Supplementary Materials for Structure Determination of Symmetric Homo-oligomers by a Complete Search of Symmetry Configuration Space Using NMR Restraints and van der Waals Packing

Shobha Potluri* Anthony K. Yan* James J. Chou† Bruce R. Donald* ‡ § ¶
Chris Bailey-Kellogg* ¶

July 14, 2006

S1 Introduction

In these Supplementary Materials, we give the mathematical derivation for the bounding regions of $G\mathbf{q}$, which are used in the subsection entitled *Bounding* which is inside the *Methods* section of the main paper. We have organized these Supplementary Materials as follows. The geometric structure of $G\mathbf{q}$ is described in Section S2. From this geometric structure, we derive properties of the convex hull of $G\mathbf{q}$ in Section S3. Next, in Section S4, we use the geometric hull of $G\mathbf{q}$ to reduce our problem to that of finding a bounding region for rotations $A\mathbf{k}$. We have two bounds for rotations. The first rotational bound, $V(A, \mathbf{k})$, is described in Section S5 and leads to a bounding region, $W(G, \mathbf{q})$, for $G\mathbf{q}$. The second rotation bound is tighter than $V(A, \mathbf{k})$, and is shown in Section S6. Because the second rotation bound is tighter, it yields a tighter bounding volume for $G\mathbf{q}$.

S2 Geometric Structure of $G\mathbf{q}$

We begin by formally specifying $G\mathbf{q}$ and then exploiting its geometric structure.

$$G = A \times T \subseteq S^2 \times \mathbb{R}^2 \tag{1}$$

$$T = [t_{x_1}, t_{x_2}] \times [t_{y_1}, t_{y_2}] \tag{2}$$

$$A = \left\{ \begin{bmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{bmatrix} \middle| \, \theta \in [\theta_1, \theta_2] \, , \phi \in [\phi_1, \phi_2] \right\}. \tag{3}$$

Let $R_{\mathbf{a}}(\alpha)$ be a rotation about the unit vector $\mathbf{a} \in S^2$ by α radians. To rotate a point $\mathbf{q} \in \mathbb{R}^3$ by α radians, about an axis that is parallel to $\mathbf{a}$ and goes through the point $\mathbf{t} \in \mathbb{R}^3$, we compute $R_{\mathbf{a}}(\alpha)(\mathbf{q} - \mathbf{t}) + \mathbf{t}$. Notice that $R_{\mathbf{a}}(\alpha)(\mathbf{q} - \mathbf{t}) + \mathbf{t} = [R_{\mathbf{a}}(\alpha) - I](\mathbf{q} - \mathbf{t}) + \mathbf{q}$, where I is the identity matrix. So we have

$$G\mathbf{q} = \{[R_{\mathbf{a}}(\alpha) - I](\mathbf{q} - \mathbf{t}) + \mathbf{q} \mid \mathbf{a} \in A \, , \mathbf{t} \in T\} \, . \tag{4}$$

Before proceeding, we generalize our operations to sets in the usual way. Let $D, F \subset \mathbb{R}^3$. Let $R_{\mathbf{A}}(\alpha)\mathbf{t} = \{R_{\mathbf{a}}(\alpha)\mathbf{t} \mid \mathbf{a} \in A\}$. Let $R_{\mathbf{a}}(\alpha)D = \{R_{\mathbf{a}}(\alpha)\mathbf{d} \mid \mathbf{d} \in D\}$. Let $R_A(\alpha)D = \{R_{\mathbf{a}}(\alpha)\mathbf{d} \mid \mathbf{d} \in D, \mathbf{a} \in A\}$. Let

*Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA
†Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
‡Department of Chemistry, Dartmouth College, Hanover, NH 03755, USA
§Departmen t of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA
¶Corresponding authors. 6211 Sudikoff Laboratory, Hanover, NH 03755. Email: {brd,cbk}@cs.dartmouth.edu.

$AD = R_A(\alpha)D$. Let $A\mathbf{t} = R_A(\alpha)\mathbf{t}$. Let $\mathbf{t}+D = \{\mathbf{t}+\mathbf{d} \mid \mathbf{d} \in D\}$. Let $D+F = \{\mathbf{d}+\mathbf{f} \mid \mathbf{d} \in D, \mathbf{f} \in F\}$. Let $-T = \{-\mathbf{t} \mid \mathbf{t} \in T\}$.

