\pi\tau} + \frac{\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}^2}{\pi\tau} + \\ & \frac{\mathfrak{cos}_{ij}}{\pi\sin(\Theta)}\big(\mathfrak{cos}_{ii} - \cos(\Theta)(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) + \cos^2(\Theta)\mathfrak{cos}_{ij}\big) + \\ & \frac{\mathfrak{cos}_{ij}^2}{2\pi\tau}\big(\mathfrak{sin}_{ik} - \tau_k\sin(\Lambda)\big) + \frac{\tau_k}{2\pi\tau\sin(\Lambda)}\big(\mathfrak{cos}_{kj} - \cos(\Lambda)\mathfrak{cos}_{ij}\big)^2 - \\ & \frac{\mathfrak{sin}_{ik}}{2\pi\tau}\mathfrak{cot}_{ik}^2\mathfrak{cos}_{ij}^2 + \frac{\mathfrak{sin}_{ii}}{2\pi\tau}\left(\mathfrak{cos}_{ij}^2 - \frac{\mathfrak{cos}_{ii}^2\mathfrak{cos}_{ij}^2}{(k-2)\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{ik}^2}\right) \end{aligned}$$

(b) $i \veebar j = p$, $i, j \neq k$, $p < k$.

$$\begin{aligned} H_{ip}^{pp} = {} & \frac{(k-2)\mathfrak{cos}_{ij}\mathfrak{cos}_{ii}}{2\pi}\left(\frac{\mathfrak{sin}_{ij}}{\tau} - \sin(\Theta)\right) + \\ & \frac{(k-3)\mathfrak{cos}_{ij}}{2\pi\sin(\Theta)}\big(\mathfrak{cos}_{ij} - \cos(\Theta)(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) + \cos^2(\Theta)\mathfrak{cos}_{ii}\big) - \\ & \frac{(k-3)\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}\mathfrak{cos}_{ii}\mathfrak{cot}_{ij}^2}{2\pi\tau} + \\ & \frac{\big(\mathfrak{cos}_{ii}\mathfrak{cos}_{ij} - \cos(\Theta)(\mathfrak{cos}_{ii}^2 + \mathfrak{cos}_{ij}^2) + \cos^2(\Theta)\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}\big)}{2\pi\sin(\Theta)} + \frac{\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}\mathfrak{cos}_{ii}}{2\pi\tau} + \\ & \frac{\mathfrak{cos}_{ij}\mathfrak{cos}_{ii}}{2\pi\tau}\left(\mathfrak{sin}_{ik} - \tau_k\sin(\Lambda)\right) - \frac{\mathfrak{sin}_{ik}}{2\pi\tau}\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}\mathfrak{cot}_{ik}^2 + \\ & \frac{\tau_k}{2\pi\tau\sin(\Lambda)}\big(\mathfrak{cos}_{kj}^2 - \cos(\Lambda)(\mathfrak{cos}_{kj}[\mathfrak{cos}_{ij} + \mathfrak{cos}_{ii}]) + \cos^2(\Lambda)\mathfrak{cos}_{ii}\mathfrak{cos}_{ij}\big) + \frac{\mathfrak{sin}_{ii}}{\pi\tau}\mathfrak{cos}_{ii}\mathfrak{cos}_{ij} \end{aligned}$$

(c) $i \veebar j = k$, $i, j \neq p$, $p < k$,

$$\begin{aligned} H_{ik}^{pp} = {} & \frac{(k-2)\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{2\pi}\left(\frac{\mathfrak{sin}_{ij}}{\tau} - \sin(\Theta)\right) + \\ & \frac{(k-3)\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{2\pi\sin(\Theta)}\big(1-\cos(\Theta)\big)^2 - \frac{(k-3)\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}\mathfrak{cot}_{ij}^2}{2\pi\tau} + \\ & \frac{\mathfrak{cos}_{ik}\big(\mathfrak{cos}_{ii} - \cos(\Theta)(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) + \cos^2(\Theta)\mathfrak{cos}_{ij}\big)}{2\pi\sin(\Theta)} + \frac{\mathfrak{sin}_{ij}\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{2\pi\tau} + \\ & \frac{\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{2\pi\tau}\big(\mathfrak{sin}_{ik} - \tau_k\sin(\Lambda)\big) + \\ & \frac{\tau_k\big(\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} - \cos(\Lambda)(\mathfrak{cos}_{kk}\mathfrak{cos}_{ij} + \mathfrak{cos}_{ik}\mathfrak{cos}_{kj}) + \cos^2(\Lambda)\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}\big)}{2\pi\tau\sin(\Lambda)} + \\ & \frac{\mathfrak{sin}_{ik}\mathfrak{cos}_{ik}\mathfrak{cos}_{ij}}{2\pi\tau} + \frac{\mathfrak{sin}_{ii}}{2\pi\tau}\left(\mathfrak{cos}_{ij}\mathfrak{cos}_{ik} - \frac{\mathfrak{cos}_{ii}^2\mathfrak{cos}_{ij}\mathfrak{cos}_{ik}}{(k-2)\mathfrak{cos}_{ij}^2 + \mathfrak{cos}_{ik}^2}\right) \end{aligned}$$

(d) $i, j \in \{p, k\}$, $p < k$,

$$\begin{aligned}
H^{pp}_{pk} &= \frac{(k-2)\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}}{2\pi}\left(\frac{\mathfrak{sin}_{ij}}{\tau} - \sin(\Theta)\right) + \\
&\quad \frac{(k-2)\mathfrak{cos}_{ik}}{2\pi\sin(\Theta)}\big(\mathfrak{cos}_{ij} - \cos(\Theta)(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) + \cos^2(\Theta)\mathfrak{cos}_{ii}\big) - \\
&\quad \frac{(k-2)\mathfrak{sin}_{ij}\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}\mathfrak{cot}^2_{ij}}{2\pi\tau} + \frac{\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}}{2\pi\tau}\big(\mathfrak{sin}_{ik} - \tau_k\sin(\Lambda)\big) + \\
&\quad \frac{\tau_k\big(\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} - \cos(\Lambda)(\mathfrak{cos}_{ii}\mathfrak{cos}_{kk} + \mathfrak{cos}_{ik}\mathfrak{cos}_{kj}) + \cos^2(\Lambda)\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}\big)}{2\pi\tau\sin(\Lambda)} + \\
&\quad \frac{\mathfrak{sin}_{ik}\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}}{2\pi\tau} + \frac{\mathfrak{sin}_{ii}\mathfrak{cos}_{ii}\mathfrak{cos}_{ik}}{\pi\tau}
\end{aligned}$$

