STABILITY AND SYMMETRY IN ENERGY MINIMIZING POINT CONFIGURATIONS

COLLIN CADEMARTORI
ADVISOR: PROF. GOVIND MENON

Abstract. We investigate the relationship between stability and symmetry for point configurations on S^2 which minimize an energy functional. We first present some known results which allow us to obtain bounds on stability in terms of symmetry. We then describe a computational experiment which tested whether the bound is always achieved. Finally, we prove a simple inequality that allows us to determine that the bound is not achieved in some concrete cases.

Contents

1. Notation

Throughout this paper, we will consider collections of points in $\{x_1, \dots x_n\} \subset \mathbb{R}^3$. In order to simplify some notations, we will often treat such a collection as a point in the space $\mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \cdots \oplus \mathbb{R}^3$ or some subspace thereof. This amount to viewing $\{x_1, \dots x_n\}$ as the matrix

$$\bigoplus_{i=1}^{n} x_i^T = \begin{bmatrix} x_1^T & & & \\ & x_2^T & & \\ & & \ddots & \\ & & & x_n^T \end{bmatrix}$$

where the x_i are thought of as row vectors. If $T_i : \mathbb{R}^3 \to \mathbb{R}$ are linear maps, then if we think of the T_i as matrices, we can represent their direct sum as

$$T = \bigoplus_{i=1}^{n} T_i = \begin{bmatrix} T_1 & & & \\ & T_2 & & \\ & & \ddots & \\ & & & T_n \end{bmatrix}$$

If S and T are such maps, then

$$ST = \begin{bmatrix} S_1T_1 & & & \\ & S_2T_2 & & \\ & & \ddots & \\ & & & S_nT_n \end{bmatrix}$$

Sometimes $\mathcal{C} = \{x_1, \dots, x_n\}$ will be regarded as the set of $x_1, \dots, x_n$ rather than as a point in $\mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \cdots \oplus \mathbb{R}^3$. This should be clear from context, as in such cases when we write $x \in \mathcal{C}$ or $|\mathcal{C}|$ to indicate set inclusion and set size respectively. (Norms in the direct sum space will be denoted with double bars $\|\cdot\|$).

We will also often suppress the indices of the points x_i and write an expression of the form

$$\bigoplus_{x \in \mathcal{C}} T_x(x) \tag{1.1}$$

If $\mathcal{C}' \subset \mathcal{C}$, then we will want to regard the element

$$\bigoplus_{x \in \mathcal{C}'} T_x(x) \tag{1.2}$$

as a member of the same space that 1.1 occupies. Thus, 1.2 will be taken to represent

$$\bigoplus_{x \in \mathcal{C}} T_x^*(x)$$

where

$$T_x^*(x) = \begin{cases} T_x(x), & x \in \mathcal{C}' \\ 0, & x \in \mathcal{C} \setminus \mathcal{C}' \end{cases}$$

2. Background

In this paper, we investigate the relationship between certain stability and symmetry properties of energy minimal point configurations on the sphere S^2. A point configuration is any finite $\mathcal{C} \subset S^2$. In order to study the properties of energy minimal configurations, we will need to appeal at times to the idea of varying a quantity over all possible force laws. Thus we need to define a formal notion of force law for this idea to make sense. In doing so, we will want our notion of law to satisfy two general requirements:

- The collecton of force laws should resemble the actual force laws found in physics
- The collection should be large enough to allow for significant variation in the behavior of different forces

In order to satisfy the first requirement, we will choose force laws that act pairwise between points of our configurations and which depend only on the distances between those points. (In our case, we will choose our forces to be functions of the squared distance, since this will make computations must more tractable later). Thus, the total energy of the configuration will will be given as

$$\sum_{\substack{x,y\in\mathcal{C}\\ x\neq y}} f\left(|x-y|^2\right)$$

for some appropriate choice of f. In the case of the classical Coulomb potential from physics, we have $f(x) = x^{-1/2}$. And, in general, compatibility with physics suggests that we should at least choose f to be decreasing and convex. In fact, we will follow [Coh16] and require something more strict, but which naturally extends these conditions. Specifically, we will require that $f \in \mathrm{CM}((0,4])$ where

$$\mathrm{CM}((0,4]) = \left\{ f \in \mathcal{C}^\infty((0,\infty)) \;\middle|\; (-1)^k f(x)^{(k)} \geq 0 \text{ for all } x \in (0,4] \text{ and all } k \geq 0 \right\}$$

is the class of completely monotonic functions. Then for any configuration $\mathcal{C}$ and force potential $f \in \mathrm{CM}((0,4])$, we can define the total energy of the configuration as

$$E_f(\mathcal{C}) = \sum_{\substack{x,y\in\mathcal{C}\\ x\neq y}} f\left(|x-y|^2\right) \tag{2.1}$$

The collection $\mathrm{CM}((0,4])$ maintains significantly compatibility with the idea of a force law from physics. It is also diverse enough to support significant variation in force behavior. In particular, it contains every inverse power law $f(x) = x^{-a}$ for $a > 0$ and all positive linear combinations thereof. It also contains a large collection of polynomials, which we will see later gives us independent control over the strength of the force at different distances.

Further motivation for this choice of force laws comes from the fact that some configurations have been shown to minimize E_f for all completely monotonic f, but not for all decreasing and convex f (see [CK06]). This suggests that the additional constraints of complete monotonicity make a meaningful difference in the study of energy minimization.

2.1. Measuring the Stability of Point Configurations.

Let $\mathcal{C}$ be some point configuration, and fix $x \in \mathcal{C}$. Then we denote the tangent space at x by $\mathcal{T}_x = x^\perp$.

Considering $\mathcal{C}$ as a point in $S^2 \oplus S^2 \oplus \cdots \oplus S^2$, we can express the tangent space at $\mathcal{C}$ as

$$T_{\mathcal{C}} = \mathcal{T}_{x_1} \oplus \mathcal{T}_{x_2} \oplus \cdots \oplus \mathcal{T}_{x_N}$$

For any $f \in \mathrm{CM}((0,4])$, we can define the surface gradient of E_f at $\mathcal{C}$ as

$$\nabla^s E_f(\mathcal{C}) = (1 - P_{\mathcal{C}}) \nabla E_f(\mathcal{C})$$

where $\nabla E_f(\mathcal{C}) = \bigoplus_{x \in \mathcal{C}} \nabla_x E_f(\mathcal{C})$, and $P_{\mathcal{C}}$ is projection onto the unit normal vectors x_i, and is given by

$$P_{\mathcal{C}} = \begin{bmatrix} x_1 x_1^T & & & \\ & x_2 x_2^T & & \\ & & \ddots & \\ & & & x_N x_N^T \end{bmatrix} = \bigoplus_{i=1}^{N} x_i x_i^T$$

Thus $(1 - P_{\mathcal{C}})$ is just projection onto $\mathcal{T}_{\mathcal{C}}$.

As in [BBC$^+$09], we will adopt the following terminology:

Definition 2.1. A configuration $\mathcal{C} \subset S^2$ is **energy minimal** if there exists some $f \in \mathrm{CM}((0,4])$ such that $\mathcal{C}$ is a local minimum for E_f. We say $\mathcal{C}$ is in **equilibrium** if

$$\nabla^s E_f(\mathcal{C}) = 0$$

and we say $\mathcal{C}$ is **balanced** if it is in equilibrium for *all* choices of $f \in \mathrm{CM}((0,4])$.

If $\mathcal{C}$ is balanced, then as we vary f over all force laws, the points in $\mathcal{C}$ remain fixed in place. In other words, they have zero degrees of freedom for movement as f varies. While very few $\mathcal{C}$ satisfy this strong stability condition, we can quantify how far from balanced a configuration $\mathcal{C}$ is. We do this by counting the number of degrees of freedom that the points in $\mathcal{C}$ have for movement as the force law f varies.

To formalize this idea, we will need some additional terminology. We denote the collection of surface gradients of $\mathcal{C}$ as

$$\nabla_E(\mathcal{C}) = \left\{ \nabla^s E_f(\mathcal{C}) \in \mathcal{T}_{\mathcal{C}} \;\middle|\; f \in \mathrm{CM}((0,4]) \right\}$$

Let $\nabla^s E_f(\mathcal{C}), \nabla^s E_g(\mathcal{C}) \in \nabla_E(\mathcal{C})$ and $\alpha, \beta \geq 0$. Then $f, g \in \mathrm{CM}((0,4])$, and it follows by definition that $\alpha f + \beta g \in \mathrm{CM}((0,4])$. Furthermore, we have

$$\alpha \nabla^s E_f(\mathcal{C}) + \beta \nabla^s E_g(\mathcal{C}) = \nabla E_{\alpha f + \beta g}(\mathcal{C}) \in \nabla_E(\mathcal{C})$$

which follows from the fact that ∇^s is linear and E_f is linear in f. Thus $\nabla_E(\mathcal{C})$ is a convex cone in $T_{\mathcal{C}}$.