We can now rewrite equation (4) as

$$G\mathbf{q} = \big\{[R_{\mathbf{a}}(\alpha) - I](\mathbf{q} + (-T)) + \mathbf{q} \;\mid\; \mathbf{a} \in A\big\}. \tag{5}$$

Notice that $\mathbf{q} + (-T)$ is a solid rectangle $-T$ that has been translated in $\mathbb{R}^3$ by $\mathbf{q}$. We further note that $[R_{\mathbf{a}}(\alpha) - I]$ is a linear mapping. Under a linear mapping, any solid rectangle becomes a solid parallelogram. Consequently, $G\mathbf{q}$ is the union of infinitely many solid parallelograms. Let our solid rectangle be $D = \mathbf{q} + (-T)$. Let the solid parallelograms be $F_{\mathbf{a}} = [R_{\mathbf{a}}(\alpha) - I]D + \mathbf{q}$. Hence,

$$G\mathbf{q} = \big\{[R_{\mathbf{a}}(\alpha) - I]D + \mathbf{q} \;\mid\; \mathbf{a} \in A\big\} = \bigcup_{\mathbf{a}\in A} F_{\mathbf{a}}. \tag{6}$$

Equation (6) represents the geometric structure of $G\mathbf{q}$ and we will use it below.

S3 Geometric Structure of the Convex Hull of $G\mathbf{q}$

In this section, we consider the convex hull of $G\mathbf{q}$. As shown above, $G\mathbf{q}$ is the union of infinitely many solid parallelograms. Let Q be the set of all the corners of all the parallelograms $F_{\mathbf{a}}$ where $\mathbf{a} \in A$. Intuitively, one can see that the convex hull of Q is the same as the convex hull of $G\mathbf{q}$. In this section, we formally derive this result.

To begin, we make the trivial observation that a solid parallelogram is the convex hull of its four corners. Let $H(P)$ denote the convex hull of $P \subset \mathbb{R}^3$. Let $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4$ be the four corners of the solid rectangle $D = \mathbf{q} + (-T)$. Then we have

$$D = H(\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4\}). \tag{7}$$

We now consider what happens to each corner $\mathbf{k}_i$ after the linear mapping $R_{\mathbf{a}}(\alpha) - I$. From equation (6), we can write $G\mathbf{q}$ as

$$G\mathbf{q} = \bigcup_{\mathbf{a}\in A} [R_{\mathbf{a}}(\alpha) - I]D + \mathbf{q} = [R_A(\alpha) - I]D + \mathbf{q}. \tag{8}$$

Before continuing, we would like to provide a simple intuitive explanation of the results we are deriving. See Figure 1. Let $M_{\mathbf{a}}(\alpha) = [R_{\mathbf{a}}(\alpha) - I]$. Let $M_A(\alpha) = [R_A(\alpha) - I]$.

$$\text{Let } C_i = M_A\mathbf{k}_i + \mathbf{q}. \tag{9}$$

C_i is the set of all possible images of the ith corner of D under the mapping in Equation (8). From Section S2 we know that $G\mathbf{q}$ is the union of infinitely many solid parallelograms in $\mathbb{R}^3$. Because a solid parallelogram is simply the convex hull of its four corners, it follows that *any* convex hull which contains the corners of the parallelogram must completely enclose the entire solid parallelogram. As a result, any convex hull which covers C_i for $i = 1, 2, 3, 4$ will cover all corners of all parallelograms, and therefore cover all of $G\mathbf{q}$. In fact, the convex hull $H(\bigcup_{i=1}^4 C_i)$ is equal to the convex hull $H(G\mathbf{q})$. We now apply the following property of convex hulls: $H\left(\bigcup_{i=1}^n P_i\right) = H\left(\bigcup_{i=1}^n H(P_i)\right)$ for any $P_i \subset \mathbb{R}^3$. Using this property, we obtain $H(G\mathbf{q}) = H(\bigcup_{i=1}^4 H(C_i))$. This is the main result we derive in this section (see equation 14). Below, we give the formal derivation.

