LOCATING PERSPECTIVE THREE-POINT PROBLEM SOLUTIONS IN SPATIAL REGIONS

Michael Q. Rieck[1]
[1]*Mathematics and Computer Science Department, Drake University, Des Moines, IA 50310, USA*
mrieck@drake.edu

Bo Wang[2†]
[2]*Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100190*
bo.wang@ia.ac.cn

Keywords: P3P: perspective: pose: camera: tracking: danger cylinder: deltoid: toroid

Abstract: The distribution of the solutions to the Perspective 3-Point Problem (P3P) has been studied for a few decades, and some understanding of this issue has emerged. However, the present article is the first to comprehensively describe, for a given location p in space, the number of other points that solve the same P3P setup that p solves, and where to find these related points. A dynamic approach is employed to solve this problem. The related points are restricted to certain regions in space, defined by certain "basic" toroids, and by a surface called the "companion surface to the danger cylinder." The nature of this surface is explored in detail, along with its intersections with the basic toroids, and the pairwise intersection of these toroids. The cubic polynomial introduced by Finsterwalder in his analysis of P3P is also related to the companion surface.

1 INTRODUCTION

The *Perspective 3-Point Problem,* whose purpose is to determine the pose of an idealized pinhole camera, based only on an image showing three known *control points* in space, is a rather old problem. It has long been understood that although the information provided as input to the problem is insufficient to exactly determine the location of the camera, it does limit the possibilities to at most four locations. However, until recently, it has been a mystery how these points are geometrically related.

In (Gao et al., 2003), a significant step was taken in gaining an understanding of these matters, by classifying the various possible solutions to the system of equations introduced in (Grunert, 1841). Another such algebraic approach to the issue was reported in (Faugère et al., 2008). Yet these works did not reveal the significant geometric aspects inherent in the P3P problem that have steadily emerged in the last couple decades, and that have proven to be essential in gaining a deeper understanding of P3P.

The so-called *danger cylinder* is a very important surface, and while this has been widely known, for some time, certain key facts concerning it have only recently been firmly established. See (Rieck, 2014) and (Zhang and Hu, 2006). It has also become clear that certain special toroids aid in understanding the number and relative positions of the various solution points to the P3P problem. See (Sun and Wang, 2010) and (Wang, et al, 2, 2019).

Quite recently, another important surface has been revealed, and shown to also play a crucial role in understand the number of solutions and their locations. In (Rieck, 2018), this surface was called the "deltoidal surface," but in (Wang, et al, 1, 2019), it was

† Research supported by National Key R&D Program of China under Grant 2020AAA0109600 and by National Natural Science Foundation of China under Grant 61873264.

called the *companion surface to the danger cylinder (CSDC)*. We will here adopt this latter terminology, as it stresses an important relationship to the danger cylinder.

Together, these various surfaces act to partition all of real three-dimensional space into various regions. One of the primary goals of this paper is to precisely determine the number of solutions associated with each region. This goal is very much achieved, in Section 5, via a series of lemmas, culminating in Theorem 3, which counts the number of solutions for various cases.

One aspect of our analysis that is rather novel is the reliance on *particles*, that is, continuously moving points. We introduce, in Section 2, the concepts of *weakly related* and *strongly related* particles, as well as fixed points, and exploit these ideas to produce a rather complete analysis of the P3P problem. Strongly related points are simply points that solve the P3P problem for the same parameter values, so this is of primary concern.

However, in allowing particles to continuously move around, and in developing a rigorous mathematical analysis of this, it becomes necessary to permit particles to pass through control points. In doing so, any strong relationship between particles breaks down. However, weak relationships are maintained, which is part of the justification in also focusing on this relationship.

Section 3 introduces (in this paper) the companion surface of the danger cylinder, explains its importance, and investigates many of its properties. Section 4 introduces (in this paper) the *basic toroids* and *double toroids* that, together with the *CSDC*, partition space for the analysis developed in Section 5. Section 4 also discusses the algebraic geometrical notion of a point *blowup* and a natural *double cover* of such a blowup. The immediate neighborhoods of the control points are investigated using these tools. Section 6 provides precise formulas that describe the pairwise intersections of the various important surfaces related to the P3P problem. Section 7 draws a connection between the *CSDC*, repeated solutions to Grunert's system of equations, and the discriminant of a cubic polynomial introduced by Finsterwalder, and described in (Haralick et al., 1994).

2 STRONGLY AND WEAKLY RELATED POINTS

As was the practice in (Rieck, 2015) and (Rieck, 2018), the analysis of the P3P problem will be greatly simplified by making some easily accommodated assumptions that in no way restrict the general utility of the results. Specifically, it will be assumed that three-dimensional real space is equipped with three orthogonal coordinate axes (x,y,z), that these are chosen, translated, rotated and scaled together as needed, so as to cause the three control points to be located on the unit circle in the xy-plane.