Now denote the set of distances occuring between points in $\mathcal{C}$ as

$$D_{\mathcal{C}} = \left\{ d \in \mathbb{R}_+ \;\middle|\; \exists x, y \in \mathcal{C},\ |x - y| = d \right\}$$

The following definition is the primary tool for understanding the stability of point configurations.

Definition 2.2. For each $d \in D_{\mathcal{C}}$, define the **d−perturbation** of $\mathcal{C}$ as

$$\nabla^{(d)}\mathcal{C} = (1 - P_{\mathcal{C}}) \left[\bigoplus_{x \in \mathcal{C}} \sum_{\substack{y \in \mathcal{C} \\ |x-y|=d}} (x - y) \right]$$

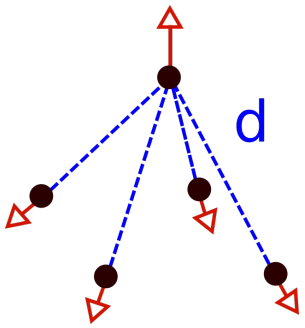

FIGURE 1. An example of a $d-$perturbation for five points arranged as a square pyramid. There are three $d-$perturbations for this configuration corresponding to the following distances: the distance between each point on the base and each adjacent base point, the distance between each base point and the opposite base point, and the distance occurring between each base point and the apex point (pictured above).

In other words, the $d-$perturbation of $\mathcal{C}$ is the element of $\mathcal{T}_{\mathcal{C}}$ that is obtained when each pair of points that are distance d apart repel each other with unit force. The usefulness of this concept becomes immediately apparent in the following proposition.

Proposition 2.3. *Every element of $\nabla_E(\mathcal{C})$ can be written as a linear combination of the $d-$perturbations of $\mathcal{C}$.*

Proof. This follows directly from the definition of the surface gradient. Indeed, we can write

$$\begin{aligned}
\nabla^s E_f(\mathcal{C}) &= (1-P_{\mathcal{C}})(\nabla_{x_1} \oplus \nabla_{x_2} \oplus \cdots \oplus \nabla_{x_N}) E_f(\mathcal{C}) \\
&= (1-P_{\mathcal{C}}) \left(\bigoplus_{x\in\mathcal{C}} \frac{1}{2} \sum_{y,z\in\mathcal{C}} \nabla_x f(|y-z|^2) \right) \\
&= (1-P_{\mathcal{C}}) \left(\bigoplus_{x\in\mathcal{C}} \sum_{y\in\mathcal{C}} f'(|x-y|^2)(x-y) \right) \\
&= (1-P_{\mathcal{C}}) \left(\bigoplus_{x\in\mathcal{C}} \sum_{d\in D_{\mathcal{C}}} \sum_{|y-x|=d} f'\left(d^2\right)(x-y) \right) \\
&= (1-P_{\mathcal{C}}) \left(\bigoplus_{x\in\mathcal{C}} \sum_{d\in D_{\mathcal{C}}} f'\left(d^2\right) \left(\sum_{|y-x|=d} (x-y) \right) \right) \\
&= \sum_{d\in D_{\mathcal{C}}} f'\left(d^2\right) \left((1-P_{\mathcal{C}}) \left(\bigoplus_{x\in\mathcal{C}} \sum_{|y-x|=d} (x-y) \right) \right) \\
&= \sum_{d\in D_{\mathcal{C}}} f'\left(d^2\right) \nabla^{(d)}\mathcal{C}
\end{aligned}$$

which shows that the surface gradient can be written as a linear combination of the $d-$perturbations.

□

Note that this does *not* show that $\nabla^{(d)}\mathcal{C} \in \nabla_E\mathcal{C}$. Indeed, for $d_0 \in D_\mathcal{C}$, there does not generally exist $f \in \mathrm{CM}((0,4])$ with the property that $f'(d_0^2) = 1$ and $f'(d^2) = 0$ for $d \in D_\mathcal{C} \setminus \{d_0\}$. However, $\mathrm{CM}((0,4])$ is flexible enough to allow us to set the values $f'(d^2)$ independently.

Proposition 2.4. *For any finite $\{d_i\}_{i=0}^n \subset (0,4]$ and any $f \in \mathrm{CM}((0,4])$, there is an infinite family $\{g_j\}_{j\in J} \subset \mathrm{CM}((0,4])$ such that $g_j'(d_i) = f'(d_i)$ for all $j \in J$ and $1 \le i \le n$ but $g_j'(d_0) \neq g_k'(d_0)$ for $j \neq k$.*

Proof. First suppose that $f, g \in \mathrm{CM}((0,4])$ with $f'(d_i) = g'(d_i)$ for $1 \le i \le n$ and $f'(d_0) \neq g'(d_0)$. If $0 \le \alpha \le 1$, then we already saw that

$$g_\alpha = \alpha f + (1-\alpha)g \in \mathrm{CM}((0,4])$$

and clearly every pair of functions in $\{g_\alpha'\}_{\alpha\in[0,1]}$ agree on d_i for $1 \le i \le n$ and disagree on d_0. Thus if we can find a single such pair of $f, g \in \mathrm{CM}((0,4])$, then we can construct an infinite family.

We can construct such a pair of functions directly as follows. We claim that if $\{\alpha_k\}_{k=1}^n \subset [4,\infty)$ are distinct, then we have

$$P(x) = (-1)^n \prod_{k=1}^n (x - \alpha_k) \in \mathrm{CM}((0,4])$$

To see this, observe that, for any $k \ge 0$, we have

$$P^{(k)}(x) = (-1)^n C_k \sum_{j_1<j_2<\cdots<j_k} \prod_{j\notin\{j_1,\ldots,j_k\}} (x - \alpha_j)$$

for some $C_k > 0$ depending only on k. Now for $x \in (0,4)$, $(x - \alpha_i) < 0$ for all $1 \le i \le n$. So if k is even, for all $j_1 < j_2 < \cdots, j_k$, we have that

$$\operatorname{sign}\left(\prod_{j\notin\{j_1,\ldots,j_k\}} (x - \alpha_j)\right) = (-1)^{n-k}$$

since there are $n-k$ nonpositive terms in every such product, and so we have

$$\operatorname{sign}\left(P^{(k)}(x)\right) = (-1)^n(-1)^{n-k} = (-1)^{-k} = 1$$

and similarly $\operatorname{sign}\left(P^{(k)}(x)\right) = -1$ for all k odd. Since clearly $P \in \mathcal{C}^\infty((0,\infty))$, we have $P \in \mathrm{CM}$, as claimed.

Now take $4 < \alpha_1 < \alpha_2 < \cdots < \alpha_n$ and form the polynomial P as above. Let

$$p(x) = x^n + a_1x^{n-1} + \cdots + a_n = P'(x)$$

and set $\beta_i = p(d_i)$ for $0 \le i \le n$. Now if $p_0 = x^n + a_1^0x^{n-1} + \cdots + a_n^0$ interpolates the points (d_i, β_i), then we must have

$$\begin{pmatrix} d_0^n & d_0^{n-1} & \cdots & d_0 & 1 \\ d_1^n & d_1^{n-1} & \cdots & d_1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_n^n & d_n^{n-1} & \cdots & d_n & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a_1^0 \\ \vdots \\ a_n^0 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

Taking D to be the lefthand matrix above, we have the Vandermonde determinant formula:

$$\det(D) = \prod_{0 \leq i < j \leq n} (d_i - d_j)$$

which immediately shows that $\det(D) > 0$ since the d_i are all distinct. Thus D is invertible, which shows that the solution p_0 is unique (and therefore $p_0 = p$ and $a_i^0 = a_i$ since p interpolates (d_i, β_i) by definition). And since D^{-1} is continuous (as a linear map between finite-dimensional normed spaces), we also see that the coefficients of the intpolating polynomial are a continuous function of the β_i.

Thus if we fix $\delta > 0$, we can find $\delta' > 0$ such that if

$$|(\gamma_0 - \beta_0, \gamma_1 - \beta_1, \ldots, \gamma_n - \beta_n)| < \delta'$$

and if $q = x^n + b_1 x^{n-1} + \cdots + b_n$ is the polynomial which interpolates (d_i, γ_i), then

$$|(a_0 - b_0, a_1 - b_1, \ldots, a_n - b_n)| < \delta$$

where $a_0 = 1$. In particular, we can choose γ_i such that $\gamma_i = \beta_i$ for $1 \leq i \leq n$ and

$$\gamma_0 \in (\beta_0 - \delta', \beta_0 + \delta')$$

Now we have

$$P = A_0 x^{n+1} + A_1 x^n + \cdots + A_n x + C$$

where $A_j = \frac{a_j}{n+1-j}$. And if we define

$$Q = x^{n+1} + B_1 x^n + \cdots + B_n x + C$$

with $B_j = \frac{b_j}{n+1-j}$, then $Q' = q$, and

$$|A_j - B_j| = \frac{1}{n+1-j} |a_j - b_j| < \frac{\delta}{n+1-j} \leq \delta$$

for all $0 \leq j \leq n$.