We use the fact that, if M is a linear operator, then $H(MP) = MH(P)$ for any $P \subset \mathbb{R}^3$. From

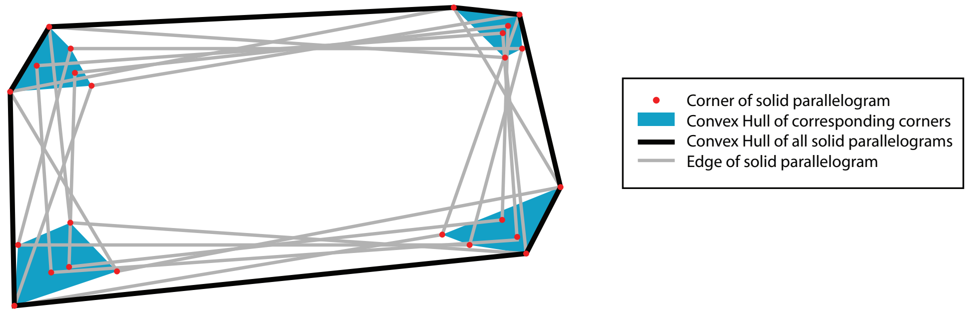

Figure 1: **Cartoon Diagram of the Convex Hull of $G\mathbf{q}$.** As described in Section S2, we know that $G\mathbf{q}$ is the union of infinitely many solid parallelograms. In this figure, we have drawn a cartoon of this fact in two dimensions for the sake of clarity. In addition, we have simplified our picture by drawing only six parallelograms instead of infinitely many. In actuality, there are infinitely many solid parallelograms that live in three dimensions, and the geometric property we are deriving is true in three dimensions. In Section S3 we have defined $C_i = [R_A(\alpha) - I]\mathbf{k}_i + \mathbf{q}$ where $\mathbf{k}_i$ are corners of our rectangle D (see Section S3 for details). The red dots represents the corners of the parallelograms, and the gray lines represent the edges of the rectangles. The blue areas are $H(C_i)$, namely the convex hulls of the corners. The outer solid black line represents the convex hull of $G\mathbf{q}$. Notice that the convex hull for all of the $H(C_i)$ is equal to the convex hull of $G\mathbf{q}$.

equations (7) and (8), we have

$$G\mathbf{q} = M_A(\alpha)H(\{\mathbf{k}_1, \mathbf{k_t}, \mathbf{k}_3, \mathbf{k}_4\}) + \mathbf{q} = \bigcup_{\mathbf{a}\in A} M_{\mathbf{a}}(\alpha) H\left(\bigcup_{i=1}^{4}\{\mathbf{k}_i\}\right) + \mathbf{q} \tag{10}$$

$$= \bigcup_{\mathbf{a}\in A} H\left(M_{\mathbf{a}}(\alpha)\bigcup_{i=1}^{4}\{\mathbf{k}_i\}\right) + \mathbf{q} = \bigcup_{\mathbf{a}\in A} H\left(\bigcup_{i=1}^{4}\{M_{\mathbf{a}}(\alpha)\mathbf{k}_i\}\right) + \mathbf{q}. \tag{11}$$

We now consider the convex hull.

$$H(G\mathbf{q}) = H\left(\bigcup_{\mathbf{a}\in A} H\left(\bigcup_{i=1}^{4}\{M_{\mathbf{a}}(\alpha)\mathbf{k}_i\}\right)\right) + \mathbf{q} = H\left(\bigcup_{\mathbf{a}\in A}\bigcup_{i=1}^{4}\{M_{\mathbf{a}}(\alpha)\mathbf{k}_i\}\right) + \mathbf{q} \tag{12}$$

$$= H\left(\bigcup_{i=1}^{4}\bigcup_{\mathbf{a}\in A}\{M_{\mathbf{a}}(\alpha)\mathbf{k}_i\}\right) + \mathbf{q} = H\left(\bigcup_{i=1}^{4} M_A(\alpha)\mathbf{k}_i + \mathbf{q}\right) \tag{13}$$

$$= H\left(\bigcup_{i=1}^{4} C_i\right) = H\left(\bigcup_{i=1}^{4} H(C_i)\right) \tag{14}$$

Equation (14) represents the geometric structure of the convex hull of $G\mathbf{q}$. We use this result in the next section to derive a bounding volume for $H(G\mathbf{q})$.

S4 Bounding Regions for $H(G\mathbf{q})$

Equation (14) is a bounding volume for $G\mathbf{q}$. Instead of computing the convex hull exactly, we approximate it. We do this by replacing each $H(C_i)$ with an axis-aligned bounding box (AABB), B_i, that completely encloses $H(C_i)$. The details for how to compute B_i are given in Section S5 and Section S6. Because $H(C_i) \subset B_i$, we have

$$G\mathbf{q} \subset H(G\mathbf{q}) = H\left(\bigcup_{i=1}^{4} H(C_i)\right) \subseteq H\left(\bigcup_{i=1}^{4} B_i\right). \tag{15}$$

From Equation (15), one can see that we have an approximation of the convex hull $H(G\mathbf{q})$.