Mathematically, we will be concerned with solutions to the following system of equations, which we call the *extended Grunert system:*

$$\left\{\begin{array}{rcl} s_2^2+s_3^2-2\cos\alpha\, s_2 s_3 &=& (x_2-x_3)^2+(y_2-y_3)^2 \\ s_3^2+s_1^2-2\cos\beta\, s_3 s_1 &=& (x_3-x_1)^2+(y_3-y_1)^2 \\ s_1^2+s_2^2-2\cos\gamma\, s_1 s_2 &=& (x_1-x_2)^2+(y_1-y_2)^2 \\ (x-x_1)^2+(y-y_1)^2+z^2 &=& s_1^2 \\ (x-x_2)^2+(y-y_2)^2+z^2 &=& s_2^2 \\ (x-x_3)^2+(y-y_3)^2+z^2 &=& s_3^2. \end{array}\right. \tag{1}$$

Here $(x_1,y_1,0)$, $(x_2,y_2,0)$ and $(x_3,y_3,0)$ are the known coordinates of the control points, and α, β and γ are known *view angles*. The camera's *optical center* (x,y,z) is unknown, as are the distances s_1, s_2 and s_3 between the optical center and the control points. The view angles are the angles between the rays from the optical center that pass through the control points, taken in pairs. All angles discussed in this paper have radian values between zero and π. It will be handy to let d_1 denote the distance $[(x_2-x_3)^2+(y_2-y_3)^2]^{1/2}$, and similarly for d_2 and d_3.

In practice, the quantities (x,y,z,s_1,s_2,s_3) for the actual optical center of the camera give only one possible solution to the system, for the corresponding parameters $(x_1,y_1,x_2,y_2,x_3,y_3,\alpha,\beta,\gamma)$. Though there may perhaps be some potential for confusion, we will often let the notation (x,y,z,s_1,s_2,s_3) refer to any of the several solutions to the system (1). Downplaying the distances s_1,s_2,s_3, the corresponding triple (x,y,z) will be called a *solution point* for the system (1).

The subsystem consisting of just the first three equations of (1) is called the *Grunert system*. A solution (s_1,s_2,s_3) to this might involve only positive real numbers, in which case, it is always possible to obtain a corresponding real solution point (x,y,z) to the extended system. Solutions of this sort are physically meaningful, *viz.* the P3P problem, of course.

However, there is also a physical significance that can be associated with a solution (s_1, s_2, s_3) involving only real numbers, at least one of which is negative. To understand this, imagine, say, that $s_1 > 0$, $s_2 > 0$, but $s_3 < 0$. Notice the mathematical fact that if we alter the system of equations by replacing α with $\pi - \alpha$, and replacing β with $\pi - \beta$, we find that $(s_1, s_2, |s_3|)$ is a solution to the resulting Grunert system, and that this involves only positive numbers. If we now obtain a corresponding solution point (x, y, z), then we might think that of it as being a "solution point" for the original system, *except* with the ray between the optical center and the third control point pointing in the reverse direction.

Because of the intimate relationship between such systems, and the importance of this when we later allow the optical center to move dynamically around in space, we will say that a Grunert system of equations obtained from the original Grunert system, by replacing exactly *two* of the angles α, β and γ with their supplementary angles $\pi - \alpha$, $\pi - \beta$ and $\pi - \gamma$, is *related* to the original Grunert system. Similarly for the extended Grunert system.

Given two solution points (x, y, z) and (x', y', z') for the extended Grunert system, using the same parameter values, we say that these two points are *strongly related* (to each other). More generally, given a solution point (x, y, z) for an extended Grunert system, and a point (x', y', z') that is either a solution point for the same system or for a related system, we will say that these two points are *weakly related* (to each other). Notice that strongly related points are also weakly related. While the notion of weakly related points may seem at first to be an unnecessary contrivance, it is not. Rather, it has a very important theoretical advantage over the notion of strongly related points, concerning the issue of continuity, as we will discover later.

In Section 5, a dynamic approach to analyzing P3P will prove to be very helpful. There we will be interested in continuously moving *particles*. Clearly the concepts of weakly and strongly related points can be immediately extended to provide us with a notion of weakly and strongly related particles.

Throughout this paper, it will be assumed that the control point locations $(x_1, y_1, 0)$, $(x_2, y_2, 0)$ and $(x_3, y_3, 0)$ are fixed. On the other hand, we will, at times, vary the other three parameters of the extended Grunert system, namely, the view angles α, β and γ. In fact, it is quite useful to consider the reverse of solving the extended Grunert system. Consider an arbitrary point in space (x, y, z) that is not one of the control points. This point is clearly a solution point to the extended Grunert system for precisely one choice of values for the parameters α, β and γ. These are simply the view angle values for this particular point. The following claims are straightforward to check.

Lemma 1. *Given two non-control points p and p' in space, with corresponding view angles (α, β, γ) and $(\alpha', \beta', \gamma')$, the following are equivalent:*

1. *p and p' are strongly related (i.e. solution points for the same extended Grunert system);*
2. $(\alpha, \beta, \gamma) = (\alpha', \beta', \gamma')$*;*
3. $(\cos\alpha, \cos\beta, \cos\gamma) = (\cos\alpha', \cos\beta', \cos\gamma')$*;*
4. *The configuration consisting of the three rays emanating from p, in the directions of the control points, can be rigidly transformed into the corresponding configuration of rays emanating from p'.*

Likewise, the following are equivalent:

1. *p and p' are weakly related (i.e. solution points for the same or related extended Grunert systems);*
2. (α, β, γ) *and* $(\alpha', \beta', \gamma')$ *are either equal, or one of the corresponding pairs of angles are equal, and each of the other two pairs consists of supplementary angles;*
3. $(\cos^2\alpha,\ \cos^2\beta,\ \cos^2\gamma,\ \cos\alpha\cos\beta\cos\gamma) = (\cos^2\alpha',\ \cos^2\beta',\ \cos^2\gamma',\ \cos\alpha'\cos\beta'\cos\gamma')$*;*
4. *The configuration consisting of the three lines through p and a control point, can be rigidly transformed into the corresponding configuration of lines through p'.*

