Divide and Conquer: A Distributed Approach to Five Point Energy Minimization

Richard Evan Schwartz

January 22, 2023

1 Introduction

The purpose of this work is to rigorously verify the phase-transition for 5 point energy minimization first observed in [**MKS**], in 1977, by T. W. Melnyk, O, Knop, and W. R. Smith. Our results contain, as special cases, solutions to Thomson's 5-electron problem [**S1**] and Polya's 5-point problem [**HS**]. This work is an updated version of my monograph from 6 years ago. I simplified the proof significantly and also I wrote this version in an experimental style designed to facilitate the verification process. This work is less than half as long as the original.

I wrote the proof in a tree-like form. Thus, the Main Theorem is an immediate consequence of Lemma A, Lemma B, and Lemma C. These three Lemmas are independent from each other. Lemma A is an immediate consequence of Lemma A1 and Lemma A2. And so on. All the "ends" of the tree, such as Lemma B313, either have short and straightforward proofs or are computer calculations which I will describe in enough detail that a competent programmer could reproduce them. At the same time, all my computer programs are available to download and use. Figure 0 below maps out the complete logical structure of the proof of the Main Theorem.

The rest of this introduction states the results and explains how to divide the verification of the proof into small pieces. Following this, §2 contains a discussion of the history and context of the results, a high-level discussion of the ideas in the proof, a discussion of the computer experiments I did, and a guide to the relevant software I wrote. Following this we get to the proof.

Results: Let S^2 be the unit sphere in $\boldsymbol{R}^3$. Given a configuration $\{p_i\} \subset S^2$ of N distinct points and a function $F : (0,2] \to \boldsymbol{R}$, define

$$\mathcal{E}_F(P) = \sum_{1 \leq i < j \leq N} F(\|p_i - p_j\|). \tag{1}$$

This quantity is commonly called the *F-potential* or the *F-energy* of P. A configuration P is a *minimizer* for F if $\mathcal{E}_F(P) \leq \mathcal{E}_F(P')$ for all other N-point configurations P'.

We are interested in the *Riesz potentials*:

$$R_s(d) = d^{-s}, \qquad s > 0. \tag{2}$$

R_s is also called a *power law potential*, and R_1 is specially called the *Coulomb potential* or the *electrostatic potential*. The question of finding the N-point minimizers for R_1 is commonly called *Thomson's problem*.

We consider the case $N = 5$. The *Triangular Bi-Pyramid* (TBP) is the 5 point configuration having one point at the north pole, one point at the south pole, and 3 points arranged in an equilateral triangle on the equator. A *Four Pyramid* (FP) is a 5-point configuration having one point at the north pole and 4 points arranged in a square equidistant from the north pole.

Define

$$15_+ = 15 + \frac{25}{512}. \tag{3}$$

Theorem 1.1 (Main) *There exists $\mathfrak{w} \in (15, 15_+)$ such that:*

1. *For $s \in (0, \mathfrak{w})$ the TBP is the unique minimizer for R_s.*
2. *For $s = \mathfrak{w}$ the TBP and some FP are the two minimizers for R_s.*
3. *For each $s \in (\mathfrak{w}, 15_+)$ some FP is the unique minimizer for R_s.*

The number $\mathfrak{w}$ is a new constant of nature. Its decimal expansion starts

$$\mathfrak{w} = 15.0480773927797...$$

This constant is *computable* in the sense that an ideal computing machine can rigorously compute as many digits of $\mathfrak{w}$ as desired in finite time. See §3.5.

In §3.5 I will also explain the main details of the following theorem.

Theorem 1.2 (Auxiliary) *Let $F_s(d) = -d^{-s}$ be the Fejes-Toth potential. The TBP is the unique minimizer for F_s for all $s \in (-2, 0)$.*

Logic Tree: Figure 0 shows all the results we prove and how they contribute to the proof of the main theorem. The color coding indicates different independent parts of the proof. There are 7 colors. Each color (so to speak) may be read independently from the others.

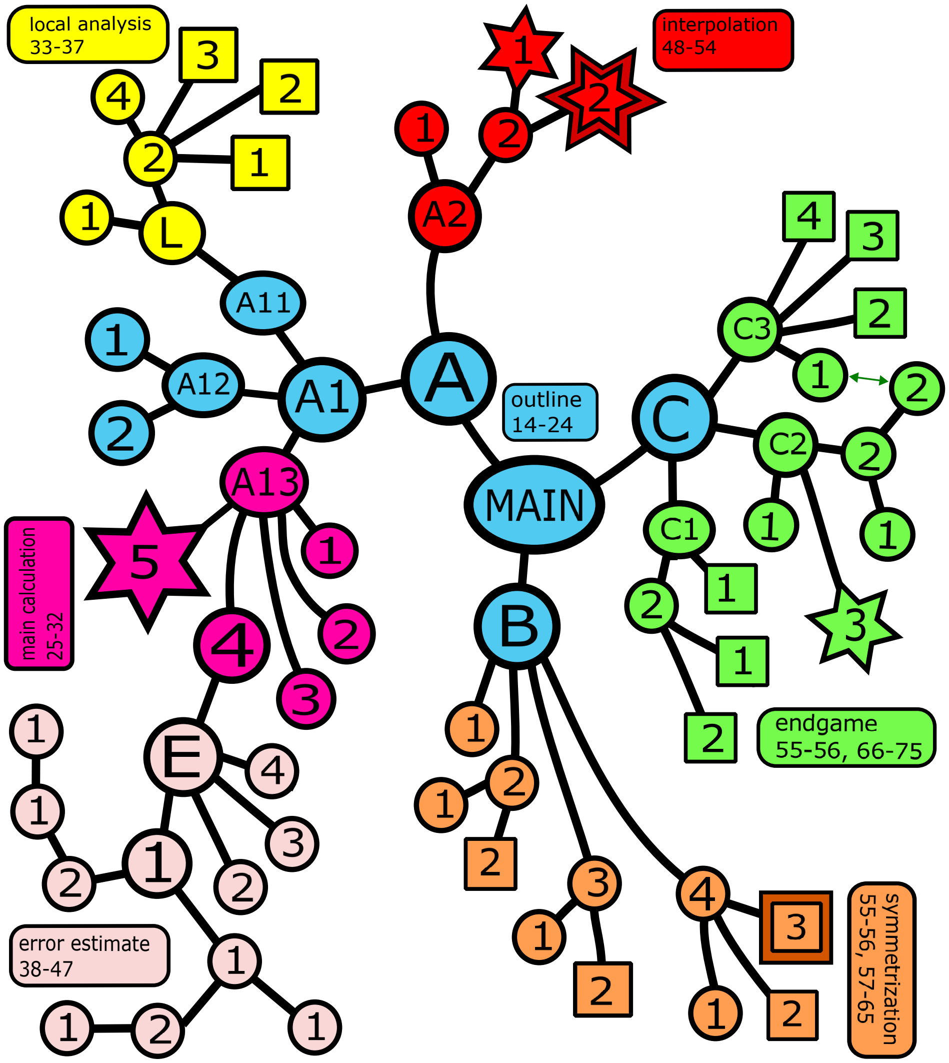

Figure 0: The tree of implications.

I have indicated the nature of each colored part and the page numbers containing that part. In the interest of space I have not given the full name of every lemma. Thus, the specially starred red lemma at the top is A222. The green arrow indicates that Lemma C31 and Lemma C222 are the same.

The starred lemmas are the divide-and-conquer computer calculations. The big one, for $A13$, is done with interval arithmetic. The others are done with exact integer arithmetic. The square nodes indicate sizeable but exact calculations done with rational polynomials in Mathematica. The 2 two-tone vertices (A222 and B43) indicate really lucky computer proofs I am especially proud of. For instance, I prove Lemma B43 by showing that a 4-variable polynomial with over 100000 terms (about half of which are negative) is positive on $(0,1)^4$. For these two lemmas my computer code also has alternate proofs (with documentation) that do not rely on pure good luck.

Verification: As Figure 0 indicates, a team of 7 readers could check the mathematical part of the proof, with the team-leader reading the blue outline and the other 6 people reading the other colors. Each color involves at most 12 pages of material. Some readers might also want to consult the discussion on pp 5-13 to get insight into where the ideas come from but this discussion is logically independent from the proof.

I wrote the computer code in such a way that the programs for each part are independent from the programs for each other part. It is probably easier in each case to reproduce the code rather than check that mine is correct. (Much of my code is wrapped inside graphical user interfaces that let the user see it working.) Each reader could team up with a strong computer programmer who could reproduce the relevant code. This would be a serious job only for Lemma A135, which a good programmer could probably recreate in few days. For the rest of the parts, the code could be recreated in a day. §2.4 discusses my software further.

Acknowledgements: I thank Henry Cohn, Javier Gomez-Serrano, Doug Hardin, John Hughes, Ian Jauslin, Alex Kontorovich, Abhinav Kumar, Curtis McMullen, Stephen D. Miller, Jill Pipher, Ed Saff, Sergei Tabachnikov, and Alexander Tumanov for discussions and/or encouragement related to this work. I thank I.C.E.R.M. for facilitating this project. My interest in this problem was rekindled during discussions about point configuration problems at the I.C.E.R.M. Science Advisory Board meeting in Nov. 2015. After writing the original version of the monograph around 2016, I let the thing languish on my website for some years. After giving the Lewis Lectures at Rutgers University in Nov, 2022, I got inspired to re-write my monograph and make it easier to understand. Finally, I thank the National Science Foundation for their continued support, currently grant DMS-2102802.

2 Discussion

This chapter discusses the history and context for the result, and the high level ideas in the proof. Following this, I explain some of my experimental methods and discuss the computer parts of the proof. None of this is logically needed for the proof, but it will shed light on why the proof looks like it does.

2.1 History and Context

We take up the question discussed in the introduction: Which configurations of points on the sphere minimize a given potential function $F : (0,2] \to \boldsymbol{R}$. The classic choice for this question is $F = R_s$, the *Riesz potential*, given by $R_s(d) = d^{-s}$. The Riesz potential is defined when $s > 0$. When $s < 0$ the corresponding function $R_s(d) = -d^{-s}$ is called the *Fejes-Toth potential*. The main difference is the minus sign out in front.

The case $s = 1$ is specially called the *Coulomb potential* or the *electrostatic potential*. This case of the energy minimization problem is known as *Thomson's problem*. See [**Th**]. The case of $s = -1$, in which one tries to maximize the sum of the distances, is known as *Polya's problem*.

There is a large literature on the energy minimization problem. See [**Fö**] and [**C**] for some early local results. See [**MKS**] for a definitive numerical study on the minimizers of the Riesz potential for n relatively small. The website [**CCD**] has a compilation of experimental results which stretches all the way up to about $n = 1000$. The paper [**SK**] gives a nice survey of results, with an emphasis on the case when n is large. See also [**RSZ**]. The paper [**BBCGKS**] gives a survey of results, both theoretical and experimental, about highly symmetric configurations in higher dimensions.

When $n = 2, 3$ the problem is fairly trivial. In [**KY**] it is shown that when $n = 4, 6, 12$, the most symmetric configurations – i.e. vertices of the relevant Platonic solids – are the unique minimizers for all R_s with $s \in (-2, \infty) - \{0\}$. See [**A**] for just the case $n = 12$ and see [**Y**] for an earler, partial result in the case $n = 4, 6$. The result in [**KY**] is contained in the much more general and powerful result [**CK**, Theorem 1.2] concerning the so-called sharp configurations.

The case $n = 5$ has been notoriously intractable. There is a general feeling that for a wide range of energy choices, and in particular for the power law potentials (when $s > -2$) the global minimizer is either the TBP or an FP. Here is a run-down on what is known so far:

- The paper [**HS**] has a rigorous computer-assisted proof that the TBP is the unique minimizer for the potential $F(r) = -r$. (Polya's problem).

- My paper [**S1**] has a rigorous computer-assisted proof that the TBP is the unique minimizer for R_1 (Thomson's problem) and R_2. Again $R_s(d) = d^{-s}$.

- The paper [**DLT**] gives a traditional proof that the TBP is the unique minimizer for the logarithmic potential.

- In [**BHS**, Theorem 7] it is shown that, as $s \to \infty$, any sequence of 5-point minimizers w.r.t. R_s must converge (up to rotations) to the FP having one point at the north pole and the other 4 points on the equator. In particular, the TBP is not a minimizer w.r.t R_s when s is sufficiently large.

- In 1977, T. W. Melnyk, O. Knop, and W. R. Smith, [**MKS**] conjectured the existence of the phase transition constant, around $s = 15.04808$, at which point the TBP ceases to be the minimizer w.r.t. R_s. This is the phase transition which our Main Theorem estabishes.

- Define
$$G_k(r) = (4 - r^2)^k, \qquad k = 1, 2, 3, ... \tag{4}$$
In [**T**], A. Tumanov proves that the TBP is the unique minimizer for G_2. The minimizers for G_1 are those configurations whose center of mass is the origin. The TBP is included amongst these.

Tumanov points out that the G_2 potential does not have an obvious geometric interpretation, but it is amenable to a traditional analysis. He also mentions that his result might be a step towards proving that the TBP minimizes a range of power law potentials. Inspired by similar material in [**CK**], he observes that if the TBP is the unique minimizer for G_2, G_3 and G_5, then the TBP is the unique minimizer for R_s provided that $s \in (0, 2]$.

We will establish implications like this during the course of our proof of the Main Theorem. The family of potentials $\{G_k\}$ behaves somewhat like the Riesz potentials. The TBP is the unique minimizer for G_3, G_4, G_5, G_6 (as a consequence of our work here) but not a minimizer for any of G_7, G_8, G_9, G_{10}. I checked up to about $k = 100$ that the TBP does not mininize G_k when $k > 10$ and I am sure this pattern persists.

2.2 Ideas in the Proof

Here are the three ingredients in the proof of the Main Theorem.

- The divide-and-conquer approach taken in [**S1**].
- Elaboration of Tumanov's observation.
- A symmetrization trick that works on a small domain.

Divide and Conquer: For certain choices of F, we are interested in searching through the moduli space of all 5-point configurations and eliminating those which have higher F-potential than the TBP. We win if we eliminate everything but the TBP. For the functions we consider, most of the configurations have much higher energy than the TBP and we can eliminate most of the configuration space just by crude calculations. What is left is just a small neighborhood Ω_0 of the TBP. The TBP is a critical point for $\mathcal{E}_F$, and (it turns out) that the function $\mathcal{E}_F$ is convex in Ω_0. In this case, we can say that the TBP must be the unique global minimizer.

To implement this, we normalize so that $(0,0,1)$ is a point of the configuration, and then we map the other 4 points into $\boldsymbol{R}^2$ using stereographic projection:

$$\Sigma(x,y,z) = \Big(\frac{x}{1-z}, \frac{y}{1-z}\Big). \tag{5}$$

We call the 4-point planar configuration the *avatar*. We use crude *a priori* estimates to produce a subset Ω of a 7-dimensional rectangular solid that (up to symmetry) contains all avatars that could have lower potential than the TBP for all the relevant functions. Inside Ω the divide-and-conquer algorithm is easy to manage. Our basic object is a *block*, a rectangular solid subset of Ω. A main feature of our proof is a result which gives a lower bound on the energy of any configuration in a block based on the energies of the configurations corresponding to the vertices, and an error term.

Having an efficient error term makes the difference between a feasible calculation and one which would outlast the universe. Our error term is fairly sharp, and also the error term is a rational function of the vertices of the block. For the potentials we end up using, we could run all our computer programs using exact integer arithmetic. Such integer calculations are too slow (in this century). I implemented the big calculations using interval arithmetic. Since everything in sight is rational, our calculations only involve the operations plus, minus, times, divide, min, and max.

Elaborations of Tumanov's Observation: So far we have discussed one function at a time, but we are interested in a 1-parameter family of power laws and we can only run our program finitely many times. Using the divide and conquer approach we show that the TBP is the unique global minimizer for G_k when $k = 3, 4, 5, 6$ and also for the wierd *energy hybrids* $G_5 - 25G_1$ and $G_{10}^{\sharp\sharp} = G_{10} + 28G_5 + 102G_2$. Converting these results to statements about the power law potentials comes down to variants of Tumanov's observation. After a lot of experimenting I found variants which cover large ranges of exponents. The results for the potentials above combine to prove that the TBP is the unique minimizer for R_s as long as $s \in (-2, 0) \cup (0, 13]$.

Symmetrization: The methods above cannot be sharp enough to arrive at the exact statement of our Main Theorem, because of the phase transition. We get around this problem as follows. First, we use the divide and conquer approach to identify a small subset Υ of the configuration space such that every configuration not in Υ, and not the TBP, has higher $G_{10}^{\sharp}$-energy, where $G_{10}^{\sharp} = G_{10} + 13G_5 + 68G_2$. This combines with the previous calculations to show that every configuration not in Υ, and not the TBP, has higher s potential than the TBP whenever $s \in [13, 15_+]$. The configurations in Υ very nearly have 4-fold symmetry.

To analyze configurations in Υ we use a symmetrization operation which maps Υ to the subset $\Upsilon_4 \subset \Upsilon$ consisting of configurations having 4-fold symmetry. This retraction turns out to reduce the s-potential for exponent values $s \in [13, 15_+]$. Finally (and slightly simplifying) we produce a retraction from Υ_4 to a subset $\Upsilon_8 \subset \Upsilon_4$ consisting entirely of FPs. This new retraction reduces the s-potential when $s \in [15, 15_+]$. Now we are left with an analysis of the s-potential on a 1-dimensional set.

Extending the Range: My techniques run out just past the value ש. The obvious conjecture, already observed by Melnyk, Knop, and Smith, is some FP is the minimizer for any Riesz potential with exponent larger than ש. The original version of my monograph contained a proof that some FP beat the TBP for all exponents up to 100. I omitted this material because it didn't seem like such a strong result and mostly it was just a tedious calculation. I would say that the main bottleneck to proving that some FP is the minimizer for R_s for all $s >$ ש is the delicacy of the symmetrization process which I discuss above and also below.

2.3 Experimentation

The proof of the Main Theorem is mostly just a verification of the things I discovered experimentally using the software I created. I will follow 3 main lines of experimental investigation in this discussion.

Experiments with Interpolation: This discussion has to do with what I called “Tumanov’s observation” in the preceding section. These kinds of methods go under the name of *interpolation*.

For the purpose of giving results about the Riesz potentials, the functions G_k lose their usefulness at $k = 7$ because the TBP is not a minimizer for $G_7, G_8, ...$ At the same time, the general method requires G_k for k large in order to *extend* all the way to the phase transition, a phenomenon that occurs at $ש = 15.04...$

I built a graphical user interface which allows me to explore combinations of the form $\sum c_k G_k$ and see whether various lists of these *energy hybrids* produce the desired results. The computer program takes a quadruple of hybrids, $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$, and then solves a linear algebra problem to find a linear combination

$$\Lambda_s = a_0 + \sum_{i=1}^{4} a_i(s)\Gamma_i \tag{6}$$

which matches the values of R_s at the values $\sqrt{2}, \sqrt{3}, \sqrt{4}$, the distances involved in the TBP. (I will usually write 2 as $\sqrt{4}$ because then the distances involved in the TBP are easier to remember.)

Concerning Equation 6, what we need for the quadruple to “work” on the interval (s_0, s_1) is that the functions $a_1(s), a_2(s), a_3(s), a_4(s)$ are nonnegative for $s \in (s_0, s_1)$ and that simultaneously the *comparison function* $1 - (\Lambda_s / R_s)$ is positive on $(0, 2) - \{\sqrt{2}, \sqrt{3}, \sqrt{4}\}$. So, my computer program lets you manipulate the coefficients defining the energy hybrids and then see plots of the functions just mentioned.

At the same time as this, my program computes the energy hybrid evaluated on the space of FPs to see how it compares to the value on the TBP. I call this the *TBP/FP competition*. On intervals $(s_0, s_1) \subset (0, ש)$ we want the TBP to win the competition, as judged by the given energy hybrids. Repeatedly running these competitions and looking at the plots of the coefficients and the comparison function, I eventually arrived at the energy hybrids mentioned in the previous section.

You can use this program too. If you actually get my Java program to run on your computer, you can get the same intuition I eventually got about what works and what doesn't. If you don't play around with the software, then choices like

$$G_5^{\flat} = G_5 - 25G_1, \qquad G_{10}^{\sharp\sharp} = G_{10} + 28G_5 + 102G_2$$

will just seem like random lucky guesses. In fact they are practically the unique (at most 3 term) energy hybrids which do the job!

To extend all the way to ש, I had to accept an energy hybrid for which the TBP would lose the TBP/FP competition. At the same time, the TBP would still do well in the overall competition, beating most of the other configurations. Eventually I hit upon the energy hybrid $G_{10}^{\sharp}$ and the small neighborhood Υ mentioned above and defined precisely in the next chapter. The quadruple $(G_1, G_2, G_5^{\flat}, G_{10}^{\sharp})$ extends a bit past ש, up to 15_+, and $G_{10}^{\sharp}$ is a pretty kind judge: With respect to this judge, the TBP wins against all configurations outside the tiny Υ.

The intuition I came away with is that you need to use some G_k for fairly large k, to get enough *extension*, and then you need to tune it by *sharpening* and *flattening*. To sharpen means to add in more of the lower G_ks. To flatten means to do the opposite. When you sharpen, you get an energy hybrid which is a kinder but less extensive judge: It works better but on a smaller range of exponents. When you flatten, you get a harsher but more extensive judge. The final quadruple $(G_1, G_2, G_5^{\flat}, G_{10}^{\sharp})$ extends to the neighborhood $[13, 15_+]$. The TBP is the minimizer for the first 3 potentials, and for $G_{10}^{\sharp}$ the TBP wins outside of Υ.

Experiments with Symmetrization: Most successful energy minimization results are about symmetry. The work culminating in that of Cohn-Kumar [**CK**] shows how to exploit the extreme symmetry of some special configurations, like the Leech cell, to show that they are the energy minimizers with respect to a wide range of potentials. These methods only work for very special numbers of points. The number $N = 5$ is not special in this way, because there are no Platonic solids with 5 points.

For $N = 5$ the TBP and the FPs are competitors for the most symmetric configurations. They have different symmetries. However, they do have one thing in common: 4 fold dihedral symmetry. One dream for proving the Main Theorem is to use a kind of symmetrization operation which replaces an arbitrary configuration with one having 4-fold dihedral symmetry and lower potential energy. This would reduce the overall problem to an exploration of a 2-dimensional moduli space and would possibly bring the result within the range of rather ordinary calculus.

Such a symmetrization operation *in general* will surely fail due to the vast range of possible configurations. However, certain operations might work well in very specific parts of the configuration space and for very specific functions. Fortunately, the divide-and-conquer-plus-interpolation method rules out everything of interest except the magical domain Υ and the exponent range $[13, 15_+]$. What I did is test various symmetrizations and various choices of Υ until I found a pair that worked.

Once I found a symmetrization operation which worked, the question became: How to prove it? Proving that symmetrization lowers the energy seems to involve studying what happens on the tiny but still 7-dimensional moduli space Υ. The secret to the proof is that, within Υ, the symmetrization operation is so good that it reduces the energy in pieces. What I mean is that the 10 term sum for the energy can be written as

$$e_1 + ... + e_{10} = (e_1 + e_2) + (e_3 + e_4) + (e_5 + e_6 + e_7) + (e_8 + e_9 + e_{10})$$

so that the symmetrization operation decreases each bracketed sum separately. This replaces one big verification by a bunch of smaller ones, conducted over lower dimensional configuration spaces.

I use a second symmetrization which improves a configuration with 4-fold symmetry to one with 8-fold symmetry. This symmetrization, though rather simple, is extremely delicate. It works on a tiny domain $\widehat{\Psi}_4 \subset \Upsilon$ and only for power laws with potential greater than about 13.53. At the same time, this symmetrization has great algebraic properties when restricted to the tiny domain where I use it. I found this operation, once again, by experimentation, and then the algebraic properties took me by surprise.