Moreover, it is known that if all the roots of P are real and simple (i.e. multiplicity 1), then there is a real-valued function ϕ defined in an open neighborhood U of $(A_0, A_1, \ldots, A_n)$ which takes $(u_0, u_1, \ldots, u_n) \in U$ to the roots of the polynomial with coefficients u_i. And in fact, ϕ is continuous at $(A_0, A_1, \ldots, A_n)$ (see [Ale13] for details).

Now fix some $\epsilon < \min_{1 \leq k \leq n} |\alpha_k - 4|$. Then the above just says that there is some $\delta > 0$ such that if $|B_j - A_j| < \delta$ for all $0 \leq j \leq n$, then the roots of Q are within ϵ of the roots of P. But we already saw that this is possible as long as $|\gamma_0 - \beta_0| < \delta'$. By our choice of ϵ, this establishes that the roots of Q are all greater than 4, and thus $Q \in \mathrm{CM}((0, 4])$, and Q' interpolates (d_i, γ_i). □

This shows that we have independent control over each of the coefficients in

$$\sum_{d \in D_{\mathcal{C}}} f'\left(d^2\right) \nabla^{(d)} \mathcal{C}$$

Moreover, we know that $\nabla_E(\mathcal{C}) \subset \operatorname{span}\left(\left\{\nabla^{(d)}\mathcal{C}\right\}_{d \in D_{\mathcal{C}}}\right)$, and the above shows that this is the smallest subspace of $\mathcal{T}_{\mathcal{C}}$ which contains $\nabla_E(\mathcal{C})$. This observation leads naturally to the following definition.

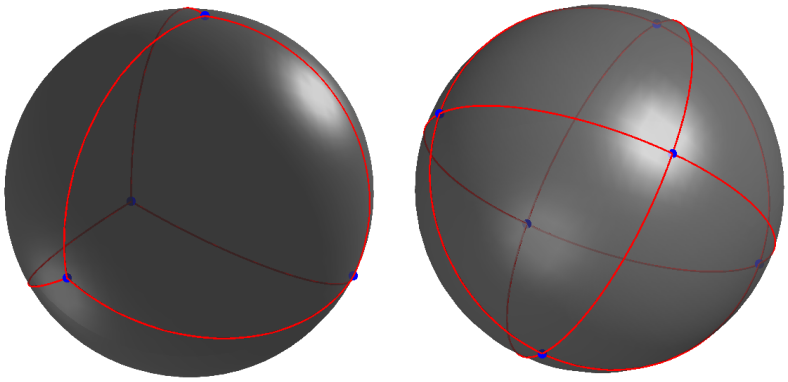

FIGURE 2. Point configuration consisting of the vertices of a regular tetrahedron (left) and a regular octahedron (right).

Definition 2.5. For any finite $\mathcal{C} \subset S^2$, we say that the **parameter count** of $\mathcal{C}$ is given by

$$P(\mathcal{C}) = \dim \operatorname{span}\left(\left\{\nabla^{(d)}\mathcal{C}\right\}_{d\in D_{\mathcal{C}}}\right)$$

Recall that $\nabla^s E_f(\mathcal{C})$ indicates the directions that the points in $\mathcal{C}$ will move under the action of the force law f. The parameter count quantifies the number of degrees of freedom that the points in $\mathcal{C}$ have to move in under arbitary force laws in $\mathrm{CM}((0,4])$. A low parameter count indicates a highly stable configuration of points, with $P(\mathcal{C}) = 0$ indicating that the configuration is balanced (in equilibrium for all potentials).

Consider, for example the point configuration $\mathcal{C}$ consisting of the vertices of a regular tetrahedron. There is only one nonzero distance d occurring in this configuration. Thus this $d-$perturbation is just the element of $\mathcal{T}_{\mathcal{C}}$ which is obtained when every pair of points in the configuration repel each other with unit force.

To compute this $d-$perturbation, fix $u \in \mathcal{C}$. Then the three other points x, y, z have $3-$fold rotational symmetry about the axis formed by u and the origin. If R is rotation by $\frac{2\pi}{3}$ around this axis, then

$$R(x+y+z) = R(x+Rx+R^2x) = x+Rx+R^2x = x+y+z$$

so $x+y+z = cu$ for some $c \in \mathbb{R}$. And if $P_u = uu^T$,

$$(1-P_u)\left(\sum_{\substack{w\in\mathcal{C}\\|u-w|=d}}(u-w)\right) = (1-P_u)(3u-(x+y+z)) = (3-c)(1-P_u)(u) = 0$$

Thus we have

$$(1-P_{\mathcal{C}})\left[\bigoplus_{x\in\mathcal{C}}\sum_{\substack{y\in\mathcal{C}\\|x-y|=d}}(x-y)\right] = \bigoplus_{u\in\mathcal{C}}(1-P_u)\left(\sum_{\substack{w\in\mathcal{C}\\|u-w|=d}}(u-w)\right) = 0$$

So in this case we have $P(\mathcal{C}) = 0$.

Now consider the configuration $\mathcal{C}$ consisting of the vertices of a regular octahedron. There are two nonzero distances in this configuration: the distance of 2 between each point and the opposite point and the distance of $\sqrt{2}$ between each point and the four adjacent points.

Then the $2-$perturbation of $\mathcal{C}$ is clearly 0 since for any $x \in \mathcal{C}$, the only point that is distance 2 from x is $-x$, and thus

$$(1 - P_x)\left(\sum_{\substack{y\in\mathcal{C}\\|x-y|=2}} (x-y)\right) = (1-P_x)(2x) = 0$$

Now the four points which are distance $\sqrt{2}$ from x have $4-$fold rotational symmetry about the axis formed by x and the origin. Thus, the same argument used for the terahedron shows that the $\sqrt{2}-$perturbation is also 0, and so we again have $P(\mathcal{C}) = 0$.

This shows that the tetrahedral and octrahedral configurations are both balanced (in equilibrium under every force law). However, this only tells us that they are critical points of E_f for all $f \in \mathrm{CM}((0,4])$, not that they are energy minimal for all such f.

2.2. **Bounding Stability Using Symmetry.** The configurations $\mathcal{C}$ which are energy minimal for some potential f almost always exhibit structure in the form of nontrivial symmetries. It is natural to expect that those configurations with more symmetric structure would be more stable in the sense of having a low parameter count. Indeed, this idea can be formalized.

Definition 2.6. The **symmetry group** of $\mathcal{C}$ is given by

$$S(\mathcal{C}) = \left\{ T \in O(3) \;\middle|\; T(x) \in \mathcal{C} \text{ for all } x \in \mathcal{C} \right\}$$

Let $P_x = xx^T$, we have the following proposition.

Proposition 2.7. *For any $T \in S(\mathcal{C})$, we have*

$$T((1-P_x)\nabla_x E_f(\mathcal{C})) = (1 - P_{T(x)})\nabla_{T(x)} E_f(\mathcal{C})$$

Proof. We have

$$T((1-P_x)\nabla_x E_f(\mathcal{C})) = T(\nabla_x E_f(\mathcal{C})) - T(P_x \nabla_x E_f(\mathcal{C}))$$

First we see that

$$\begin{aligned} T(\nabla_x E_f(\mathcal{C})) &= T\left(\sum_{y\in\mathcal{C}} f'(|x-y|^2)(x-y)\right) \\ &= \sum_{y\in\mathcal{C}} f'(|x-y|^2)(Tx - Ty) \\ &= \sum_{y\in\mathcal{C}} f'(|Tx-Ty|^2)(Tx-Ty) \\ &= \sum_{z\in\mathcal{C}} f'(|Tx-z|^2)(Tx-z) \qquad (2.2) \\ &= \nabla_{T(x)} E_f(\mathcal{C}) \end{aligned}$$

where 2.2 follows from the substitution $y = T^{-1}z$, which preserves the equality since $T^{-1}(\mathcal{C}) = \mathcal{C}$.

Next observe that

$$P_{T(x)}T = (Tx)(Tx)^T T = Txx^T T^T T = Txx^T = TP_x$$

where we used the fact that the inverse of an orthogonal transform is its transpose. Combining this with the last part gives

$$T(P_x \nabla_x E_f(\mathcal{C})) = P_{T(x)}T(\nabla_x E_f(\mathcal{C})) = P_{T(x)}\nabla_{T(x)} E_f(\mathcal{C})$$

Putting everything together, we get

$$T((1 - P_x)\nabla_x E_f(\mathcal{C})) = (1 - P_{T(x)})\nabla_{T(x)} E_f(\mathcal{C})$$

as needed. □

Now we recall a few definitions from group actions.