Henceforth, it will be convenient to let c_1, c_2 and c_3 denote $\cos\alpha$, $\cos\beta$ and $\cos\gamma$, respectively. As in (Rieck, 2018), let

$$\eta^2 = 1 - c_1^2 - c_2^2 - c_3^2 + 2c_1c_2c_3.$$

For our purposes, it turns out sometimes to be quite convenient to identify the xy-plane with the complex plane, by identifying a point (x, y) with the complex number $\zeta = x + iy$. The preferred coordinate system is easily chosen so that the three control points ζ_1, ζ_2, and ζ_3 satisfy these simple requirements:

$$|\zeta_1| \;=\; |\zeta_2| \;=\; |\zeta_3| \;=\; \zeta_1\zeta_2\zeta_3 \;=\; 1.$$

Accommodating the last of these conditions just amounts to applying a rotation to the *xy*-plane (complex plane), which imposes no significant restriction. Letting ϕ_α denote the signed angle subtended at the origin, between the positive *x*-axis and the ray through the control point (x_α, y_α), the requirement that $z_1 z_2 z_3 = 1$ just means that $\phi_1 + \phi_2 + \phi_3 = 0$, the convention followed in (Rieck, 2015) and (Rieck, 2018).

Though complex numbers were not used there, it was essentially shown in (Rieck, 2018) that the orthocenter of the triangle that has ζ_1, ζ_2 and ζ_3 as vertices is $\zeta_H = x_H + i\,y_H = \zeta_1 + \zeta_2 + \zeta_3$. This orthocenter of this *control points triangle* plays a vital role in the P3P problem. A general point (x,y,z) in three-dimensional real space will be identified with an ordered pair (ζ, z) where $\zeta = x + i\,y$.

It is easily seen that each non-control point (x,y,z) is strongly related to its reflection about the *xy*-plane, $(x,y,-z)$. When dealing with the problem of counting related points (strongly or weakly), and describing their locations, it is usually convenient to restrict attention to the *upper-half space*, consisting of points (x,y,z) for which $z > 0$. However, when studying a smoothly moving *particle,* moving around in three-dimensional real space, and asking how its weakly related particles move around, it becomes more natural to work in all of the three-dimensional real space, and, very importantly, to allow particles to pass through the control points!

3 THE COMPANION SURFACE TO THE DANGER CYLINDER

Using our setup, the *danger cylinder* is the surface in 3-dimensional real space given by the equation $x^2 + y^2 = 1$, or alternatively, $\zeta\overline{\zeta} = 1$. In (Rieck, 2018) and (Wang, et al, 1, 2019), another surface was introduced and shown to be of great importance in classifying points according to their number of strongly related (real) points, *i.e.* to the number of other real points that have the same view angles. Following the practice of the latter article, this surface will be called the *companion surface to the danger cylinder (CSDC).* Actually, *CSDC* provides a very simple way to classify points according to their number of weakly related (real) points, as follows. *CSDC* partitions the rest of three-dimensional real space into two parts, an inside and an outside.

To simplify the discussion, let us focus here only on the upper-half space, and so assume that $z > 0$. Each point outside *CSDC* has exactly one weak relative (*i.e.* weakly related point), which is also outside *CSDC*, and which might be a strong relative (*i.e.* strongly related point). Similarly, each point inside *CSDC* has exactly three weak relatives, all inside *CSDC*, with the caveat that each point on the danger cylinder, which is inside *CSDC*, counts as two points, due to the well-known phenomenon that these points correspond to repeated solutions of the Grunert system. These "counting facts" have been alluded to previously in (Rieck, 2018) and (Wang, et al, 1, 2019), but they are also presented here, as part of a far more comprehensive description of P3P solution points, and their distribution.

To set the stage for some new results, very straightforward formulas will now be presented for *CSDC* that can then be used to show that it indeed serves the role of separating points in space according to their number of weakly related points. The approach taken here is more direct than those taken in the past, yet somewhat tedious calculations are still required.

Recall that we are identifying a point in space (x,y,z) with an ordered pair (ζ, z), where $\zeta = x + i\,y$. Now, for each such point, assign two other complex numbers as follows:

$$\zeta_L \;=\; \overline{\zeta_H} \;+\; \frac{1}{\eta^2}\sum_{\alpha=1}^{3}\left(\zeta_\alpha^2 + 2\,\overline{\zeta_\alpha} - \overline{\zeta_H}\right)\left(1 - c_\alpha^2\right)$$

and

$$\zeta_L' = \zeta^2 - 2\,\overline{\zeta} + \left(\zeta\overline{\zeta} - 1\right)\left(\zeta^2 - \zeta_H\zeta - \overline{\zeta} + \overline{\zeta_H}\right)/z^2. \tag{2}$$

Notice that ζ_L depends only on the parameters of the Grunert system ($x_1, y_1, x_2, y_2, x_3, y_3, c_1, c_2$ and c_3), and so is independent of which solution point plays the role of (x,y,z) here. Indeed, since any two weakly related points have the same values for c_1^2, c_2^2, c_3^2 and η^2, they must also have the same value for ζ_L.