Experiments with Local Analysis: Another part of the proof deals with configurations that are very near the TBP. Here we are fortunate because the functions $\mathcal{E}_F$ are convex near the TBP. That means that the TBP is the unique minimizer in a small neighborhood around the TBP. I want to emphasize that what we need for convexity is not just a calculation *at* the TBP. In order to use this information effectively in a computational proof, we need an explicit neighborhood of convexity. Proving this sets up a recursive problem.

Consider the simpler situation where we would like to show that some function f is positive on some interval $I = [0, \epsilon]$. Let's say that we have free access to the values $f(0), f'(0), f''(0), ...$ and we can also look at the explicit expressions for f and its derivatives. If we had some information about $\max_I |f'|$ we could combine it with information about $f(0)$ to perhaps complete the job. But how do we get information about $\max_I |f'|$? Well, if we had information about $\max_I |f''|$ we could combine it with information about $f'(0)$ to perhaps complete the job. And so on.

This is the situation we find ourselves in. We can compute all the partial derivatives of $\mathcal{E}_F$ at the TBP, though we have a function of 7 variables, and so eventually it gets expensive to compute them *all*. However, no matter how many derivatives we compute, it seems that we need to compute more of them to get the bounds we need.

There is something that saves us: The error multiplier in Taylor's Theorem with Remainder. This multiplier is essentially $\epsilon^N/N!$, a number that becomes tiny as N increases. If we can get any kind of reasonable bounds on high derivatives of our function, then we get pretty good bounds when we multiply through by the tiny number. I eventually found a combinatorial trick for getting reasonable bounds on high dimensional derivative. The magic formula is Equation 46.

2.4 Guide to the Software

The software for my proof can be dowloaded from

http//www.math.brown.edu/ ∼ res/Java/TBP.tar

Once you untar this program, you get a directory with a suite of smaller programs. There are 5 subdirectories, corresponding to 5 of the 7 parts of the monograph. The outline and the part having to do with the error estimate do not rely on any computer assists. The reader who is interested in verifying any part of the monograph need only look at the programs for that part.

Main: This does the interval arithmetic calculation for the main divide-and-conquer result, Lemma A135. This program is quite extensive, and spread out in about 20 Java files, but almost all the length comes from the visual/experimental part. I show the calculations in action, allow the user to experiment with fairly arbitrary energy hybrids, and also give detailed written instructions on the operation of the program.

The reason for the extensive program is debugging. The big infrastructure is designed to prevent errors in the actual computation. I have also included a stripped down version that runs without all the bells and whistles. I try to explain the main computation in §5 in enough detail that a reasonably good programmer would be able to reproduce it.

Interpolation: The main program in this section, contained in the directory `JavaMain`, does all the experimentation with the energy hybrids discussed above, and also formally proves that the given energy hybrids extend to their advertised ranges. However, the code I actually use in this version of the proof is different. The subdirectory `Proof` has this shorter method. Why both? I only discovered the method in `Proof` recently, and the method in `JavaMain` is more robust. It doesn't require the kind of *ad hoc* argument I give in §9.5 and also (a very small part of) it is still needed logically for the Auxiliary Theorem. I include a PDF file which explains the method in `JavaMain`. Independent from all this, I also include Mathematics files such as `LemmaA221.m` which generate all the plots for the corresponding lemmas. One can compare the Mathematica and Java plots and see that they are the same.

Local Analysis: This directory has 3 Mathematica files, `LemmaL21.m` and `LemmaL22.m` and `Lemma23.m`, which perform the straightforward and exact calculations needed for these lemmas. The calculations involve manipulating rational polynomials and evaluating them at special points.

Symmetrization: This directory has 4 Mathematica files, with names like `Lemma B22.m` which do the calculations for the corresponding lemmas in this part of the monograph. These short files essentially just manipulate rational polynomials using standard operations in Mathematica. I also include things in other subdirectories, such as a Java program that lets the user experiment with the symmetrization operation.

Endgame: This directly contains a program which is a baby version of the main divide-and-conquer program. This program does the calculation for Lemma C2. I explain it in detail in §13. The calculation is done with exact integer arithmetic over a 3-dimensional space. We also include a number of Mathematica files such as `LemmaC3.m`.

3 Main Theorem: Proof Outline

3.1 Preliminaries

Stereographic Projection: Let $S^2 \subset \boldsymbol{R}^3$ be the unit 2-sphere. *Stereographic projection* is the map $\Sigma : S^2 \to \boldsymbol{R}^2 \cup \infty$ given by the following formula.

$$\Sigma(x,y,z) = \Big(\frac{x}{1-z}, \frac{y}{1-z}\Big). \tag{7}$$

Here is the inverse map:

$$\Sigma^{-1}(x,y) = \Big(\frac{2x}{1+x^2+y^2}, \frac{2y}{1+x^2+y^2}, 1 - \frac{2}{1+x^2+y^2}\Big). \tag{8}$$

Σ^{-1} maps circles in $\boldsymbol{R}^2$ to circles in S^2 and $\Sigma^{-1}(\infty) = (0,0,1)$.

Avatars: Stereographic projection gives us a correspondence between 5-point configurations on S^2 having $(0,0,1)$ as the last point and planar configurations:

$$\widehat{p}_0, \widehat{p}_1, \widehat{p}_2, \widehat{p}_3, (0,0,1) \in S^2 \iff p_0, p_1, p_2, p_3 \in \boldsymbol{R}^2, \qquad \widehat{p}_k = \Sigma^{-1}(p_k). \tag{9}$$

We call the planar configuration the *avatar* of the corresponding configuration in S^2. By a slight abuse of notation we write $\mathcal{E}_F(p_1,p_2,p_3,p_4)$ when we mean the F-potential of the corresponding 5-point configuration.

Figure 3.1 shows the two possible avatars (up to rotations) of the triangular bi-pyramid, first separately and then superimposed. We call the one on the left the *even avatar*, and the one in the middle the *odd avatar*. The points for the even avatar are $(\pm 1, 0)$ and $(0, \pm\sqrt{3}/3)$. When we superimpose the two avatars we see some extra geometric structure that is not relevant for our proof but worth mentioning. The two circles respectively have radii 1/2 and 1 and the 6 segments shown are tangent to the inner one.

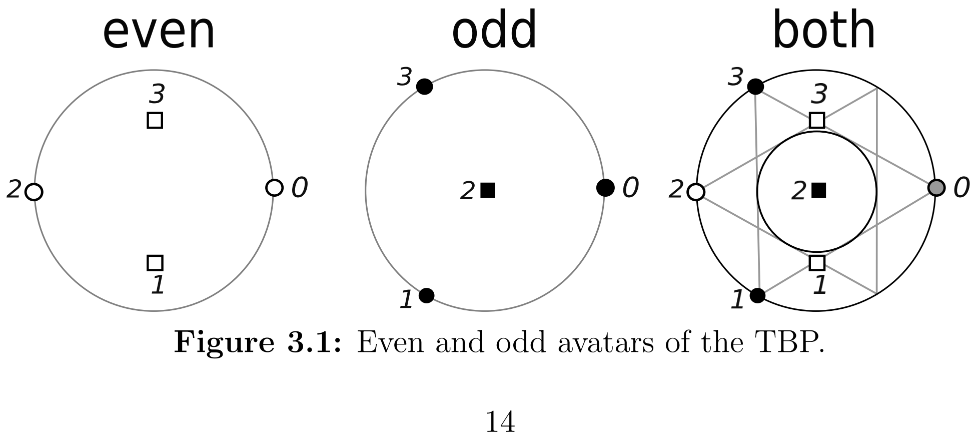

Figure 3.1: Even and odd avatars of the TBP.

The Special Domain: We let $\Upsilon \subset (\boldsymbol{R}^2)^4$ denote those avatars p_0, p_1, p_2, p_3 such that

1. $\|p_0\| \geq \|p_k\|$ for $k = 1, 2, 3$.
2. $512p_0 \in [433, 498] \times [0, 0]$. (That is, $p_0 \in [433/512, 498/512] \times \{0\}$.)
3. $512p_1 \in [-16, 16] \times [-464, -349]$.
4. $512p_2 \in [-498, -400] \times [0, 24]$.
5. $512p_3 \in [-16, 16] \times [349, 464]$.

We discuss the significance of Υ extensively in §2.3. In brief, the set Υ contains the avatars that compete with the TBP near the exponent ש.

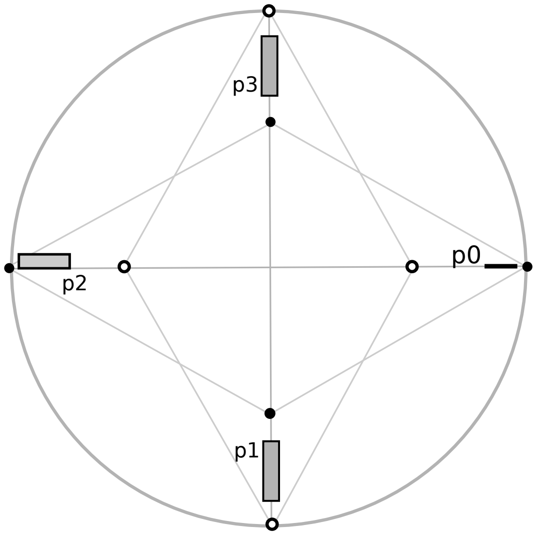

Figure 3.2: The sets defining Υ compared with two TBP avatars.

Symmetrization: Let (p_0, p_1, p_2, p_3) be an avatar with $p_0 \neq p_2$. We define

$$d_{02} = 2\|p_0 - p_2\|, \qquad d_{13} = 2\|\pi_{02}(p_1 - p_3)\|. \tag{10}$$

Here π_{02} is the projection onto the subspace perpendicular to the vector $p_0 - p_2$. Finally, we define

$$p_0^* = (d_{02}, 0), \quad p_1^* = (0, -d_{13}), \quad p_2^* = (-d_{02}, 0), \quad p_3^* = (0, d_{13}). \tag{11}$$

The avatar $p_1^*, p_2^*, p_3^*, p_4^*$ is invariant under reflections in the coordinate axes.

3.2 Reduction to Three Lemmas

We now reduce the Main Theorem to Lemmas A, B, C. Let

$$15_+ = 15 + \frac{25}{512}$$

as in the Main Theorem. Let Υ be as in §3.1. Recall that R_s is the Riesz potential.

Lemma 3.1 (A) *For $s \in (0, 13]$ the TBP uniquely minimizes the R_s-potential. For $s \in (13, 15_+]$, any R_s-potential minimizer is either the TBP or else isometric to a configuration whose avatar lies in Υ.*

Lemma A focuses our attention on the small domain Υ and the parameter range $[13, 15_+]$. Now we bring in the symmetrization operation from §3.1.

Lemma 3.2 (B) *Let $s \in [12, 15_+]$ and $(p_0, p_1, p_2, p_3) \in \Upsilon$. Then*

$$\mathcal{E}_{R_s}(p_0^*, p_1^*, p_2^*, p_3^*) \leq \mathcal{E}_{R_s}(p_0, p_1, p_2, p_3)$$

with equality if and only if the two avatars are equal.

Let Υ_4 denote the subset of Υ consisting of avatars which are invariant under reflections in the coordinate axes. Lemma B (which also works for $s \in [12, 13]$) focuses our attention on the same small parameter range $[13, 15_+]$ and on the symmetric avatars living in Υ_4.

Lemma 3.3 (C) *Let ξ_0 denote a avatar of the TBP. There exist $\mathfrak{w} \in (15, 15_+)$ such that the following is true.*

1. *For $s \in (13, \mathfrak{w})$ we have $\mathcal{E}_s(\xi_0) < \mathcal{E}_s(\xi)$ for all $\xi \in \Upsilon_4$.*
2. *For $s \in (\mathfrak{w}, 15_+)$ we have $\mathcal{E}_s(\xi_0) > \mathcal{E}_s(\xi)$ for some $\xi \in \Upsilon_4$.*

Also, for $s \in [15, 15_+]$ the restriction of $\mathcal{E}_s$ to Υ_4 has a unique minimum, and this minimum represents an FP.

The Main Theorem is an obvious consequence of Lemma A (§3.3), Lemma B (§10), and Lemma C (§3.4.) As a matter of convention we will point the reader to where the proof of the given lemma starts. Thus, the proof of Lemma A starts in §3.3.

3.3 Proof of Lemma A

Define

$$G_k(r) = (4 - r^2)^k. \tag{12}$$

Also define

$$\begin{aligned} G_5^\flat &= G_5 - 25G_1, \\ G_{10}^{\sharp\sharp} &= G_{10} + 28G_5 + 102G_2, \\ G_{10}^\sharp &= G_{10} + 13G_5 + 68G_2 \end{aligned} \tag{13}$$

Lemma 3.4 (A1) *The following is true.*

1. *The TBP is the unique minimizer for $G_4, G_5^\flat, G_6$.*
2. *The TBP is the unique minimizer for $G_{10}^\sharp$ among 5-point configurations which are not isometric to ones which have avatars in Υ.*
3. *The TBP is the unique minimizer for $G_{10}^{\sharp\sharp}$ among 5-point configurations which have avatars in Υ.*

We note two implications of Lemma A1:

- Since G_5 is a positive combination of $G_5^\flat$ and G_1, Lemma A1 immediately implies that the TBP is the unique minimizer for G_5.
- Since $G_{10}^{\sharp\sharp}$ is a positive combination of $G_{10}^\sharp$ and G_5 and G_2, Lemma A1 immediately implies that the TBP is the unique minimizer for $G_{10}^{\sharp\sharp}$.

Interpolation: Let T_0 be the TBP. We say that a pair (Γ_3, Γ_4) of functions *forces* the interval I if the following is true: If T is another 5-point configuration such that $\Gamma_k(T_0) < \Gamma_k(T)$ for $k = 3, 4$ then $\mathcal{E}_s(T_0) < \mathcal{E}_s(T)$ for all $s \in I$.

Lemma 3.5 (A2) *The following is true.*

1. *The pair (G_4, G_6) forces $(0, 6]$.*
2. *The pair $(G_5, G_{10}^{\sharp\sharp})$ forces $[6, 13]$.*
3. *The pair $(G_5^\flat, G_{10}^\sharp)$ forces $[13, 15_+]$.*

Lemma A is an immediate consequence of Lemma A1 (§4) and Lemma A2 (§9).

3.4 Proof of Lemma C

Let Ψ_4 denote the set of avatars of the form

$$(x,0), \qquad (0,-y), \qquad (-x,0), \qquad (0,y), \qquad 64(x,y) \in [43,64]. \tag{14}$$

We have $\Upsilon_4 \subset \Psi_4$. We like Ψ_4 better because it is more symmetric. We identify Ψ_4 with the square $[43/64,1]^2$ and we think of $\mathcal{E}_{R_s}$ as a function on this square. We usually write $\mathcal{E}_s = \mathcal{E}_{R_s}$. Again, the point (a,b) corresponds to the avatar with points $-p_2 = p_0 = (a,0)$ and $-p_1 = p_3 = (0,b)$. Though the TBP does not lie in Ψ_4, it corresponds to $(1, \sqrt{3}/3)$.

The FP avatars in Ψ_4 lie along the main diagonal. We call this diagonal Ψ_8. We define a smaller square $\widehat{\Psi}_4 \subset \Psi_4$ such that

$$64\widehat{\Psi}_4 = [55,56]. \tag{15}$$

We think of $\widehat{\Psi}_4$ as the sweet spot, the place where all the action happens.

We now define another symmetrization.

$$\sigma(x,y) = (z,z), \qquad z = \frac{x+y+(x-y)^2}{2}. \tag{16}$$

We have $\sigma : \widehat{\Psi}_4 \to \Psi_8$.

Lemma 3.6 (C1) *If $s \in [14,16]$ and $p \in \widehat{\Psi}_4$ Then $\mathcal{E}_s(\sigma(p)) \leq \mathcal{E}_s(p)$ with equality if and only if $\sigma(p) = p$.*

Remark: The operation σ is delicate. If we take the exponent $s = 13$, the operation actually seems to *increase* the energy for all points of $\widehat{\Upsilon}_4 - \widehat{\Upsilon}_8$. The magic only kicks in around exponent 13.53.

Our next result eliminates exponents and avatars not in $[13, 15_+] \times \widehat{\Psi}_4$.

Lemma 3.7 (C2) *Let ξ_0 be the point in the plane representing the TBP.*

1. *If $s \in [13,15]$ and $p \in \Psi_4$ then $\mathcal{E}_s(\xi_0) < \mathcal{E}_s(\xi)$.*

2. *If $s \in [15,15_+]$ and $p \in \Psi_4 - \widehat{\Psi}_4$ then $\mathcal{E}_s(\xi_0) < \mathcal{E}_s(\xi)$.*

3. *If $s \in [15,15_+]$ the restriction of $\mathcal{E}_s$ to $\widehat{\Psi}_8$ has a unique minimum.*

Statements 1 and 2 of Lemma C2 imply that for any $s \in [13, 15_+]$, any minimizer ξ of $\mathcal{E}_s$, not equal to the TBP avatar, lies in $\widehat{\Psi}_4$. Furthermore, such a ξ can only exist when $s \in [15, 15_+]$. Lemma C1 now says that ξ in fact lies in $\widehat{\Psi}_8$. Statement 3 of Lemma C2 adds the information that ξ is the unique minimizer in $\widehat{\Psi}_8$.

Lemma 3.8 (C3) *For any $\xi \in \widehat{\Psi}_8$ let $\Theta(s, \xi) = \mathcal{E}_s(\xi) - \mathcal{E}_s(\xi)$. Then for $s \in [15, 15_+]$ we have $\partial\Theta/\partial s < 0$.*

Here is how to deduce Lemma C from Lemma C1 (§12), Lemm C2 (§13) and Lemma C3 (§14). By Lemma C2, we have $\Theta(15, *) > 0$ on $\widehat{\Psi}_8$. We compute that $\Theta(15_+, x, x) < 0$ for $x = 445/512 \in [55, 56]/64$. Combining this with Lemma C3, we see that there exists a smallest parameter $\mathfrak{w} \in (15, 15_+)$ such that $\Theta(\mathfrak{w}, p^*) = 0$ for some $p^* \in \widehat{\Psi}_8$. For $s > \mathfrak{w}$, Lemma C3 now says that $\Theta(s, p^*) < 0$. This establishes Lemma C.

3.5 Afterword

(1) Lemma C22 (§13), Lemma C3, and Equation 133 give a basis for an algorithm to rigorously compute as many digits of $\mathfrak{w}$ as we like, up to the practical limits of the computer. If we pick $s \in (15, 15_+)$ and exhibit $x \in I$ such that $\Theta(s, x, x) < 0$ then $\mathfrak{w} < s$ by Lemma C3. By Equation 133, the function Θ is convex on $\widehat{\Psi}_8$ so it is easy to find the minimum with as much accuracy as we like. At the same time, if $\Theta(s, x, x) > 0$ for all $x \in I$ then $s < \mathfrak{w}$. Lemma C22 gives us the computational tools to rigorously prove a result like this. Running these two calculations simultaneously and updating the choices of s as we go, we can theoretically get as close an approximation to $\mathfrak{w}$ as we like.

(2) Here I explain the main details of the proof of the Auxiliary Theorem: All we need here is Lemma A for the interval $(-2, 0)$. Lemma A1 also applies to G_3. The only differences in the proof are discussed in the remarks in §4.5 and §6.4. Once these details are in place, our software gives a computational proof of Lemma A1 for G_3 in the same way it does for the other potentials. A variant of Lemma A2, with the same kind of proof, shows that the pair (G_3, G_5) forces $(-2, 0)$ with respect to F_s. The main extra detail needed here is the matrix of power combos given in §9.4. My software also shows this case, and gives a rigorous positivity proof as well.

4 Main Theorem: Proof of Lemma A1

4.1 Odd and Even Avatars

We call a pair of points $\widehat{p}, \widehat{q} \in S^2$ *far* if $\|\widehat{p} - \widehat{q}\| \geq 4/\sqrt{5}$. Note that $(\widehat{p}, \widehat{q})$ is a far pair if and only if $(\widehat{q}, \widehat{p})$ is a far pair. Our rather strange definition has a more natural interpretation in terms of the avatars. If we rotate S^2 so that $\widehat{p} = (0,0,1)$ then $q = \Sigma(\widehat{q})$ lies in the disk of radius $1/2$ centered at the origin if and only if $(\widehat{p}, \widehat{q})$ is a far pair.

We say that a point in a 5-point configuration is *odd* or *even* according to the parity of the number of far pairs it makes with the other points in the configuration. Correspondingly, define the parity of the avatar to be the parity of the number of points which are contained in the closed disk of radius $1/2$ about the origin. This extends our definition for the TBP avatars.

We call 2 avatars *isomorphic* if the corresponding 5-point configurations on S^2 are isometric. Every avatar is isomorphic to an even avatar. To see this, we form a graph by joining two points in a 5-point configuration by an edge if and only if they make a far pair. As for any graph, the sum of the degrees is even. Hence there is some vertex having even degree. When we rotate so that this vertex is $(0,0,1)$, the corresponding avatar is even. By focusing on the even avatars, and further using symmetry, we arrive at a configuration space where there is just one TBP avatar.

4.2 The Domains

Given an avatar $\xi = (p_0, p_1, p_2, p_3)$, we write $p_k = (p_{k1}, p_{k2})$. We define a domain $\Omega \subset \boldsymbol{R}^7$ to be the set of avatars ξ satisfying the following conditions.

1. ξ is even.

2. $\|p_0\| \geq \max(\|p_1\|, \|p_2\|, \|p_3\|)$.

3. $p_{12} \leq p_{22} \leq p_{32}$ and $p_{22} \geq 0$.

4. $p_{01} \in [0,2]$ and $p_{01} = 0$.

5. $p_j \in [-3/2, 3/2]^2$ for $j = 1,2,3$.

6. $\min(p_{1k}, p_{2k}, p_{3k}) \leq 0$ for $k = 1,2$.

We define $\Omega^\flat$ (to be used specially with $G_5^\flat$) by the same conditions except that we leave off Condition 6.

Closed Versus Open Conditions: If we want to check for inclusion in the interior of Ω or $\Omega^\flat$, all the inequalities above must be strict. We find it useful to work with the interior of Ω and $\Omega^\flat$ because we won't need to specially treat some boundary cases. This will make for a cleaner calculation.

A Tiny Cube: We specially treat avatars very near the TBP. When we string out the points of ξ_0, we get $(1,\ 0,-u,-1,0,0,u)$ where $u=\sqrt{3}/3$. See Figure 3.1. The space indicates that we do not record $p_{02}=0$. We let Ω_0 denote the cube of side-length 2^{-17} centered at ξ_0.

4.3 Reduction to Simpler Lemmas

Recall that we mean $\mathcal{E}_F(\xi)$ to be the F-potential of the 5-point configuration on the sphere corresponding to a planar avatar ξ. Lemma A1 makes 3 claims:

1. When $F=G_4,G_5^\flat,G_6$, the TBP avatar uniquely minimizes $\mathcal{E}_F$.
2. When $F=G_{10}^\sharp$, the TBP avatar uniquely minimizes $\mathcal{E}_F$ among avatars which are not isomorphic to ones in Υ.
3. When $F=G_{10}^{\sharp\sharp}$, the TBP avatar has smaller $\mathcal{E}_F$ value than all avatars in Υ.