(e) $i, j \neq p$, $p = k$,

$$\begin{aligned}
H^{kk}_{ij} &= \frac{(k-1)\mathfrak{cos}^2_{kj}}{2\pi\tau_k}\big(\mathfrak{sin}_{kj} - \tau\sin(\Lambda)\big) + \frac{(k-3)\tau}{2\pi\tau_k\sin(\Lambda)}\big(\mathfrak{cos}_{ij} - \cos(\Lambda)\mathfrak{cos}_{kj}\big)^2 - \\
&\quad \frac{(k-3)\mathfrak{sin}_{kj}\mathfrak{cot}^2_{kj}\mathfrak{cos}^2_{kj}}{2\pi\tau_k} + \frac{\mathfrak{sin}_{kj}\mathfrak{cos}^2_{kj}}{\pi\tau_k} + \\
&\quad \frac{\tau}{\pi\tau_k\sin(\Lambda)}\big(\mathfrak{cos}_{ii}\mathfrak{cos}_{ij} - \cos(\Lambda)\mathfrak{cos}_{kj}(\mathfrak{cos}_{ii} + \mathfrak{cos}_{ij}) + \cos^2(\Lambda)\mathfrak{cos}^2_{kj}\big) + \\
&\quad \frac{\mathfrak{sin}_{kk}}{2\pi\tau_k}\left(\mathfrak{cos}^2_{kj} - \frac{\mathfrak{cos}^2_{kk}}{k-1}\right)
\end{aligned}$$

(f) $i \veebar j = p$, $p = k$,

$$\begin{aligned}
H^{kk}_{ik} &= \frac{(k-1)\mathfrak{cos}_{kk}\mathfrak{cos}_{kj}}{2\pi\tau_k}\big(\mathfrak{sin}_{kj} - \tau\sin(\Lambda)\big) - \frac{(k-2)\mathfrak{sin}_{kj}}{2\pi\tau_k}\mathfrak{cot}^2_{kj}\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \\
&\quad \frac{(k-2)\tau\big(\mathfrak{cos}_{ij}\mathfrak{cos}_{ik} - \cos(\Lambda)(\mathfrak{cos}_{kk}\mathfrak{cos}_{ij} + \mathfrak{cos}_{ik}\mathfrak{cos}_{kj}) + \cos^2(\Lambda)\mathfrak{cos}_{kk}\mathfrak{cos}_{kj}\big)}{2\pi\tau_k\sin(\Lambda)} + \\
&\quad \frac{\mathfrak{sin}_{kj}}{2\pi\tau_k}\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \frac{\mathfrak{sin}_{kk}}{\pi\tau_k}\mathfrak{cos}_{kk}\mathfrak{cos}_{kj} + \\
&\quad \frac{\tau}{2\pi\tau_k\sin(\Lambda)}\big(\mathfrak{cos}_{ii}\mathfrak{cos}_{ik} - \cos(\Lambda)(\mathfrak{cos}_{kk}\mathfrak{cos}_{ii} + \mathfrak{cos}_{ik}\mathfrak{cos}_{kj}) + \cos^2(\Lambda)\mathfrak{cos}_{kk}\mathfrak{cos}_{kj}\big)
\end{aligned}$$

D Estimating the Hessian spectrum

The stage is now set for deriving the estimates for the Hessian spectrum. We first present a detailed derivation of the spectrum of type II minima which follows along the same lines of Example 22, and then briefly state the adjustments needed for the analysis of types A and I.

D.1 The spectrum at type II minima

First, we use the infinite series representation for type II minima given in Lemma 5 to obtain the following estimates (notations as in Proposition 26 and Proposition 27):

1. $\cos(\Theta) = (2e_4 + 4)k^{-2} + 2e_5 k^{-\frac{5}{2}}$.
2. $\sin(\Theta) = 1 + O(k^{-4})$
3. $\cos(\Lambda) = -(e_4 + 2)k^{-1} - e_5 k^{-\frac{3}{2}}$.
4. $\sin(\Lambda) = 1 - \frac{(e_4+2)^2}{2k^2} - (e_4 + 2)e_5 k^{-\frac{5}{2}}$.
5. $\mathfrak{cos}_{ii} = 1 - \frac{2}{k^2}$.
6. $\mathfrak{sin}_{ii} = 2k^{-1} + (\frac{e_4^2}{4} + 2 - d_2)k^{-2}$.

7. $\mathfrak{cos}_{ij} = \frac{e_4}{k^2} + e_5 k^{-\frac{5}{2}}$.

8. $\mathfrak{sin}_{ij} = 1 + O(k^{-4})$.

9. $\mathfrak{cos}_{ik} = 2k^{-1} + (2 - d_2)k^{-2}$.

10. $\mathfrak{sin}_{ik} = 1 - 2k^{-2}$.

11. $\mathfrak{cos}_{kk} = -1 + \frac{e_4^2}{2k} + e_4 e_5 k^{-\frac{3}{2}}$.

12. $\mathfrak{sin}_{kk} = -\frac{e_4}{\sqrt{k}} - e_5 k^{-1}$.

13. $\mathfrak{cos}_{kj} = -e_4 k^{-1} - e_5 k^{-\frac{3}{2}}$.

14. $\mathfrak{sin}_{kj} = 1 - \frac{e_4^2}{2k^2}$.

15. $\tau = 1 + (c_4 + 2)k^{-2} + c_5 k^{-\frac{5}{2}}$

16. $\tau^{-1} = 1 - (c_4 + 2)k^{-2} - c_5 k^{-\frac{5}{2}}$.

17. $\tau_k = 1 + \frac{e_4^2 - 2d_2}{2} k^{-1} + (e_4 e_5 - d_3)k^{-\frac{3}{2}}$

18. $\tau_k^{-1} = 1 - \frac{e_4^2 - 2d_2}{2} k^{-1} - (e_4 e_5 - d_3)k^{-\frac{3}{2}}$

Next, we use these 'primitive' estimates to compute the Hessian entries. The estimates for the entries of off-diagonal blocks are obtained through the respective expressions in Proposition 26, see Table 2.