Definition 2.8. For any $x \in \mathcal{C}$, the **stabilizer** of x under $S(\mathcal{C})$ is the subgroup

$$S_x = \left\{ T \in S(G) \;\middle|\; T(x) = x \right\}$$

and the **orbit** of x under $S(\mathcal{C})$ is

$$O_x = \left\{ y \in \mathcal{C} \;\middle|\; y = Tx \text{ for some } T \in S(\mathcal{C}) \right\}$$

The following subspaces are the key to extracting stability information from symmetry.

Definition 2.9. For any $x \in \mathcal{C}$, the **fixed subspace** of x is

$$V_x = \left\{ y \in \mathcal{T}_x \;\middle|\; T(y) = y \text{ for all } T \in S_x \right\} = \bigcap_{T \in S_x} \left\{ y \in \mathcal{T}_x \;\middle|\; T(y) = y \right\}$$

Then clearly V_x is a subspace of T_x.

Let $\mathcal{O} = \{O_x \mid x \in \mathcal{C}\}$ be the partition of $\mathcal{C}$ into orbits under $S(\mathcal{C})$. Then we claim that we have the following upper bound on the parameter count.

Proposition 2.10. *For any point configuration $\mathcal{C}$, we have*

$$P(\mathcal{C}) \leq \sum_{O_x \in \mathcal{O}} \dim V_x$$

Remark 2.11.

- This sum is well-defined. Indeed, if $O_x = O_y$, then there is some $T \in S(\mathcal{C})$ such that $T(y) = x$. Then it follows directly that $S_y = TS_xT^{-1}$, and so $U(z) = z$ for all $U \in S_x$ if and only if $V(Tz) = Tz$ for all $V \in S_y$. Since T is a linear bijection, it follows that $\dim V_x = \dim V_y$, and so the above sum is well-defined.
- The above sum depends only on knowing the symmetry group of $\mathcal{C}$, so this result formalizes the intuition that configurations with more symmetric structure are more dynamically stable.

Proof. To prove the above bound, first observe that, for all $T \in S_x$, we have $T((1 - P_x)\nabla_x E_f(\mathcal{C})) = (1 - P_x)\nabla_x E_f(\mathcal{C})$ by proposition 2.7. Therefore we have

$$(1 - P_x)\nabla_x E_f(\mathcal{C}) \in V_x$$

for all $x \in \mathcal{C}$. Now we can break up the surface gradient over orbits.

$$\nabla E_f(\mathcal{C}) = (1-P_{\mathcal{C}}) \bigoplus_{x\in\mathcal{C}} \nabla_x E_f(\mathcal{C}) = \bigoplus_{x\in\mathcal{C}} (1-P_x)\nabla_x E_f(\mathcal{C}) = \sum_{O_x\in\mathcal{O}} \left(\bigoplus_{y\in O_x} (1-P_y)\nabla_y E_f(\mathcal{C}) \right)$$

Now for each $x \in \mathcal{C}$ and $y \in O_x$, let $T_x^y \in S(\mathcal{C})$ be such that $T_x^y(x) = y$ (with $T_x^x = I$). Then, by the above claim,

$$\begin{aligned} \bigoplus_{y\in O_x} (1-P_y)\nabla_y E_f(\mathcal{C}) &= \bigoplus_{y\in O_x} (1-P_{T_x^y(x)})\nabla_{T_x^y(x)} E_f(\mathcal{C}) \\ &= \bigoplus_{y\in O_x} T_x^y \left((1-P_x)\nabla_x E_f(\mathcal{C})\right) \\ &\in \Phi_x(V_x) \end{aligned}$$

where $\Phi_x : \mathcal{T}_x \to \mathcal{T}_{\mathcal{C}}$ is the map

$$\Phi_x(z) = \bigoplus_{y\in O_x} T_y^x(z)$$

Plugging this in to the above, we have

$$\nabla E_f(\mathcal{C}) = \sum_{O_x\in\mathcal{O}} \left[\bigoplus_{y\in O_x} T_x^y \left((1-P_x)\nabla_x E_f(\mathcal{C})\right) \right] \in \bigoplus_{O_x\in\mathcal{O}} \Phi_x(V_x)$$

Now observe that Φ_x clearly has inverse given by the map $P_{\mathcal{T}_x} : \mathcal{T}_{\mathcal{C}} \to \mathcal{T}_x$ which is projection onto $\mathcal{T}_x$. Thus we know that

$$\dim(\Phi_x(V_x)) = \dim(V_x)$$

for all $x \in \mathcal{C}$. And by proposition 2.4, we know that $\operatorname{span}\left(\left\{\nabla^{(d)}\mathcal{C}\right\}_{d\in D_{\mathcal{C}}}\right)$ is the smallest linear subspace of $\mathcal{T}_{\mathcal{C}}$ containing every $\nabla E_f(\mathcal{C})$, so we must have

$$\operatorname{span}\left(\left\{\nabla^{(d)}\mathcal{C}\right\}_{d\in D_{\mathcal{C}}}\right) \subset \bigoplus_{O_x\in\mathcal{O}} \Phi_x(V_x)$$

and thus

$$P(\mathcal{C}) \leq \dim\left(\bigoplus_{O_x\in\mathcal{O}} \Phi_x(V_x) \right) = \sum_{O_x\in\mathcal{O}} \dim(\Phi_x(V_x)) = \sum_{O_x\in\mathcal{O}} \dim V_x$$

as claimed.

□

For simplicity, we will henceforth denote the quantity $\sum_{O_x\in\mathcal{O}} \dim V_x$ by $B(\mathcal{C})$.

Consider the tetrahedral configuration $\mathcal{C}$ from above. As we noted when we computed the parameter count of $\mathcal{C}$, any $x \in \mathcal{C}$ is fixed by a 3−fold rotatation $R_x \in S(\mathcal{C})$. But R_x fixes only the subspace $\operatorname{span}(\{x\})$, which clearly intersects $\mathcal{T}_x$ only at the origin. Thus $\dim V_x = 0$ for all $x \in \mathcal{C}$, and $B(\mathcal{C}) = 0$.

For the octahedral configuration $\mathcal{C}$, we again have that each $x \in \mathcal{C}$ is fixed by a 4−fold rotation $R_x \in S(\mathcal{C})$. But then we again have $\dim V_x = 0$ for all $x \in \mathcal{C}$, and so $B(\mathcal{C}) = 0$.

The calculations of $P(\mathcal{C})$ and $B(\mathcal{C})$ above both exploited the symmetry of the configurations to constrain the degrees of freedom. The question therefore arises

as to whether there is anything constraining the points other than symmetry. In other words, is it the case that

$$P(\mathcal{C}) = B(\mathcal{C})$$

for all $\mathcal{C}$ which are in equilibrium under the action of some potential $f \in \mathrm{CM}((0,4])$? We investigate this question through a series of computational experiments.

3. Computational Experiments

In order to investigate this question, we found configurations of 4 to 55 points on S^2 which minimized energy for the Coulomb potential ($f(x) = x^{-1/2}$). We then computed the parameter count and symmetry bound for each of these configurations in order to study the relationship between them.

3.1. **Minimizing Energy with Gradient Descent.**

The first step in this process involved applying a gradient descent algorithm to find good approximations of energy minimal configurations. Traditional gradient descent cannot be directly applied to this problem since the domain is

$$\underbrace{S^2 \oplus S^2 \oplus \cdots \oplus S^2}_{N \text{ times}}$$

which is closed and nonconvex. Instead, we use an algorithm that searches for the direction of steepest descent in the tangent space

$$\mathcal{T}_{\mathcal{C}} = \mathcal{T}_{x_1} \oplus T_{x_2} \oplus \cdots \oplus \mathcal{T}_{x_N}$$

In other words, we solve the following optimization problem:

$$\operatorname*{argmin}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \langle \nabla E_f(\mathcal{C}), u \rangle \tag{3.1}$$

Now observe that by homogeneity of the inner product, we can replace $\nabla E_f(\mathcal{C})$ by $K\nabla E_f(\mathcal{C})$ for any $K > 0$ without changing the solution. In order to solve this problem, first observe that

$$\begin{aligned}\operatorname*{argmax}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \langle K\nabla E_f(\mathcal{C}), u \rangle &= \operatorname*{argmin}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \left(\|K\nabla E_f(\mathcal{C})\|^2 - 2\langle K\nabla E_f(\mathcal{C}), u \rangle + \|u\|^2 \right) \\ &= \operatorname*{argmin}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \|K\nabla E_f(\mathcal{C}) - u\|^2\end{aligned}$$