Lemma 4.1 (A11) *Let F be any of $G_4,G_5^\flat,G_6,G_{10}^\sharp$. Then ξ_0 is the unique minimizer for $\mathcal{E}_F$ inside Ω_0.*

Lemma 4.2 (A12) *The following is true:*

1. *Let $F=G_4,G_6,G_{10}^\sharp$. If ξ is not equivalent to any avatar in Ω then then ξ does not minimize $\mathcal{E}_F$.*
2. *Let $F=G_5^\flat$. If ξ is not equivalent to any avatar in $\Omega^\flat$ then then ξ does not minimize $\mathcal{E}_F$.*

Let $[F]$ be the $\mathcal{E}_F$ value of the TBP avatars.

Lemma 4.3 (A13) *The following is true.*

1. *The infimum of* $\mathcal{E}_{G_4}$ *on* interior$(\Omega) - \Omega_0$ *is at least* $[G_4] + 2^{-50}$.
2. *The infimum of* $\mathcal{E}_{G_6}$ *on* interior$(\Omega) - \Omega_0$ *at at least* $[G_6] + 2^{-50}$.
3. *The infimum of* $\mathcal{E}_{G_5^\flat}$ *on* interior$(\Omega) - \Omega_0$ *is at least* $[G_5^\flat] + 2^{-50}$.
4. *The infimum of* $\mathcal{E}_{G_{10}^\sharp}$ *on* interior$(\Omega) - \Upsilon - \Omega_0$ *is at least* $[G_{10}^\sharp] + 2^{-50}$.
5. *The infimum of* $\mathcal{E}_{G_{10}^{\sharp\sharp}}$ *on* Υ *is at least* $[G_{10}^{\sharp\sharp}] + 2^{-50}$.

Lemma A13 is the main calculation. It follows from continuity that Lemma A13 remains true if we replace the interior of Ω by Ω itself. But then Lemma A1 follows immediately from Lemma A11 (§4.4), Lemma A12 (§4.5), and Lemma A13 (§5). Our choice of 2^{-50} is somewhat arbitrary.

4.4 Proof of Lemma A11

Recall that Ω_0 is the cube of side length 2^{-17} centered at ξ_0. For all our choices of F, the function $\mathcal{E}_F$ is a smooth function on $\boldsymbol{R}^7$. We check first of all that the gradient of $\mathcal{E}_F$ vanishes at ξ_0. This probably follows from symmetry, but to be sure we make a direct calculation in all cases.

Recall that the *Hessian* of a function is its matrix of second partial derivatives.

Lemma 4.4 (A111) *For each* $F = G_4, G_6, G_5^\flat, G_{10}^\sharp$*, the Hessian of* $\mathcal{E}_F$ *is positive definite at every point of* Ω_0.

Let $\xi \in \Omega_0$ be other than ξ_0. Lemma A111 (§6) together with the vanishing gradient implies that the restriction of $\mathcal{E}_F$ to the line segment γ joining ξ_0 to ξ is convex and has 0 derivative at ξ_0. Hence $\mathcal{E}_F(\xi) > \mathcal{E}_F(\xi_0)$. This proves Lemma A11

4.5 Proof of Lemma A12

Recall that ξ_0 is the avatar of the TBP. Let $[F] = \mathcal{E}_F(\xi_0)$. Since the TBP has 6 bonds of length $\sqrt{2}$, and 3 of length $\sqrt{3}$, and 1 of length $\sqrt{4}$, we have $[G_k] = 6 \times 2^k + 3$. Using this result, and Equation 13, we compute

$$[G_4] = 99, \qquad [G_6] = 387, \qquad [G_5^\flat] = -180, \qquad [G_{10}^\sharp] = 10518. \quad (17)$$

Let $\xi = p_0, p_1, p_2, p_3$ some other avatar.

Lemma 4.5 (A121) *Let* $F = G_6, G_5^\flat, G_{10}^\sharp$. *If* $\|p_0\| > 3/2$ *then* ξ *does not minimize* $\mathcal{E}_F$. *If* $F = G_4$ *then* ξ *does not minimize* $\mathcal{E}_F$ *provided that either* $\|p_0\| > 2$ *or* $\|p_0\|, \|p_j\| > 3/2$ *for some* $j = 1, 2, 3$.

Proof: Let τ_j be the term in $\mathcal{E}_F$ corresponding to the pair (p_j, p_4). Rather than work with $G_5^\flat$ we work with $G_5^* = G_5^\flat + 30$ so that all our functions are non-negative on $(0, 2]$. We have $[G_5^*] = 120$. When $\|p_0\| > 3/2$ we check that $\tau_0 > 450, 123, 26909$, which respectively exceeds $[G_6], [G_5^*], [G_{10}^\sharp]$. (We check this by computing that the distance involved is at most $d_0 = 4/\sqrt{13}$ and that F is monotone decreasing on $[0, d_0]$. Then we evaluate $F(d_0)$ in each case.)

Now we treat the case $F = G_4$. When $\|p_0\| > 2$ we have $\tau_0 > 104 > [G_4]$. When $\|p_0\|, \|p_i\| > 3/2$ we have $\tau_0 + \tau_j > 58 + 58 > [G_4]$. ♠

Lemma 4.6 (A122) *Let* F *be any strictly monotone decreasing potential. If* $\min(p_{1k}, p_{2k}, p_{3k}) > 0$ *for one of* $k = 1, 2$ *then* ξ *does not minimize* $\mathcal{E}_F$.

Proof: The corresponding 5-point configuration in S^2 is contained in a hemisphere H, and at least 3 of the points are in the interior of H. If we reflect one of the interior points across ∂H then we increase at least 2 of the distances in the configuration and keep the rest the same. ♠

Assume ξ is a minimizer for $\mathcal{E}_F$. As discussed in §4.1, we normalize so that ξ is even. Reordering p_0, p_1, p_2, p_3 and rotating, about the origin, we make $\|p_0\| \geq \|p_i\|$ for $i = 1, 2, 3$ and we move p_0 into the positive x-axis. Reflecting in the x-axis if necessary and reordering the points p_1, p_2, p_3 if necessary, we arrange that $p_{12} \leq p_{22} \leq p_{32}$ and $p_{22} \geq 0$. Lemma A121 tells us that, in all cases, $p_{01} \in [0, 2]$ and $p_j \in [-3/2, 3/2]^2$ for $j = 1, 2, 3$. We have also arranged that $p_{02} = 0$. For $F = G_5^\flat$ we have nothing left to check. Otherwise, Lemma A122 shows that ξ satisfies $\min(p_{1k}, p_{2k}, p_{3k}) \leq 0$ for $k = 1, 2, 3$.

Remark: (Auxiliary Theorem) We need to analyze G_3 for this case. We have $[G_3] = 51$. When $\|p_0\| > 2$ we have $\tau_0 \in [32, 33]$, which is not helpful. When $\|p_0\| > 4$ we have $\tau_0 > 53$, which works. So, we take 4 in place of 2 in our computer calculation for G_3. When $\|p_0\|, \|p_j\| > 3/2$ we have $\tau_0 + \tau_j > 21 + 21$, which is not quite enough, but here we can scramble to overcome the difference. The G_3-potential of the 4 point configuration on S^2 not involving $(0, 0, 1)$ is at least that of the regular tetrahedron, $14 + \frac{2}{9}$. Since $14 + 42 > 51$ we see that a minimizer for G_3 cannot have $\|p_0\|, \|p_j\| > 3/2$.

5 Main Calculation: Lemma A13

5.1 Blocks

We first list the ingredients in our main calculation and then explain the calculation itself.

Dyadic Subdivision: The *dyadic subdivision* of a D-dimensional cube is the list of 2^D cubes obtained by cutting the cube in half in all directions. We sometimes blur this terminology and say that any one of these 2^D smaller cubes is a *dyadic subdivision* of the big cube.

Blocks: We define a *block* to be a product of the form

$$B = Q_0 \times Q_1 \times Q_2 \times Q_3 \subset \square := [0,2] \times [-2,2]^2 \times [-2,2]^2 \times [-2,2]^2. \quad (18)$$

where Q_0 is a segment and Q_1, Q_2, Q_3 are squares, each obtained by iterated dyadic subdivision respectively of $[0,2]$ and $[-2,2]^2$.

We call B *acceptable* if Q_0 has length at most 1 and Q_1, Q_2, Q_3 have sidelength at most 2. If B is not acceptable we let the *offending index* be the lowest index where the condition fails.

The kth subdivision of a block amounts to performing dyadic subdivision to the kth factor and leaving the others alone. We call these operations S_0, S_1, S_2, S_3. Thus S_0 cuts B into two pieces and each other S_k cuts B into 4 pieces for $k = 1,2,3$. We let $S_k(B)$ denote the list of the blocks obtained by performing S_k on B. All the blocks our algorithm produces come from iterated subdivision of $\square$.

Rational Block Calculations: We say that a *rational block computation* is a finite calculation, only involving the arithmetic operations and min and max. The output of a rational block computation will be one of two things: **yes**, or an integer. A return of an integer is a statement that the computation does not definitively answer to the question asked of it. If the integer is -1 then there is no more information to be learned. If the integer lies in $\{0,1,2,3\}$ we use this integer as a guide in our algorithm. For example, we might ask if the block is acceptable. If not, then we would return the offending index, and our algorithm would subdivide the block along this index.

5.2 The Main Calculation

Recall that

$$\xi_0 = (1, 0, -\sqrt{3}/3, -1, 0, 0, \sqrt{3}/3) \in \Omega$$

is the avatar of the TBP and Ω_0 is the cube of side length 2^{-17} around ξ_0. Recall also that Υ is the special domain defined in §3.1.

Lemma 5.1 (A131) *There exists a rational block computation C_1 such that an output of* **yes** *for a block B implies that $B \subset \Omega_0$.*

Lemma 5.2 (A132) *There exists a rational block computation C_3 such that an output of* **yes** *for an acceptable block B implies that B is disjoint from the interior of Ω. The same goes for $\Omega^\flat$.*

Lemma 5.3 (A133) *There exists a rational block computation $C_3^\sharp$ such that an output of* **yes** *for a block B implies that $B \subset \Upsilon$. Likewise, there exists a rational block computation $C_3^{\sharp\sharp}$ such that an output of* **yes** *for a block B implies that B is disjoint from Υ.*

The proofs of the Lemmas A131, A132, A133, given below, just amount to checking the conditions in a fairly straightforward way. The final ingredient is the main ingredient. It is much more involved. All the energy potentials we consider are what we call *energy hybrids*. They have the form

$$F = \sum_{k=1}^{m} c_k G_k, \qquad G_k(r) = (4 - r^2)^k, \qquad c_1 \in \boldsymbol{Q}, \qquad c_2, ..., c_k \in \boldsymbol{Q}_+. \quad (19)$$

With some modification of Lemma E below we could also handle the case when some of $c_2, ..., c_k$ are negative. See Remark (1) after the statement of Lemma E.

Lemma 5.4 (A134) *For any function F given by Equation 19, there exists a rational block computation $C_{4,F}$ such that an output of* **yes** *for an acceptable block B implies that the minimum of $\mathcal{E}_F$ on B is at least $\mathcal{E}_F(\xi_0) + 2^{-50}$. Otherwise $C_{4,F}(B)$ is an integer in $\{0, 1, 2, 3\}$.*

Here is the main calculation.

1. We start with the list $L = \{\square\}$.

2. If $L = \emptyset$ then **HALT**. Otherwise let $B = Q_0 \times Q_1 \times Q_2 \times Q_3$ be the last block of L.

3. If B is not acceptable we delete B from L and append to L the subdivision of B along the offending index. We then return to Step 2. Any blocks considered beyond this step are acceptable.

4. If $C_1(B) =$ **yes** or $C_2(B) =$ **yes** we remove B from L and go to Step 2. Here we are eliminating blocks disjoint from the interior of Ω or else contained in Ω_0.

5. If $F = G_{10}^{\sharp}$ and $C_3^{\sharp}(B) =$ **yes** we remove B from L and go to Step 2. If $F = G_{10}^{\sharp\sharp}$ and $C_3^{\sharp\sharp}(B) =$ **yes** we remove B from L and go to Step 2.

6. If $C_{4,F}(B) =$ **yes** then we remove B from L and go to Step 2. Here we have verified that the F-energy of any avatar in B exceeds $[F] + 2^{-50}$.

7. If $C_{4,F}(B) = k \in \{0, 1, 2, 3\}$ then we delete B from L and append to L the blocks of the subdivision $S_k(B)$ and return to step 2.

If the algorithm reaches the **HALT** state for a given choice of F, this constitutes a proof that the corresponding statement of Lemma A13 is true.

Lemma 5.5 (A135) *The Main Computation reaches the* **HALT** *state for each choice of F listed in Lemma A13.*

Lemma A13 follows from Lemma A131 (§5.3), Lemma A132 (§5.4), Lemma A133 (§5.5), Lemma A134 (§5.6) and Lemma A135 (§5.8).

5.3 Proof of Lemma A131

Define intervals $I_0, I_1, I_{\sqrt{3}/3}$ such that

$$I_0 = [-2^{-17}, 2^{-17}], \quad I_1 = [1 - 2^{-17}, 1 + 2^{-17}] \quad 2^{30} I_{\sqrt{3}/3} = [619916940, 619933323] \quad (20)$$

$I_{\sqrt{3}/3}$ is a rational interval that is just barely contained inside the interval of length 2^{-17} centered at $\sqrt{3}/3$. Define

$$\Omega_{00} = (I_1 \times \{0\}) \times (I_0 \times -I_{\sqrt{3}/3}) \times (-I_1 \times I_0) \times (I_0 \times I_{\sqrt{3}/3}). \qquad (21)$$

We have $\Omega_{00} \subset \Omega_0$, though just barely. There are 128 vertices of B. We simply check whether each of these vertices is contained in Ω_{00}. If so then we return **yes**. In practice our program scales up all the coordinates by 2^{30} so that this test just involves integer comparisons.

5.4 Proof of Lemma A132

Let $B = Q_0 \times Q_1 \times Q_2 \times Q_3$ be an acceptable block. These blocks are such that the squares Q_1, Q_2, Q_3 do not cross the coordinate axes. For such squares, the minimum and maximum norm of a point in the square is realized at a vertex. Thus, we check that a square lies inside (respectively outside) a disk of radius r centered at the origin by checking that the square norms of each vertex is at most (respectively at least) r^2.

We check whether there is an index $j \in \{1,2,3\}$ such that all vertices of Q_j have norm at least $\max Q_0$. We return **yes** if this happens, because then all avatars in the interior of B will have some p_j with $\|p_j\| > \|p_0\|$.

We check whether there is an index $j \in \{1,2,3\}$ such that all vertices of Q_j have norm at least $3/2$. If so, we return **yes**. If this happens then $\|p_0\|, \|p_j\| > 3/2$ for all avatars in the interior of B.

We count the number a of indices j such that the vertices of Q_j all have norm at most $1/2$. We then count the number b of indices j such that all vertices of Q_j have norm at least $1/2$. We return **yes** if a is odd and $a+b=4$. In this case, every avatar in the interior of B is odd.

We write $I \leq J$ to indicate that all values in an interval I are less or equal to all values in an interval J. We also allow I and J to be single points in this notation. For each $j = 0,1,2,3$ we let Q_{jk} be the projection of Q_j onto the kth factor. Thus Q_{j1} and Q_{j2} are both line segments in $\boldsymbol{R}$.

We return **yes** for any of the following reasons.

- If $Q_{jk} \leq -3/2$ or $Q_{jk} \geq 3/2$ for any $j = 1,2,3$ and $k = 1,2$.
- $Q_{12} \geq Q_{22}$ or $Q_{12} \geq Q_{32}$ or $Q_{22} \geq Q_{32}$ or $Q_{22} \leq 0$.
- $Q_{j1} \geq 0$ for $j = 1,2,3$ or $Q_{j2} \geq 0$ for $j = 1,2,3$.

If any of these things happens, all avatars in Q violate some condition for membership in the interior of Ω. We don't check the last item for $\Omega^\flat$. ♠

5.5 Proof of Lemma A133

For $C_3^\sharp$ we return **yes** if all the vertices of B lie in Υ. For $C_3^{\sharp\sharp}$ we return **yes** if one of the factors of B is disjoint from the corresponding factor of Υ. This amounts to checking whether a pair of rational squares in the plane are disjoint. We do this using the projections defined for Lemma A132.

5.6 Proof of Lemma A134

We say that an acceptable block $B = Q_0 \times Q_1 \times Q_2 \times Q_3$ is *good* if we have $Q_j \in [-3/2, 3/2]^2$ for all $j = 1, 2, 3$. We first test whether B is a good block. If not, we return the lowest index i such that Q_i has a vertex outside $[-3/2, 3/2]^2$. Otherwise we proceed as follows.

We let $\mathcal{Q}$ denote the set of components of good blocks – either segments or squares. We also let $\{\infty\}$ be a member of $\mathcal{Q}$. We first define some measurements we take of members in $\mathcal{Q}$.

0. The Flat Approximation: Let Σ^{-1} be inverse stereographic projection, as in Equation 8. Given $Q \in \mathcal{Q}$ we define

$$Q^\bullet = \text{Convex Hull}(\Sigma^{-1}(v(Q)). \tag{22}$$

The set $Q^\bullet$ is either the point $(0, 0, 1)$, a chord of S^2 or else a convex planar quadrilateral with vertices in S^2 that is inscribed in a circle. We let $d_\bullet$ be the diameter of $Q_\bullet$. The quantity $d_\bullet^2$ is a rational function of the vertices of Q.

1. The Hull Approximation Constant: We think of $Q^\bullet$ as the linear approximation to

$$\widehat{Q} = \Sigma^{-1}(Q). \tag{23}$$

The constant we define here turns out to measure the distance between $\widehat{Q}$ and $Q^\bullet$. When $Q = \{\infty\}$ we define $\delta(Q) = 0$. Otherwise, let

$$\chi(D, d) = \frac{d^2}{4D} + \frac{(d^2)^2}{4D^3}. \tag{24}$$

This wierd function turns out to be an upper bound to a more geometrically meaningful non-rational function that computes the distance between an chord of length d of a circle of radius D and the arc of the circle it subtends.

When Q is a dyadic segment we define

$$\delta(Q) = \chi(2, \|\widehat{q}_1 - \widehat{q}_2\|). \tag{25}$$

Here q_1, q_2 are the endpoints of Q. When Q is a dyadic square we define

$$\delta(Q) = \max(s_0, s_2) + \max(s_1, s_3), \qquad s_j = \chi(1, \|q_j - q_{j+1}\|). \tag{26}$$

Here q_1, q_2, q_3, q_4 are the vertices of Q and the indices are taken cyclically. These are rational computations because $\chi(2, d)$ is a polynomial in d^2.

2. The Dot Product Estimator: By way of motivation, we point out that if $V_1, V_2 \in S^2$ then $G_k(\|V_1 - V_2\|) = (2 + 2V_1 \cdot V_2)^k$.

Now suppose that Q_1 and Q_2 are two dyadic squares. We set $\delta_j = \delta(Q_j)$. Given any $p \in \boldsymbol{R}^2 \cup \infty$ let $\widehat{p} = \Sigma^{-1}(p)$. Define

$$Q_1 \cdot Q_2 = \max_{i,j}(\widehat{q}_{1i} \cdot \widehat{q}_{2j}) + (\tau) \times (\delta_1 + \delta_2 + \delta_1\delta_2). \tag{27}$$

Here $\{q_{1i}\}$ and $\{q_{2j}\}$ respectively are the vertices of Q_1 and Q_2. The constant τ is 0 if one of Q_1 or Q_2 is $\{\infty\}$ and otherwise $\tau = 1$. Finally, we define

$$T(Q_1, Q_2) = 2 + 2(Q_1 \cdot Q_2). \tag{28}$$

3. The Local Error Term: For $Q_1, Q_2 \in \mathcal{Q}$ and $k \geq 1$ we define

$$\epsilon_k(Q_1, Q_2) = \frac{1}{2}k(k-1)T^{k-2}d_1^2 + 2kT^{k-1}\delta_1, \tag{29}$$

where

$$d_1 = d_\bullet(Q_1), \qquad \delta_1 = \delta(Q_1), \qquad T = T(Q_1, Q_2).$$

One of the terms in the error estimate comes from the analysis of the flat approximation and the second term comes from the analysis of the difference between the flat approximation and the actual subset of the sphere. The quantity is not symmetric in the arguments and $\epsilon_k(\{\infty\}, Q_2) = 0$.

4. The Global Error Estimate: Given a block $Q_0 \times Q_1 \times Q_2 \times Q_3$ we define

$$\mathbf{ERR}_k(B) = \sum_{i=0}^{N} \mathbf{ERR}_k(B, i), \qquad \mathbf{ERR}_k(B, i) = \sum_{j \neq i} \epsilon(Q_i, Q_j). \tag{30}$$

More generally, when $F = \sum c_k G_k$ is as in Equation 19, we define

$$\mathbf{ERR}_F(B) = \sum_{k=0}^{N} \mathbf{ERR}_F(B, i), \qquad \mathbf{ERR}_F(B, i) = \sum |c_k| \; \mathbf{ERR}_k(B, i) \tag{31}$$

Now we state the main error estimate, proved in §7. For the most part we only care about the (+) case of the lemma. We only need the (−) case when we deal with the potential $G_5 - 25G_1$.

Lemma 5.6 (E) *Let B be a good block. Let $F = G_k$ for any $k \geq 1$ or $F = -G_1$. Then*

$$\min_{p \in B} \mathcal{E}_F(v) \geq \min_{p \in v(B)} \mathcal{E}_k(v) - \mathbf{ERR}_k(B)$$

Proof of Lemma A134: Now we can prove lemma A134. Let B be an acceptable block. Once again, we mention that we immediately return an integer if our block B is not a good block. So, assume B is a good block. Let F be an energy hybrid. Let $[F]$ denote the F-potential of the TBP. If

$$\min_{p \in v(B)} \mathcal{E}_F(v) - \mathbf{ERR}_k(B) \geq [F] + 2^{-50} \tag{32}$$

we return **yes**. Otherwise we return the index i such that $\mathbf{ERR}_F(B, i)$ is the largest. In case of a tie, which probably never happens, we pick the lowest such index. ♠

Remarks: (1) Lemma E is true more generally for $F = \pm G_k$ but we do not need the general result and so (in the interest of simplicity) we ignore it. (2) The integer we return is designed to be a recommendation for the subdivision that is most likely to speed the computation along. We try to subdivide in such a way as to decrease the error term as fast as possible.

5.7 Discussion of the Implementation

Representing Blocks: We represent the coordinates of blocks by `longs`, which have 31 digits of accuracy. What we list are 2^{30} times the coordinates. Our algorithm never does so many subdivisions that it defeats this method of representation. In all but the main step (Lemma A134) in the algorithm below we compute with exact integers. When the calculation (such as squaring a `long`) could cause an overflow error, we first recast the `longs` as a `BigIntegers` in Java and then do the calculations.

Interval Arithmetic: For the main step of the algorithm we use interval arithmetic. We use the same implementation as we did in [**S1**], where we explain it in detail. Here is how it works in brief. If we have a calculation involving numbers $r_1, ..., r_n$, and we produce intervals $I_1, ..., I_n$ with dyadic rational numbers represented exactly by the computer such that $r_i \in I_i$ for

$i = 1, ..., n$. We then perform the usual arithmetic operations on the intervals, rounding outward at each step. The final output of the calculation, an interval, contains the result of the actual calculation.

In our situation here, the numbers $r_1, ..., r_n$ are, with one exception, dyadic rationals. (The exception is that the coordinates of the point representing the TBP are quadratic irrationals.) In principle we could do the entire computation, save for this one small exception, with expicit integer arithmetic. However, the complexity of the rationals involved, meaning the sizes of their numerators and denominators, qets quite large this way and the calculation is too slow.