Case	Sub-case	Hessian entry	Estimate
A	a	H_{ii}^{pq}	$\frac{1}{4} + O\left(k^{-2}\right)$
A	b	H_{ii}^{pq}	$\frac{1}{4} + O\left(k^{-2}\right)$
A	c	H_{kk}^{pq}	$\frac{1}{4} + O\left(k^{-2}\right)$
B	a	H_{ii}^{pq}	$\frac{1}{4} - \frac{e_4}{\pi} k^{-1} - \frac{e_5 k^{-1.5}}{\pi} + O\left(k^{-2}\right)$
B	b	H_{ii}^{pq}	$\frac{1}{4} + \left(-\frac{e_4}{2\pi} - \frac{1}{\pi}\right) k^{-1} - \frac{e_5 k^{-1.5}}{2\pi} + O\left(k^{-2}\right)$
B	c	H_{ii}^{pq}	$\frac{1}{4} - \frac{2}{\pi} k^{-1} + O\left(k^{-2}\right)$
C	a	H_{ij}^{pq}	$O\left(k^{-2}\right)$
C	b	H_{ij}^{pq}	$O\left(k^{-2}\right)$
C	c	H_{ij}^{pq}	$\frac{1}{2\pi} + O\left(k^{-2}\right)$
C	d	H_{ij}^{pq}	$\frac{k^{-1}}{\pi} + O\left(k^{-2}\right)$
C	e	H_{ij}^{pq}	$O\left(k^{-2}\right)$
D	a	H_{ij}^{pq}	$O\left(k^{-2}\right)$
D	b	H_{ij}^{pq}	$-\frac{e_5 k^{-1.5}}{2\pi} - \frac{e_4 k^{-1}}{2\pi} + O\left(k^{-2}\right)$
D	c	H_{ij}^{pq}	$-\frac{1}{2\pi} + \frac{e_4 e_5 k^{-1.5}}{2\pi} + \frac{e_4^2 k^{-1}}{4\pi} + O\left(k^{-2}\right)$
D	d	H_{ij}^{pq}	$O\left(k^{-2}\right)$

Table 2: Estimates for the entries of off-diagonal blocks of the Hessian using the expression derived in Proposition 26.

Similarly, the estimates for the entries of the diagonal blocks are obtained through the relevant expressions in Proposition 27.

Hessian entry	Estimate
H_{ii}^{pp}	$0.5 + k^{-1.5}\left(\frac{0.5c_5}{\pi} - \frac{0.5d_3}{\pi} + \frac{0.5e_4}{\pi}e_5\right) + k^{-1.0}\left(\frac{0.5c_4}{\pi} - \frac{0.5d_2}{\pi} + \frac{0.25e_4^2}{\pi}\right) + \mathcal{O}\left(k^{-2}\right)$
H_{pp}^{pp}	$0.5 + \mathcal{O}\left(k^{-2}\right)$
H_{kk}^{pp}	$0.5 + k^{-1.5}\left(\frac{0.5c_5}{\pi} - \frac{1.0d_3}{\pi}\right) + k^{1.0}\left(\frac{0.5c_4}{\pi} - \frac{1.0d_2}{\pi} + \frac{1.0}{\pi}\right) + \mathcal{O}\left(k^{-2}\right)$
H_{ii}^{kk}	$0.5 + k^{-1.5}\left(\frac{0.5c_5}{\pi} + \frac{0.5d_2}{\pi}e_4 - \frac{0.25e_4^3}{\pi} - \frac{0.5e_4}{\pi}e_5 + \frac{e_4}{2\pi} - \frac{1.0e_5}{\pi}\right)$
	$+k^{-1.0}\left(\frac{0.5c_4}{\pi} - \frac{1.0e_4}{\pi} + \frac{e_5}{2\pi}\right) + \frac{e_4k^{-0.5}}{2\pi} + \mathcal{O}\left(k^{-2}\right)$
H_{kk}^{kk}	$0.5 + \frac{e_4^3k^{-1.5}}{\pi} + \mathcal{O}\left(k^{-2}\right)$
H_{ij}^{pp}	$\mathcal{O}\left(k^{-2}\right)$
H_{ip}^{pp}	$\mathcal{O}\left(k^{-2}\right)$
H_{ik}^{pp}	$k^{-1.0}\left(\frac{0.5e_4}{\pi} + \frac{1}{\pi}\right) + \frac{0.5e_5}{\pi}k^{-1.5} + \mathcal{O}\left(k^{-2}\right)$
H_{pk}^{pp}	$\mathcal{O}\left(k^{-2}\right)$
H_{ij}^{kk}	$\frac{e_4k^{-1.5}}{2\pi} + \mathcal{O}\left(k^{-2}\right)$
H_{ik}^{kk}	$-\frac{e_4^2k^{-1.5}}{\pi} + \mathcal{O}\left(k^{-2}\right)$

Table 3: Estimates for the entries of diagonal blocks of the Hessian using Proposition 27.

Our next goal is to compute the product of the Hessian of type II minima by the representative vectors described in Section B.4.

The eigenvalues $\lambda_{\mathfrak{x}}$ and $\lambda_{\mathfrak{y}}$. Computing $H(\mathfrak{X}^{k-1,1})_{12} = \langle \mathbf{r}_2, \mathfrak{X}^{k-1,1}\rangle$, we find that

$$H(\mathfrak{X}^{k-1,1})_{12} = H_{22}^{11} - H_{ij}^{11} - H_{12}^{12} + 2H_{1j}^{12} - H_{33}^{12} = \frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2}).$$

Since $\mathfrak{X}_{12}^{k-1,1} = 1$, $\lambda_{\mathfrak{x}} = \frac{1}{4} - \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2})$. Along similar lines. we find that

$$\begin{aligned} H(\mathfrak{Y}^{k-1,1})_{12} =&(k-4)H_{22}^{11} - (k-4)H_{ij}^{11} + (k-4)H_{12}^{12} \\ &- 2(k-4))H_{1j}^{12} - (k-4)H_{33}^{12} + 2(k-4)H_{ij}^{12} \\ =&\frac{1}{4} + \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2}), \end{aligned}$$

by which we conclude $\lambda_{\mathfrak{y}} = \frac{1}{4} + \frac{1}{2\pi} - \frac{1}{\pi k} + O(k^{-2})$.

Remarks 28. (1) The $1/k$ term that occurs for both eigenvalues appears to be a correction in going from k to $k-1$—to the $(k-1)^2$ block in $M(k,k)$.