Also define the following real number (which also is the same for weakly related points):

$$\mathcal{D} = \zeta_L^2\,\overline{\zeta_L^2} \,-\, 4\,(\zeta_L^3 + \overline{\zeta_L^3}) \,+\, 18\,\zeta_L\,\overline{\zeta_L} \,-\, 27. \qquad (3)$$

The quantity $\mathcal{D}$ is actually seen to be the same as that given in (Rieck, 2018)[Theorem 2], up to a constant factor, though the use of ζ_L here makes its expression more compact. There, it is made clear that the vanishing of $\mathcal{D}$ is necessary and sufficient in order for Grunert's system, *i.e.* the first three equations in (1), to have a repeated solution. These formulas also essentially agree with (Wang, et al, 1, 2019)[Equation (25)].

This fact is then easily obtained: The extended Grunert system (1), with given parameters $x_1, y_1, x_2, y_2, x_3, y_3, c_1, c_2$ and c_3, has a repeated solution point if and only if $\mathcal{D} = 0$. Later, in Section 7, this fact will be reestablished as Theorem 7, using a substantially more direct approach than was used in (Rieck, 2018).

(Rieck, 2018)[Theorem 1] is subsumed in the following result, which will shortly be proven here, in a far more direct manner, using complex number theory.

Theorem 1. *The complex numbers ζ_L and ζ_L', defined in (2), are equal.*

In all that follows, $\mathfrak{C}$ denotes the unit circle, and $\mathfrak{D}$ denotes the *standard deltoid curve*, both in the complex plane. These are defined respectively by $\zeta\overline{\zeta} = 1$ and $\zeta^2\,\overline{\zeta^2} - 4\,(\zeta^3 + \overline{\zeta^3}) + 18\,\zeta\overline{\zeta} - 27 = 0$. Also, $Z = z^2$ henceforth.

The equation $\mathcal{D} = 0$ means that the complex number ζ_L is on the standard deltoid curve $\mathfrak{D}$. We will see that this equation is satisfied when (x, y, z) is on the danger cylinder, and also when it is on another connected surface that will become our definition of *CSDC*. We will come to see that it is *CSDC*, more so than the danger cylinder, that is crucial in separating the points in space into very different types, *viz.* the P3P problem.

Although it is not needed for the goals set out in this paper, it is perhaps worth briefly noting that the projection of a curve in space having a constant value of ζ_L (= ζ_L'), onto the xy-plane, is a cubic curve that passes through the three control points and the orthocenter, and that has asymptotes that are parallel to the altitudes of the control points triangle. This is left as an exercise for the interested reader.

In order to prove Theorem 1, a somewhat tedious, but somewhat interesting, algebraic fact, *i.e.* the next lemma, will be used. To prove it, one can simply expand, expressing each side as a polynomial in σ and Z, and then compare the coefficients of powers of σ times powers Z. Admittedly, this is a bit tedious. Algebraic manipulation software can be used to verify the equation much faster.

Lemma 2. *For any complex numbers ζ_1, ζ_2, ζ_3, ζ, σ and Z, the following identity holds:*

$$\begin{aligned}
&\zeta_1(\zeta_2-\zeta_3)\,[Z+(\zeta-\zeta_1)\,\zeta_2\zeta_3\,(\zeta_1\sigma-1)]\\
&\quad\cdot\left[4Z+\zeta_1(\zeta+\zeta_2\zeta_3\sigma-\zeta_2-\zeta_3)^2\right] +\\
&\zeta_2(\zeta_3-\zeta_1)\,[Z+(\zeta-\zeta_2)\,\zeta_3\zeta_1\,(\zeta_2\sigma-1)]\\
&\quad\cdot\left[4Z+\zeta_2(\zeta+\zeta_3\zeta_1\sigma-\zeta_3-\zeta_1)^2\right] +\\
&\zeta_3(\zeta_1-\zeta_2)\,[Z+(\zeta-\zeta_3)\,\zeta_1\zeta_2\,(\zeta_3\sigma-1)]\\
&\quad\cdot\left[4Z+\zeta_3(\zeta+\zeta_1\zeta_2\sigma-\zeta_1-\zeta_2)^2\right] \quad =
\end{aligned}$$

$$\begin{aligned}
&-(\zeta_2-\zeta_3)(\zeta_3-\zeta_1)(\zeta_1-\zeta_2)\cdot\\
&\{\,(\zeta^2-2\zeta_1\zeta_2\zeta_3\sigma-\zeta_2\zeta_3-\zeta_3\zeta_1-\zeta_1\zeta_2)\,Z\\
&\quad+\zeta_1\zeta_2\zeta_3\,[\zeta^2-(\zeta_1+\zeta_2+\zeta_3)\,\zeta-\zeta_1\zeta_2\zeta_3\sigma\\
&\quad+\zeta_2\zeta_3+\zeta_3\zeta_1+\zeta_1\zeta_2]\,(\zeta\sigma-1)\,\}.
\end{aligned}$$

To continue towards a proof of Theorem 1, it will help to introduce additional notation. Recall that d_1 denotes the triangle side length $[(x_2-x_3)^2+(y_2-y_3)^2]^{1/2}$, and similarly for d_2 and d_3. Let θ_1, θ_2 and θ_3 denote the view angles α, β and γ, respectively. In (Rieck, 2014)[Lemma 2], it was proven that $\eta^2 = d_1^2 d_2^2 d_3^2 Z/(4s_1^2 s_2^2 s_3^2)$. Now, $\sin^2\theta_1 = 1 - c_1^2 = (2s_2^2 s_3^2 + 2d_1^2 s_2^2 + 2d_1^2 s_3^2 - d_1^4 - s_2^4 - s_3^4)/(4s_2^2 s_3^2)$, which is an immediate consequence of $c_1 = (s_2^2 + s_3^2 - d_1^2)/(2s_2 s_3)$. Similarly for $\sin^2\theta_2$ and $\sin^2\theta_3$. The claim being made in Theorem 1 is therefore the same as the claim made in the next lemma.