One way to think about the difference between our explicitly defined exact integer arithmetic and interval arithmetic is that the integer arithmetic interrupts the calculation at each step and rounds outward so as to keep the complexity of the rational numbers from growing too large.

Guess and Check: Here is how we speed up the calculation. When we do Steps 6-7, we first do the calculation $C_{4,F}$ using floating point operations. If the floating version returns an integer, we use this integer to subdivide the box and return to step 2. If $C_{4,F}$ says **yes** then we retest the box using the interval arithmetic. In this way, we only pass a box for which the interval version says **yes**. This way of doing things keeps the calculation rigorous but speeds it up by using the interval arithmetic as sparingly as possible.

Parallelization: We also make our calculation more flexible using some parallelization. We classify each block $B = Q_0 \times Q_1 \times Q_2 \times Q_3$ with a number in $\{0, ..., 7\}$ according to the formula

$$\text{type}(B) = \sigma(c_{01} - 1) + 2\sigma(c_{11}) + 4\sigma(c_{31}) \in \{0, ..., 7\}.$$

Here c_{j1} is the first coordinate of the center of B_j and $\sigma(x)$ is 0 if $x < 0$ and 1 if $x > 0$. Step 3 of our algorithm guarantees that $\sigma(\cdot)$ is always applied to nonzero numbers.

We wrote our program so that we can select any subset $S \subset \{0, ..., 7\}$ we like and then (after Step 3) automatically pass any block whose type is not in S. Running the algorithm in parallel over sets which partition $\{0, ..., 7\}$ is logically the same as running the basic algorithm without any parallelization. To be able to do the big calculations in pieces, we run the program for various subsets of $\{0, ..., j\}$, sometimes in parallel.

5.8 Proof of Lemma A135

Here I give an account of one time I ran the computations to completion during January 2023. I used a 2017 iMac Pro with a 3.2 GHz Intel Zeon W processor, running the Mojave operating system. I ran the programs using Java 8 Update 201. (The Java version I use is not the latest one. The graphical parts of my program use some methods in the Applet class in a very minor but somehow essential way that I find hard to eliminate.) In listing the calculations I will give the approximate time and the exact number of blocks passed. Since we use floating point calculations to guide the algorithm, the sizes of the partitions can vary slightly with each run.

For G_4 : 2 hrs 14 min, 10848537 blocks.

For G_6: 5 hr 11 min, 25159337 blocks.

For $G_5^\flat$ types 1&2: 2 hr 31 min, 6668864 blocks.
For $G_5^\flat$ types 3&4: 1 hr 55 min, 4787489 blocks.
For $G_5^\flat$ types 5&6: 5 hr 33 min, 14160332 blocks.
For $G_5^\flat$ types 7&8: 3 hr 49 min, 9219550 blocks.

For $G_{10}^\sharp$ type 1: 4 hr 23 min, 6885912 blocks.
For $G_{10}^\sharp$ type 2: 9 hr 47 min, 15982122 blocks.
For $G_{10}^\sharp$ type 3: 3 hr 47 min, 5872029 blocks.
For $G_{10}^\sharp$ type 4: 7 hr 59 min, 13475260 blocks.
For $G_{10}^\sharp$ type 5: 8 hr 30 min, 13313492 blocks.
For $G_{10}^\sharp$ type 6: 15 hr 16 min, 24110457 blocks.
For $G_{10}^\sharp$ type 7: 5 hr 19 min, 7862780 blocks.
For $G_{10}^\sharp$ type 8: 8 hr 33 min, 13478467 blocks.

For $G_{10}^{\sharp\sharp}$ (on the domain Υ): 28 minutes, 805242 blocks.

6 Local Analysis: Proof of Lemma A111

6.1 Reduction to Simpler Statements

We set L=A111, so that we are trying to prove Lemma L. We consider F to be any of the 4 functions

$$G_4, \quad G_6, \quad G_5^\flat = G_5 - 25G_1, \quad 2^{-5}G_{10}^\sharp = 2^{-5}(G_{10} + 13G_5 + 68G_2).$$

Scaling the last function by 2^{-5} makes our estimates more uniform.

Recall that Ω_0 is the cube of side length 2^{-17} centered at the point

$$\xi_0 = \Big(1, 0, \frac{-1}{\sqrt{3}}, -1, 0, 0, \frac{1}{\sqrt{3}}\Big) \in \boldsymbol{R}^7 \tag{33}$$

In general, the point $(x_1, ..., x_7)$ represents the avatar

$$p_0 = (x_1, 0),\ p_1 = (x_2, x_3),\ p_2 = (x_4, x_5),\ p_3 = (x_6, x_7). \tag{34}$$

The quantity $\mathcal{E}_F(x_1, ..., x_7)$ is the F-potential of the 5-point configuration associated to the avatar under inverse stereographic projection Σ^{-1}.

$$\mathcal{E}_F(x_1, ..., x_7) = \sum_{i<j} F(\|\widehat{p}_i - \widehat{p}_j\|), \qquad \widehat{p} = \Sigma^{-1}(p). \tag{35}$$

Equation 8 gives the formula for Σ^{-1}.

Let $H\mathcal{E}_F$ be the Hessian of $\mathcal{E}_F$. Lemma L says $H\mathcal{E}_F$ is positive definite in Ω_0. Let $\partial_J \mathcal{E}_F$ be the (iterated) partial derivative of $\mathcal{E}_F$ with respect to a multi-index $J = (j_1, ..., j_7)$. Let $|J| = j_1 + ... + j_7$. Let

$$M_N = \sup_{|J|=N} M_J, \qquad M_J = \sup_{\xi \in \Omega_0} |\partial_J \mathcal{E}_F(\xi)|, \tag{36}$$

Let $\lambda(M)$ be the smallest eigenvalue of a real symmetric matrix M. Lemma L is an immediate consequence of the following two lemmas.

Lemma 6.1 (L1) *If $M_3(\mathcal{E}_F) < 2^{12}\lambda(H\mathcal{E}_F(\xi_0))$ then $\lambda(H\mathcal{E}_F(\xi)) > 0$ for all points $\xi \in \Omega_0$.*

Lemma 6.2 (L2) *$M_3(\mathcal{E}_F) < 2^{12}\lambda(H\mathcal{E}_F(\xi_0)))$ in all cases.*

6.2 Proof of Lemma L1

Let

$$H_0 = H\mathcal{E}_F(\xi_0), \qquad H = H\mathcal{E}_F(\xi), \qquad \Delta = H - H_0. \tag{37}$$

For any real symmetric matrix X define the L_2 matrix norm:

$$\|X\|_2 = \sqrt{\sum_{ij} X_{ij}^2} = \sup_{\|v\|=1} \|Xv\|. \tag{38}$$

Given a unit vector $v \in \boldsymbol{R}^7$ we have $H_0 v \cdot v \geq \lambda$. Hence

$$Hv \cdot v = (H_0 v + \Delta v) \cdot v \geq H_0 v \cdot v - |\Delta v \cdot v| \geq \lambda - \|\Delta v\| \geq \lambda - \|\Delta\|_2 > 0.$$

So, to prove Lemma L1 we just need to establish the implication

$$M_3 < 2^{12}\lambda(H_0) \quad \Longrightarrow \quad \|\Delta\|_2 < \lambda(H_0).$$

Let $t \to \gamma(t)$ be the *unit speed parametrized* line segment connecting p_0 to p in Ω_0. Note that γ has length $L \leq \sqrt{7} \times 2^{-18}$. We write $\gamma = (\gamma_1, ..., \gamma_7)$. Let H_t denote the Hessian of $\mathcal{E}_F$ evaluated at $\gamma(t)$. Let D_t denote the directional derivative along γ.

Now $\|D_t(H_t)\|_2$ is the speed of the path $t \to H_t$ in $\boldsymbol{R}^{49}$, and $\|\Delta\|_2$ is the Euclidean distance between the endpoints of this path. Therefore

$$\|\Delta\|_2 \leq \int_0^L \|D_t(H_t)\|_2 \, dt. \tag{39}$$

Let $(H_t)_{ij}$ denote the ijth entry of H_t. From the definition of directional derivatives, and from the Cauchy-Schwarz inequality, we have

$$(D_t H_t)_{ij}^2 = \Big(\sum_{k=1}^{7} \frac{d\gamma_k}{dt}\frac{\partial H_{ij}}{\partial k}\Big)^2 \leq 7M_3^2. \qquad \|D_t(H_t)\|_2 \leq 7^{3/2} M_3. \tag{40}$$

The second inequality follows from summing the first one over all 7^2 pairs (i, j) and taking the square root. Equation 39 now gives

$$\|\Delta\|_2 \leq L \times 7^{3/2} M_3 = 49 \times 2^{-18} M_3 < 2^{-12} M_3 < \lambda(H_0). \tag{41}$$

This completes the proof of Lemma L1.

6.3 Proof of Lemma L2

Let F be any of our functions. Let $H_0 = H\mathcal{E}_F(\xi_0)$.

Lemma 6.3 (L21) $\lambda(H_0) > 39$.

Proof: Let χ be the characteristic polynomial of H_0. This turns out to be a rational polynomial. We check in Mathematica that the signs of the coefficients of $\chi(t+39)$ alternate. Hence $\chi(t+39)$ has no negative roots. The file we use is `LemmaL21.m`. ♠

Recalling that $\xi_0 \in \boldsymbol{R}^7$ is the point representing the TBP, we define

$$\mu_N(\mathcal{E}_F) = \sup_{|I|=N} |\partial_I \mathcal{E}_F(\xi_0)|. \tag{42}$$

Lemma 6.4 (L22) *For any of our functions we have the bound*

$$\mu_3 < 45893, \qquad \frac{(7\times 2^{-18})^j}{j!}\mu_{j+3} < 38, \qquad j = 1,2,3. \tag{43}$$

Proof: We compute this in Mathematica. The file we use is `LemmaL22.m`. ♠

Lemma 6.5 (L23) *For any of our functions we have the bound*

$$\frac{(7\times 2^{-18})^4}{4!}M_7 < 2354.$$

Lemma 6.6 (L24) *We have*

$$M_3 \leq \mu_3 + \sum_{j=1}^{3} \frac{(7\times 2^{-18})^j}{j!}\mu_{j+3} + \frac{(7\times 2^{-18})^4}{4!}M_7 \tag{44}$$

Proof: Choose any multi-index J with $|J| = 3$. Let γ be the line segment connecting ξ_0 to any $\xi \in \Omega$. We parametrize γ by unit speed and furthermore set $\gamma(0) = \xi_0$. Let

$$f(t) = \partial_J \mathcal{E}_F \circ \gamma(t).$$

The bound for $|M_J|$ follows from Taylor's Theorem with remainder once we notice that

$$0 \leq t \leq \sqrt{7} \times 2^{-18}, \qquad \Big|\frac{\partial^n f(0)}{\partial t^n}\Big| \leq (\sqrt{7})^n \mu_n \qquad \Big|\frac{\partial^n f}{\partial t^n}\Big| \leq (\sqrt{7})^n M_n.$$

Since this works for all J with $|J| = 3$ we get the same bound for M_3. ♠

Lemmas L21 - L23 and Equation 43 imply

$$M_3 < 45893 + 3 \times 38 + 2354 \leq 65536 = 2^{16} \leq 2^{12}\lambda(H_0).$$

This completes the proof of Lemma L2.

6.4 Proof of Lemma L23

Now we come to the interesting part of the proof, the one place where we need to go beyond specific evaluations of our functions. When $r, s \geq 0$ and $r + s \leq 2d$ we have

$$\sup_{(x,y)\in \boldsymbol{R}^2} \frac{x^r y^s}{(1 + x^2 + y^2)^d} \leq (1/2)^{\min(r,s)}. \tag{45}$$

One can prove Equation 45 by factoring the expression into pieces with quadratic denominators. Here is a more general version. Say that a function $\phi : \boldsymbol{R}^4 \to \boldsymbol{R}$ is *nice* if it has the form

$$\sum_i \frac{C_i a^{\alpha_i} b^{\beta_i} c^{\gamma_i} d^{\delta_i}}{(1 + a^2 + b^2)^{u_i}(1 + c^2 + d^2)^{v_i}}, \quad \alpha_i, \beta_i, \gamma_i, \delta_i \geq 0, \quad \alpha_i+\beta_i \leq 2u_i, \qquad \gamma_i+\delta_i \leq 2v_i.$$

It follows from Equation 45 that

$$\sup_{\boldsymbol{R}^4} |\phi| \leq \langle\phi\rangle, \qquad \langle\phi\rangle = \sum_i |C_i|(1/2)^{\min(\alpha_i,\beta_i)+\min(\gamma_i,\delta_i)}. \tag{46}$$

Equation 46 is useful to us because it allows us to bound certain kinds of functions without having to evaluate then anywhere. We also note that if ϕ is nice, then so is any iterated partial derivative of ϕ. Indeed, the nice functions form a ring that is invariant under partial differentiation. This fact makes it easy to identify nice functions.

For any $\phi : \boldsymbol{R}^n \to \boldsymbol{R}$ we define

$$\overline{M}_7(\psi) = \sup_{|J|=7} \overline{M}_J(\psi), \qquad \overline{M}_J(\psi) = \sup_{\xi \in \boldsymbol{R}^n} |\partial_J(\phi)|. \tag{47}$$

We obviously have

$$M_7(\mathcal{E}_F) \leq \overline{M}_7(\mathcal{E}_F). \tag{48}$$

Recall that $\widehat{p} = \Sigma^{-1}(p)$, the inverse stereographic image of p. Define

$$f(a,b) = 4 - \|\widehat{(a,b)} - (0,0,1)\|^2 = \frac{4(a^2+b^2)}{1+a^2+b^2}. \tag{49}$$

$$g(a,b,c,d) = 4 - \|\widehat{(a,b)} - \widehat{(c,d)}\|^2 = \frac{4(1+2ac+2bd+(a^2+b^2)(c^2+d^2))}{(1+a^2+b^2)(1+c^2+d^2)}. \tag{50}$$

Notice that g is nice. Hence g^k is nice and $\partial_I g^k$ is nice for any multi-index. That means we can apply Equation 46 to $\partial_I g^k$.

$\mathcal{E}_{G_k}$ is a 10-term expression involving 4 instances of f^k and 6 of g^k. However, each variable appears in at most 4 terms. So, as soon as we take a partial derivative, at least 6 of the terms vanish. Moreover, $\partial_I f$ is a limiting case of $\partial_I g$ for any multi-index I. From these considerations, we see that

$$\overline{M}_7(\mathcal{E}_{G_k}) \leq 4 \times \overline{M}_7(g^k). \tag{51}$$

The function $\partial_I(g^k)$ is nice in the sense of Equation 46. Therefore

$$4 \times \overline{M}_7(g^k) \leq 4 \times \max_{|I|=7} \langle \partial_I g^k \rangle. \tag{52}$$

Using this estimate, and the Mathematica file `LemmaL23.m`, we get

$$\max_{k \in \{1,2,3,4,5,6\}} \frac{(7 \times 2^{-18})^4}{4!} \times 4 \times \overline{M}_7(g^k) \leq \frac{1}{1000}.$$

$$2^{-5} \times \frac{(7 \times 2^{-18})^4}{4!} \times 4 \times \overline{M}_7(g^{10}) \leq 2353. \tag{53}$$

The bounds in Lemma L23 follow directly from Equations 51 - 53 and from the definitions of our functions.

Remark: (For the Auxiliary Theorem) The analysis above works easily for G_3. In this case, the minimum eigenvalue satisfies $\lambda_0 > 14$. The bounds satisfy $\mu_3 \leq 316$ and $\mu_j < 1$ for $j = 4,5,6$. Again, $M_7 < 1/1000$ in this case.

7 Error Estimate: Proof of Lemma E

7.1 Guide to the Proof

Lemma E is stated in §5.6. It is the main error estimate that feeds into Lemma A134, which in turn feeds into Lemma A13, our main computation.

Our proof of Lemma E splits into two halves, an algebraic part and a geometric part. The algebraic part, which we do in this chapter, simply promotes a "local" result to a "global result". The geometric part, done in the next chapter, explains the meaning of the local error term $\epsilon_k(Q_1, Q_2)$ for $Q_1, Q_2 \in \mathcal{Q}$. Here $\mathcal{Q}$ is the space of components of good blocks, and also the point ∞.

The algebraic part involves what we call an *averaging system*. For the purpose of giving a uniform treatment, we treat every member of $\mathcal{Q}$ as a quadrilateral by the trick of repeating vertices. Thus, if we have a dyadic segment with vertices q_1, q_2 we will list them as q_1, q_1, q_2, q_2. For the point $\{\infty\}$ we will list the single vertex $q_1 = \infty$ as q_1, q_1, q_1, q_1. We say that an *averaging system* for a member of $\mathcal{Q}$ is a collection of maps $\lambda_1, \lambda_2, \lambda_3, \lambda_4 : Q \to [0,1]$ such that

$$\sum_{i=1}^{4} \lambda_i(z) = 1, \qquad \forall\ z \in Q.$$

The functions need not vary continuously. In case Q is a segment, we would have $\lambda_1 = \lambda_2$ and $\lambda_3 = \lambda_4$. In case $Q = \{\infty\}$ we would have $\lambda_j = 1/4$ for $j = 1, 2, 3, 4$.

We say that an *averaging system* for $\mathcal{Q}$ is a choice of averaging system for each member Q of $\mathcal{Q}$. The averaging systems for different members need not have anything to do with each other. In this chapter we will posit some additional properties of an averaging system and then prove Lemma E under the assumption such such an averaging system exists. In the next chapter we will prove the existence of the desired averaging system.

7.2 Reduction to a Local Result

We fix the function $F = G_k$ for some $k \geq 1$ or else $F = -G_1$. We write $\mathcal{E} = \mathcal{E}_F$. We let $\epsilon = \epsilon_k$, as in Equation 29. Our algebraic argument would work for any choice of F, but we need to use the choices above to actually get the averaging system we need. Let $q_{1,1}, q_{1,2}, q_{1,3}, q_{1,4}$ be the vertices of Q_1.

Lemma 7.1 (E1) *There exists an averaging system on $\mathcal{Q}$ with the following property: Let Q_1, Q_2 be distinct members of $\mathcal{Q}$. Given any $z_1 \in Q_1$ and $z_2 \in Q_2$ we have*

$$\sum_{i=1}^{4} \lambda_i(z_1) F(\|\widehat{q}_{1,i} - \widehat{z}_2\|) - F(\|\widehat{z}_1 - \widehat{z}_2\|) \leq \epsilon(Q_1, Q_2). \tag{54}$$

See §8 for the proof.

We are interested in 5-point configurations but we will work more generally so as to elucidate the general structure of the argument. We suppose that we have the good dyadic block $B = Q_0 \times ... \times Q_N$. The vertices of B are indexed by a multi-index

$$I = (i_0, ..., i_n) \in \{1, 2, 3, 4\}^{N+1}.$$

Given such a multi-index, which amounts to a choice of vertex of in each component member of the block. We define (as always, *via* inverse stereographic projection) the energy of the corresponding vertex configuration:

$$\mathcal{E}(I) = \mathcal{E}(q_{0,i_0}, ..., q_{N,i_N}) \tag{55}$$

Here is one more piece of notation. Given $z = (z_0, ..., z_n) \in B$ and a multi-index I we define

$$\lambda_I(z) = \prod_{i=0}^{N} \lambda_{i_j}(z_j). \tag{56}$$

Here λ_{i_j} is defined relative to the averaging system on Q_j.

Now we are ready to state our main global result. The global result uses the existence of an efficient averaging system. That is, it relies on Lemma E1.

Lemma 7.2 (E2) *Let $z = (z_0, ..., z_N) \in B$. Then*

$$\sum_I \lambda_I(z) \mathcal{E}(I) - \mathcal{E}(z) \leq \sum_{i=0}^{N} \sum_{j=0}^{N} \epsilon(Q_i, Q_j). \tag{57}$$

The lefthand sum is taken over all multi-indices. In the righthand sum, we set $\epsilon(Q_i, Q_i) = 0$ for all i.

Now let us deduce Lemma E from Lemma E2. Notice that

$$\sum_I \lambda_I(z) = \prod_{j=0}^{N} \Big(\sum_{a=1}^{4} \lambda_a(z_j) \Big) = 1. \tag{58}$$

Choose some $(z_1, ..., z_N) \in B$ which minimizes $\mathcal{E}$. We have

$$0 \leq \min_{p \in v(B)} \mathcal{E}(v) - \min_{v \in B} \mathcal{E}(v) = \min_{p \in v(B)} \mathcal{E}(v) - \mathcal{E}(z) \leq^*$$

$$\sum_I \lambda_I(z) \mathcal{E}(I) - \mathcal{E}(z) \leq \sum_{i=0}^{N} \sum_{j=0}^{N} \epsilon(Q_i, Q_j). \tag{59}$$

The starred inequality comes from the fact that a minimum is less or equal to a convex average. The last expression is $\mathbf{ERR}(B)$ when $N = 4$ and $Q_4 = \infty$.

7.3 From Local to Global

Now we deduce the global Lemma E2 from the local Lemma E1.

Lemma 7.3 (E21) *Lemma E2 holds when $N = 1$.*

Proof: In this case, we have a block $B = Q_0 \times Q_1$. Setting $\epsilon_{ij} = \epsilon(Q_i, Q_j)$, Lemma E1 gives us

$$F(\|z_0 - z_1\|) \geq \sum_{\alpha=1}^{4} \lambda_\alpha(z_0) F(\|q_{0\alpha} - z_1\|) - \epsilon_{01}. \tag{60}$$

Applying Lemma E1 to the pair of points $(z_1, q_{0\alpha}) \in Q_1 \times Q_0$ we have

$$F(\|z_1 - q_{0\alpha}\|) \geq \sum_{\beta=1}^{4} \lambda_\beta(z_1) F(\|q_{1\beta} - q_{0\alpha}\|) - \epsilon_{10}. \tag{61}$$

Plugging the second equation into the first and using $\sum \lambda_\alpha(z_0) = 1$, we have

$$F(\|z_0 - z_1\|) \geq \sum_{\alpha,\beta} \lambda_\alpha(z_0)[\lambda_\beta(z_1) F(\|q_{1\beta} - q_{0\alpha}\|) - \epsilon_{10}] - \epsilon_{01} =$$

$$\sum_{\alpha,\beta} \lambda_\alpha(z_0)\lambda_\beta(z_1) F(\|q_{1\beta} - q_{0\alpha}\|) - (\epsilon_{10} + \epsilon_{01}). \tag{62}$$

Equation 62 is equivalent to Equation 57 when $N = 1$. ♠

Now we do the general case.

Lemma 7.4 (E22) *Lemma E2 holds when $N \geq 2$.*

Proof: We rewrite Equation 62 as follows:

$$F(\|z_0 - z_1\|) \geq \sum_A \lambda_{A_0}(z_0)\lambda_{A_1}(z_1)\ F(\|q_{0A_0} - q_{1A_1}\|) - (\epsilon_{01} + \epsilon_{10}). \tag{63}$$

The sum is taken over multi-indices A of length 2.