(2) Using the results in [24] it is not difficult to compute the coefficient of k^{-2} in power series (in $1/\sqrt{k}$) for both eigenvalues. Although the coefficient of $k^{-\frac{3}{2}}$ is zero, the coefficients of higher order fractional powers of $1/k$ are typically non-zero. ✠

The eigenvalues associated to $\mathfrak{s}_k$. As in the case of $\boldsymbol{W} = \boldsymbol{V}$ (Example Example 22), we denote the matrix associated to the factor $3\mathfrak{s}_k$ by $B_{\mathfrak{s}} \in M(3,3)$, and use Table 4 to show that modulo $o(1)$ terms,

$$B_{\mathfrak{s}} = \begin{pmatrix} 0.5 & -0.25k & 0.25k & 0 & 0.25 \\ -0.125 & 0.125k+0.5 & -0.125k & 0 & -0.125 \\ 0.125 & -0.125k-0.25 & 0.125k+0.25 & 0 & 0.125 \\ 0 & 0 & 0 & 0.25 & -0.5/\pi \\ 0.25 & -0.25k & 0.25k & -0.5/\pi & 0.5 \end{pmatrix}$$

This allows us to compute the coefficients of the linear term for eigenvalues of the form $a + bk + o(1)$. Indeed, taking the limit of $B_{\mathfrak{s}}/k$ for $k \to \infty$, we have that the b-coefficients are the eigenvalues of

$$\begin{pmatrix} 0 & -0.25 & 0.25 & 0 & 0 \\ 0 & 0.125 & -0.125 & 0 & 0 \\ 0 & -0.125 & 0.125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.25 & 0.25 & 0 & 0 \end{pmatrix}$$

which are zero, except for a single eigenvalue which equals 0.25.

	$\mathfrak{S}_1^{k-1,k-1}$
$H(\cdot)_{11}$	$H_{11}^{11} - H_{12}^{12}$
$H(\cdot)_{12}$	$H_{1j}^{11} - H_{11}^{12}$
$H(\cdot)_{13}$	$H_{1j}^{11} - H_{1j}^{12}$
$H(\cdot)_{1k}$	$H_{1k}^{11} - H_{1k}^{12}$
$H(\cdot)_{k1}$	$H_{11}^{1k} - H_{1j}^{1k}$
	$\mathfrak{S}_2^{k-1,k-1}$
$H(\cdot)_{11}$	$(k-1)H_{1j}^{11} - (k-1)H_{11}^{12}$
$H(\cdot)_{12}$	$2H_{22}^{11} + (k-3)H_{ij}^{11} - 2H_{12}^{12} - 2(k-3)H_{1j}^{12} + (k-3)H_{33}^{12}$
$H(\cdot)_{13}$	$H_{22}^{11} + (k-2)H_{ij}^{11} - H_{12}^{12} - (k-3)H_{1j}^{12} - H_{33}^{12}$
$H(\cdot)_{1k}$	$(k-1)H_{ik}^{11} - (k-1)H_{1k}^{12}$
$H(\cdot)_{k1}$	$(k-1)H_{1j}^{1k} - (k-1)H_{22}^{1k}$
	$\mathfrak{S}_3^{k-1,k-1}$
$H(\cdot)_{11}$	$(k-3)H_{1j}^{11} + (k-3)H_{11}^{12} - 2(k-3)H_{1j}^{12}$
$H(\cdot)_{12}$	$(k-3)H_{ij}^{11} - (k-3)H_{33}^{12}$
$H(\cdot)_{13}$	$H_{22}^{11} + (k-4)H_{ij}^{11} + H_{12}^{12} + (k-5)H_{1j}^{12} - H_{33}^{12} - 2(k-4)H_{ij}^{12}$
$H(\cdot)_{1k}$	$(k-3)H_{ik}^{11} + (k-3)H_{1k}^{12} - 2(k-3)H_{ik}^{12}$
$H(\cdot)_{k1}$	$(k-3)H_{1j}^{1k} + (k-3)H_{22}^{1k} - 2(k-3)H_{ij}^{1k}$
	$\mathfrak{S}_4^{k-1,1}$
$H(\cdot)_{11}$	$H_{1k}^{11} - H_{1k}^{12}$
$H(\cdot)_{12}$	$H_{ik}^{11} - H_{1k}^{12}$
$H(\cdot)_{13}$	$H_{ik}^{11} - H_{ik}^{12}$
$H(\cdot)_{1k}$	$H_{kk}^{11} - H_{kk}^{12}$
$H(\cdot)_{k1}$	$H_{1k}^{1k} - H_{ik}^{1k}$
	$\mathfrak{S}_5^{1,k-1}$
$H(\cdot)_{11}$	$H_{11}^{1k} - H_{1j}^{1k}$
$H(\cdot)_{12}$	$H_{1j}^{1k} - H_{22}^{1k}$
$H(\cdot)_{13}$	$H_{1j}^{1k} - H_{ij}^{1k}$
$H(\cdot)_{1k}$	$H_{1k}^{1k} - H_{ik}^{1k}$
$H(\cdot)_{k1}$	$H_{11}^{kk} - H_{ij}^{kk}$

Table 4: Product of the Hessian matrix by the representative vectors of the standard representation, as described in Section B.4

The eigenvalues associated to $\mathfrak{t}_k$. The matrix associated to the factor $3\mathfrak{t}_k$, $B_{\mathfrak{t}} = [\beta_{ij}] \in M(3,3)$ is computed through Table Table 5. Modulo $o(1)$ terms, we have

$$B_{\mathfrak{t}} = \begin{pmatrix} 0.5k/\pi + 0.5 & 0.25k & 0.25 & 0 & -0.5/\pi \\ 0.25 & 0.25k + 0.5 & 0.25 & 0 & 0 \\ 0.25 & 0.25k & 0.5 & -0.5/\pi & 0 \\ 0 & 0 & -0.5/\pi & 0.25k + 0.5 & 0.25 \\ -0.5k/\pi + 1.0 & 1.0 & 1.0 & 0.25k + 1.0 & 0.5 \end{pmatrix}$$

This allows us to easily compute the coefficients of the linear term for eigenvalues of the form $a + bk + o(1)$. Indeed, this follows by computing the spectrum of $B_{\mathfrak{t}}/k$ where $k \to \infty$,

$$\begin{pmatrix} 0.5/\pi & 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ -0.5/\pi & 0 & 0 & 0.25 & 0 \end{pmatrix}$$

which is 0, $\frac{1}{4}$ and $\frac{1}{2\pi}$ of multiplicity $2, 1, 2$, respectively.