Lemma 3. *For the control points ζ_1, ζ_2 and ζ_3, the optical center (ζ, z), and the quantities ζ_H, d_1, d_2, d_3, s_1, s_2 and s_3 based on these, the following equation holds:*

$$
\begin{aligned}
&(\zeta_1^2+\overline{\zeta_1}-\overline{\zeta_2}-\overline{\zeta_3})\, s_1^2\cdot\\
&\quad(2s_2^2s_3^2+2d_1^2s_2^2+2d_1^2s_3^2-d_1^4-s_2^4-s_3^4)\;+\\
&(\zeta_2^2+\overline{\zeta_2}-\overline{\zeta_3}-\overline{\zeta_1})\, s_2^2\cdot\\
&\quad(2s_3^2s_1^2+2d_2^2s_3^2+2d_2^2s_1^2-d_2^4-s_3^4-s_1^4)\;+\\
&(\zeta_3^2+\overline{\zeta_3}-\overline{\zeta_1}-\overline{\zeta_2})\, s_3^2\cdot\\
&\quad(2s_1^2s_2^2+2d_3^2s_1^2+2d_3^2s_2^2-d_3^4-s_1^4-s_2^4)\;=\\
&\\
&d_1^2d_2^2d_3^2\,\{\,(\zeta^2-2\overline{\zeta}-\overline{\zeta_H})Z+\\
&\quad(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)\,\}.
\end{aligned}
$$

Proof. We will make repeated use of these facts: $\overline{\zeta_1}=1/\zeta_1$, $\overline{\zeta_2}=1/\zeta_2$, $\overline{\zeta_3}=1/\zeta_3$ and $\zeta_1\zeta_2\zeta_3=1$. By means of these facts, one discovers that $\zeta_1^2+\overline{\zeta_1}-\overline{\zeta_2}-\overline{\zeta_3} = (\zeta_1-\zeta_2)(\zeta_1-\zeta_3)$, etcetera. Likewise, $2s_2^2s_3^2+2d_1^2s_2^2+2d_1^2s_3^2-d_1^4-s_2^4-s_3^4$ equals $-\zeta_1(\zeta_2-\zeta_3)^2\{4Z+\zeta_1[\zeta+\zeta_2\zeta_3\overline{\zeta}-(\zeta_2+\zeta_3)]^2\}$, etcetera. So $(\zeta_1^2+\overline{\zeta_1}-\overline{\zeta_2}-\overline{\zeta_3})\,s_1^2\,(2s_2^2s_3^2+2d_1^2s_2^2+2d_1^2s_3^2-d_1^4-s_2^4-s_3^4)=\zeta_1(\zeta_2-\zeta_3)^2(\zeta_3-\zeta_1)(\zeta_1-\zeta_2)\{Z+(\zeta-\zeta_1)\zeta_2\zeta_3(\zeta_1\overline{\zeta}-1)\}\{4Z+\zeta_1[\zeta+\zeta_2\zeta_3\overline{\zeta}-(\zeta_2+\zeta_3)]^2\}$.

Adding this and the other two similar expressions yields $-(\zeta_2-\zeta_3)^2(\zeta_3-\zeta_1)^2(\zeta_1-\zeta_2)^2[(\zeta^2-2\overline{\zeta}-\overline{\zeta_H})Z+(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)]=-\zeta_1\zeta_2\zeta_3(\zeta_2-\zeta_3)^2(\zeta_3-\zeta_1)^2(\zeta_1-\zeta_2)^2[(\zeta^2-2\overline{\zeta}-\overline{\zeta_H})Z+(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)]=(\zeta_2-\zeta_3)(\overline{\zeta_2}-\overline{\zeta_3})(\zeta_3-\zeta_1)(\overline{\zeta_3}-\overline{\zeta_1})(\zeta_1-\zeta_2)(\overline{\zeta_1}-\overline{\zeta_2})[(\zeta^2-2\overline{\zeta}-\overline{\zeta_H})Z+(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)]=d_1^2d_2^2d_3^2\{(\zeta^2-2\overline{\zeta}-\overline{\zeta_H})Z+(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)\}$.

□

Having proven Lemma 3, we have also proven Theorem 1 as a consequence. We will next strive to gain a better understanding of the quantity $\mathcal{D}$.

Lemma 4. *Fix a non-control point (ζ,z). Then*

$$\mathcal{D} \;=\; (\zeta\overline{\zeta}-1)^2\,\mathcal{P}/Z^4\,,$$

where $\mathcal{P}$ is a polynomial of degree four in Z, having coefficients that are polynomials in ζ and $\overline{\zeta}$ (or alternatively in x and y). Moreover, when treated as a polynomial in ζ, $\overline{\zeta}$ and z (or in x, y and z), it has degree twelve.