We also observe that

$$\sum_{I'} \lambda_{I'}(z') = 1, \qquad z' = (z_2, ..., z_N). \tag{64}$$

The sum is taken over all multi-indices $I' = (i_2, ..., i_N)$. Therefore, if we hold $A = (A_0, A_1)$ fixed, we have

$$\lambda_{A_0}(z_0)\lambda_{A_1}(z_1) = \sum_{I''} \lambda_{I''}(z). \tag{65}$$

The sum is taken over all multi-indices of length $N + 1$ which have $I_0 = A_0$ and $I_1 = A_1$. Combining these equations, we have

$$F(\|z_0 - z_1\|) \geq \sum_I \lambda_I(z) F(\|q_{0I_0} - q_{1I_1}\|) - (\epsilon_{01} + \epsilon_{10}). \tag{66}$$

The same argument works for other pairs of indices, giving

$$F(\|z_i - z_j\|) \geq \sum_I \lambda_I(z) F(\|q_{iI_i} - q_{jI_j}\|) - (\epsilon_{ij} + \epsilon_{ji}). \tag{67}$$

Let us restate this as $X_{ij} - Y_{ij} \geq Z_{ij}$, where

$$X_{ij} = \sum_I \lambda_I(z) F(\|q_{iI_i} - q_{jI_j}\|), \quad Y_{ij} = F(\|z_i - z_j\|), \quad Z_{ij} = \epsilon_{ij} + \epsilon_{ji}.$$

When we sum Y_{ij} over all $i < j$ we get the second term in Equation 57. When we sum Z_{ij} over all $i < j$ we get the third term in Equation 57. When we sum X_{ij} over all $i < j$ we get

$$\sum_{i<j} \Big(\sum_I \Lambda_I(z) F(\|q_{iI_i} - q_{jI_j}\|) \Big) = \sum_I \sum_{i<j} \Lambda_I(z)\ F(\|q_{iI_i} - q_{jI_j}\|) =$$

$$\sum_I \Lambda_I(z) \bigg(\sum_{i<j} F(\|q_{iI_i} - q_{jI_j}\|) \bigg) = \sum_I \lambda_I(z) \mathcal{E}(I).$$

This is the first term in Equation 57. This proves Lemma E2. ♠

8 Error Estimate: Proof of Lemma E1

8.1 The Efficient Averaging System

Lemma E1 posits the existence of what we call an efficient averaging system. Here we define it. Recall that $Q^\bullet$ is the convex hull of the vertices $\widehat{q}_1, \widehat{q}_2, \widehat{q}_3, \widehat{q}_4$ of $\widehat{Q} = \Sigma^{-1}(Q)$. What we want from the system is that for any $z^\bullet \in Q^\bullet$

$$z^\bullet = \sum_{i=1}^{4} \lambda_i(z^\bullet)\widehat{q}_i. \tag{68}$$

If $z^\bullet$ lies in the convex hull of $\widehat{q}_1$, $\widehat{q}_2$, $\widehat{q}_3$, then we let $\lambda_1(z^\bullet)$, $\lambda_2(z^\bullet)$, $\lambda_3(z^\bullet)$ be barycentric coordinates on this triangle and we set $\lambda_4(z^\bullet) = 0$. If $z^\bullet$ lies in the convex hull of $\widehat{q}_1$, $\widehat{q}_2$, $\widehat{q}_4$, then we let $\lambda_1(z^\bullet)$, $\lambda_2(z^\bullet)$, $\lambda_4(z^\bullet)$ be barycentric coordinates on this triangle and we set $\lambda_3(z^\bullet) = 0$. This definition agrees on the overlap, which is the line segment joining $\widehat{q}_3$ to $\widehat{q}_4$.

To get our averaging system on $Q \in \mathcal{Q}$ we define

$$\lambda_j(z) = \lambda_j(z^\bullet), \tag{69}$$

where $z^\bullet$ is some choice of point in $Q^\bullet$ which is <u>closest</u> to $\widehat{z}$. If there are several closest points we pick the one (say) which has the smallest first coordinate. We prove Lemma E1 with respect to the averaging system we have just defined.

8.2 Reduction to Simpler Statements

Let F be either G_k for some $k \geq 1$ or else $F = -G_1$. For convenience we expand out the statement of Lemma E1.

Lemma 8.1 (E1) *The efficient averaging system on $\mathcal{Q}$ has the following property. Let Q_1, Q_2 be distinct members of $\mathcal{Q}$. Given any $z_1 \in Q_1$ and $z_2 \in Q_2$ we have*

$$\sum_{i=1}^{4} \lambda_i(z_1)F(\|\widehat{q}_{1,i} - \widehat{z}_2\|) - F(\|\widehat{z}_1 - \widehat{z}_2\|) \leq \frac{1}{2}k(k-1)T^{k-2}d_1^2 + 2kT^{k-1}\delta_1. \tag{70}$$

Here (as in §5.6) δ_1 and d_1 respectively are the Hull Approximation constant and diameter of Q_1, and

$$T = 2 + 2(Q_1 \cdot Q_1), \qquad Q_1 \cdot Q_2 = \max_{i,j}(\widehat{q}_{1,i} \cdot \widehat{q}_{2,j}) + (\tau) \times (\delta_1 + \delta_2 + \delta_1\delta_2). \tag{71}$$

$\tau = 0$ or $\tau = 1$ depending on whether one of Q_1, Q_2 is $\{\infty\}$. We are maximizing over the dot product of the vertices and then either adding an error term or not.

Define

$$X_\bullet = F(z_1^\bullet - \widehat{z}_2) = (2 + 2z_1^\bullet \cdot \widehat{z}_2)^k \quad \text{or} \quad -2 - 2z_1^\bullet \cdot \widehat{z}_2. \tag{72}$$

Lemma E1 is an immediate consequence of the following two results.

Lemma 8.2 (E11) $\sum_{i=1}^4 \lambda_i(z_1) F(\|\widehat{q}_{1,i} - \widehat{z}_2\|) - X_\bullet \leq \frac{1}{2}k(k-1)T_\bullet^{k-2} d_1^2.$

Lemma 8.3 (E12) $X_\bullet - F(\|\widehat{z}_1 - \widehat{z}_2\|) \leq 2kT^{k-1}\delta.$

8.3 Proof of Lemma E11

Suppose first $F = -G_1$. We hold $\widehat{z}_2$ fixed and define

$$L(\widehat{q}) = F(\|\widehat{q} - \widehat{z}_2\|) = -2 - 2\widehat{q} \cdot \widehat{z}_2.$$

Lemma E2, in this special case, says that

$$\sum_{i=1}^4 \lambda_i(z_1) L(\widehat{q}_{1,i}) - L(z_1^\bullet) = 0.$$

But this follows from Equation 69 and the (bi) linearity of the dot product.

Now we deal with the case where $F = G_k$ for $k \geq 1$. We prove the following two lemmas at the end of the chapter.

Lemma 8.4 (E111) *For $j = 1, 2$ let γ_j be a point on a line segment connecting a point of $\widehat{Q}_j$ to a closest point on $Q_j^\bullet$. Then $\gamma_1 \cdot \gamma_2 \leq Q_1 \cdot Q_2$.*

Lemma 8.5 (E112) *Let $M \geq 2$ and $k = 1, 2, 3....$ Suppose*

- $0 \leq x_1 \leq ... \leq x_M$
- $\sum_{i=1}^M \lambda_i = 1$ *and* $\lambda_i \geq 0$ *for all* i.

Then

$$0 \leq \sum_{i=1}^M \lambda_i x_i^k - \Big(\sum_{I=1}^M \lambda_i x_i\Big)^k \leq \frac{1}{8}k(k-1)x_M^{k-2}\ (x_M - x_1)^2. \tag{73}$$

Recall that $q_{1,1}, q_{1,2}, q_{1,3}, q_{1,4}$ are the vertices of Q_1. Let $\lambda_i = \lambda_i(z_1)$. We set

$$x_i = 4 - \|\widehat{q}_{1,i} - \widehat{z}_2\|^2 = 2 + 2\widehat{q}_{1,i} \cdot \widehat{z}_2, \qquad i = 1,2,3,4. \tag{74}$$

Note that $x_i \geq 0$ for all i. We order so that $x_1 \leq x_2 \leq x_3 \leq x_4$. We have

$$\sum_{i=1}^{4} \lambda_i(z) F(\|q_{1,i} - z_2\|) = \sum_{i=1}^{4} \lambda_i x_i^k, \tag{75}$$

$$X_\bullet = (2 + 2z_1^\bullet \cdot \widehat{z}_2)^k = \Big(\sum_{i=1}^{4} \lambda_i \times (2 + \widehat{q}_i \cdot \widehat{z}_2)\Big)^k = \Big(\sum_{i=1}^{4} \lambda_i x_i\Big)^k. \tag{76}$$

By Equation 75, Equation 76, and the case $M = 4$ of Lemma E112, we have

$$\sum_{i=1}^{4} \lambda_i(z) F(\|q_{1,i} - z_2\|) - X_\bullet = \sum_{i=1}^{4} \lambda_i x_i^k - \Big(\sum_{i=1}^{4} \lambda_i x_i\Big)^k \leq \frac{1}{8} k(k-1) x_4^{k-2} (x_4 - x_1)^2. \tag{77}$$

By Lemma E111

$$x_4 = 2 + 2(\widehat{q}_4 \cdot \widehat{z}_2) \leq T. \tag{78}$$

Since d_1 is the diameter of $Q_1^\bullet$, and $\widehat{z}_2$ is a unit vector,

$$x_4 - x_1 = 2\widehat{z}_2 \cdot (\widehat{q}_4 - \widehat{q}_1) \leq 2\|\widehat{q}_4 - \widehat{q}_1\| \leq 2d_1 \tag{79}$$

Plugging Equations 78 and 79 into Equation 77, we get Lemma E12.

8.4 Proof of Lemma E12

Let $\delta(Q)$ be the hull approximation constant for $Q \in \mathcal{Q}$, as defined (depending on Q) in Equation 25 or Equation 26.

Lemma 8.6 (E121) *Let Q be any good dyadic square or segment. Then every point of $\widehat{Q}$ is within $\delta(Q)$ of the quadrilateral $Q^\bullet$.*

Lemma E121 implies that $\|\widehat{z}_1 - z_1^\bullet\| < \delta(Q)$. Let γ_1 denote the unit speed line segment connecting $z_1^\bullet$ to $\widehat{z}_1$. The length L of γ_1 is at most δ_1, by Lemma E11. So, $\gamma_1(0) = z_1^\bullet$ and $\gamma_1(L) = \widehat{z}_1$. Define

$$f(t) = \Big(2 + 2\widehat{z}_2 \cdot \gamma_1(t)\Big)^k \text{ or } -2 - 2\widehat{z}_2 \cdot \gamma_1(t), \tag{80}$$

depending on the case. The argument we give works equally well more generally when we use $F = \pm G_k$.

We have $f(0) = X_\bullet$ and $f(L) = F(\|\widehat{z}_1 - \widehat{z}_2\|)$. Hence

$$X_\bullet - F(\|\widehat{z}_1 - \widehat{z}_2\|) = f(0) - f(L), \qquad L \leq \delta_1. \tag{81}$$

Combining the Chain Rule, the Cauchy-Schwarz inequality, and Lemma E111, we have

$$|f'(t)| = \left|(2\widehat{z}_2 \cdot \gamma_1'(t)) \times k\Big(2 + 2\widehat{z}_2 \cdot \gamma_1(t)\Big)^{k-1}\right| \leq$$

$$2k\Big|(2 + 2\widehat{z}_2 \cdot \gamma_1(t))\Big|^{k-1} \leq 2k(2 + 2(Q_1 \cdot Q_2))^{k-1} = 2kT^{k-1}.$$

In short

$$|f'(t)| \leq 2kT^{k-1}. \tag{82}$$

Lemma E13 follows Equation 82, Equation 81, and integration.

8.5 Proof of Lemma E111

See Equation 71 (or §5.6) for the definition of $Q_1 \cdot Q_2$. We first treat the case $\tau = 1$, meaning that neither Q_1 nor Q_1 is $\{\infty\}$. Since the dot product is bilinear,

$$q_1^\bullet \cdot q_2^\bullet \leq \max_{i,j}(\widehat{q}_{1i} \cdot \widehat{q}_{2j}). \tag{83}$$

By Lemma E11, and by hypothesis, we can find points $z_1^\bullet$ and $z_2^\bullet$ such that

$$\gamma_j = z_1^\bullet + h_1, \qquad \gamma_2 = z_2^\bullet + h_2, \qquad \|h_j\| \leq \delta_j.$$

But then by the triangle inequality and the Cauchy-Schwarz inequality

$$|(\gamma_1 \cdot \gamma_2) - (z_1^\bullet \cdot z_2^\bullet)| \leq |z_1^\bullet \cdot h_2| + |z_2^\bullet \cdot h_1| + |h_1 \cdot h_2| \leq \delta_1 + \delta_2 + \delta_1\delta_2.$$

This combines with Equation 83 to complete the proof when $\tau = 1$.

Suppose $\tau = 0$. Without loss of generality assume that $Q_2 = \{\infty\}$. The maximum of $\widehat{q}_1 \cdot (0,0,1)$, for $q_1 \in Q_1$, is achieved when q_1 is vertex of Q_1. At the same time, the maximum of $q_1^\bullet \cdot (0,0,1)$, for $q_1^\bullet \in Q_1^\bullet$ is achieved when $q_1^\bullet$ is a vertex of $Q_1^\bullet$. But then our lemma is true for the endpoints of the segment containing γ. Since the dot product with $(0,0,1)$ varies linearly along this line segment, the same result is true for all points on the line segment.

8.6 Proof of Lemma E112

Lemma 8.7 (E1121) *Suppose $a, x \in [0,1]$ and $k \geq 2$. Then $f(x) \leq g(x)$, where*

$$f(x) = (ax^k + 1 - a) - (ax + 1 - a)^k; \qquad g(x) = \frac{1}{8}k(k-1)(1-x)^2. \quad (84)$$

Proof: Since $f(1) = g(1) = f'(1) = g'(1) = 0$ the Cauchy Mean Value Theorem (applied twice) tells us that for any $x \in (0,1)$ there are values $y < z \in [x,1]$ such that

$$\frac{f(x)}{g(x)} = \frac{f'(y)}{g'(y)} = \frac{f''(z)}{g''(z)} = 4az^{k-2}\Big[1-a\Big(a+\frac{1-a}{z}\Big)^{k-2}\Big] \leq 4a(1-a) \leq 1. \quad (85)$$

This completes the proof. ♠

Remark: The above proof, suggested by an anonymous referee of [**S4**], is better than my original proof.

Now we prove the main inequality The lower bound is a trivial consequence of convexity, and both bounds are trivial when $k = 1$. So, we take $k = 2, 3, 4, ...$ and prove the upper bound. Suppose first that $M \geq 3$. We have one degree of freedom when we keep $\sum \lambda_i x_i$ constant and try to vary $\{\lambda_j\}$ so as to maximize the left hand side of the inequality. The right hand side does not change when we do this, and the left hand side varies linearly. Hence, the left hand size is maximized when $\lambda_i = 0$ for some i. But then any counterexample to the lemma for $M \geq 3$ gives rise to a counter example for $M - 1$. Hence, it suffices to prove the inequality when $M = 2$.

In the case $M = 2$, we set $a = \lambda_1$. Both sides of the inequality in Lemma E112 are homogeneous of degree k, so it suffices to consider the case when $x_2 = 1$. We set $x = x_1$. Our inequality then becomes exactly the one treated in Lemma E1121. This completes the proof.

8.7 Proof of Lemma E121

We remind the reader of the wierd function $\chi(D)$ and we introduce a more geometrically meaningfun function

$$\chi(D,d) = \frac{d^2}{4D} + \frac{d^4}{4D^3}, \qquad \chi^*(D,d) = \frac{1}{2}(D - \sqrt{D^2 - d^2}). \qquad (86)$$

Lemma 8.8 (E1211) *$\chi^*(D,d) \leq \chi(D,d)$ for all $d \in [0,D]$.*

Proof: By homogeneity, it suffices to prove the result when $D = 1$. To simpify the algebra we define $A = 2\chi(1,d) - 1$ and $A^* = 2\chi^*(1,d) - 1$. We compute $4A^2 - 4(A^*)^2 = d^4(d-1)(d+1)(d^2+3)$. Hence, the sign of $A - A^*$ does not change on $(0,1)$. We check that $A > A^*$ when $d = 1/2$. Hence $A > A^*$ on $(0,1)$. This implies the inequality. ♠

Segment Case: Let Q be dyadic segment. Here $\widehat{Q}$ is the arc of a great circle and $Q^\bullet$ is the chord of the arc joining the endpoints of this arc. Let d be the length of $Q^\bullet$. The point of $\widehat{Q}$ farthest from $Q^\bullet$ is the midpoint of this $\widehat{Q}$. Let x be the distance between the midpoint of $\widehat{Q}$ and the midpoint of $Q^\bullet$. From elementary geometry, $x(D-x) = (d/2)^2$. Solving for x we find that $x = \chi^*(2,d)$. Lemma E1211 finishes the proof.

Square Case: Let Q be a dyadic square and let $z \in Q$ be a point. Let L be the vertical line through x and let z_{01}, z_{23} be the endpoints of the segment $L \cap Q$. We label the vertices of Q (in cyclic order) so that z_{01} lies on the edge joining q_0 to q_1 and z_{23} lies on the edge joining q_2 to q_3.

If M is a horizontal line intersecting Q then the circle $\Sigma^{-1}(M \cup \infty)$ has diameter at least 1. The point is that this circle contains $(0,0,1)$ and also $\Sigma^{-1}(0,y)$ for some $|y| \leq 3/2$. In fact the diameter is at least $4/\sqrt{13}$. The same goes for vertical lines intersecting Q.

Define $d_j = \|\widehat{p}_j - \widehat{p}_{j+1}\|$ with the indices taken cyclically. The length of the segment σ joining the endpoints of $\Sigma^{-1}(L \cap Q)$ varies monotonically with the position of L. Hence, σ has length at most $\max(d_1, d_3)$. At the same time, $\Sigma^{-1}(L \cap Q)$ is contained in a circle of diameter at least 1. The same argument as in the segment case now shows that there is a point $z^* \in \sigma$ which is within $t_{13} = \max(\chi(1,d_1), \chi(1,d_3))$ of $\widehat{z}$.

The endpoints of σ respectively are on the spherical arcs obtained by mapping the top and bottom edge of Q onto S^2 via Σ^{-1}. Hence, one endpoint of σ is within $\chi(1,d_0)$ of a point on the corresponding edge of $\partial Q^\bullet$ and the other endpoint of σ is within $\chi(1,d_2)$ of a point on the opposite edge of $\partial Q^\bullet$. But that means that either endpoint of σ is within $t_{02} = \max(\chi(1,d_0), \chi(1,d_2))$ of a point in $Q^\bullet$. But then every point of the segment σ is within t_{02} of some point of the line segment joining these two points of $Q^\bullet$. In particular, there is a point $z^\bullet \in Q^\bullet$ which is within t of z^*. The triangle inequality completes the proof of Lemma E121.

9 Interpolation: Proof of Lemma A2

9.1 Reduction to Smaller Results

Recall that $15_+ = 15 + \frac{25}{512}$. Referring to Equations 12 and 13, we define

$$P_1 = (G_4, G_6), \qquad P_2 = (G_5, G_{10}^{\sharp\sharp}), \qquad P_3 = (G_5^{\flat}, G_{10}^{\sharp}), \tag{87}$$

$$I_1 = (0, 6], \qquad I_2 = [6, 13], \qquad I_3 = [13, 15_+]. \tag{88}$$

Lemma A2 says that the pair P_j forces the interval I_j for $j = 1, 2, 3$. The beginning of our proof of Lemma A21 will recall what this means.

Let R_s be the Riesz s-potential. We say that a pair of functions (Γ_3, Γ_4) *specially forces* $s \in \boldsymbol{R} - \{0\}$ if there are constants $a_0, ..., a_4$ (depending on s) such that

$$\Lambda_s = a_0 + a_1 G_1 + a_2 G_2 + a_3 \Gamma_3 + a_4 \Gamma_4, \tag{89}$$

1. $\Lambda_s(x) = R_s(x)$ for $x = \sqrt{2}, \sqrt{3}, \sqrt{4}$.

2. $a_1, a_2, a_3, a_4 > 0$.

3. $\Lambda_s(x) \leq R_s(x)$ for all $x \in (0, 2]$.

We say that (Γ_3, Γ_4) *specially forces* the interval I if this pair specially forces all $s \in I$.

Lemma 9.1 (A21) *If* (Γ_3, Γ_4) *specially forces* I *then* Γ *forces* I*.*

Proof: Let T_0 be the TBP and let T be some other 5-point configuration. We simplify the notation and write $F(T) = \mathcal{E}_F(T)$. We assume $\Gamma_j(T_0) < \Gamma_j(T)$ for $j = 3, 4$ and we want to show that that $R_s(T_0) < R_s(T)$ for all $s \in I$. It is well known that $\Gamma_1(T_0) \leq \Gamma_1(T)$ and, by Tumanov's result, $\Gamma_2(T_0) \leq \Gamma_2(T)$. Let $a_j = a_j(s)$ for $s \in I$. The quantities $\sqrt{2}, \sqrt{3}, \sqrt{4}$ are the distances which appear between pairs of points in T_0. Therefore $\Lambda_s(T_0) = R_s(T_0)$. But then

$$R_s(T) \geq \Lambda_s(T) = a_0 + \sum_{j=1}^{4} a_j \Gamma_j(T) > a_0 + \sum_{j=1}^{4} a_j \Gamma_j(T_0) = \Lambda_s(T_0) = R_s(T_0).$$

This completes the proof. ♠

Lemma 9.2 (A22) *For each* $i = 1, 2, 3$ *the pair* P_i *specially forces* I_i*.*

Lemma A2 is an immediate consequence of Lemma A21 and Lemma A22.

9.2 Proof of Lemma A22

Referring to Equation 89 we solve the equations

$$\Lambda_s(\sqrt{m}) = R_s(\sqrt{m}), \quad m = 2,3,4, \qquad \Lambda_s'(\sqrt{m}) = R_s'(\sqrt{m}), \quad m = 2,3. \quad (90)$$

Here f' denotes the derivative of f, a function defined on $(0,2]$. We don't need to constrain $f'(2)$. For each s this gives us a linear system with 5 variables and 5 equations. In all cases, our solutions have the following structure

$$(a_0,a_1,a_2,a_3,a_4) = M(2^{-s/2}, 3^{-s/2}, 4^{-s/2}, s2^{-s/2}, s3^{-s/2}) \qquad (91)$$

We will list M below for each of the 3 cases.

Lemma 9.3 (A221) *For each $i = 1,2,3$ the following is true. When M is defined relative to the pair P_i then the coefficients a_1,a_2,a_3,a_4 are positive functions on the interval I_i.*

We want to see that the function

$$H_s = 1 - \frac{\Lambda_s}{R_s}. \qquad (92)$$

takes its minima at $r = \sqrt{2}, \sqrt{3}$ on $(0,2]$. Differentiating with respect to $r \in (0,2]$ we have

$$H_s'(r) = r^{s-1}(s\Lambda_s(r) + r\Lambda_s'(r)). \qquad (93)$$

Using the general equation $rG_k'(r) = 2kG_k(r) - 8kG_{k-1}(r)$, we see that

$$\psi_s = s\Lambda_s(r) + r\Lambda_s'(r) \qquad (94)$$

is a polynomial in $t = 4 - r^2$.

Lemma 9.4 (A222) *For each choice P_j and each $s \in I_j$ the following is true. The function ψ_s has 4 simple roots in $[0,4]$. Two of the roots are 1 and 2 and the other two respectively lie in $(0,1)$ and $(1,2)$.*

Let us deduce Lemma A2. Our construction and Lemma A221 immediately take care of Conditions 1 and 2 of special forcing. Condition 3: The roots of ψ_s in $[0,4)$ are in bijection with the roots of H_s' in $(0,2]$ and their nature (min, max, simple) is preserved under the bijection. We check for one parameter in each of the three cases that the roots 1 and 2 correspond to local minima and the other two roots correspond to local maxima. Since these roots remain simple for all s in the relevant interval, the nature of the roots cannot change as s varies. Hence H_s has exactly 2 local minima in $(0,2]$, at $r = \sqrt{2}, \sqrt{3}$. But then $H_s \geq 0$ on $(0,2]$. This completes the proof.