	$\mathfrak{D}_1^{k-1,k-1}$
$H(\cdot)_{11}$	$H^{11}_{11} + (k-2)H^{12}_{12}$
$H(\cdot)_{12}$	$H^{11}_{1i} + H^{12}_{11} + (k-3)H^{12}_{1j}$
$H(\cdot)_{1k}$	$H^{11}_{1k} + (k-2)H^{12}_{1k}$
$H(\cdot)_{k1}$	$H^{1k}_{11} + (k-2)H^{1k}_{1j}$
$H(\cdot)_{kk}$	$(k-1)H^{1k}_{1k}$
	$\mathfrak{D}_2^{k-1,k-1}$
$H(\cdot)_{11}$	$(k-2)H^{11}_{1j} + (k-2)H^{12}_{11} + (k-2)(k-3)H^{12}_{1j}$
$H(\cdot)_{12}$	$H^{11}_{22} + (k-3)H^{11}_{ij} + H^{12}_{12} + 2(k-3)H^{12}_{1j} + (k-3)H^{12}_{33} + (k^2-7k+8)H^{12}_{ij}$
$H(\cdot)_{1k}$	$(k-2)H^{11}_{ik} + (k-2)H^{12}_{1k} + (k-2)(k-3)H^{12}_{ik}$
$H(\cdot)_{k1}$	$(k-2)H^{1k}_{1j} + (k-2)H^{1k}_{22} + (k-2)(k-3)H^{1k}_{ij}$
$H(\cdot)_{kk}$	$(k-1)(k-2)H^{1k}_{ik}$
	$\mathfrak{D}_3^{k-1,1}$
$H(\cdot)_{11}$	$H^{1k}_{11} + (k-2)H^{1k}_{1j}$
$H(\cdot)_{12}$	$H^{1k}_{1j} + H^{1k}_{22} + (k-3)H^{1k}_{ij}$
$H(\cdot)_{1k}$	$H^{1k}_{1k} + (k-2)H^{1k}_{ik}$
$H(\cdot)_{k1}$	$H^{kk}_{11} + (k-2)H^{kk}_{ij}$
$H(\cdot)_{kk}$	$(k-1)H^{kk}_{ik}$
	$\mathfrak{D}_3^{1,k-1}$
$H(\cdot)_{11}$	$H^{11}_{1k} + (k-2)H^{12}_{1k}$
$H(\cdot)_{12}$	$H^{11}_{ik} + H^{12}_{1k} + (k-3)H^{12}_{ik}$
$H(\cdot)_{1k}$	$H^{11}_{kk} + (k-2)H^{12}_{kk}$
$H(\cdot)_{k1}$	$H^{1k}_{1k} + (k-2)H^{1k}_{ik}$
$H(\cdot)_{kk}$	$(k-1)H^{1k}_{kk}$
	$\mathfrak{D}_1^{1,1}$
$H(\cdot)_{11}$	H^{1k}_{1k}
$H(\cdot)_{12}$	H^{1k}_{ik}
$H(\cdot)_{1k}$	H^{1k}_{kk}
$H(\cdot)_{k1}$	H^{kk}_{ik}
$H(\cdot)_{kk}$	H^{kk}_{kk}

Table 5: Product of the Hessian matrix by the representative vectors of the trivial representation, as described in Section B.4

D.2 The spectrum of type A and type I minima

The computation of the Hessian spectrum at types A and I uses the estimates derived in [24], which we provide here for convenience. Modulo high-order terms, we have (notations as in Lemma 5)

$$\text{Type A:}\quad \xi_1,\xi_5 \sim -1 + 2k^{-1} + \left(\frac{8}{\pi} - 4\right)k^{-2},\quad \xi_2,\xi_3,\xi_4 \sim 2k^{-1} + \left(\frac{4}{\pi} - 2\right)k^{-2}.$$

$$\text{Type I:}\quad \xi_1 = -1 + \sum_{n=2}^{\infty} c_n k^{-\frac{n}{2}},\quad \xi_2 = \sum_{n=2}^{\infty} e_n k^{-\frac{n}{2}},\quad \xi_5 = 1 + \sum_{n=2}^{\infty} d_n k^{-\frac{n}{2}},$$

$$\xi_3 = \sum_{n=2}^{\infty} f_n k^{-\frac{n}{2}},\quad \xi_4 = \sum_{n=4}^{\infty} g_n k^{-\frac{n}{2}},$$

where

$c_2 =$	2	$d_2 =$	$\frac{8(\pi-1)}{\pi^2}$	$e_2 =$	2	$f_2 =$	0	$g_2 =$	$2 - \frac{4}{\pi}$
$c_3 =$	0	$d_3 =$	-4.798751	$e_3 =$	0	$f_3 =$	0	$g_3 =$	$\frac{32}{\pi^2}\left(\frac{1}{\pi} - 1\right)$
$c_4 =$	$\frac{16}{\pi} - 4$			$e_4 =$	$\frac{8}{\pi} - 2$	$f_4 =$	$\frac{16}{\pi^2} - \frac{12}{\pi}$		
$c_5 =$	4.441691			$e_5 =$	$\frac{8(\pi^2+4(\pi-1))}{\pi^3}$	$f_5 =$	6.205827		

The rest of the derivation follows along the same lines of type II minima.

Let us show how to compute the 3 distinct eigenvalues of type A which are related to the standard representation. Here, the matrix associated with the $3\mathfrak{s}$ factor is

$$M = \begin{pmatrix} -0.5/\pi + 0.5 & 2.0/\pi - 0.25k & -0.5 + 0.25k \\ -0.125 & 0.5/\pi + 0.25 + 0.125k & 0.25 - 0.125k \\ 0.125 & -0.125k & -0.5/\pi + 0.125k \end{pmatrix}.$$

We now express the 3 eigenvalues by $a_ik + b_i + o(1)$. The coefficients of the linear terms can be computed by taking $k \to \infty$ in M/k, which gives

$$\begin{pmatrix} 0 & -0.25 & 0.25 \\ 0 & 0.125 & -0.125 \\ 0 & -0.125 & 0.125 \end{pmatrix},$$

whose eigenvalues are easily shown to be $0, 0, 0.25$. It remains to compute the constant terms b_i. To this end, note that

$$\begin{aligned} b_1 + b_2 + b_3 &= \text{the constant term of trace}(M) = -0.5/\pi + 0.75, \\ 2a_1b_1 &= \text{the coefficient of } k \text{ in trace}(M^2) = 0.125, \\ a_1b_1b_2 &= \text{the coefficient of } k \text{ in } \det(M) = 0.03125/\pi + 0.015625. \end{aligned}$$

The system of equations yields $b_1 = \frac{1}{4}, b_2 = \frac{1}{4}, b_3 = \frac{1}{4} - \frac{1}{2\pi}$.