Proof. Let $Z_L = Z\zeta_L = (\zeta^2-2\overline{\zeta})Z+(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\zeta\overline{\zeta}-1)$. Let $\widehat{\mathcal{D}} = Z^4\,\mathcal{D} = Z_L^2\overline{Z_L}^2-4Z(Z_L^3+\overline{Z_L}^3)+18Z^2Z_L\overline{Z_L}-27Z^4$, assuming always that Z is real-valued. The equation in the statement of the lemma then amounts to the equation $\widehat{\mathcal{D}} = (\zeta\overline{\zeta}-1)^2\,\mathcal{P}$. Clearly, $\widehat{\mathcal{D}}$ can be expressed as a polynomial in ζ, $\overline{\zeta}$ and Z, where the highest power of Z that occurs is the fourth power. Replacing Z with z^2, it is also immediate that as polynomials in ζ, $\overline{\zeta}$ and z, the polynomial Z_L has degree four, and the polynomial $\widehat{\mathcal{D}}$ has degree sixteen. It remains only to show that $(\zeta\overline{\zeta}-1)^2$ is a factor of $\widehat{\mathcal{D}}$.

Momentarily, assume that $|\zeta|=1$, and so $\overline{\zeta}=1/\zeta$. It can be readily checked that Z_L becomes $(\zeta^2-2\zeta^{-1})Z$, $\overline{Z_L}$ becomes $(\zeta^{-2}-2\zeta)Z$, and $\widehat{\mathcal{D}}$ vanishes. For general ζ, it follows that $\zeta\overline{\zeta}-1$ is a factor of $\widehat{\mathcal{D}}$.

Also, upon setting $\overline{\zeta}=1/\zeta$ again, we see that $(\partial\widehat{\mathcal{D}}/\partial Z_L\,,\,\partial\widehat{\mathcal{D}}/\partial\overline{Z_L})$ becomes $(-4Z^3(\zeta+1)^3(\zeta^2-\zeta+1)^3\zeta^{-5}\;,\;-4Z^3(\zeta+1)^3(\zeta^2-\zeta+1)^3\zeta^{-4})$, and $(\partial Z_L/\partial\overline{\zeta}\,,\,\partial\overline{Z_L}/\partial\overline{\zeta})$ becomes $(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-2Z-1\;,\;-\zeta^2+\zeta_H\zeta-\overline{\zeta_H}+(2Z+1)\zeta^{-1})$. The dot product of these two vectors is zero, telling us that $\partial\widehat{\mathcal{D}}/\partial\overline{\zeta}$ becomes zero. This then guarantees that $(\zeta\overline{\zeta}-1)^2$ is a factor of $\widehat{\mathcal{D}}$.

To be a bit more rigorous perhaps, let ω be a new indeterminate, and let $g\in\mathbb{Q}(\zeta,Z)[\omega]$ (a polynomial in ω with coefficients in the rational function field $\mathbb{Q}(\zeta,Z)$) be defined by replacing each of the occurrences of $\overline{\zeta}$ in $\mathcal{D}$ with ω. Since $g(1/\zeta)=g'(1/\zeta)=0$, $(\omega-1/\zeta)^2$ must be a factor of the polynomial g in the polynomial ring $\mathbb{Q}(\zeta,Z)[\omega]$, and so $(\zeta\omega-1)^2$ must be a factor of the same polynomial in the polynomial ring $\mathbb{Q}[\zeta,\omega,Z]$, and so $(\zeta\overline{\zeta}-1)^2$ must divide $\mathcal{D}$.

□

By means of a series of preliminary lemmas, found in the appendix, the following result can be established:

Lemma 5. *Let $\mathcal{P}$ be the polynomial introduced in Lemma 4. Consider its expression as a polynomial in real variables x, y and Z (rather than ζ, $\overline{\zeta}$ and Z), with real coefficients. This real polynomial is irreducible. That is, it is irreducible as an element of the polynomial ring $\mathbb{R}[x,y,Z]$.*

The important role played by the polynomial $\mathcal{P}$, and the surface defined by $\mathcal{P}=0$, will now be investigated.

Lemma 6. *The following are equivalent for a point $p\in\mathbb{R}^3$, not on the unit circle $\mathfrak{C}$:*

1. *p is either on the danger cylinder (and so a double solution point for the extended Grunert system), or is weakly related to a point on the danger cylinder;*
2. *p is on the surface of* $\mathbb{R}^3$ *defined by the polynomial equation, in x, y and* $Z\,(=z^2)$*, given by* $\widehat{D} = 0$*;*
3. *p either satisfies the equation* $x^2 + y^2 = 1$ *or the polynomial equation* $\mathcal{P} = 0$*.*

Proof. Let *DC* be the danger cylinder. If *p* is on *DC*, then it is clearly a solution point for the equation $\widehat{\mathcal{D}} = 0$, since $\zeta_L = \zeta^2 - 2\overline{\zeta}$ easily makes $\mathcal{D}$ zero. But by the Theorem 1, $\mathcal{D}$ can be expressed as a rational function of $c_1c_2c_3$, c_1^2, c_2^2, c_3^2, *not* involving the coordinates of *p* in any other way. Therefore, still assuming that *p* is on *DC*, any point weakly related to it is also a solution point for the equation $\widehat{\mathcal{D}} = 0$. This is due to the fact that two points are weakly related if and only if they have the same values for the four quantities $c_1c_2c_3$, c_1^2, c_2^2, c_3^2 (see Lemma 1). This makes it clear that item 1 in the lemma implies item 2. The equivalence of items 2 and 3 is clear from Lemma 4.