Note that the B_i are bounding regions for $H(C_i)$. If we are given *any* bounding regions for each $H(C_i)$, then we can construct a bounding region for $G\mathbf{q}$. To get a bounding region for C_i, we consider its detailed structure (Equation 9).

$$C_i = M_A\mathbf{k}_i + \mathbf{q} = [R_A(\alpha) - I]\mathbf{k}_i + \mathbf{q} = R_A(\alpha)\mathbf{k}_i + (\mathbf{q} - \mathbf{k}_i). \tag{16}$$

The points $\mathbf{q}$ and $\mathbf{k}_i$ are fixed constants (for a given NOE constraint and a given grid cell G). From equation (16), we see that C_i is identical to $R_A(\alpha)\mathbf{k}_i$ except that it has been translated by $(\mathbf{q}-\mathbf{k}_i)$. Consequently, finding a bounding region for C_i is equivalent to finding a bounding region for $R_A(\alpha)\mathbf{k}_i$. *Given a method for bounding $R_A(\alpha)\mathbf{k}_i$, we can construct a bounding region for $G\mathbf{q}$.*

We define our bounding volume $W(G, \mathbf{q})$ as follows. Let $V(A, \mathbf{k})$ be a bounding region for $R_A(\alpha)\mathbf{k}$ so that $R_A(\alpha)\mathbf{k} \subset V(A, \mathbf{k})$.

$$G\mathbf{q} \subset W(G, \mathbf{q}) = H\left(\bigcup_{i=1}^{4} B_i\right) = H\left(\bigcup_{i=1}^{4} V(A, \mathbf{k}_i) + \mathbf{q} - \mathbf{k}_i\right). \tag{17}$$

If we replace $V(A, \mathbf{k}_i)$ with a tighter bound for $R_A(\alpha)\mathbf{t}$, then equation (17) yields a tighter bounding volume for $G\mathbf{q}$.

S5 Quick Rotation Bound $V(A, \mathrm{k})$ for $A\mathrm{k}$

In this section, we present a quick way to compute the bounding volume $V(A, \mathbf{k})$ for $R_A(\alpha)\mathbf{k}$. Consider Figure 2. Let $\mathbf{a} \in S^2$ be an arbitrary axis of rotation ($\|\mathbf{a}\| = 1$). Let $\mathbf{k} \in \mathbb{R}^3$ be an arbitrary point. Let $\mathbf{u} = R_{\mathbf{a}}(\alpha)\mathbf{k}$. Let $\epsilon_0 \in [0, 2]$ be a constant which represents the maximal variation in the direction of the axis of rotation. Let $\boldsymbol{\epsilon} \in \mathbb{R}^3$ be a vector where $\|\boldsymbol{\epsilon}\| \leq \epsilon_0$ and $\|\mathbf{a} + \boldsymbol{\epsilon}\| = 1$. We use $\boldsymbol{\epsilon}$ to represent a small "perturbation" of the direction of the axis of rotation. Let $\mathbf{w} = R_{\mathbf{a}+\boldsymbol{\epsilon}}(\alpha)\mathbf{k}$. Without any perturbation, a rotation about axis $\mathbf{a}$ gives us $\mathbf{u}$. If we perturb the axis by $\boldsymbol{\epsilon}$, we get a (possibly) different point $\mathbf{w}$. (Figure 2.)