E Completion of the proof of Theorem 2

E.1 Extension to the case $d > k$

Given $d > k$, append $d - k$ zeros to the end of each row of $\boldsymbol{W} \in M(k,k)$ to define $\widetilde{\boldsymbol{W}} \in M(k,d)$. Similarly, extend the target $\mathbf{V}$ to $\widetilde{\mathbf{V}} \in M(k,d)$. Denote the associated objective function by $\widetilde{\mathcal{F}}$ and note that if $\boldsymbol{W} \in M(k,k)$ is a critical point of $\mathcal{F}$, then $\widetilde{\boldsymbol{W}} \in M(k,d)$ is a critical point of $\widetilde{\mathcal{F}}$.

We make use of the following result, adapted from Lemma 8 in [22]

Lemma 29. *(Notation and assumptions as above.) Assume $d > k$ and set $m = d - k$. Let $\boldsymbol{W}$ be a critical point of $\mathcal{F}$ which has no parallel rows. Then the Hessian $\widetilde{H}$ of $\widetilde{\mathcal{F}}$ at $\widetilde{\boldsymbol{W}}$ may, after a permutation of rows and columns, be written in block diagonal form $[\widetilde{H_{ii}}]_{i\in[m+1]}$ where $\widetilde{H_{11}} = H \in M(k^2,k^2)$ is the Hessian of $\mathcal{F}$, and for $i > 1$, the matrices $\widetilde{H_{ii}}$ are all equal to the $k \times k$-matrix $M = [m_{ij}]$ defined by*

$$m_{ij} = \begin{cases} \frac{1}{2} + \frac{1}{2\pi}\sum_{\ell\in[k]}\left(\frac{\sin(\theta_{\boldsymbol{w}_i,\boldsymbol{w}_\ell})\|\boldsymbol{w}_\ell\|}{\|\boldsymbol{w}_i\|} - \frac{\sin(\theta_{\boldsymbol{w}_i,\mathbf{v}_\ell})}{\|\boldsymbol{w}_i\|}\right), & i = j \\ \frac{1}{2\pi}(\pi - \theta_{\boldsymbol{w}_i,\boldsymbol{w}_j}), \quad i \neq j \end{cases}$$

Theorem 30. *(Assumptions and notation of Theorem 2) Let $d > k$.*

1. *Suppose $\widetilde{\boldsymbol{W}} = \widetilde{\mathbf{V}}$. In addition to the eigenvalues described in Theorem 2, there will be an 2 additional eigenvalues: one equal to $\frac{1}{4}$, multiplicity $m(k-1)$, the other to $\frac{k+2}{4}$, multiplicity m.*
2. *Suppose $\boldsymbol{W}$ is of type A. Then $\widetilde{\boldsymbol{W}}$ will have an additional 2 eigenvalues. One equal to $\frac{1}{4} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$, multiplicity $m(k-1)$, the other to $\frac{k+1}{4} - \frac{1}{\pi\sqrt{k}} + O(k^{-1})$, multiplicity m.*
3. *Suppose $\boldsymbol{W}$ is of type II. Then $\widetilde{\boldsymbol{W}}$ will have an additional 3 eigenvalues. One equal to $\frac{1}{4} + \frac{4}{\pi^2 k} + O(k^{-2})$ of multiplicity $m(k-2)$, and two eigenvalues of multiplicity m, one equal to $\frac{k+1}{4} + O(k^{-\frac{1}{2}})$, the other to $\frac{1}{2} + O(k^{-\frac{1}{2}})$.*

In particular, type A and type II spurious minima exist for all $d \geq k \geq 6$.

Proof Suppose $\widetilde{\boldsymbol{W}} = \widetilde{\mathbf{V}}$. The matrix $M \in M(k,k)$ defines an S_k-map of $\mathbb{R}^k$. Computing M, we find that $m_{ii} = \frac{1}{2}$ and $m_{ij} = \frac{1}{4}$, $i,j \in [k]$, $i \neq j$. Write $(\mathbb{R}^k, S_k)$ uniquely as the orthogonal direct sum $(H_{k-1}, S_k) \oplus (T, S_k)$ Since M is an S_k-map, $M : H_{k-1} \to H_{k-1}$ and $M : T \to T$. Taking $X = [1, -1, 0, \cdots, 0] \in H_{k-1}$, $M(X) = (m_{11} - m_{12})X$, giving the eigenvalue $\frac{1}{4}$. Similarly, for the eigenvalue associated to (T_k, S_k) is $\frac{k+2}{4}$. The argument for Type A critical points is similar: both the diagonal and off-diagonal entries are easily computed given the estimates on the critical points used in the proof of Theorem 2. Finally, for type II critical points, we use the S_{k-1}-representation $(\mathbb{R}^k, S_{k-1})$ which has isotypic decomposition $\mathfrak{s}_{k-1} + 2\mathfrak{t}$. The eigenvalue associated to the $\mathfrak{s}_{k-1}$ factor is found exactly as for type A critical points and only uses the m_{11} and m_{12} entries of M. For the eigenvalues associated to the factor $2\mathfrak{t}$, we use the realizations spanned by the basis vector $\mathbf{v}_k$ and the vector $\sum_{i \in [k-1]} \mathbf{v}_i$. However, $\sin(\theta_{\boldsymbol{w}_k, \mathbf{v}_k})$ appears in the expression for m_{kk} and this leads to the presence of terms in $k^{-\frac{1}{2}}$ since $\sin(\theta_{\boldsymbol{w}_k, \mathbf{v}_k}) = \frac{4}{\pi\sqrt{k}}$ [24]. $\square$

F Empirical results

F.1 Perturbing the trained model

Our analysis shows that local minima exhibit a small number of distinct eigenvalues, independent of the number inputs d and hidden neurons k. However, during the training processes we expect to see a small number clusters of eigenvalues forming upon convergence. Below, we perturb the type II local minima of $k = 20$ by adding an independent zero-mean Gaussian noise per entry for different choices of variance.