It is easily established that there exists some point in $\mathbb{R}^3$, not on *DC*, that is weakly related to a point on *DC*. Since the first point does not satisfy $x^2 + y^2 = 1$, but does satisfy $\widehat{\mathcal{D}} = 0$, it must satisfy $\mathcal{P} = 0$. Thus, $\mathcal{P} = 0$ has real solutions. Its algebraic set (solution surface) $\mathcal{S}$ includes all of the points that are weakly related to points on *DC*, but not on *DC* themselves. By "elimination theory," the set of points $\mathcal{S}'$ that are either on *DC* or are weakly related to points on *DC* is also an algebraic set, *i.e.* the solution set of a system of polynomial equations. Now, $\mathcal{S} \cap \mathcal{S}' \neq \phi$, and $\mathcal{S}$ is irreducible, since $\mathcal{P}$ is an irreducible polynomial (see Lemma 5), so $\mathcal{S} \subseteq \mathcal{S}'$. Therefore, the entire surface given by $\mathcal{P} = 0$ must consist only of points that are either on *DC* or are weakly related to points on *DC*. Consequently, item 3 in the lemma implies item 1.

□

The following is a crucial result for determining, given a point, its number of weakly related points, and also its number of strongly related points.

Corollary 1. *Consider a point* $p = (x, y, z) \in \mathbb{R}^3$ *with* $z > 0$*, and* $\mathcal{D} \neq 0$*. If* $\mathcal{P} > 0$*, then p is weakly related to exactly one other point in the upper-half space. If instead* $\mathcal{P} < 0$*, then p is weakly related to exactly three other points in the upper-half space.*

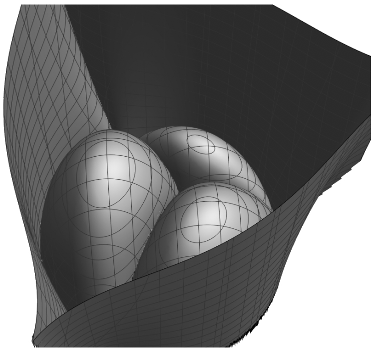

Figure 1: *CSDC* near the *xy*-plane

Proof. The quantity ζ_L in the current paper is the same as $-1 - \mathcal{R} + \mathcal{L}\,i$ with $\mathcal{L}$ and $\mathcal{R}$ as defined in (Rieck, 2018). Therefore, $\mathcal{D}$ here is the same as $\mathcal{D}$ there, and $\widehat{\mathcal{D}}$ here is the same as $\widehat{\mathcal{D}}$ there. Theorem 2 there amounts to the claim that the a point $p \in \mathbb{R}^3$ is a solution point for an extended Grunert system having a repeated solution if and only if $\widehat{\mathcal{D}} = 0$. By Lemma 4, this is so if and only if *p* is on the danger cylinder, or on the surface defined by $\mathcal{P} = 0$.

Treat *p* as a particle now, and consider moving it continuously in the upper-half space. Since the formula for the danger cylinder in squared in the formula for $\widehat{\mathcal{D}} = 0$, the number of real solutions to the extended Grunert system does not change when *p* crosses the danger cylinder. On the other hand, the irreducible polynomial $\mathcal{P}$ only occurs to the first power, and so when *p* crosses the surface $\mathcal{P} = 0$, this does affect the number of real solutions to the extended Grunert system for which *p* is a solution point. In other words, this does affect the number of points that are weakly related to *p*.

When $\mathcal{P} > 0$, the Grunert system is found to have two real solutions (along with two other complex solutions), and so the solution point *p* has one weak relative in the upper-half space. But, when $\mathcal{P} < 0$, the Grunert system has four real solutions, and so the solution point *p* has three weak relative in the upper-half space.

□

Note that Proposition 2 in (Wang, et al, 1, 2019) further supports the corollary. The surface defined by $\mathcal{P} = 0$ is thus critical in determining the number of weakly related points that a given point has. It is this surface that we will call the *companion surface to the* *danger cylinder (CSDC)*. A point p for which $\mathcal{P} > 0$ is outside *CSDC*, but if instead $\mathcal{P} < 0$ then it is inside *CSDC*.

When the control points triangle is acute, it was first observed in (Rieck, 2018) that *CSDC* consists of an unbounded region that is outside the danger cylinder, wrapping around the danger cylinder, and tending towards a deltoid curve as z goes to infinity, together with a bounded region inside the danger cylinder. This inside portion consists of three "humps," as seen in Figure 1. We close this section by proving that this portion of *CSDC* is actually contained inside the unit sphere, when the control points triangle is acute.