But we know that from linear algebra that the problem

$$\operatorname*{argmin}_{u \in T_{\mathcal{C}}} \|K\nabla E_f(\mathcal{C}) - u\|^2$$

is solved by taking u to be the orthogonal projection of $K\nabla E_f(\mathcal{C})$ onto the space $\mathcal{T}_{\mathcal{C}}$. But this is just $u = (1 - P_{\mathcal{C}})K\nabla E_f(\mathcal{C}) = K\nabla^s E_f(\mathcal{C})$. By taking $K = \|\nabla^s E_f(\mathcal{C})\|^{-1}$, we get $\|u\| = 1$, and thus

$$\operatorname*{argmin}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \langle \nabla E_f(\mathcal{C}), u \rangle = -\operatorname*{argmax}_{\substack{u \in T_{\mathcal{C}} \\ \|u\|=1}} \langle K\nabla E_f(\mathcal{C}), u \rangle = -\frac{(1 - P_{\mathcal{C}})\nabla E_f(\mathcal{C})}{\|(1 - P_{\mathcal{C}})\nabla E_f(\mathcal{C})\|} \tag{3.2}$$

Once we have found the direction of steepest descent in the tangent space, we would like to follow this direction to obtain our updated point configuration. However, this point will be an element of the ambient space $\mathbb{E} = \mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \cdots \oplus \mathbb{R}^3$ rather

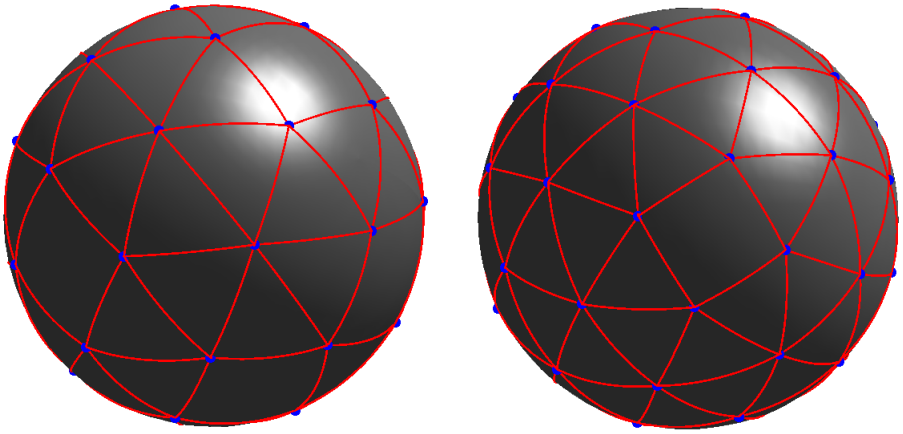

FIGURE 3. Computed energy minimal configurations for the Coloumb force law with 35 and 48 points respectively

than $\mathcal{C}$ itself, so we must perform the additional step of projecting our new point back into our domain.

If $v \in \mathbb{E}$, then let $v_i \in \mathbb{R}^3$ denote the i^{th} component of v, so that $v = \oplus_{i=1}^N v_i$. We consider $\mathbb{E}$ as a normed space with norm

$$\|v\| = \sum_{i=1}^{N} \|v_i\|_{\mathbb{R}^3}$$

where $\|\cdot\|_{\mathbb{R}^3}$ is the usual Euclidean norm. (We note that this norm is equivalent to the usual 2−norm on $\mathbb{R}^{3N}$.) Then v' solves the optimization problem

$$\underset{w \in S^2 \oplus \cdots \oplus S^2}{\operatorname{argmin}} \|w - v\| \tag{3.3}$$

if and only if v_i' solves the problem

$$\underset{w_i \in S^2}{\operatorname{argmin}} \|w_i - v_i\|_{\mathbb{R}^3} \tag{3.4}$$

for $i = 1, \ldots, N$. But this optimization is clearly solved by taking $v_i' = \frac{v_i}{\|v_i\|}$, and so our original problem is solved by taking $v' = \oplus_{i=1}^N \frac{v_i}{\|v_i\|}$. With these two modifications (choosing the descent direction in the tangent space and projecting back into the domain), we can describe the full gradient descent algorithm:

(1) Choose an initial point configuration $\mathcal{C}_0$, and set $k = 0$.

(2) Compute $\nabla E_f(\mathcal{C}_k)$ and the matrix $P_{\mathcal{C}_k}$. Set $D_k \leftarrow (1 - P_{\mathcal{C}_k})\nabla E_f(\mathcal{C}_k)$.

(3) Search $(0, 1]$ for the optimal step size α_k.

(4) Set $\mathcal{C}_k' \leftarrow \mathcal{C}_k - \alpha_k D_k$, then $\mathcal{C}_k' = \oplus_{i=1}^N y_i^k \in \mathbb{E}$.

(5) Set $x_i^{k+1} \leftarrow \frac{y_i^k}{\|y_i^k\|}$ and $\mathcal{C}_{k+1} \leftarrow \oplus_{i=1}^N x_i^{k+1}$.

(6) If $\|D_k\| < \epsilon$, return $\mathcal{C}_{k+1}$. Otherwise set $k \leftarrow k + 1$ and return to step 2.

Step 1 was performed by generating a random N point configuration in S^2. Step 3 was performed using a backtracking line search (i.e. starting with $\alpha_k = 1$ and decrementing at each step) that stopped when α_k satisfied the following condition:

$$E_f\left(\mathcal{C}_k - \alpha_k D_k\right) \leq E_f\left(\mathcal{C}_k\right) - c\alpha_k \mathrm{tr}\left(D_k^T \nabla E_f\left(\mathcal{C}_k\right)\right) \tag{3.5}$$

where $0 < c < 1$ is some tolerance parameter. This condition is referred to as the Armijo condition. The quantity $\mathrm{tr}\left(D_k^T \nabla E_f\left(\mathcal{C}_k\right)\right)$ is the directional derivative of E_f in the direction D_k, and so applying the Armijo condition amounts to searching for the largest step size such that E_f decreases by the amount predicated by the directional derivative (up to some tolerance c).

It should be noted that there is no guarantee that the computed approximate energy minima are *global* minima of E_f. In general, we do not know enough about the structure of the minima of such functions to be able to rule out the possibility that the global minimum has a very small basin of attraction. In fact, [BBC+09] present evidence that there may be local minima with very small basins of attraction for some E_f. However, as long as our computed configurations are close to some *local* minimum of E_f, the following computations remain sensible and informative.

3.2. **Computing The Parameter Count and Symmetry Bound.** Our next step is to compute $P(\mathcal{C})$ and $B(\mathcal{C})$ for each of the approximately energy minimal configurations generated by gradient descent. While the true energy minimal configurations often have enough symmetry to contain many repeated distances, the distances in the approximate, computed configurations are rarely exactly equal. Thus the following computations of $P(\mathcal{C})$ and $B(\mathcal{C})$ are based on a suitable rounding of the distances in the computed configurations.

Recall that the $d-$perturbation of $\mathcal{C}$ is given as

$$\nabla^{(d)}\mathcal{C} = \left(1 - P_{\mathcal{C}}\right)\left[\bigoplus_{x\in\mathcal{C}} \sum_{\substack{y\in\mathcal{C} \\ |x-y|=d}} (x-y)\right]$$

which we computed directly from $\mathcal{C}$ for each $d \in D_{\mathcal{C}}$. Then the parameter count was computed by taking the rank of the matrix $\left(\nabla^{(d)}\mathcal{C}\right)_{d\in D_{\mathcal{C}}}$.

In order to compute the symmetry bound, we first had to compute the symmetry group $S(\mathcal{C})$. Following [BBC+09], we solve an equivalent problem of computing the symmetry group of a graph.

Definition 3.1. An **edge-labelled graph** $G = (E, V, \ell)$ is a graph $G = (V, E)$ along with a function $\ell : E \to \mathbb{R}$ which is called the edge labelling of G.

An **isomorphism** of edge-labelled graphs $G_1 = (V_1, E_1, \ell_1)$ and $G_2 = (V_2, E_2, \ell_2)$ is an invertible mapping $\phi : V_1 \to V_2$ such that

- $(u, v) \in V_1$ if and only if $(\phi(u), \phi(v)) \in V_2$
- $\ell_1(u, v) = \ell_2(\phi(u), \phi(v))$ for all $u, v \in V_1$

Let $\mathcal{C}$ be a point configuration. We can assocate the edge-labelled graph $G_{\mathcal{C}} = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, \ell)$ where

$$\ell(x, y) = \langle x, y\rangle$$

Then in order to compute the symmetry group of $\mathcal{C}$, it suffices to compute the automorphism group of $G_{\mathcal{C}}$.

Proposition 3.2. *Suppose* $\dim \mathrm{span}(\mathcal{C}) = 3$*, then* $S(\mathcal{C})$ *is isomorphic to* $\mathrm{Aut}(G_{\mathcal{C}})$.