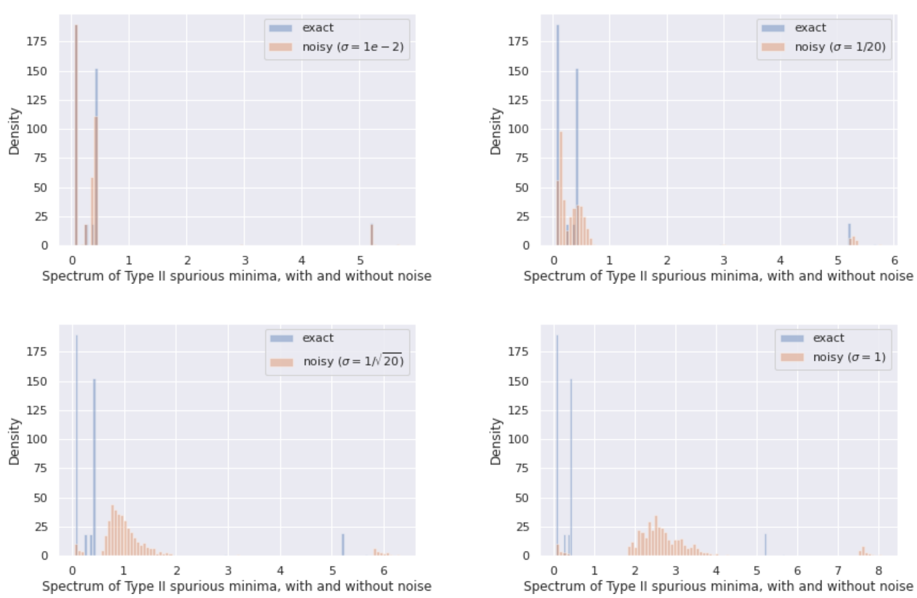

Figure 3: The spectrum of the Hessian at type II spurious minima where the entries are perturb by adding independent zero-mean Gaussian random entries with different variance values. As expected, the eigenvalues accumulate in clusters around the eigenvalues of the type II minima.

F.2 Eigenvalue data for type A, I, II spurious minima

In the sequel, we provide numerical estimates for the Hessian spectrum at types A, I and II minima. The Hessian is computed using the expressions given in Section C, and evaluated using the estimates of the spurious minima. The spectrum is then approximated numerically using LinAlg, a linear algebra package of Python.

k	4	5	6	7	8	9	10	11	12	13	14	15	20	50	100
xi_1	-0.4898600208	-0.5940757868	-0.6633973683	-0.7126281816	-0.7493249249	-0.7777074711	-0.8003040345	-0.8187167157	-0.8340080421	-0.8469096664	-0.857941499	-0.8674829742	-0.9007460136	-0.9602572702	-0.9800854958
xi_2	0.4954785342	0.3967285644	0.3307097833	0.2834951026	0.2480699062	0.220516412	0.1984765225	0.1804470696	0.1654251182	0.1527161193	0.1418239656	0.132384904	0.09935065305	0.03983992056	0.01995135856
Trivial Rep (times 1)	1.4560394	0.70990586	0.8480067	0.990511	1.135952	1.2834839	1.4325787	1.5828876	1.7341673	1.8862413	2.0389771	2.1922746	2.9648666	7.6802244	15.608578
	0.57904404	1.7242408	1.9897203	2.2535357	2.5160947	2.7776048	3.0382068	3.298006	3.5570922	3.8155413	4.0734167	4.3307796	5.611412	13.199169	25.748272
Standard Rep (times k-1)	1.109484036	1.375470284	1.635817171	1.893362863	2.149283757	2.404151115	2.658279921	2.911859526	3.165013547	3.417828734	3.670365601	3.922670617	5.181727777	12.70512831	25.2178866
	0.2203240728	0.2049980342	0.1946480554	0.1875084205	0.1824611761	0.1788291887	0.1761892425	0.1742664477	0.1728748546	0.1718852454	0.1712038256	0.1707629932	0.1707309861	0.08774232198	0.08977066797
	0.04751754505	0.0514193864	0.05510416677	0.05842876435	0.0613882308	0.06401088927	0.06633332964	0.06839218345	0.07021950498	0.07184625776	0.07329648491	0.07459396964	0.07936788816	0.1834150619	0.1974143104
X Rep (times (k-1)(k-2)/2)	-0.01753896847	-0.005831507035	0.001503654756	0.006847392768	0.01105740014	0.01453184709	0.01748837531	0.02005898207	0.0223294831	0.02436031401	0.02619439736	0.02786448598	0.0344661884	0.05203547329	0.06214639544
Y Rep (times k(k-3)/2)	0.298961103	0.3122170269	0.3202357292	0.3258672953	0.3301824629	0.3336762488	0.3366119266	0.3391433954	0.3413672745	0.3433507383	0.3451385498	0.3467643261	0.3531934023	0.3704492748	0.3804876208

Figure 4: The spectrum of type A spurious minima.

k	5	6	7	8	9	10	11	12	13	14	15	20	50	100
xi_1	-0.4830355462	-0.5877302017	-0.6579648893	-0.70808019	-0.7455234023	-0.7745103399	-0.7975913049	-0.8163930175	-0.8319990065	-0.8451574585	-0.8564009402	-0.8946139295	-0.959294303	-0.9798436861
xi_2	0.4884077711	0.3911538654	0.3262295256	0.2798049673	0.2449666066	0.2178625868	0.1961761796	0.178430816	0.1636412895	0.1511252882	0.1403954144	0.1036641823	0.04046198792	0.02009879575
xi_3	0.01440735973	0.01677031362	0.01818315076	0.01896456522	0.01932369293	0.0194022927	0.01929576152	0.01906863983	0.01876427588	0.0184120735	0.01803202798	0.01606495388	0.009090335323	0.005277459527
xi_4	-0.01527369217	-0.01379889472	-0.01202812067	-0.01040931892	-0.009029099865	-0.007876361661	-0.006917063442	-0.006116172395	-0.00544350044	-0.004874521939	-0.00438973409	-0.002787831404	-0.0005807633564	-0.0001645277529
xi_5	1.063150521	1.068395642	1.069424581	1.06852328	1.066734455	1.06456357	1.062264857	1.059968247	1.057739533	1.055610406	1.053594085	1.045172924	1.023263207	1.013111353
Trivial Rep (times 1)	1.74485985	2.00748071	2.268785083	2.52929655	2.789167588	3.048453427	3.307187958	3.565399005	3.823122832	4.080388805	4.33723197	5.616135615	13.20110106	25.74931705
	1.396092281	1.648778854	1.901337223	2.153887011	2.406412581	2.658877863	2.911255781	3.163529273	3.415692853	3.667744066	3.919685917	5.177942127	12.70217173	25.21607128
	0.7221266194	0.8566065873	0.9968903313	1.140882465	1.287421413	1.43580951	1.585596641	1.736480223	1.888246349	2.040738104	2.193838276	2.965833076	7.680449743	15.60864623
	0.2547984971	0.2436021821	0.2360487686	0.230795479	0.2270566597	0.2243502517	0.2223668139	0.2209034859	0.2198216631	0.2190235799	0.218442759	0.2174068768	0.2223895252	0.2285558823
	0.07739439357	0.07768427594	0.07813568317	0.07863686523	0.07914477647	0.07964019484	0.08011357874	0.080561551	0.08098252367	0.08137570871	0.08174390423	0.08325409544	0.08709450714	0.08880356756
Standard Rep (times k-2)	1.385696445	1.643764231	1.899867182	2.154735278	2.408795079	2.66228793	2.915358842	3.168098934	3.420573561	3.672825958	3.924890952	5.183170997	12.70550562	25.21805575
	0.345675456	0.3513956535	0.3544633906	0.3565888066	0.358286142	0.3597465232	0.3610517106	0.3622431587	0.3633437286	0.3643675924	0.3653260152	0.3693468508	0.3814014159	0.3887547905
	0.2140349063	0.201678982	0.1931210066	0.187036907	0.1826232661	0.17937935	0.1769781035	0.1752005927	0.1738939428	0.1729486752	0.17228538	0.1715226146	0.1832385386	0.1971764903
	0.06294336939	0.06429450196	0.0659687698	0.06766043518	0.06927782766	0.0707903436	0.0721901269	0.07347733273	0.07465752584	0.07573748682	0.07672530439	0.08053494072	0.08788777421	0.08981091997
	0.0257839133	0.0278933091	0.03042886615	0.03293615983	0.03527622583	0.03741438549	0.03935304206	0.04110846721	0.04270044768	0.04414838909	0.04547086142	0.05067136473	0.06384494842	0.07103627267
X Rep (times (k-2)(k-3)/2)	-0.01777676493	-0.005825300235	0.001661930699	0.007101179101	0.01137061138	0.0148812597	0.01785925776	0.02043984085	0.02271434665	0.02474402636	0.02657379396	0.03366916254	0.05186842754	0.06213029474
Y Rep (times (k-1)(k-4)/2)	0.2987003326	0.3126812577	0.3211107254	0.3269506097	0.3313557804	0.3348731101	0.3377957046	0.3402929604	0.3424740732	0.3444090188	0.3461468816	0.3528721631	0.3703959286	0.3805052638