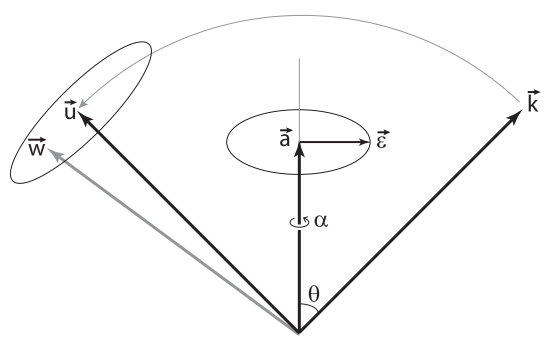

Figure 2: **Applying a set of rotations.** We consider rotating an arbitrary vector $\mathbf{k}$ by a rotation $R_{\mathbf{a}}(\alpha)$, where the angle of rotation α is fixed, but the axis of rotation $\mathbf{a}$ can range over a small spherical cap of size ϵ_0. We can think of this as perturbing $\mathbf{a}$ by a small vector $\boldsymbol{\epsilon}$ where we require $\|\mathbf{a}\| = \|\mathbf{a} + \boldsymbol{\epsilon}\| = 1$, and $\|\epsilon\| \leq \epsilon_0$. Let $\mathbf{u} = R_{\mathbf{a}}(\alpha)\mathbf{k}$. Let $\mathbf{w} = R_{\mathbf{a}+\epsilon}(\alpha)\mathbf{k}$. We require an upper bound on $\|\mathbf{u} - \mathbf{w}\|$.

An upper bound on $\|\mathbf{w} - \mathbf{u}\|$ can be used to find a bounding volume for $R_A(\alpha)\mathbf{k}$. Let $r \geq \|\mathbf{w} - \mathbf{u}\|$ be an upper bound. If we choose $\mathbf{a} \in A$ and find an ϵ_0 such that every element in A is ϵ_0 close to $\mathbf{a}$, then $R_A(\alpha)\mathbf{k}$ will be completely contained within a sphere of radius r centered at $\mathbf{u}$.

We can reduce the bounding volume even farther, because rotations cannot change the length of a vector. Therefore, $\mathbf{w}$ and $\mathbf{u}$ must lie on a sphere centered at the origin with a radius of $\|\mathbf{k}\|$. The intersection of a solid ball of radius r centered at $\mathbf{u}$ and a spherical shell centered at the origin with radius $\|\mathbf{k}\|$ gives us a filled circular patch of the spherical shell, that is a "spherical cap." In terms of latitude and longitude, the spherical cap is the area around a "north" pole which has latitude greater than or equal to some fixed value.

The basis for our bound is a vector equation for rotations. A derivation of Equation (18) can be found in [1], Chapter 4, pages 164-165.

$$\mathbf{u} = R_{\mathbf{a}}(\alpha)\mathbf{k} = (\mathbf{k}\cdot\mathbf{a})\mathbf{a} + (\sin\alpha)(\mathbf{a}\times\mathbf{k}) + (\cos\alpha)\left[\mathbf{k} - (\mathbf{k}\cdot\mathbf{a})\mathbf{a}\right]. \tag{18}$$

We write down the vector equations for $\mathbf{u} = R_{\mathbf{a}}(\alpha)\mathbf{k}$ and for $\mathbf{w} = R_{\mathbf{a}+\boldsymbol{\epsilon}}(\alpha)\mathbf{k}$, and then subtract to get an exact expression for $\mathbf{w}-\mathbf{u}$. Finally, we find an upper bound on the magnitude $\|\mathbf{w}-\mathbf{u}\|$.

$$\mathbf{u} = R_{\mathbf{a}}(\alpha)\mathbf{k} = (\mathbf{k}\cdot\mathbf{a})\mathbf{a} + (\sin\alpha)(\mathbf{a}\times\mathbf{k}) + (\cos\alpha)\left[\mathbf{k} - (\mathbf{k}\cdot\mathbf{a})\mathbf{a}\right] \tag{19}$$

$$\mathbf{w} = R_{\mathbf{a}+\boldsymbol{\epsilon}}(\alpha)\mathbf{k} = [\mathbf{k}\cdot(\mathbf{a}+\boldsymbol{\epsilon})](\mathbf{a}+\boldsymbol{\epsilon}) + (\sin\alpha)[(\mathbf{a}+\boldsymbol{\epsilon})\times\mathbf{k}] + (\cos\alpha)\Big[\mathbf{k} - \big(\mathbf{k}\cdot[\mathbf{a}+\boldsymbol{\epsilon}]\big)(\mathbf{a}+\boldsymbol{\epsilon})\Big] \tag{20}$$

After subtracting $\mathbf{u}$ from $\mathbf{w}$ and performing some algebra, we obtain

$$\mathbf{w}-\mathbf{u} = (\sin\alpha)(\boldsymbol{\epsilon}\times\mathbf{k}) + (1-\cos\alpha)\Big[(\mathbf{k}\cdot\mathbf{a})\boldsymbol{\epsilon} + (\mathbf{k}\cdot\epsilon)(\mathbf{a}+\boldsymbol{\epsilon})\Big]. \tag{21}$$