Proof. First let ϕ be an automorphism of $G_{\mathcal{C}}$. Then we claim there is a unique $T \in S(\mathcal{C})$ such that $T \mid_{\mathcal{C}} = \phi$. To see this, fix $x_1, x_2, x_3 \in \mathcal{C}$ linearly independent, which we can choose in view of our assumption that the points in $\mathcal{C}$ span $\mathbb{R}^3$. Then there is a unique linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that $T(x_i) = \phi(x_i)$ for $i = 1, 2, 3$.

For any $y, z \in \mathbb{R}^3$, we can write $y = a_1x_1 + a_2x_2 + a_3x_3$ and $z = b_1x_1 + b_2x_2 + b_3x_3$, and then we have

$$\begin{aligned}
\langle Ty, Tz\rangle &= \sum_{i,j=1}^{3} \langle a_i Tx_i, b_j Tx_j\rangle \\
&= \sum_{i,j=1}^{3} a_i b_j \langle Tx_i, Tx_j\rangle \\
&= \sum_{i,j=1}^{3} a_i b_j \langle \phi(x_i), \phi(x_j)\rangle \\
&= \sum_{i,j=1}^{3} a_i b_j \ell(\phi(x_i), \phi(x_j)) \\
&= \sum_{i,j=1}^{3} a_i b_j \ell(x_i, x_j) \\
&= \sum_{i,j=1}^{3} a_i b_j \langle x_i, x_j\rangle \\
&= \langle y, z\rangle
\end{aligned}$$

So T is orthogonal, and thus also invertible. Therefore, Tx_1, Tx_2, Tx_3 form a basis of $\mathbb{R}^3$. And if $x \in \mathcal{C}$, then we have

$$\langle Tx, Tx_i\rangle = \langle x, x_i\rangle = \langle \phi(x), \phi(x_i)\rangle = \langle \phi(x), Tx_i\rangle \tag{3.6}$$

for $i = 1, 2, 3$. Now if we have $Tx - \phi(x) = c_1Tx_1 + c_2Tx_2 + c_3Tx_3$, then we have that

$$\|Tx - \phi(x)\|^2 = \langle Tx - \phi(x), c_1Tx_1 + c_2Tx_2 + c_3Tx_3\rangle = 0$$

by linearity of the inner product and 3.6. Thus we must have $Tx = \phi(x)$ for all $x \in \mathcal{C}$. But this shows both that $T \mid_{\mathcal{C}} = \phi$ and that $T \in S(\mathcal{C})$. Clearly any such T must satisfy $Tx_i = \phi(x_i)$ for $i = 1, 2, 3$, and thus T is unique.

Let T_ϕ denote this transformation, and let $\Psi : \mathrm{Aut}(G_c) \to S(\mathcal{C})$ be given by $\Psi(\phi) = T_\phi$. If $\phi_1, \phi_2 \in \mathrm{Aut}(G_{\mathcal{C}})$, then for any $x \in S(\mathcal{C})$,

$$T_{\phi_1\phi_2}(x) = \phi_1(\phi_2(x)) = \phi_1(T_{\phi_2}(x)) = T_{\phi_1}T_{\phi_2}(x)$$

and so $\Psi(\phi_1\phi_2) = \Psi(\phi_1)\Psi(\phi_2)$, which shows that Ψ is a homomorphism. Now if $T \in S(\mathcal{C})$, then $T \mid_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ is an automorhism of $G(\mathcal{C})$ since $G_{\mathcal{C}}$ is complete and T preserves inner products and hence ℓ. We claim that $\Psi^{-1}(T) = T \mid_{\mathcal{C}}$. To see this, observe that

$$\Psi^{-1}(\Psi(\phi)) = T_\phi \mid_{\mathcal{C}} = \phi$$

by construction of T_ϕ. Likewise

$$\Psi(Psi^{-1}(T)) = \Psi(T \mid_{\mathcal{C}}) = T$$

since T agrees with $T\mid_{\mathcal{C}}$ on $\mathcal{C}$, and we showed that this extension was unique. Thus, $S(\mathcal{C}) \simeq \mathrm{Aut}(G_{\mathcal{C}})$, as claimed. □

The method of constructing T_ϕ in the above proof allows us to compute $S(\mathcal{C})$ from $\mathrm{Aut}\,(G_{\mathcal{C}})$. Thus, we can replace the problem of direclty constructing the $T \in O(3)$ which are symmetries of $\mathcal{C}$ with the combinatorial problem of finding $\mathrm{Aut}\,(G_{\mathcal{C}})$.

In order to carry out this computation, we first converted $G_{\mathcal{C}}$ from an edge-labelled graph to a vertex-labelled graph with the same automorphism group. As in [MP16], we use the vertex-labelled graph $H_{\mathcal{C}} = (V_H, E_H, \ell_H)$ which is constructed from $G_{\mathcal{C}}$ as follows. Let ℓ_G be a labelling of $G_{\mathcal{C}}$ which is equivalent to the inner-product labelling (i.e. gives the same automorphism group) but takes values in the positive integers, let $\{v_i\}_{i=1}^N$ be the vertices of $G_{\mathcal{C}}$, and let $M = \lceil \log_2(|D_{\mathcal{C}}| + 1)\rceil$.

For each $1 \le k \le M$ we form a row of $N = |\mathcal{C}|$ vertices $\{v_i^k\}_{i=1}^N$ and label all the vertices in this row as $\ell_H\left(v_i^k\right) = k$. We then form the edges (v_i^k, v_i^{k+1}) for all $1 \le i \le N$ and $1 \le k \le M-1$, connecting the rows. Then for each $1 \le k \le M$, and $1 \le i < j \le N$, we form the edge (v_i^k, v_j^k) if the k^{th} bit in the binary expansion of $\ell_G(v_i, v_j)$ is 1 (hence why we took the number of rows to be the base 2 logarithm of 1 plus the number of labels in $G_{\mathcal{C}}$).

The resulting graph $H_{\mathcal{C}}$ encodes all of the information that was contained in $G_{\mathcal{C}}$ since the edge colors can be recovered by observing which rows of $H_{\mathcal{C}}$ contain a given edge. It is easy to see that the coloring of the rows and the paths connecting the vertices between rows remove any additional degrees of freedom that adding vertices could have introduced, and indeed $\mathrm{Aut}(H_{\mathcal{C}}) \simeq \mathrm{Aut}(\mathrm{G}_{\mathcal{C}})$.

Once we formed $H_{\mathcal{C}}$, we then used the software package NAUTY, which can compute a collection of permutations that generate the automorphism group of any vertex-labelled graph.

Now in order to compute $B(\mathcal{C})$ from $S(\mathcal{C})$, we constructed a basis for each fixed subspace V_x. To do this, we used the software package GAP, which can perform computations on finitely-generated permutation groups. In particular we supplied the generators found by NAUTY and, for each $O_x \in \mathcal{O}$, we used GAP to find the stabilizer S_x.

For each stabilizer S_x, we converted the permutation representation to a representation in terms of matrices (using the same method as the proof of proposition 3.2). For each such matrix M, let $E_M(\lambda)$ be the eigenspace of M corresponding to eigenvalue λ. Finding V_x then requires finding

$$V_x = \bigcap_{M \in S_x} E_M(1) = \left(\bigoplus_{M \in S_x} E_M(1)^\perp \right)^\perp \tag{3.7}$$

Now if A is a matrix and U is a vector space, we write $A \sim U$ when the columns of A form a basis for U. We also let $\{M_i\}_{i=1}^k$ enumerate S_x. To find a basis for the right hand side of 3.7, we started by computing for each M_i a matrix A_i^x such that $A_i^x \sim E_{M_i}(1)$. Next we computed $B_i^x \sim E_{(A_i^x)^T}(0)$ and formed the matrix $B_x = [B_1^x\ B_2^x\ \cdots\ B_k^x]$, which has as colums a basis for $\oplus_{M \in S_x} E_M(1)^\perp$. Finally we computed $C_x \sim E_{B_x^T}(0)$, which has as columns a basis for V_x. With bases for each V_x, we then directly computed

$$B(\mathcal{C}) = \sum_{O_x \in \mathcal{O}} \dim(V_x) = \sum_{O_x \in \mathcal{O}} \mathrm{rank}(C_x)$$

3.3. **Computational Results.** For 43 of the 52 energy minimal configurations $\mathcal{C}$ we generated, we found that $P(\mathcal{C}) = B(\mathcal{C})$. However, for all of the remaining 9 configurations we found instead that $P(\mathcal{C}) = B(\mathcal{C}) - 1$.