9.3 A Positivity Algorithm

In our proofs of Lemmas A221 and A222 we need to deal with expressions of the following form:

$$F(s)=\sum c_i s^{t_i} b_i^{s/2}, \tag{95}$$

where $b_i, c_i \in \boldsymbol{Q}$ and $t_i \in \boldsymbol{Z}$ and $b_i > 0$. Here we explain how we deal with such expressions.

For each summand we compute a floating point value, x_i. We then consider the floor and ceiling of $2^{32}x_i$ and divide by 2^{32}. This gives us rational numbers x_{i0} and x_{i1} such that $x_{i0} \leq x_i \leq x_{i1}$. Since we don't want to trust floating point operations without proof, we formally check these inequalities with what we call the *expanding out method*.

Expanding Out Method: Suppose we want to establish an inequality like $(\frac{a}{b})^{\frac{p}{q}} < \frac{c}{d}$, where every number involved is a positive integer. This inequality is true iff $b^p c^q - a^p d^q > 0$. We check this using exact integer arithmetic. The same idea works with $(>)$ in place of $(<)$.

To check the positivity of F on some interval $[s_0, s_1]$ we produce, for each term, the 4 rationals $x_{i00}, x_{i10}, x_{i01}, x_{i01}$. Where x_{ijk} is the approximation computed with respect to s_k. We then let y_i be the minimum of these expressions. The sum $\sum y_i$ is a lower bound for Equation 95 for all $s \in [s_0, s_1]$. On any interval exponent I where we want to show that Equation 95 is positive, we pick the smallest dyadic interval $[0, 2^k]$ that contains I and then run the following subdivision algorithm.

1. Start with a list L of intervals. Initially $L = \{[0, 2^k]\}$.
2. If L is empty, then **HALT**. Otherwise let Q be the last member of L.
3. If either $Q \cap I = \emptyset$ or the method above shows that Equation 95 is positive on Q we delete Q from L and go to Step 2.
4. Otherwise we delete Q from L and append to L the 2 intervals obtained by cutting Q in half. Then we ago to to Step 2.

If this algorithm halts then it constitutes a proof that $F(s) > 0$ for all $s \in I$.

9.4 Proof of Lemma A221 and part of Lemma A222

Referring to Equation 91 we first list out the matrices in all 3 cases. For P_1 we get

$$792M = \begin{bmatrix} 0 & 0 & 792 & 0 & 0 \\ 792 & 1152 & -1944 & -54 & -288 \\ -1254 & -96 & 1350 & 87 & 376 \\ 528 & -312 & -216 & -39 & -98 \\ -66 & 48 & 18 & 6 & 10 \end{bmatrix} \tag{96}$$

For P_2 and P_3 we list $368536M$ in each case.

$$\begin{bmatrix} 0 & 0 & 268536 & 0 & 0 \\ 88440 & 503040 & -591480 & -4254 & -65728 \\ -77586 & -249648 & 327234 & 2361 & 65896 \\ 41808 & -19440 & -22368 & -2430 & -9076 \\ -402 & 264 & 138 & 33 & 68 \end{bmatrix} \tag{97}$$

$$\begin{bmatrix} 0 & 0 & 268536 & 0 & 0 & 0 \\ 982890 & 116040 & -1098930 & -52629 & -267128 & 0 \\ -91254 & -240672 & 331926 & 3483 & 68208 & 0 \\ 35778 & -15480 & -20298 & -1935 & -8056 & 0 \\ -402 & 264 & 138 & 33 & 68 & 0 \end{bmatrix} \tag{98}$$

Remark: (Auxiliary Theorem) We also list the matrix we get for the Auxiliary theorem. Here the interval is $(-2, 0)$ and the pair is (G_3, G_5).

$$144M = \begin{bmatrix} 0 & 0 & -144 & 0 & 0 \\ -312 & -96 & 408 & 24 & 80 \\ 684 & -288 & -396 & -54 & -144 \\ -402 & 264 & 138 & 33 & 68 \\ 30 & -24 & -6 & -3 & -4 \end{bmatrix}.$$

The proof of Lemmas A221 and A22 for this pair is very much like our Case 1 above. Our computer code does it rigorously.

Now we turn to the analysis of the coefficients. For Cases 2 and 3 (meaning $j = 2, 3$) we get Lemma A22 by running the positivity algorithm for a_1, a_2, a_3, a_4 on the intervals I_j. The algorithm halts and we are done. For $j = 1$ the situation is trickier because these coefficients vanish at the endpoint $s = 0$ of the interval $I_1 = (0, 6]$.

Before we launch into Case 1, we add two quantities we test, namely $\psi_s(0)$ and $\psi_s(4)$. We have

$$11\psi_s(0) = \begin{bmatrix} -88 \\ -128 \\ +216 \\ +6 \\ +32 \\ +11 \end{bmatrix} \cdot \begin{bmatrix} 2^{-s/2} \\ 3^{-s/2} \\ 4^{-s/2} \\ s2^{-s/2} \\ s3^{-s/2} \\ s4^{-s/2} \end{bmatrix}, \qquad \frac{11}{s}\psi_s(4) = \begin{bmatrix} -2112 \\ +1664 \\ +459 \\ +219 \\ 288 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 2^{-s/2} \\ 3^{-s/2} \\ 4^{-s/2} \\ s2^{-s/2} \\ s3^{-s/2} \\ s4^{-s/2} \end{bmatrix}$$

In other words, these quantities have the same form as the functions $a_j(s)$ for $j = 1, 2, 3, 4$. We run the positivity algorithm and show that all 6 quantities are positive on $[1/4, 6]$.

Now we deal with the interval $(0, 1/4]$. Note that

$$\sup_{m=2,3,4} \sup_{s\in[0,1]} \left|\frac{\partial^6}{\partial s^6} m^{-s/2}\right| < \frac{1}{8}. \tag{99}$$

All our (scaled) expressions have the form $Y \cdot V(s)$,

$$V(s) = (2^{-s/2}, 3^{-s/2}, 4^{-s/2}, s2^{-s/2}, s3^{-s/2}, s4^{-s/2}).$$

For an integer vector Y. Moreover the sum of the absolute values of the coefficients in each of the Y vectors is at most 5000. This means that, when we take the 5th order Taylor series expansion for $Y \cdot V(s)$, the error term is at most

$$5000 \times \frac{1}{8} \times \frac{1}{6!} < 1.$$

We compute each Taylor series, set all non-leading positive terms to 0, and crudely round down the other terms:

$$\begin{aligned}
792a_1(s): &\qquad 98s - 69s^2 + 0s^3 - 6s^4 + 0s^5 - 1s^6 \\
792a_2(s): &\qquad 14s - 3s^2 - 2s^3 + 0s^4 - 1s^5 - 1s^6. \\
792a_3(s): &\qquad 1s + 0s^2 - 1s^3 + 0s^4 + 0s^5 - 1s^6. \\
792a_4(s): &\qquad .03s + 0s^2 + 0s^3 - .01s^4 + 0s^5 - 1s^6. \\
11\psi_s(0): &\qquad .08s + 0s^2 - .02s^3 + 0s^4 - .01s^5 - 1s^6. \\
(11/s)\psi_s(4): &\qquad 11 + 0s + 0s^2 - 1s^3 - 1s^4 + 0s^5 - 1s^6.
\end{aligned}$$

These under-approximations are all easily seen to be positive on $(0, 1/4]$. My computer code does these calculations rigorously with interval arithmetic, but it hardly seems necessary.

9.5 Proof of Lemma A222

Case 1: In Case 1 we compute that

$$\psi_s(t) = t^6 - \frac{48}{12+s}t^5 + ... \tag{100}$$

We don't care about the other terms. Since ψ_s has degree 6 we conclude that ψ_s has at most $N = 6$ roots, counting multiplicity. By construction $H_s(\sqrt{m}) = H'_s(\sqrt{m}) = 0$ for $m = 2, 3$ and $H_s(\sqrt{4}) = 0$. This means that H_s has extrema at $r_2 = \sqrt{2}$ and $r_3 = \sqrt{3}$ and at points $r_{23} \in (\sqrt{2}, \sqrt{3})$ and $r_{34} \in (\sqrt{3}, \sqrt{4})$. Correspondingly ψ_s has roots $t_1 = 1$ and $t_2 = 2$ and $t_{01} \in (0, 1)$ and $t_{12} \in (1, 2)$. The sum of all the roots of ψ_s is $48/(12+s) < 4$. Since $t_1 + t_2 + t_{01} + t_{12} > 4$ we see that not all roots can be positive. Hence $N < 6$. Since ψ_s is positive at $t = 0, 4$ we see that N is even. Hence $N = 4$. This means that the only roots of ψ_s in $(0, 4)$ are the 4 roots we already know about. Since these roots are distinct, they are simple roots.

Cases 2 and 3: First of all, the functions H_s are the same in Cases 2 and 3. This is not just a computational accident. In both cases we are building H_s from the functions G_1, G_2, G_5, G_{10}. So, we combine Cases 2 and 3 by proving that the common polynomial ψ_s just has 4 roots for each $s \in [6, 16]$. I will describe a proof which took me quite a lot of experimentation to find. One tool I will use is *positive dominance*. This concept is discussed (with proofs) in more generality in §10.2. Here I will just explain the easy case we need in this section: A real polynomial $a_0 + a_1 t + ... a_n t^n$ is positive on $[0, 1]$ provided that the sums $a_0, a_0 + a_1, a_0 + a_1 + a_2, ..., a_0 + ... + a_n$ are all positive.

The same analysis as in Case 1 shows that ψ_s has roots at $1, 2$, and in $(0, 1)$ and in $(1, 2)$. We just want to see that there are no other roots.

We can factor ψ_s as $(t-1)(t-2)\beta_s$ where β_s is a degree 8 polynomial. Taking derivatives with respect to t, we notice that

1. $\gamma_s = 268536 \times 12^{s/2} \times (\beta''_s - \beta'_s)$ is positive for $s \times t \in [6, 16] \times [0, 4]$.
2. $-\beta'_s(0) > 0$ for all $s \in [6, 16]$.
3. $\beta'_s(4) > 0$ for all $s \in [6, 16]$.

Statement 1 shows in particular that β'_s never has a double root. This combines with Statements 2 and 3 to show that the number of roots of β'_s in $[0, 4]$

is independent of $s \in [6,16]$. We check explicitly that β_6' has only one root in $[0,4]$. Hence β_s' always has just one root. But this means that β_s has at most 2 roots in $[0,4]$. This, in turn, means that ψ_s has at most 4 roots in $[0,4]$. This completes the proof modulo the 3 statements.

Now we establish the 3 statements. We first give a formula for γ_s. Define matrices M_3, M_4, M_6 respectively as:

$$\begin{bmatrix} -546840 & -1800480 & 99720 & -397440 & -234600 & -33120 & 173880 & -22080 \\ 18366 & 17112 & 80766 & 24288 & 18630 & 11592 & 4830 & -1104 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} -345600 & -1576320 & -509760 & -760320 & -448800 & -63360 & 332640 & -42240 \\ -199296 & -698784 & 75216 & -149376 & -79960 & 5856 & 94920 & -12992 \\ 7104 & 8432 & 33960 & 11968 & 9180 & 5712 & 2380 & -544 \end{bmatrix}$$

$$\begin{bmatrix} 892440 & 3376800 & 410040 & 1157760 & 683400 & 96480 & -506520 & 64320 \\ -73350 & -246888 & -228942 & -165792 & -110370 & -41688 & 27510 & -2064 \\ 1473 & 4092 & 10557 & 5808 & 4455 & 2772 & 1155 & -264 \end{bmatrix}$$

Define 3 polynomials P_3, P_4, P_6 by the formula:

$$P_k(s,t) = (1, s, s^2) \cdot M_k \cdot (1, ..., t^7) = \sum_{i=0}^{2} \sum_{j=0}^{7} (M_k)_{ij} s^i t^j, \qquad k = 3, 4, 6. \quad (101)$$

We have

$$\gamma = P_3 3^{s/2} + P_4 4^{s/2} + P_6 6^{s/2}. \quad (102)$$

To check the positivity of γ_s we check that each of the 16 functions

$$\gamma_s(v/4 + 1/4) = a_{v,0} + a_{v,1}t + ... a_{v,7}t^7 \quad (103)$$

satisfies the following condition: $A_{v,k} = a_{v,0} + ... + a_{v,k}$ is positive for all $k = 0, ..., 7$ and all $s \in [6,16]$. This shows that the corresponding polynomial is positive on $[0,1]$.

For each $v = 0, ..., 15$ and each $k = 0, ..., 7$ we have a 3×3 integer matrix $\mu_{v,k}$ such that

$$A_{v,k} = (1, s, s^2) \cdot \mu_{v,t} \cdot (3^{s/2}, 4^{s/2}, 6^{s,2}). \quad (104)$$

This gives 128 matrices to check. We get two more such matrices from the conditions $-\beta_s'(0) > 0$ and $\beta_s'(4) > 0$. All in all, we have to check that 130 expressions of the form in Equation 104 are positive for $s \in [6,16]$. These expressions are all special cases of Equation 95, and we use the method discussed above to show positivity in all 130 cases. The program runs in several hours.

10 Symmetrization: Preliminaries

In this part of the monograph we prove Lemma B. In this preliminary chapter we discuss a few useful lemmas.

10.1 Exponential Sums

We begin with two easy and well-known lemmas about exponential sums. The first is an exercise with Lagrange multipliers.

Lemma 10.1 (Convexity) *Suppose that $\alpha, \beta, \gamma \geq 0$ have the property that $\alpha + \beta \geq 2\gamma$. Then $\alpha^s + \beta^s \geq 2\gamma^s$ for all $s > 1$, with equality iff $\alpha = \beta = \gamma$.*

Lemma 10.2 (Descartes) *Let $0 < r_1 \leq r_1 ... \leq r_n < 1$ be a sequence of positive numbers. Let $c_1, ..., c_n$ be a sequence of nonzero numbers. Define*

$$E(s) = \sum_{i=1}^{n} c_i \; r_i^s. \tag{105}$$

Let K denote the number of sign changes in the sequence $c_1, ..., c_n$. Then E changes sign at most K times on $\boldsymbol{R}$.

Proof: Suppose we have a counterexample. By continuity, perturbation, and taking mth roots, it suffices to consider a counterexample of the form $\sum c_i t^{e_i}$ where $t = r^s$ and $r \in (0,1)$ and $e_1 > ... > e_n \in \boldsymbol{N}$. As s ranges in r, the variable t ranges in $(0, \infty)$. But $P(t)$ changes sign at most K times on $(0, \infty)$ by Descartes' Rule of Signs. This gives us a contradiction. ♠

10.2 Positive Dominance

See [**S2**] and [**S3**] for more details about the material here. Let $G \in \boldsymbol{R}[x_1, ..., x_n]$ be a multivariable polynomial:

$$G = \sum_I c_I X^I, \qquad X^I = \prod_{i=1}^{n} x_i^{I_i}. \tag{106}$$

Given two multi-indices I and J, we write $I \preceq J$ if $I_i \leq J_i$ for all i. Define

$$G_J = \sum_{I \preceq J} c_I, \qquad G_\infty = \sum_I c_I. \tag{107}$$

We call G *weak positive dominant* (WPD) if $G_J \geq 0$ for all J and $G_\infty > 0$. We call G *positive dominant* if $G_J > 0$ for all J.

Lemma 10.3 (Weak Positive Dominance) *If G is weak positive dominant then $G > 0$ on $(0,1]^n$. If G is positive dominant then $G > 0$ on $[0,1]^n$.*

Proof: We prove the first statement. The second one has almost the same proof. Suppose $n = 1$. Let $P(x) = a_0 + a_1 x + ...$ Let $A_i = a_0 + ... + a_i$. The proof goes by induction on the degree of P. The case $\deg(P) = 0$ is obvious. Let $x \in (0,1]$. We have

$$P(x) = a_0 + a_1 x + x_2 x^2 + \cdots + a_n x^n \geq$$

$$x(A_1 + a_2 x + a_3 x^2 + \cdots a_n x^{n-1}) = xQ(x) > 0$$

Here $Q(x)$ is WPD and has degree $n-1$.

Now we consider the general case. We write

$$P = f_0 + f_1 x_k + ... + f_m x_k^m, \qquad f_j \in \boldsymbol{R}[x_1, ..., x_{n-1}]. \tag{108}$$

Since P is WBP so are the functions $P_j = f_0 + ... + f_j$. By induction on the number of variables, $P_j > 0$ on $(0,1]^{n-1}$. But then, when we arbitrarily set the first $n-1$ variables to values in $(0,1)$, the resulting polynomial in x_n is WPD. By the $n = 1$ case, this polynomial is positive for all $x_n \in (0,1]$. ♠

Polynomial Subdivision: Let $P \in \boldsymbol{R}[x_1, ..., x_n]$ as above. For any x_j and $k \in \{0,1\}$ we define

$$S_{x_j,k}(P)(x_1, ..., x_n) = P(x_1, ..., x_{j-1}, x_j^*, x_{j+1}, ..., x_n), \qquad x_j^* = \frac{k}{2} + \frac{x_j}{2}. \tag{109}$$

If $S_{x_j,k}(P) > 0$ on $(0,1]^n$ for $k = 0,1$ then we also have $P > 0$ on $(0,1]^n$.

Positive Numerator Selection: If $f = f_1/f_2$ is a bounded rational function on $[0,1]^n$, written in so that f_1, f_2 have no common factors, we always choose f_2 so that $f_2(1, ..., 1) > 0$. If we then show, one way or another, that $f_1 > 0$ on $(0,1]^n$ we can conclude that $f_2 > 0$ on $(0,1]^n$ as well. The point is that f_2 cannot change sign because then f blows up. But then we can conclude that $f > 0$ on $(0,1]^n$. We write $\text{num}_+(f) = f_1$.

11 Symmetrization: Proof of Lemma B

11.1 Reduction to Smaller Steps

Now we define the domains involved in our proof. Recall that the domain Υ is defined in §3.1 and shown in Figure 3.1. In this section we describe the domains that play a role in the proof of Lemma B. The various transformations we make start in Υ but then move us into slightly different domains.

Rotation: We let (p_1', p_2', p_3', p_4') be the planar configuration which is obtained by rotating X about the origin so that p_0' and p_2' lie on the same horizontal line, with p_0' lying on the right. We call this operation *rotation*. Rotation does not quite map Υ into itself. To find a suitable image, let Υ' denote the domain of avatars p_0', p_1', p_2', p_3' such that

1. $\|p_0'\| \geq \|p_k'\|$ for $k = 1, 2, 3$.
2. $512p_0' \in [432, 498] \times [-16, 16]$. (Compare $[433, 498] \times [0, 0]$.)
3. $512p_1' \in [-32, 32] \times [-465, -348]$. (Compare $[-16, 16] \times [-464, -349]$.)
4. $512p_2' \in [-498, -400] \times [-16, 16]$. (Compare $[-498, -400] \times [0, 24]$.)
5. $512p_3' \in [-32, 32] \times [348, 465]$. (Compare $[-16, 16] \times [349, 464]$.)
6. $p_{02}' = p_{22}'$. (Compare $p_{02} = 0$.)

The comparisons are with Υ.

Lemma 11.1 (B21) *Rotation gives a map from Υ to Υ' where*

Horizontal Symmetrization: Given an avatar $X' = (p_0', p_1', p_2', p_3') \in \Upsilon'$, there is a unique configuration $X'' = (p_0'', p_1'', p_2'', p_3'')$, invariant under under reflection in the y-axis, such that p_j' and p_j'' lie on the same horizontal line for $j = 0, 1, 2, 3$ and $\|p_0'' - p_2''\| = \|p_0' - p_2'\|$. We call this *horizontal symmetrization*. In a straightforward way we see that horizontal symmetrization maps Υ' into Υ'', the set of avatars $p_0'', p_1'', p_2'', p_3''$ such that

1. $512p_0'' \in [416, 498] \times [-16, 16]$ and $(p_{21}'', p_{22}'') = (-p_{01}'', p_{02}'')$.
2. $-512p_1, 512p_3'' \in [0, 0] \times [348, 465]$.

Vertical Symmetrization: Given a configuration $X'' = (p''_0, p''_1, p''_2, p''_3) \in \Upsilon''$ there is a unique configuration $X''' = (p'''_0, p'''_1, p'''_2, p'''_3) \in \mathbf{K4}$ such that p''_j and p'''_j lie on the same vertical line for $j = 0, 1, 2, 3$. The configuration X''' coincides with the configuration X^* defined in Lemma B. We call this operation *vertical symmetrization*.

In summary (and using obvious abbreviations) we have

$$\Upsilon \quad \overrightarrow{\text{Rot}} \quad \Upsilon' \quad \overrightarrow{\text{HS}} \quad \Upsilon'' \quad \overrightarrow{\text{VS}} \quad \mathbf{K_4}.$$

Symmetrization, as an operation on Υ', is the composition of vertical and horizontal symmetrization.

Each avatar corresponds to a 5-point configuration on S^2 *via* stereographic projection. The energy of the 5 point configuration involves 10 pairs of points. A typical term is:

$$R_s(p_i, p_j) = \frac{1}{\|\Sigma^{-1}(p_i) - \Sigma^{-1}(p_j)\|^s}. \qquad (110)$$

Given a list L of pairs of points in the set $\{0, 1, 2, 3, 4\}$ we define $\mathcal{E}_s(P, L)$ to be the sum of the R_s-potentials just over the pairs in L. Thus, for instance

$$L = \{(0,2), (0,4), (2,4)\} \implies \mathcal{E}_s(P, L) = R_s(p_0, p_2) + R_s(p_0, p_4) + R_s(p_2, p_4).$$

We call the subset L *good* for the parameter s, and with respect to one of the operations, if the operation does not increase the value of $\mathcal{E}_s(P, L)$. We call L *great* if the operation strictly lower $\mathcal{E}_s(P, L)$ unless the operation fixes P. When we make this definition we mean to take the appropriate domains.

Lemma 11.2 (B2) *The lists* $\{(0,2), (0,4), (2,4)\}$ *and* $\{(1,3), (1,4), (3,4)\}$ *are both great for all* $s \geq 2$ *and with respect to symmetrization.*

Lemma 11.3 (B3) *The lists* $\{(0,1), (1,2)\}$ *and* $\{(0,3), (3,2)\}$ *are both good for all* $s \geq 2$ *and with respect to horizontal symmetrization.*

Lemma 11.4 (B4) *The lists* $\{(0,1), (0,3)\}$ *and* $\{(2,1), (2,3)\}$ *are both good for all* $s \geq 12$ *and with respect to vertical symmetrization.*

Lemma B follows from Lemma B1 (§11.2), Lemma B1 (§11.3), Lemma B3 (§11.4) and Lemma B4 (§11.5). Lemma B2 is pretty robust and Lemma B4 is very delicate. Lemma B3 is in the middle.

11.2 Proof of Lemma B1

The proof of Lemma B1 is a tedious exercise in trigonometry and arithmetic.

Let $P \in \Upsilon$ and let P' be the rotation of P. Rotation about the origin does not change the norms, so P' satisfies Condition 1. Moreover, Condition 6 holds by construction. Now we verify the other properties.