We now bound the magnitude $\|\mathbf{w}-\mathbf{u}\|$ using very simple upper bounds. We apply the triangle inequality, the Pythagorean theorem, and elementary bounds on the magnitude of cross and dot products. Notice that in equation (21) that the term $(\sin\alpha)(\boldsymbol{\epsilon}\times\mathbf{k})$ is perpendicular to the term $(1-\cos\alpha)(\mathbf{k}\cdot\mathbf{a})\boldsymbol{\epsilon}$. Recall that we restrict $\boldsymbol{\epsilon}$ so that we have $\|\mathbf{a}+\boldsymbol{\epsilon}\| = 1$. We don't simplify $(\mathbf{k}\cdot\mathbf{a})$ because both $\mathbf{k}$ and $\mathbf{a}$ are known, so there is no need to approximate.

$$\|\mathbf{w}-\mathbf{u}\| \leq r(\epsilon_0, \mathbf{a}, \alpha, \mathbf{k}) \tag{22}$$

where $r(\epsilon_0, \mathbf{a}, \alpha, \mathbf{k}) = \epsilon_0\Big(\sqrt{(\sin\alpha)^2\|\mathbf{k}\|^2 + (1-\cos\alpha)^2(\mathbf{k}\cdot\mathbf{a})^2} + |1-\cos\alpha|\,\|\mathbf{k}\|\Big)$.

We now have a geometric bounding shape for $R_A(\alpha)\mathbf{k}$. Choose $\epsilon_0 > 0$ and $\mathbf{a}\in A$ so that every element of A is within ϵ_0 of $\mathbf{a}$. Let $B(\mathbf{a}, r)$ be a solid ball centered at $\mathbf{a}$ with radius r. Let $S(\mathbf{a}, r)$ be a sphere centered at $\mathbf{a}$ with radius r.

$$R_A(\alpha)\mathbf{k} \subseteq U(A,\mathbf{k}), \quad \text{where} \quad U(A,\mathbf{k}) = B\big(R_{\mathbf{a}}(\alpha)\mathbf{k}, r(\epsilon_0,\mathbf{a},\alpha,\mathbf{k})\big) \cap S(\mathbf{0}, \|\mathbf{k}\|). \tag{23}$$

The vector $\mathbf{0}$ is the origin. The bounding volume (23) is a spherical cap. It is straight-forward to compute the smallest axis-aligned bounding box (AABB) which completely encloses the spherical cap (23).

$$R_A(\alpha)\mathbf{k} \subset V(A,\mathbf{k}), \quad \text{where} \quad V(A,\mathbf{k}) = \text{ the axis-aligned bounding box of } U(A,\mathbf{k}) \tag{24}$$

Our bounding volumes $U(A,\mathbf{k})$ and $V(A,\mathbf{k})$ can be computed very rapidly using equation (22). Unfortunately, the bound is fairly loose. In practice, the bound (22) is too large by about a factor of one to two for the majority of cases with a distribution extending out to about a factor of five. Because this bound is loose, we find a tighter bound below, although it is somewhat more expensive to compute.

S6 Tighter Rotation Bound for $A\mathbf{k}$

In this Section, our goal is to find the smallest axis-aligned bounding box (AABB) which contains $A\mathbf{k} =$

$R_A(\alpha)\mathbf{k}$. This is equivalent to finding the extrema

$$\max_{\mathbf{a}\in A} u_i(\mathbf{a}) \quad \text{and} \quad \min_{\mathbf{a}\in A} u_i(\mathbf{a}), \quad \text{where} \quad i = 1,2,3 \quad \text{and} \quad \mathbf{u}(\mathbf{a}) = \begin{bmatrix} u_1(\mathbf{a}) \\ u_2(\mathbf{a}) \\ u_3(\mathbf{a}) \end{bmatrix} = R_{\mathbf{a}}(\alpha)\mathbf{k}. \tag{25}$$

The extrema of $u_i(\mathbf{a})$ specify the smallest possible axis-aligned bounding box (AABB) for $A\mathbf{k}$. To find these extrema, we form a grid over A which has a mesh-resolution of at most γ radians (in practice, we set $\gamma = 0.01$ radians), and for each grid point, we perform numerical gradient ascent and descent to find local extrema. We then find global extrema by taking the maximum and minimum of the local extrema.

References

[1] Goldstein H. Classical Mechanics. Addison-Wesley, Reading, Massachusetts, second edition, 1980.