TABLE 1. Parameter count and symmetry information for the 9 configurations for which the parameter count and bound disagreed. The column $S(\mathcal{C})$ lists whether the symmetry group was generated by a reflection or a rotation, and the orbits column lists information about the orbit structure of $\mathcal{C}$ in the format n(number of orbits of size n).

$\lvert\mathcal{C}\rvert$	$P(\mathcal{C})$	$B(\mathcal{C})$	$\lvert S(\mathcal{C})\rvert$	$S(\mathcal{C})$	Orbits
25	24	25	2	reflection	1(5) 2(10)
26	25	26	2	rotation	1(0) 2(13)
33	32	33	2	reflection	1(7) 2(13)
35	33	34	2	rotation	1(1) 2(17)
47	46	47	2	reflection	1(9) 2(19)
49	31	32	3	rotation	1(1) 3(16)
52	33	34	3	rotation	1(1) 3(17)
54	53	54	2	rotation	1(0) 2(27)
55	53	54	2	rotation	1(1) 2(27)

The most important observations about these 9 discrepant configurations are presented in table 1 and the following list:

- Every symmetry group for these configurations has order either 2 or 3. By contrast, every symmetry group of the configurations for which we found agreement between parameter count and bound has order at least 4.
- Every orbit is contained in a proper affine subspace of $\mathbb{R}^3$.
- Those cases in which $B(\mathcal{C}) < |\mathcal{C}|$ are exactly the cases in which $S(\mathcal{C})$ is a rotation group and some point is fixed by the whole group.

These observations indicate that when the symmetry of the configuration degenerates sufficiently, the configuration may contain some additional structure that is not captured by its global symmetries. However, we note that both the parameter count and symmetry group calculations are sensitive to the accuracy of the inner product calculations for our configurations. Because these inner products can be close, it is difficult to know a priori that our approximations of the energy minimal configurations are accurate enough to give exactly the right parameter count and bound in every case. The fact that the count and bound agreed in our computations for most configurations, and that the other discrepancies were small, indicates that some independent verification of these gaps is needed. Indeed, the inequality presented in the next section allows us to rule out the possibility that these gaps are due to some small computational error.

4. A Necessary Condition for $P(\mathcal{C}) = B(\mathcal{C})$

In order to verify these gaps, we prove a simple inequality that allows us to bound the product of $P(\mathcal{C})$ and $|S(\mathcal{C})|$ whenever $P(\mathcal{C}) = B(\mathcal{C})$. We use this necessary condition as a way to check whether the discrepancies listed in table 1 could be computational errors.

4.1. **A Basic Inequality from Symmetry.** We begin by establishing a simple lemma. Define

$$U_x = \left\{ x \in \mathbb{R}^3 \;\middle|\; Tx = x \text{ for all } T \in S_x \right\} \tag{4.1}$$

If V_x is the subspace of $\mathcal{T}_x$ fixed by S_x as in definition 2.9, then clearly $V_x = U_x \cap \mathcal{T}_x$. Using this relation and the properties of orthogonal transformations, we can infer $\dim(V_x)$ from the size of the stabilizer of x.

Lemma 4.1. *Fix $x \in \mathcal{C}$. If V_x, U_x, and S_x are as above, then*

- $\dim(V_x) = \dim(U_x) - 1$
- *If $|S_x| > 2$, then* $\dim(V_x) = 0$.

Proof. Clearly we have $x \in U_x$ by definition. Then we also have

$$\left\{ y \in U_x \;\middle|\; \langle y, x \rangle = 0 \right\} = \operatorname{span}(x)^\perp \cap U_x = \mathcal{T}_x \cap U_x = V_x$$

which shows that

$$U_x = \operatorname{span}(x) \oplus V_x$$

and thus $\dim(V_x) = \dim(U_x) - \dim(\operatorname{span}(x)) = \dim(U_x) - 1$.

Now if $|S_x| > 2$, then we can find two distinct non-identity $T_1, T_2 \in S_x$. Let W_1 and W_2 be the subspaces of $\mathbb{R}^3$ fixed by T_1 and T_2 respectively. We must have $\dim(W_i) < 3$ for $i = 1, 2$ since otherwise $W_i = \mathbb{R}^3$ and $T_i = I$ which contradicts our choice of T_1 and T_2.

If $\dim(W_1) = 2$, then let $w \in W_1^\perp$ with $\|w\| = 1$. Then the fact that T_1 is orthogonal implies that $\|T_1(w)\| = 1$ and $T_1(w) \in W_1^\perp$. But since $T_1 \neq I$, we must have $T_1(w) = -w$, and so T_1 is uniquely determined by W_1 when $\dim(W_1) = 2$ and likewise for T_2. Since T_1 and T_2 are distinct, and since $W_1, W_2 \supset \operatorname{span}(\{x\})$, we therefore must have

$$\dim(W_1 \cap W_2) = 1$$

but then

$$\dim(V_x) = \dim(U_x) - 1 \leq \dim(W_1 \cap W_2) - 1 = 0$$

and we are done. □

The basic tool we need to prove the following inequality is the orbit-stabilizer theorem which we restate here for $S(\mathcal{C})$.

Theorem 4.2. *(Orbit-Stabilizer Theorem) Let $\mathcal{C}$ be a point configuration with symmetry group $S(\mathcal{C})$. For $x \in \mathcal{C}$, let O_x and S_x be the orbit and stabilizer of x as before, then we have*

$$|O_x||S_x| = |S(\mathcal{C})|$$

Now using this and the preceeding lemma, we can prove a basic inequality that relates the symmetry bound to the size of the configuration and the size of its symmetry group.

Proposition 4.3. *(Symmetry Inequality) Let $N = |\mathcal{C}|$, $B = B(\mathcal{C})$, and $M = |S(\mathcal{C})|$, then if $M \geq 2$, we have*

$$\frac{MB}{2} \leq N \quad \textit{and} \quad MB = 0 \pmod 2$$

and if $M = 2$ and $S(\mathcal{C})$ is either generated by a reflection or generated by a rotation which fixes no point in $\mathcal{C}$, we have

$$B = N$$

Proof. Fix $x \in \mathcal{C}$. If $|O_x| = M$, then by the orbit-stabilizer theorem, we have $|S_x| = 1$. But then lemma 4.1 implies that

$$\dim(V_x) = \dim(U_x) - 1 = \dim\left(\mathbb{R}^3\right) - 1 = 2$$

If $|O_x| = M/2$, then again by the orbit-stabilizer theorem, $|S_x| = 2$, and so $\dim(U_x) \leq 2$. Again by the lemma, we then know that $\dim(V_x) \leq 1$. Finally if $|O_x| < M/2$, then $|S_x| > 2$, so $\dim(V_x) = 0$ by the lemma.

Now let n_1 be the number of orbits of size M, and let n_2 be the number of orbits of size $M/2$ for which $\dim(V_x) \neq 0$. Then the above conclusions imply the formula

$$B = \sum_{O_x \in \mathcal{O}} \dim(V_x) = 2n_1 + n_2 \tag{4.2}$$

On the other hand, we know that the collection of orbits $\mathcal{O}$ partitions $\mathcal{C}$. Thus we must also have

$$Mn_1 + \frac{M}{2}n_2 \leq N \tag{4.3}$$

But the left side of 4.3 is just $M/2$ times the right side of 4.2. This shows that

$$\frac{MB}{2} \leq N \tag{4.4}$$

as claimed.

If M is even, then we automatically have $MB = 0 \pmod 2$. If M is odd, then there are no orbits of size $M/2$, so $n_2 = 0$ in 4.2 above, which shows that B is even and therefore again $MB = 0 \pmod 2$.

Now suppose that $M = 2$ and $S(\mathcal{C})$ is generated by a reflection. Then for any $x \in \mathcal{C}$, we have either $|O_x| = M$ or $|O_x| = M/2$. And in the latter case, we must always have $\dim(V_x) = 1$ (since U_x is the plane of reflection). But then we have

$$B = 2n_1 + n_2 = Mn_1 + \frac{M}{2}n_2 = N$$

which is enough to make the inequality 4.4 into an equality. If $M = 2$ and $S(\mathcal{C})$ is generated by a rotation which does not fix any point in $\mathcal{C}$, then $|O_x| = M$ for all $x \in \mathcal{C}$, and then we again have

$$B = 2n_1 = N$$

as claimed. □

4.2. **A Test for Discrepancies Between Parameter Count and Bound.** We now apply these constraints on the symmetries of point configurations to restrict the possibilities for errors in the calculations presented in table 1. Suppose that $\mathcal{C}^*$ is a true energy minimal configuration and $\mathcal{C}$ is our computed approximation to $\mathcal{C}^*$. Recall that $D_{\mathcal{C}}$ is the collection of distances occurring between points in $\mathcal{C}$. We begin by noting that errors introduced by approximating $\mathcal{C}^*$ by $\mathcal{C}$ could result in an error of the form $|D_{\mathcal{C}}| > |D_{\mathcal{C}^*}|$. On the other hand, approximation error almost certainly would never lead to an error of the form $|D_{\mathcal{C}}| < |D_{\mathcal{C}^*}|$. For this to occur, there would have to be $x_1^*, x_2^*, y_1^*, y_2^* \in \mathcal{C}^*$ such that

$$|x_1^* - x_2^*| \neq |y_1^* - y_2^*|$$

while the points $x_1, x_2, y_1, y_2 \in \mathcal{C}$ which approximate the x_i^* and y_i^* would have the property that $|x_1 - x_2|$ would agree with $|y_1 - y_2|$ to many decimal places (over 8 in our case). Any such error can be ruled out by noting that multiple

runs of our gradient descent search with random starting configurations resulted in configurations that had the same computed parameter counts and bounds.