Let ρ_θ denote the counterclockwise rotation through the angle θ. Since p_0 lies on the x axis and p_2 lies on or above it, we have to rotate by a small amount counterclockwise to get p'_0 and p'_2 on the same horizontal line. That is, the rotation moves the right point up and the left one down. Hence $\theta \geq 0$. This angle is maximized when p_0 is an endpoint of its segment of constraint and p_2 is one of the two upper vertices of rectangle of constaint. Not thinking too hard which of the 4 possibilities actually realizes the max, we check for all 4 pairs (p_0, p_2) that the second coordinate of $\rho_{1/34}(p_0)$ is larger than the second coordinate of $\rho_{1/34}(p_0)$. From this we conclude that $\theta < 1/34$. This yields

$$512\cos(\theta) \in [0,1], \qquad 512\sin(\theta) \in [0,16]. \tag{111}$$

From Equation 111, the map $512p_0 \to 512p'_0$ changes the first coordinate by $512\delta_{01} \in [0,16]$ and $512\delta_{02} \in [-1,0]$. This gives (something stronger than) Condition 2 for Υ'. (We are symmetrizing Υ' for the purposes of later steps in the arguement and that is why we have weakened the conditions needed for inclusion.) At the same time, we have $p'_{21} = p'_{01}$ and the change $512p_2 \to 512p'_2$ changes the second coordinate by $512\delta_{21} \in [0,1]$. This gives Condition 4 for Υ' once we observe that $|p'_{21}| \leq |p'_{01}|$.

For Condition 3 we just have to check (using the same notation) that $512\delta_{11} \in [0,16]$ and $512\delta_{12} \in [-1,1]$. The first bound comes from the inequality $512\sin(\theta) < 16$. For the second bound we note that the angle that p_1 makes with the y-axis is maximized when p_1 is at the corners of its constraints in Υ. That is,

$$p_1 = \Big(\frac{\pm 16}{512}, \frac{349}{512}\Big).$$

Since $\tan(1/21) > 16/349$ we conclude that this angle is at most $1/21$. Hence

$$|512\delta_{12}| \leq \max_{|x|\leq 1/21} \Big|\cos\Big(x + \frac{1}{34}\Big) - \cos(x)\Big| < 1.$$

This gives Condition 3. The same argument gives Condition 5.

11.3 Proof of Lemma B2

Let (u,v) stand for either $(0,2)$ or $(1,3)$. Also all rotations we consider fix the origin. For the points associated with $\{(u,v),(u,4),(v,4)\}$ our symmetrization operation on Υ' is a special case of the following general operation.

1. Start with p_u, p_v so that $\|p_u\|, \|p_v\| < 1$ and $2d := \|p_u - p_v\| > 2\sqrt{3}/3$.
2. Replace p_u, p_v with points $q_u = (-d,0)$ and $q_v = (d,0)$.
3. Let $\lambda \in (0,1)$ satisfy $\lambda d > \sqrt{3}/3$. Let $r_u = (-\lambda d, 0)$ and $r_v = (\lambda d, 0)$.
4. Let p_u^*, p_v^* be respective images of r_u, r_v under any rotation.

The operaton is $(p_u, p_v) \to (p_u^*, p_v^*)$. Lemma B2 is implied by:

$$\begin{aligned}\|\widehat{r}_u - \widehat{r}_v\|^{-2} + \|\widehat{r}_u - (0,0,1)\|^{-2} + \|\widehat{r}_v - (0,0,1)\|^{-2} \leq \\ \|\widehat{p}_u - \widehat{p}_v\|^{-2} + \|\widehat{p}_u - (0,0,1)\|^{-2} + \|\widehat{p}_v - (0,0,1)\|^{-2}\end{aligned} \qquad (112)$$

with equality if and only if $r_u = p_u$ and $r_v = p_v$ up to a rotation.

Lemma 11.5 (B21) *Let $s \geq 2$ and*

$$A_s = \|\widehat{p}_u - \widehat{p}_v\|^{-s} - \|\widehat{q}_u - \widehat{q}_v\|^{-s},$$

$$B_s = \|\widehat{p}_u - (0,0,1)\|^{-2} + \|\widehat{p}_v - (0,0,1)\|^{-2} - \|\widehat{q}_u - (0,0,1)\|^{-2} - \|\widehat{q}_v - (0,0,1)\|^{-2}.$$

Then $A_s, B_s \geq 0$, with equality iff $p_u = q_u$ and $p_v = q_v$ up to a rotation.

Proof: The case $s = 2$ of this result combines with the Convexity Lemma to get the case $s > 2$. So, we take $s = 2$. We rotate so that

$$p_u = (-x+h, y), \quad p_v = (x+h, y), \quad q_u = (-x, 0), \quad q_v = (x, 0). \qquad (113)$$

We compute

$$A_2 = \frac{h^4 + y^2(2 + 2x^2 + y^2) + h^2(2 - 2x^2 + 2y^2)}{16x^2}, \qquad B_2 = \frac{y^2 + h^2}{2}. \qquad (114)$$

Since $x \in (0,1)$ we have $A_2 > 0$ unless $h = y = 0$. Likewise we have $B_2 > 0$ unless $h = y = 0$. ♠

Lemma B21 implies that Step 2 of our construction above decreases energy unless it fixes the points up to rotation. We just need to see that Step 3 of the construction does not increase the energy. So far we have not used the condition $d > \sqrt{3}/3$ but now we will. The following result is perhaps a standard result for 3-point energy minimization on a circle.

Lemma 11.6 (B22) *As long as $s \geq 2$ we have*

$$\|\widehat{r}_u - \widehat{r}_v\|^{-s} + \|\widehat{r}_u - (0,0,1)\|^{-s} + \|\widehat{r}_v - (0,0,1)\|^{-s} \leq$$
$$\|\widehat{q}_u - \widehat{p}_v\|^{-s} + \|\widehat{q}_u - (0,0,1)\|^{-s} + \|\widehat{q}_v - (0,0,1)\|^{-s} \tag{115}$$

Proof: We prove this equation under more general assumptions. Define

$$F_s(a_u, a_v) = \|\widehat{\zeta}_u - \widehat{\zeta}_v\|^{-s} + \|\widehat{\zeta}_u - (0,0,1)\|^{-s} + \|\widehat{\zeta}_v - (0,0,1)\|^{-s}, \tag{116}$$

Where

$$\zeta_u = (-\sqrt{3}/3 - a_u, 0), \quad \zeta_v = (\sqrt{3}/3 + a_v). \tag{117}$$

Our assumption $d > \sqrt{3}/3$ lets us take $a_u, a_v > 0$. Next for $0 < b_u < a_u$ and $0 < b_v \leq a_v$ define

$$E(s) = F_s(a_u, a_v) - F_s(b_u, b_v) \tag{118}$$

We will show that $E(s) \geq 0$ when $s \geq 2$. It suffices to prove this result in the intermediate case when $a_u = b_u$ or $a_v = b_v$ because then we can apply the intermediate result twice to get the general case. Without loss of generality we consider the case when $a_v = b_v$ and $b_u < a_u$. We think of ζ_u as moving from its old location defined by a_u inward to its new location defined by b_u.

With the file `LemmaB22.m` we compute that

$$\frac{\partial F_2}{\partial a_u}\bigg|_{(a_u,a_v)}, \qquad -\frac{\partial F_{-2}}{\partial a_u}\bigg|_{(a_u,a_v)}$$

are both rational functions of a_u, a_v with all positive coefficients. Hence $E(2) > 0$ and $E(-2) < 0$.

When $a_u = a_v = 0$ the points $\widehat{\zeta}_u, \widehat{\zeta}_v$ and $(0,0,1)$ make an equilateral triangle on a great circle. Hence, when $a_u, a_v, b_u, b_v > 0$ the point $\widehat{\zeta}_u$ is closer to $(0,0,1)$ than it is to $\widehat{\zeta}_v$ both in its old location and in its new location. The inward motion of the point ζ_u increases the shorter (corresponding spherical) distance and decreases the longer (corresponding spherical) distance. More to the point, our move decreases the longer inverse-distance and increases the shorter inverse-distance. Thus the sign sequence (§10.1) for $E(s)$ is $+, -.-, +$. By Descartes' Lemma, $E(s)$ changes sign at most twice and also $E(s) > 0$ when $|s|$ is sufficiently large.

Since $E(-2) < 0$ as see that E changes sign on $(-\infty, -2)$. If E changes sign on $(2, \infty)$ then in fact E changes sign twice because it starts and ends positive on this interval. But then E changes sign 3 times, a contradiction. Hence $E(s) \geq 0$ for $s \geq 2$. ♠

11.4 Proof of Lemma B3

By symmetry, it suffices to prove Lemma B3 for the list $\{(0,1),(1,2)\}$. Let D denote the set of triples of points $(q_0,q_1,q_2)\in(\boldsymbol{R}^2)^3$ such that there is some q_3 such that $q_0,q_1,q_2,q_3\in\Upsilon'$. The symmetrization operation is given by $(q_0,q_1,q_2)\to(q_1',q_2',q_3')$, where

$$q_0'=\Big(\frac{q_{01}-q_{21}}{2},q_{02}\Big),\qquad q_1'=(0,q_{21}),\qquad q_2'=\Big(\frac{q_{21}-q_{01}}{2},q_{22}\Big), \tag{119}$$

Note that $\|q_0'-q_1'\|=\|q_2'-q_1'\|$. Therefore, by the Convexity Lemma, it suffices to prove that $\{(0,1),(1,2)\}$ is good for the parameter $s=2$.

We define

$$[u,v]t=u(1-t)+vt. \tag{120}$$

The map $t\to[u,v]t$ maps $[0,1]$ to $[u,v]$.

For all 4 choices of signs we define $\phi_{\pm,\pm}:[0,1]^5\to(\boldsymbol{R}^2)^3$ as follows:

$$\phi_{\pm,\pm}(a,b,c,d,e)=q_0(a,d,\pm b),q_1(\pm e,c),q_2(a,d,\pm b), \tag{121}$$

where

$$512q_0(a,d,\pm b)=([416,498]a+49e,\pm 16b).$$
$$512q_1(\pm d,c)=(\pm 32d,[348+465]c)$$
$$512q_2(a,d,\pm b)=([-416,-498]a+49e,\pm 16b).$$

In these coordinates, horizontal symmetrization is the map $(a,b,c,d,e)\to(a,b,c,0,0)$.

Lemma 11.7 (B31) *We have*

$$D\subset\phi_{+,+}([0,1]^5)\cup\phi_{+,-}([0,1]^5)\cup\phi_{-,+}([0,1]^5)\cup\phi_{-.-}([0,1]^5).$$

Proof: Let D_{ij} denote the set of possible coordinates q_{ij} that can arise for points in D. This, for instance $D_{01}=[-16,16]/512$. Let D_{ij}^* denote the set of possible coordinates q_{ij} that can arise from the union of our parametrizations. By construction $D_{i2}\subset D_{i2}^*$ for $i=0,1,2$ and $D_{11}\subset D_{11}^*$.

Remembering that we have $q_{01}\geq|q_{21}|$, we see that the set of points pairs (q_{01},q_{21}) satisfying all the conditions for inclusion in D lies in the triangle X with vertices $(498,-498)$ and $(498,-400)$ and $(432,-400)$. At the same time, the set of pairs $(512)(p^*_{01}, p^*_{21})$ that we can reach with our parametrization is the rectangle X^* with vertices

$$(498, -498), \quad (416, -416), \quad (498, -498) + (49, 49), \quad (416, -416) + (49, 49).$$

We have $X \subset X^*$ because

$$(432, -400) = (416, -416) + (16, 16), \qquad (498, -400) = (449, -449) + (49, 49).$$

This completes the proof. ♠

Using our coordinates above, we define

$$F_{\pm,\pm}(a, b, c, d, e) = \|\widehat{p}_0 - \widehat{p}_1\|^{-2} + \|\widehat{p}_2 - \widehat{p}_1\|^{-2},$$

$$\Phi_{\pm,\pm}(a, b, c, d, e) = \text{num}_+(F_{\pm,\pm}(a, b, c, d, e) - F_{\pm,\pm}(a, b, c, 0, 0)). \tag{122}$$

Lemma 11.8 (B32) *For any sign choice $\Phi_{\pm,\pm} > 0$ on $(0, 1)^5$.*

Proof: Let Φ be any of the 4 polynomials. The file `LemmaB32.m` computes that

- F and Φ_d and Φ_e are zero when $d = e = 0$.
- $\Phi_d + 2\Phi_e$ and Φ_{dd} and Φ_{ee} are weak positive dominant and hence non-negative on $[0, 1]^5$.

Let $Q_d \subset [0, 1]^5$ be the sub-cube where $d = 0$. Let $\phi(d)$ be the restriction of Φ to a line segment which starts at some point $(a, b, c, 0, 0)$ and moves parallel to $(0, 0, 0, 1, 0)$. By Lemma B321 we have $\phi(0) = \phi'(0)$ and also $\phi''(d) \geq 0$. Hence $\phi(d) \geq 0$ for $d \geq 0$. Hence $\Phi \geq 0$ on Q_d. A similar argument shows that likewise $\Phi \geq 0$ on Q_e. Any point in $(0, 1)^5$ can be joined to a point in $Q_d \cup Q_e$ by a line segment L which is parallel to the vector $(0, 0, 0, 1, 2)$. By Lemma B3121, Φ increases along such a line segment as we move out of $Q_d \cup Q_e$. Hence $\Phi \geq 0$ on $[0, 1]^5$. ♠

Lemma B312 implies $F_\pm(a, b, c, d, e) \geq F_\pm(a, b, c, 0, 0)$. See §10.2. Lemma B3 thus follows from Lemma B31 and Lemma B32.

11.5 Proof of Lemma B4

By symmetry, it suffices to prove that $\{(0,1),(3,1)\}$ is good for all $s \geq 12$. For ease of notation set $q_k = p''_k$. Let D be the set of configurations (q_0,q_1,q_3) such that $(q_0,q_1,q_2,q_3) \in \Upsilon''$ when q_2 is the reflection of q_0 in the y-axis. By symmetry, it suffices to treat the case when $q_{02} \geq 0$. We let $D_\pm \subset D$ denote those configurations with $\pm(q_{12}+q_{32}) \geq 0$. Obviously $D = D_+ \cup D_-$.

As in Equation 120, let $[u,v]t = u(1-t)+vt$. Similar to the horizontal case, we define $\phi_\pm(a,b,c,d) = (q_0(b,d), q_1(a,c), q_3(a,c))$, where

$$512q_0(b,d) = ([416,498]b, 16d).$$

$$512q_1(a,c) = (0, -[348,465]a \pm 59c).$$

$$512q_3(a,c) = (0, +[348,465]a \pm 59c).$$

In these coordinates, the symmetrization operation is $(a,b,c,d) \to (a,b,0,0)$.

Lemma 11.9 (B41) $D_\pm \subset \phi_\pm([0,1]^4)$.

Proof: This is just like the proof of Lemma B21. The only non-obvious point is why every pair (p_{12}, p_{32}) is reached by the map $\phi_\pm$. The essential point is that for configurations in $D_\pm$ we have $512|p_{12}+p_{32}| \leq 2 \times 59$. ♠

Following the same idea as in the proof of Lemma B3, we define

$$F_{s,\pm}(a,b,c,d) = \|\Sigma^{-1}(q_0) - \Sigma^{-1}(q_1)\|^{-s} + \|\Sigma^{-1}(q_0) - \Sigma^{-1}(q_3)\|^{-s}, \quad (123)$$

$$\Phi_{s,\pm}(a,b,c,d) = \mathrm{num}_+(F_{s,\pm}(a,b,c,d) - F_{s,\pm}(a,b,0,0)). \quad (124)$$

The points on the right side of Equation 123 are coordinatized by the map $\phi_\pm$. We can finish the proof by showing that $\phi_{2,+} \geq 0$ and $\phi_{12,-} \geq 0$ on $[0,1]^4$. The Convexity Lemma then takes care of all exponents greater than 2 on D_+ and all exponents greater than 12 on D_-.

Lemma 11.10 (B42) $\Phi_{2,+} \geq 0$ *on* $[0,1]^4$.

Proof: Let $\Phi = \Phi_{2,+}$. Let $\Phi|_{c=0}$ denote the polynomial we get by setting $c=0$. Etc. Let $\Phi_c = \partial\Phi/\partial c$, etc. The Mathematica file `LemmaB42.m` computes that $\Phi|_{c=0}$ and $\Phi|_{d=0}$ and $\Phi_c + \Phi_d$ are weak positive dominant. Hence $\Phi \geq 0$ when $c=0$ or $d=0$ and the directional derivative of Φ in the direction $(0,0,1,1)$ is non-negative. This suffices to show that $\Phi \geq 0$ on $[0,1]^4$. ♠

Lemma 11.11 (B43) $\Phi_{12,-} \geq 0$ *on* $[0,1]^4$.

Proof: The file `LemmaB43.m` has the calculations for our argument. Let $\Phi = \Phi_{12,-}$. This monster has 102218 terms and we simplify it carefully.

Let M denote the maximum coefficient of Φ. We let Φ^* be the polynomial we get by taking each coefficient of c of Φ and replacing it with the greatest integer less than $10^{10}c/M$. This has the effect of killing off about half the terms of Φ, namely the positive terms that are less than $10^{-10}M$. The "small" negative coefficients are changed to -1. The polynomial $10^{10}\Phi - M\Phi^*$ has all non-negative coefficients. Hence, if $\Phi^* \geq 0$ on $[0,1]^4$ so is $M\Phi \geq 0$ and so is $10^{10}\Phi$ and finally so is Φ. Now Φ^* has 37760 monomials in which the coefficient is -1. We check that each such monomial is divisible by one of c^2 or d^2 or cd. We therefore define

$$\Psi = \Phi^{**} - 37760(c^2 + d^2 + cd),$$

where Φ^{**} is obtained from Φ^* by setting all the (-1) monomials to 0. We have $\Psi \leq \Phi^*$ on $[0,1]^4$. Hence, if Ψ is non-negative on $[0,1]^4$ then so is Φ. We have reduced the problem to showing that $\Psi \geq 0$ on $[0,1]^4$. The polynomial Ψ has 5743 terms, which is more manageable.

Again we let $F_a = \partial F/\partial a$, etc. We check that Ψ_{aaa} is weak positive dominant and hence non-negative on $[0,1]^4$. This massive calculation reduces us to showing that the restrictions $\Psi|_{a=0}$ and $\Psi_a|_{a=0}$ and $\Psi_{aa}|_{a=0}$ are all non-negative on $[0,1]^3$. Letting F be any of these 3 functions, we consider

$$F|_{c=0}, \qquad F|_{d=0} \qquad 4F_c + F_d, \tag{125}$$

We show that all three functions are weak positive dominant for $\Psi_a|_{a=0}$ and $\Psi_{aa}|_{a=0}$. This shows that $\Psi_a|_{a=0}$ and $\Psi_{aa}|_{a=0}$ are non-negative on $[0,1]^3$. Concerning the choice $F = \Psi|_{a=0}$, all that remains is showing (in some other way) that $G = 4F_c + F_d \geq 0$.

We check that G_d is weak positive dominant and hence non-negative on $[0,1]^3$. This reduces us to showing that $H = G|_{d=0}$ is non-negative on $[0,1]^2$. Here H is a 2-variable polynomial in b, c. We check that the two subdivisions $S_{b,0}(H)$ and $S_{b,1}(H)$ are weak positive dominant. This proves that H is non-negative on $[0,1]^2$. ♠

Remark: The proof of Lemma B43 is pretty crazy. Along with my computer code, I include a PDF document having an alternate (but longer) proof that involves much smaller polynomials.

12 Endgame: Proof of Lemma C1

The proof here uses the material in §10.

We also recall some notation that we use repeatedly. Σ^{-1} is inverse stereographic projection, as in Equation 8. We often write $\widehat{p} = \Sigma^{-1}(p)$ for $p \in \boldsymbol{R}^2$. The sets Ψ_4 and $\widehat{\Psi}_4$ are defined by

$$64\Psi_4 = [43, 64]^2, \qquad 64\widehat{\Psi}_4 = [55, 56]^2. \tag{126}$$

The sets Ψ_8 and $\widehat{\Psi}_8$ respectively are the main diagonals of Ψ_4 and $\widehat{\Psi}_4$. A point (x, y) in these domains defines the avatar with $-p_2 = p_0 = (x, 0)$ and $-p_1 = p_3 = (0, y)$; we define $\mathcal{E}_s(x, y)$ to be the R_s potential of the corresponding configuration in S^2.

12.1 Reduction to Two Halves

We have the symmetrization operation $\sigma : \widehat{\Psi}_4 \to \Psi_8$ given by

$$\sigma(x, y) = (z, z), \qquad z = \frac{x + y + (x - y)^2}{2}. \tag{127}$$

Lemma C1 says $\mathcal{E}_s \circ \sigma \leq \mathcal{E}_s$ on $\widehat{\Psi}_4$ for all $s \in [14, 16]$, with equality iff $x = y$. Using notation special to this chapter, we write $\mathcal{E}_s(x, y) = G_s(x, y) + H_s(x, y)$, where

$$G_s(x, y) = \|\widehat{p}_0, \widehat{p}_2\|^{-s} + \|\widehat{p}_1, \widehat{p}_3\|^{-s},$$
$$H_s(x, y) = 2\|\widehat{p}_0, (0, 0, 1)\|^{-s} + 2\|\widehat{p}_1, (0, 0, 1)\|^{-s} + 4\|\widehat{p}_0, \widehat{p}_1\|^{-s}. \tag{128}$$

Lemma C1 follows immediately from Lemmas C11 and C12.

Lemma 12.1 (C11) *$G_s(x, y) \geq G_s(z, z)$ for $s \geq 2$ and $(x, y) \in \widehat{\Psi}_4$. When $s > 2$ we get equality if and only if $x = y$.*

Proof: By the Convexity Lemma from §10.1 it suffices for us to prove that $G_2(x, y) \geq G_2(z, z)$ for all $x, y \in \Psi_4$. Let $\phi : [0, 1]^2 \to \widehat{\Psi}_4$ be the affine isomorphism whose linear part is a positive diagonal matrix. Define

$$\Phi = \text{num}_+(G_2 \circ \phi - G_2 \circ \sigma \circ \phi). \tag{129}$$

The file `LemmaC11.m` computes that $\Phi(a, b) = (a - b)^2 \Phi^*$, where Φ^* is weak positive dominant. Hence $\Phi^* > 0$ on $(0, 1)^2$. This does it. ♠

Lemma 12.2 (C12) *$H_s(x, y) \geq H_s(z, z)$ for $s \in [14, 16]$ and $(x, y) \in \widehat{\Psi}_4$.*

12.2 Proof of Lemma C12

Lemma 12.3 (C121) *The following is true:*

1. *$H_2(x,y) \leq H_2(z,z)$ for all $(x,y) \in \widehat{\Psi}_4$.*
2. *$H_{14}(x,y) \geq H_{14}(z,z)$ for all $(x,y) \in \widehat{\Psi}_4$.*
3. *$H_{16}(x,y) \geq H_{16}(z,z)$ for all $(x,y) \in \widehat{\Psi}_4$.*

We get strict inequalities for points in the interior of $\widehat{\Psi}_4 - \Psi_8$.

Proof: For integers $k = 2, 14, 16$ define

$$\Phi_k = \mathrm{num}_+(H_k \circ \phi - H_k \circ \sigma \circ \phi). \tag{130}$$

An algebraic miracle happens. The file `LemmaC121.m` computes that

1. $-\Phi_2(a,b) = (a-b)^2 \Phi_2^*(a,b)$ and Φ_2^* is weak positive dominant.
2. $\Phi_{14}(a,b) = (a-b)^2 \Phi_{14}^*(a,b)$ and Φ_{14}^* is weak positive dominant.
3. $\Phi_{16}(a,b) = (a-b)^2 \Phi_{16}^*(a,b)$ and Φ_{16}^* is weak positive dominant.