Figure 5: The spectrum type I spurious minima.

k	5	6	7	8	9	10	11	12	13	14	15	20	50	100
xi_1	0.9616068174	0.9867036362	0.9969227946	1.001434253	1.003474556	1.004354823	1.004662096	1.004678965	1.004549164	1.004348139	1.00411668	1.003003746	1.00072305	1.000207713
xi_2	-0.07908255776	-0.05041345074	-0.03508015694	-0.02588362808	-0.01991470917	-0.01581192581	-0.01286627299	-0.01067773591	-0.009006055193	-0.00769970374	-0.006659067921	-0.003655140678	-0.0005580411545	-0.0001363604554
xi_3	0.287549952	0.2245161452	0.1859910548	0.1598160433	0.1407087506	0.1260423609	0.1143656351	0.1048096181	0.09681901944	0.09002247393	0.08416022593	0.06371471024	0.02618552295	0.01318159633
xi_4	0.3760650557	0.3080111148	0.2616910791	0.2280890166	0.202525064	0.1823677558	0.1660288578	0.1524928777	0.1410788904	0.1313131566	0.1228549211	0.0931874083	0.03856576237	0.01960293703
xi_5	-0.5597784395	-0.601512325	-0.6327039183	-0.6582562628	-0.6800502369	-0.6990368175	-0.7157983022	-0.7307373008	-0.7441491733	-0.7562658311	-0.7672705594	-0.8100205756	-0.9092014274	-0.9512631475
Trivial Rep (times 1)	1.74888179	2.02373391	2.295489804	2.564593456	2.83143737	3.096356053	3.359655841	3.621578783	3.882325475	4.142062393	4.400926516	5.685553785	13.26772252	25.80391663
	1.352933708	1.617871289	1.880596512	2.141277694	2.400283634	2.657945764	2.914530035	3.170242179	3.425240691	3.6796493	3.933565354	5.198019132	12.7273582	25.23830254
	0.7281653401	0.8694840157	1.0131957	1.15874289	1.305788955	1.454087619	1.603448578	1.753719625	1.904775009	2.056513357	2.208850154	2.977355079	7.683041551	15.6084993
	0.2234922735	0.2211724347	0.2194946734	0.2181558785	0.2170622739	0.2161654018	0.2154323696	0.2148365746	0.2143550761	0.2139703923	0.2136667681	0.2129914195	0.2162712472	0.2217633259
	0.05484339971	0.06244204294	0.06834925311	0.07286984944	0.07636060821	0.07909314639	0.08126332124	0.08300942882	0.08442931397	0.08559776211	0.08656692283	0.08955913818	0.0924073933	0.09215372556
Standard Rep (times k-2)	1.408298036	1.671978277	1.931566257	2.188882771	2.444731209	2.6995429	2.953583796	3.207030307	3.460005327	3.712601009	3.964884557	5.223144208	12.73899878	25.24451896
	0.3273211384	0.3388550294	0.34772121	0.3544345581	0.3596246645	0.3637405037	0.3670837346	0.369855904	0.372193668	0.3741961337	0.375932867	0.3820679692	0.3940261326	0.3989494024
	0.2391507758	0.2383455741	0.2381737662	0.238373771	0.2387562001	0.2392149557	0.2396955197	0.2401700884	0.2406244382	0.2410522281	0.2414520331	0.2430664059	0.2468108805	0.2483286417
	0.0723950212	0.07729074833	0.08040751665	0.08253332498	0.08405573712	0.0851860203	0.08605060341	0.08672803056	0.08726831449	0.08770730756	0.08806937875	0.08918828705	0.09054433281	0.0907623885
	-0.007701481192	0.004700453437	0.01405908725	0.02134645822	0.02718223121	0.03196399919	0.03595817112	0.03934816015	0.04226265346	0.04479866355	0.047026995	0.05509339701	0.07140544709	0.07816495097
X Rep (times (k-2)(k-3)/2)	0.03828606755	0.04637753963	0.05217475817	0.05657319725	0.06004140154	0.06285340339	0.06518359482	0.06718876741	0.06882896274	0.07028251886	0.07155331969	0.07609294355	0.08469195664	0.08772839606
Y Rep (times (k-1)(k-4)/2)	0.3560564518	0.3638006747	0.3694555461	0.3738227487	0.377314508	0.3801749647	0.3825637996	0.3845902085	0.3863299191	0.387840271	0.3891634941	0.3939078748	0.4028793573	0.406003207

Figure 6: The spectrum type II spurious minima.