Thus, if we can rule out the possibility of agreement between $P(\mathcal{C})$ and $B(\mathcal{C})$ for values of $P(\mathcal{C})$ close to but not above the calculated values in table 1, we would have strong evidence that the computed discrepancies between parameter count and bound are not merely an artifact of computational error.

In order to do this, we note that if $P(\mathcal{C}) = B(\mathcal{C})$, then we can substitute $P = P(\mathcal{C})$ for B in all of the conclusions of propositon 4.3 above. Thus, if the conclusions fail to hold for given values of $B = P(\mathcal{C})$,M, and N, then we cannot have $P(\mathcal{C}) = B(\mathcal{C})$ for any configuration $\mathcal{C}$ which realizes those values.

For each configuration $\mathcal{C}$ that appears in table 1, we computed the largest value $P^* \leq P(\mathcal{C})$ such that the conlusions of proposition 4.3 are satisfied for any $M \geq 3$ and $B = P^*$. For the case $M = 2$, we note that there is no $\mathcal{C}$ in table 1 such that $P(\mathcal{C}) = |\mathcal{C}|$. Therefore, by the second part of 4.3, we can never have agreement between parameter count P and bound when $P < P(\mathcal{C})$ and $S(\mathcal{C})$ is generated by a reflection or a rotation which fixed no points in $\mathcal{C}$.

Thus the only case not covered by these tests is that of a configuation with order 2 rotational symmetry that fixes at least one point of $\mathcal{C}$. But there are only finitely many such rotations (at most one per $x \in \mathcal{C}$), so we can computationally check if the computed configuration $\mathcal{C}$ is close to a configuration with such a symmetry group. If not, then we can be confident that $\mathcal{C}^*$, the energy minimum we are approximating, does not have this size and symmetry.

In particular, for each $x \in \mathcal{C}$, we took the rotation R_x of π radians around the axis formed by x and the origin. We then computed $\mathcal{C}' = R_x(\mathcal{C})$, and for each $y \in \mathcal{C}'$, we computed

$$d_x(y) = \min_{x \in \mathcal{C}} |x - y|$$

where $|\cdot|$ is just the Euclidean norm. We then compute the value

$$D_x = \max_{y \in \mathcal{C}'} d_x(y)$$

Then D_x is the maximum of the minimal distances between a point in $\mathcal{C}'$ and $\mathcal{C}$. If $\mathcal{C}$ were close to having rotational symmetry R_x, then D_x should be small.

The values in table 2 show the computed values of P^* and $\min_{x \in \mathcal{C}} D_x$ for each discrepant configuration listed in table 1. We can split the configurations in 2 up as follows. For $|\mathcal{C}| \in \{25, 26, 33, 47, 54\}$, the values of P^* are at least 8 less than $P(\mathcal{C})$ and $\min_{x \in \mathcal{C}} D_x$ is not within approximation error of zero. The second number indicates that the true $\mathcal{C}^*$ do not have an order 2 rotational symmetry group that fixes at least one point in $\mathcal{C}$. And the first number indicates that if $P(\mathcal{C}^*) = B(\mathcal{C}^*)$, then there must have been an error of at least 8 in the computation of $P(\mathcal{C})$.

For $|\mathcal{C}| \in \{35, 55\}$, we have P^* much smaller than $P(\mathcal{C})$, but $\min_{x \in \mathcal{C}} D_x$ is 0 because $S(\mathcal{C})$ is an order two rotation group that fixes a point in both cases. If the true configurations $\mathcal{C}^*$ have this symmetry, then since 35 and 55 are odd, there must be exactly one point in $\mathcal{C}^*$ that is fixed (rather than 2). But in that case, the first result of lemma 4.1 tells us that $B(\mathcal{C}^*) = |\mathcal{C}^*| - 1$. In table 1, we have $P(\mathcal{C}) = |\mathcal{C}| - 2$, and we observed that almost certainly $P(\mathcal{C}^*) \leq P(\mathcal{C})$. Thus if $S(\mathcal{C}^*)$ has order two rotation symmetry that fixes a point, then we cannot have agreement between parameter count and bound. If $S(\mathcal{C}^*)$ has higher order, then as before there must be a large error in the calculation of $P(\mathcal{C})$.

TABLE 2. Results of testing each $|\mathcal{C}|$ that appears in table 1 for whether possibly $P(\mathcal{C}^*) = B(\mathcal{C}^*)$. The second column displays the largest value less than or equal to the computed value of $P(\mathcal{C})$ such that the conlusions of propositon 4.3 hold for some $M \geq 3$. The third column displays the minimal distances between the computed configurations $\mathcal{C}$ and their order 2 rotations $\mathcal{C}'$.

$\lvert\mathcal{C}\rvert$	P^*	$\min_{x\in\mathcal{C}} D_x$
25	16	0.327
26	16	0.355
33	22	0.331
35	22	2.446×10^{-7}
47	30	0.257
49	30	0.276
52	32	0.262
54	36	0.224
55	36	1.305×10^{-8}

For $|\mathcal{C}| \in \{49, 52\}$, the values of $\min_{x\in\mathcal{C}} D_x$ are large enough to rule out order two rotational symmetry with a fixed point. However, $P^* = P(\mathcal{C}) - 1$. But as in the last case, we can rule out some additional possibilities. If $|S(\mathcal{C}^*)| = 3$, then $S(\mathcal{C})$ must be a rotation group, and so there can be at most 2 fixed points in $\mathcal{C}^*$. But we can see directly that this only happens if there is one fixed point and 16 orbits of size 3 (for 49 points) or 17 orbits of size 3 (for 52 points). Again lemma 4.1 tells us that the bounds from symmetry must then be 32 and 34 respectively, which are larger than the respective computed values of $P(\mathcal{C})$. Since $P(\mathcal{C}^*) \leq P(\mathcal{C})$, we cannot have agreement between parameter count and bound for these configurations if $|S(\mathcal{C}^*)| = 3$.

We can then look for the largest value of $P^* \leq P$ such that the conclusions of propositon 4.3 are satisfied for some $M \geq 4$, which yields $P^* = 24$ for $|\mathcal{C}| = 49$ and $P^* = 26$ for $|\mathcal{C}| = 52$. These numbers are both 8 less than the computed $P(\mathcal{C})$, and so we can again conclude that if $P(\mathcal{C}^*) = B(\mathcal{C}^*)$, then there must have been a significant error in the calculation of $P(\mathcal{C})$.

Thus the computations in table 2 strongly indicate that the calculated discrepancies between parameter count and bound listed in table 1 are real and not merely the result of computational error.

References

[Ale13] Alex Alexanderian. On the continuous dependence of roots of polynomials on coefficients. 2013. URL: http://www4.ncsu.edu/~aalexan3/articles/polyroots.pdf.

[BBC+09] Brandon Ballinger, Grigoriy Blekherman, Henry Cohn, Noah Giansiracusa, Elizabeth Kelly, and Achill Schrmann. Experimental study of energy-minimizing point configurations on spheres. *Experimental Mathematics*, 18(3):257–283, 2009. URL: https://doi.org/10.1080/10586458.2009.10129052.

[BV09] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009. URL: https://web.stanford.edu/~boyd/cvxbook/.

[CK06] Henry Cohn and Abhinav Kumar. Universally optimal distribution of points on spheres. *Journal of the American Mathematical Society*, 20(1):99–148, 2006. URL: https://arxiv.org/abs/math/0607446.

[Coh16] Henry Cohn. Packing, coding, and ground states. 2016. URL: https://arxiv.org/abs/1603.05202.

[GAP18] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.8.7*, 2018. URL: https://www.gap-system.org.

[MP16] Brendan D. McKay and Adolfo Piperno. *nauty and Traces User's Guide*, 2016. URL: http://pallini.di.uniroma1.it/Guide.html.

[Tag11] Hement Tagare. Notes on optimization on stiefel manifolds. 2011. URL: http://noodle.med.yale.edu/hdtag/notes/steifel_notes.pdf.

Brown University, Providence, RI 02912