This does it. ♠

Now suppose there is some $(x,y) \in \widehat{\Psi}_4 - \Psi_8$ and some $s_0 \in (14,16)$ such that $H_{s_0}(x,y) < H_{s_0}(z,z)$. Perturbing, we can assume that (x,y) lies in the interior of $\widehat{\Psi}_4 - \Psi_8$. Let p_0, p_1, p_2, p_3 and p_0', p_1', p_2', p_3' respectively be the configurations corresponding to (x,y) and (z,z). Define

1. $r_{01} = \|\Sigma^{-1}(p_0) - \Sigma^{-1}(p_1)\|^{-1}$.
2. $r_0 = \|\Sigma^{-1}(p_0) - (0,0,1)\|^{-1}$ and $r_1 = \|\Sigma^{-1}(p_1) - (0,0,1)\|^{-1}$.
3. $r_{01}' = \|\Sigma^{-1}(p_0') - \Sigma^{-1}(p_1')\|^{-1}$.
4. $r_0' = \|\Sigma^{-1}(p_0') - (0,0,1)\|^{-1} = \|\Sigma^{-1}(p_1') - (0,0,1)\|^{-1}$.

Replacing (x,y) by (y,x) if necessary, we arrange that $r_0 < r_1$. The purpose of the next result is to get us into shape to apply Descartes' Lemma from §10.1.

Lemma 12.4 (C122) *$r_0, r_1, r_0' < 1/\sqrt{2} < r_{01}, r_{01}'$ and $r_{01} < r_{01}'$.*

Proof: We have $x, y, z \in (0,1)$. We compute

$$(1/2) - r_0^2 = \frac{1-x^2}{4} > 0, \quad (1/2) - r_1^2 = \frac{1-y^2}{4} > 0, \quad (1/2) - (r_0')^2 = \frac{1-z^2}{4} > 0,$$

$$(r_{01})^2 - (1/2) = \frac{(1-x^2)(1-y^2)}{4(x^2+y^2)} > 0, \qquad (r_{01}')^2 - (1/2) = \frac{(1-z^2)^2}{8z^2} > 0.$$

This proves the first statement.

For the second statement, we define $J = \|\widehat{p}_0 - \widehat{p}_1\|^{-2} = r_{01}^2$ and then define Φ in terms of J just as in Equation 130. The file LemmaC122.m computes that $\Phi(a,b) = -(a-b)^2\Phi^*(a,b)$ where Φ^* is weak positive dominant. Hence $\Phi^* > 0$ on $(0,1)^2$. Hence $\Phi < 0$ on $(0,1)^2$. Hence $J(z,z) > J(x,y)$. But this implies that $r_{01} < r_{01}'$. ♠

We now deduce Lemma C12 from Lemma C121 and Lemma C122. We fix (x,y) and $(z,z) = \sigma(x,y)$ and define

$$h(s) := H_s - H_s \circ \sigma = +2r_0^s - 4(r_0')^s + 2r_1^s + 4r_{01}^s - 4(r_{01}')^s \tag{131}$$

Now we observe the following consequences of Lemma C121:

- $h(2) < 0$ and $h(14) > 0$. Hence h changes sign in $(2,14)$.
- $h(14) > 0$ and $h(s_0) < 0$. Hence h changes sign in $(14, s_0)$.
- $h(s_0) < 0$ and $h(16) > 0$. Hence h changes sign in $(s_0, 16)$.
- $h(s) < 0$ for s sufficiently large because the term $-4(r_{01}')^s$ eventually dominates. Hence h changes sign in $(16, \infty)$.

Hence h vanishes at least 4 times. By Descartes' Lemma, the sign sequence must change signs at least 4 times. Lemma C122 implies that the sign sequence must be one of

$$-,+,+,+,-, \qquad +,-,+,+,-, \qquad +,+,-,+,-.$$

In no case does it change sign at least 4 times. This is a contradiction. The proof of Lemma C12 is done, but we remark that more analysis would show that Equation 131 has the terms in the correct order and the middle sign seqence is correct.

13 Endgame: Proof of Lemma C2

13.1 The Goal

Recall that the point $(1,\sqrt{3}/3)$ represents the TBP avatar. We define

$$\Theta(s,x,y)=\mathcal{E}_s(x,y)-\mathcal{E}(1,\sqrt{3}/3). \tag{132}$$

Lemma C2 is equivalent to the following statements

1. If $s\in[15,15_+]$ then Θ has a unique minimum in $\widehat{\Psi}_8$.
2. $\Theta>0$ on $[13,15]\times\Psi_4$ and on $[15,15_+]\times(\Psi_4-\widehat{\Psi}_4)$.

13.2 Proof of Statement 1

Let Θ_x be the partial derivative of Θ with respect to x, etc. Statement 1 is equivalent to the statement that the single variable function $f(x)=\Theta(s,t,t)$ has only one minimum for $s\in I$. Here $64I=[55,56]$.

Lemma 13.1 (C21) *For all $s\in[13,15_+]$ and $(x,y)\in\Psi_4$ the quantities $\Theta_{xx},\Theta_{yy},\Theta_{xy}$ are all positive.*

We prove this result below. By the Chain Rule,

$$f_{tt}=\Theta_{xx}+\Theta_{yy}+2\Theta_{xy}>0 \tag{133}$$

Hence f is a convex function on I. Hence f has a unique minimum in I.

13.3 Proof of Statement 2

The proof of Statement 2 is a divide-and-conquer calculation. We first explain some details of the calculation.

The Expanding Out Method: This is a repeat of the definition in §9.3. Suppose we want to establish an inequality like $(\frac{a}{b})^{\frac{p}{q}}<\frac{c}{d}$, where every number involved is a positive integer. This inequality is true iff $b^pc^q-a^pd^q>0$. We check this using exact integer arithmetic. The same idea works with $(>)$ in place of $(<)$. We call this the *expanding out method*.

Rational Approximation Method: More generally, we will want to verify inequalities like

$$\sum_{i=1}^{10} b_i^{-s} - \sum_{i=1}^{10} a_i^{-s/2} > C. \tag{134}$$

where all a_i belong to the set $\{2,3,4\}$, and b_i, c, s are all rational. more specifically $s \in [13, 15_+]$ will be a dyadic rational and c will be positive. The expression on the left will be $\mathcal{E}_s(p) - \mathcal{E}_s(p_0)$ for various choices of p, and the constant C is related to the error term we define below.

Here is how we handle expressions like this. For each index $i \in \{1, ..., 10\}$ we produce rational numbers A_i and B_i such that

$$A_i^{s/2} > a_i \qquad B_i^s < b_i. \tag{135}$$

We use the expanding out method to check these inequalities. We then check that

$$\sum_{i=1}^{10} B_i - \sum_{i=1}^{10} A_i > C. \tag{136}$$

This last calculation is again done with integer arithmetic. Equations 135 and 136 together imply Equation 134. Logically speaking, the way that we produce the rational A_i and B_i does not matter, but let us explain how we find them in practice. For A_i we compute $2^{32}a_i^{-s/2}$ and round the result up to the nearest integer N_i. We then set $A_i = N_i/2^{32}$. We produce B_i in a similar way. When we have verified Equation 134 in this manner we say that we have used the *rational approximation method* to verify Equation 134. We will only need to make verifications like this on the order of 20000 times.

Error Estimate: We say that a *block* is a rectangular solid, having the following form:

$$X = I \times Q \subset [0,16] \times [0,1]^2, \tag{137}$$

where I is a dyadic interval and Q is a dyadic square. We define $|X|_1$ to be the length of I and $|X|_2$ to be the side length of Q.

Lemma 13.2 (C22) *For any block $X \subset [13,16] \times \Psi_4$,*

$$\min_X \Theta \geq \min_{v(X)} \Theta - (|X|_1^2/512 + |X|_2^2).$$

Here $v(X)$ denotes the vertex set of X. Thus, to show that $\Theta|_X > 0$ we just need to show that

$$\Theta_{v(X)} > \frac{|X|_1^2}{512} + |X|_2^2.$$

Grading a Block: We perform the following pass/fail evaluation of X.

1. If $I \subset [0,13]$ or $I \subset [15_+, 16]$ or $Q \cap \Psi_4 = \emptyset$, we pass X because X is irrelevant to the calculation.

2. If $s_0 \geq 15$ and $Q \subset \widehat{\Psi}_4$ we pass X.

3. $s_0 < 13$ and $s_1 > 13$ we fail X because we don't want to make any computations which involve exponents less than 13.

4. If X has not been passed or failed, we try to use the rational approximation method to verify that $\Theta(v) > |X|_1^2/512 - |X|_2^2$ for each vertex v of X. If we succeed at this, then we pass X. Otherwise we fail X.

To prove Statement 2 above it suffices to find a partition of $[0,16] \times [0,1]^2$ into blocks which all pass the evaluation.

Subdivision: Let $X = I \times Q$. Here is the rule we use to subdivide X: If $16|X|_2 > |X|_1$ we subdivide X along Q dyadically, into 4 pieces. Otherwise we subdivide X along I, into two pieces. This method takes advantage of the lopsided form of Lemma C22 and produces a small partition.

The Main Algorithm: We perform the following algorithm.

1. We start with a list L of blocks. Initially L has the single member $\{0,16\} \times \{0,1\}^2$.

2. We let B be the last block on L. We grade B. If B passes, we delete B from L. If $L = \emptyset$ then **HALT**. If B fails, we delete B from L and append to L the subdivision of B. Then we go back to Step 1.

Lemma 13.3 (C23) *When the algorithm runs it halts.*

Proof: As for the other calculations, I used a 2017 iMac Pro with a 3.2 GHz Intel Zeon W processor, running the Mojave operating system. When I run the algorithm, it halts with success after 21655 steps and in about 1 minute. The partition it produces has 14502 blocks. ♠

This proves Statement 2.

13.4 Proof of Lemma C21

We will show that $\Theta_{xx} > 0$ and $\Theta_{xy} > 0$ on $[13,16] \times \Psi_4$. The case of Θ_{yy} follows from the case of Θ_{xx} and symmetry. Setting $u = s/2$ we compute

$$\mathcal{E}_s(x,y) = A(s,x) + A(s,y) + 2B(s,x) + 2B(s,y) + 4C(s,x,y), \qquad (138)$$

$$A(x) = a(x)^u, \qquad B(x) = b(x)^u, \qquad C(x) = c(x)^u,$$

$$a(x) = \frac{(1+x^2)^2}{16x^2} \qquad b(x) = \frac{1+x^2}{4} \qquad c(x,y) = \frac{(1+x^2)(1+y^2)}{4(x^2+y^2)}$$

Hence

$$\Theta_{xx} = A_{xx} + 2B_{xx} + 4C_{xx}, \qquad \Theta_{xy} = C_{xy}. \qquad (139)$$

For each choice of $F = A, B, C$ we have

$$F_{xx} = u(u-1)f^{u-2}f_x^2 + uf^{u-1}f_{xx}, \quad C_{xy} = u(u-1)c^{u-2}c_xc_y + uc^{u-1}c_{xy}. \qquad (140)$$

(The second equation is just relevant for C.) We compute

$$a_{xx} = \frac{3+x^4}{8x^4} > 0, \qquad b_{xx} = \frac{1}{2}, \qquad c_{xx} = \frac{(1-y^4)(3x^2-y^2)}{2(x^2+y^2)^3} \geq 0.$$

$$c_x = \frac{x(y^4-1)}{2(x^2+y^2)^2} < 0, \quad c_y = \frac{y(x^4-1)}{2(x^2+y^2)^2} < 0, \quad c_{xy} = \frac{2xy(1+x^2y^2)}{(x^2+y^2)^3} > 0.$$

Equation 140 combines with all this to prove that $\Theta_{xx} > 0$ and $\Theta_{xy} > 0$ on $[13,16] \times \Psi_4$.

13.5 Proof of Lemma C22

Lemma 13.4 (C221) $|\Theta_{xx}|, |\Theta_{yy}| \leq 4$ *on* $[13,16] \times \Psi_4$.

Proof: By symmetry it suffices to prove this for Θ_{xx}. We already know $\Theta_{xx} > 0$ on our domain. We use the notation from §13.4. An easy exercise in calculus shows that $f \in (0, 3/5)$ on Ψ_4 for each $f = a, b, c$. From this bound, we see that the expression in Equation 140 is decreasing as a function of u for $u \geq 6$. (Recall that $u = s/2$.) Hence it suffices to prove that $4 - \Theta_{xx} \geq 0$ on $\{12\} \times [43/64, 1]^2$.

We define $\phi(t) = (43/64)(1-t) + t$. The file `LemmaC221.m` computes that for $s = 12$ the polynomial $\Phi = \text{num}_+(4 - \Theta_{xx} \circ \phi)$ is weak positive dominant and hence non-negative on $[0,1]^2$. Hence $4 - \Theta_{xx} \geq 0$ when $s = 12$ and $(x,y) \in \Psi_4$. ♠

Lemma 13.5 (C222) $|\Theta_{ss}| \leq 1/64$ *on* $[13,16] \times \Psi_4$.

Proof: Let $\psi(s) = b^{-s}$. Let $\beta = (1.3, \sqrt{2}, \sqrt{3})$ and $\gamma = (440, 753, 4184)$. We first establish the following bound:

$$0 < \min_{b \geq \beta_j} \psi_{ss}(s,b) \leq 1/\gamma_j, \qquad j = 1,2,3, \qquad \forall s \geq 13. \tag{141}$$

As a function of s, and for $b > 1$ fixed, $\psi_{ss}(s,b) = b^{-s}\log(b)^2$ is decreasing. Hence, it suffices to prove Equation 141 when $s = 13$. Choose $b \geq 1.3$. The equation $\psi_{ssb}(13,b) = 0$ has its unique solution in $[1,\infty)$ at the value $b = \exp(2/13) < 1.3$. Moreover, the function $\psi_{ss}(13,b)$ tends to 0 as $b \to \infty$. Hence the restriction of the function $b \to \psi_{ss}(13,b)$ to $[b,\infty)$ takes its maximum value at b. Evaluating at $b = 1.3, \sqrt{2}, \sqrt{3}$ we get Equation 141.

For $x,y \in [43/64, 1]$ we easily check the inequalities

$$A(-1,x) \geq 3, \quad B(-1,x) \geq 2, \quad C(-1,x,y) \geq (1.3)^2.$$

The quantities on the left are the square distances of the various pairs of points in the corresponding configuration on S^2. From this analysis we conclude that the 10 distances associated to a 5-point configuration parametrized by a point in Ψ_4 exceed 1.3, and at least 6 of them exceed $\sqrt{2}$, and at least 2 of them exceed $\sqrt{3}$. The same obviously holds for the TBP.

Now, 10 of the 20 terms comprising $\Theta_{ss}(s,x,y)$ are positive and 10 are negative. Also, for the terms of the same sign, all 10 of them are less than 1/440, and at least 6 of them are less than 1/753, and at least 2 of them are less than 1/4184. Hence, by Equation 141, we have the final bound $|\Theta_{ss}| \leq (4/440) + (4/753) + (2/4184) < 1/64$. ♠

Write $I = [s_0, s_1]$ and $Q = [x_0, x_1] \times [y_0, y_1]$. Choose $(s,x,y) \in X = I \times Q$. Taylor's Theorem with remainder tells that for any function $f : [a,b] \to \boldsymbol{R}$ and any $x \in [a,b]$ we have

$$f(x) \geq \min(f(a), f(b)) - \frac{1}{8}\max_{[a,b]} |f''|.$$

Applying this result 3 times, using Lemmas C221 and C222, we have

$$\Theta(s,x,y) \geq \min_i \Theta(s_i,x,y) - |I|/512 \geq \min_{i,j} \Theta(s_i,x_j,y) - |I|/512 - |x_0 - x_1|/2 \geq$$

$$\min_{i,j,k} \Theta(s_i,x_j,y_j) - |I|/512 - |x_0 - x_1|/2 - |y_0 - y_1|/2 = \min_{v(X)} \Theta - |X|_1/512 - |X|_2.$$

This completes the proof of Lemma C22.

14 Endgame: Proof of Lemma C3

14.1 Reduction to a Simpler Statement

We carry over the notation from the previous two chapters. In particular, we define Θ as in Equation 132. We use the same equation for $\mathcal{E}_s(t,t)$, in terms of the functions A, B, C, as in Equation 138. Recall that $I = [55, 56]/64$. Let $t_0 = 55/64$ be the left endpoint of I. We claim that

$$\Theta_{stt}(15, t, t) < 0, \qquad \forall t \in I. \tag{142}$$

We compute that

$$\Theta_{st}(15, t_0, t_0) < 0, \qquad \Theta_s(15, t_0, t_0) < -2^{-7}, \tag{143}$$

and these conditions combine with Equation 142 to show that

$$\Theta_s(15, t, t) < -2^{-7}. \qquad \forall t \in I. \tag{144}$$

Lemma 14.1 (C31) $|\Theta_{ss}| \leq 2^{-6}$ *on* $[13, 16] \times \Psi_4$.

This is just Lemma C222. By Lemma C31 we have

$$|\Theta_{ss}| \times |15_+ - 15| \leq 2^{-6} \times \frac{25}{512} < 2^{-7}. \tag{145}$$

Hence $\Theta_s(s, t, t)$ varies by less than 2^{-7} as s ranges in $[15, 15_+]$. Hence $\Theta_s(s, t, t) < 0$ for all $s \in [15, 15_+]$ and all $t \in I$. This is Lemma C3.

Now we prove Equation 142. The file `LemmaC3.m` does the calculations for this proof. Because the s-energy of the TBP does not depend on the t-variable, we have

$$\Theta_{stt}(15, t, t) = 2A_{stt}|_{s=15} + 4B_{stt}|_{s=15} + 4C_{stt}|_{s=15}. \tag{146}$$

Call the three functions on the right $\alpha(t)$, $\beta(t)$, and $\gamma(t)$. To finish the proof, we just need to see that each of these is negative in I. We write $f \sim f^*$ if

$$\frac{f}{f^*} = 2^u t^v (1 + t^2)^w (2 + t^2 + t^{-2})^x$$

for exponents $u, v, w, x \in \boldsymbol{R}$. In this case, f and f^* have the same sign.

Lemma 14.2 (C32) *$\beta < 0$ on I.*

Proof: Taking $(u,v,w,x) = (-14, 0, 11/2, 0)$ we have $\beta \sim -\beta^*$,

$$\beta^*(t) = (-2 + 30\log(2)) + t^2(-58 + 420\log(2)) - 15(1 + 14t^2)\log(1 + t^2).$$

Noting that $\log(2) = 0.69...$ we eyeball β^* and see that it is positive for $t \in I$. The term $+420\log(2)t^2$ dominates. Hence $\beta < 0$ on I. ♠

Lemma 14.3 (C33) *$\gamma < 0$ on I.*

Proof: Taking $(u,v,w,x) = (-41/2, -16, 12, 1/2)$ we have $\gamma \sim -\gamma^*$,

$$\gamma^*(t) = (-31 + 360\log(2)) + \underline{t^2(56 - 585\log(2))} + t^4(-29 + 315\log(2)) +$$

$$15(-8 + 13t^2 - 7t^4)\log(2 + t^2 + t^{-2}).$$

We have $\gamma^*(55/64) > 2^4$ and we estimate easily that $\gamma^*_t > -2^{10}$ on I. Only the underlined term has negative derivative in I. Noting that I has length 2^{-6}, we see that γ^* cannot decrease more than 2^4 as we move from x_0 to any other point of I. Hence $\gamma^* > 0$ on I. Hence $\gamma < 0$ on I. ♠

Lemma 14.4 (C34) *$\alpha < 0$ on I.*

Proof: Taking $(u,v,w,x) = (-29, -14, 10, 3/2)$ we have $\alpha \sim -\alpha*$,

$$\alpha^*(t) = \gamma^*(t) + \delta^*(t), \qquad \delta^*(t) = 15\log 2 \times (8 - 13t^2 + 7t^4).$$

We see easily that $\delta^* > 0$ on I. So, from Lemma C33, we have $\alpha^* > 0$ on I. Hence $\alpha < 0$ on I. ♠

15 References

[**A**] A. N. Andreev, *An extremal property of the icosahedron* East J Approx **2** (1996) no. 4 pp. 459-462

[**BBCGKS**] Brandon Ballinger, Grigoriy Blekherman, Henry Cohn, Noah Giansiracusa, Elizabeth Kelly, Achill Schurmann,
Experimental Study of Energy-Minimizing Point Configurations on Spheres, arXiv: math/0611451v3, 7 Oct 2008

[**BDHSS**] P. G. Boyvalenkov, P. D. Dragnev, D. P. Hardin, E. B. Saff, M. M. Stoyanova, *Universal Lower Bounds and Potential Energy of Spherical Codes*, Constructive Approximation 2016 (to appear)

[**BHS**], S. V. Bondarenko, D. P. Hardin, E.B. Saff, *Mesh Ratios for Best Packings and Limits of Minimal Energy Configurations*,

[**C**] Harvey Cohn, *Stability Configurations of Electrons on a Sphere*, Mathematical Tables and Other Aids to Computation, Vol 10, No 55, July 1956, pp 117-120.

[**CK**] Henry Cohn and Abhinav Kumar, *Universally Optimal Distributions of Points on Spheres*, J.A.M.S. **20** (2007) 99-147

[**CCD**] online website:
http://www-wales.ch.cam.ac.uk/∼ wales/CCD/Thomson/table.html

[**DLT**] P. D. Dragnev, D. A. Legg, and D. W. Townsend, *Discrete Logarithmic Energy on the Sphere*, Pacific Journal of Mathematics, Volume 207, Number 2 (2002) pp 345–357

[**Fö**], Föppl *Stabile Anordnungen von Electron in Atom*, J. fur die Reine Agnew Math. **141**, 1912, pp 251-301.

[**HZ**], Xiaorong Hou and Junwei Zhao, *Spherical Distribution of 5 Points with Maximal Distance Sum*, arXiv:0906.0937v1 [cs.DM] 4 Jun 2009

[**I**] IEEE Standard for Binary Floating-Point Arithmetic (IEEE Std 754-1985) Institute of Electrical and Electronics Engineers, July 26, 1985

[**KY**], A. V. Kolushov and V. A. Yudin, *Extremal Dispositions of Points on the Sphere*, Anal. Math **23** (1997) 143-146

[**MKS**], T. W. Melnyk, O. Knop, W.R. Smith, *Extremal arrangements of point and and unit charges on the sphere: equilibrium configurations revisited*, Canadian Journal of Chemistry 55.10 (1977) pp 1745-1761

[**RSZ**] E. A. Rakhmanoff, E. B. Saff, and Y. M. Zhou, *Electrons on the Sphere*,
Computational Methods and Function Theory, R. M. Ali, St. Ruscheweyh, and E. B. Saff, Eds. (1995) pp 111-127

[**S1**] R. E. Schwartz, *The* 5 *Electron Case of Thomson's Problem*, Journal of Experimental Math, 2013.

[**S2**] R. E. Schwartz, *The Projective Heat Map*, A.M.S. Research Monograph, 2017.

[**S3**] R. E. Schwartz, *Lengthening a Tetrahedron*, Geometriae Dedicata, 2014.

[**S4**], R. E. Schwartz, *Five Point Energy Minimization: A Summary*, Journal of Constructive Approximation (2019)

[**SK**] E. B. Saff and A. B. J. Kuijlaars, *Distributing many points on a Sphere*, Math. Intelligencer, Volume 19, Number 1, December 1997 pp 5-11

[**Th**] J. J. Thomson, *On the Structure of the Atom: an Investigation of the Stability of the Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle with Application of the results to the Theory of Atomic Structure*. Philosophical magazine, Series 6, Volume 7, Number 39, pp 237-265, March 1904.

[**T**] A. Tumanov, *Minimal Bi-Quadratic energy of* 5 *particles on* 2*-sphere*, Indiana Univ. Math Journal, **62** (2013) pp 1717-1731.

[**W**] S. Wolfram, *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press, Champaign/Cambridge (1999)

[**Y**], V. A. Yudin, *Minimum potential energy of a point system of charges* (Russian) Diskret. Mat. **4** (1992), 115-121, translation in Discrete Math Appl. **3** (1993) 75-81