Theorem 2. *Assume that the control points triangle is acute. Consider a point $p = (x,y,z) \in \mathbb{R}^3$ that is on CSDC, and also inside the danger cylinder ($x^2+y^2 < 1$). Then $x^2+y^2+z^2 < 1$.*

Proof. A triangle is acute if and only if its orthocenter is inside its circumcircle, and so we may assume that $x_H^2 + y_H^2 < 1$. We are also assuming that $x^2+y^2 < 1$. Because we are also assuming that p is on *CSDC*, it follows that $\mathcal{D} = 0$ and so the complex number ζ_L is on the deltoid curve $\mathfrak{D}$. This curve is known to be rational, and it fact, it is straightforward to see that the mapping $\omega \to \omega^2 - 2\overline{\omega}$ maps the unit circle $\mathfrak{C}$ bijectively onto the deltoid $\mathfrak{D}$. So $\zeta_L = \omega^2 - 2\overline{\omega}$ for some ω_L with $\omega_L \overline{\omega_L} = 1$. By the formulas (2), Theorem 1, and the fact that z is real, we see that

$$Z = z^2 = \frac{(\zeta\overline{\zeta}-1)(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})}{\omega_L^2-2\overline{\omega_L}-\zeta^2+2\overline{\zeta}}$$

$$= \frac{(\zeta\overline{\zeta}-1)(\overline{\zeta}^2-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)}{\overline{\omega}_L^2-2\omega_L-\overline{\zeta}^2+2\zeta},$$

and so

$$\frac{1-\zeta\overline{\zeta}-Z}{1-\zeta\overline{\zeta}} = \frac{\omega_L^2-2\overline{\omega_L}-\zeta_H\zeta+\overline{\zeta}+\overline{\zeta_H}}{\omega_L^2-2\overline{\omega_L}-\zeta^2+2\overline{\zeta}}$$

$$= \frac{\overline{\omega_L}^2-2\omega_L-\overline{\zeta_H}\,\overline{\zeta}+\zeta+\zeta_H}{\overline{\omega}_L^2-2\omega_L-\overline{\zeta}^2+2\zeta}.$$

We need to show that this is positive, thereby proving the lemma. This will in turn follow from the assertion that for any complex numbers ζ, ζ_H and ω with $|\zeta| < 1$, $|\zeta_H| < 1$ and $|\omega| = 1$, if the quantity

$$\frac{\omega^2-2\overline{\omega}-\zeta_H\zeta+\overline{\zeta}+\overline{\zeta_H}}{\omega^2-2\overline{\omega}-\zeta^2+2\overline{\zeta}} \tag{4}$$

is real, then it is positive. This will now be established. Using the fact that $\overline{\omega} = 1/\omega$, the above formula can be rewritten as

$$\frac{2+(\zeta_H\zeta-\overline{\zeta}-\overline{\zeta_H})\omega-\omega^3}{2+(\zeta^2-2\overline{\zeta})\omega-\omega^3}$$

and its conjugate (presumed to be equal) can be rewritten as

$$\frac{1+(\zeta+\zeta_H-\overline{\zeta_H}\,\overline{\zeta})\omega^2-2\omega^3}{1+(2\zeta-\overline{\zeta}^2)\omega^2-2\omega^3}.$$

Let's begin examining quantity (5) by first asking if it can be zero. If it is zero, then $2 + (\zeta_H\zeta - \overline{\zeta} - \overline{\zeta_H})\omega - \omega^3 = 1 + (\zeta + \zeta_H - \overline{\zeta_H}\,\overline{\zeta})\omega^2 - 2\omega^3 = 0$. If we eliminate $\overline{\zeta}$ from these, we find that

$$\zeta = \frac{1-2\overline{\zeta_H}\,\omega+(\zeta_H+\overline{\zeta_H}^2)\,\omega^2-2\,\omega^3+\overline{\zeta_H}\,\omega^4}{(\zeta_H\overline{\zeta_H}-1)\,\omega^2}.$$

We are assuming here that $|\zeta| < 1$, so $|1-2\overline{\zeta_H}\,\omega + (\zeta_H + \overline{\zeta_H}^2)\,\omega^2 - 2\,\omega^3 + \overline{\zeta_H}\,\omega^4| < |\zeta_H\overline{\zeta_H} - 1|$. However, $[1 - 2\overline{\zeta_H}\,\omega + (\zeta_H + \overline{\zeta_H}^2)\,\omega^2 - 2\,\omega^3 + \overline{\zeta_H}\,\omega^4] \cdot [1 - 2\zeta_H\,\omega^{-1} + (\overline{\zeta_H} + \zeta_H^2)\,\omega^{-2} - 2\,\omega^{-3} + \zeta_H\,\omega^{-4}] - (\zeta_H\overline{\zeta_H} - 1)^2 = (\zeta_H\,\omega^{-1} - 2 + \overline{\zeta_H}\,\omega) \cdot (\omega^{-2} - \overline{\zeta_H}\,\omega^{-1} + \zeta_H\,\omega - \omega^2)^2$. On the right side of this equation, the first factor is negative, since $|\zeta_H| < 1$ and $|\omega| = 1$, and the rest is the square of a purely imaginary number, and so also negative. The right side as a whole is therefore positive. This leads us to conclude that $|1-2\overline{\zeta_H}\,\omega+(\zeta_H+\overline{\zeta_H}^2)\,\omega^2-2\,\omega^3+\overline{\zeta_H}\,\omega^4| > |\zeta_H\overline{\zeta_H}-1|$. We have a contradiction here, and so conclude that quantity (4) cannot be zero.

Now, (4) is always defined since its denominator is never zero, since $\omega^2 - 2\overline{\omega} \in \mathfrak{D}$, but $\zeta^2 - 2\overline{\zeta} \notin \mathfrak{D}$, since $\omega \in \mathfrak{C}$, but $\zeta \notin \mathfrak{C}$ and $\zeta \notin \mathfrak{D}$. (The preimage of $\mathfrak{D}$ under this mapping $\zeta \to \zeta^2 - 2\overline{\zeta}$ is $\mathfrak{C} \cup \mathfrak{D}$.) Now, the quantity (4) equals one, which is positive, when $\zeta = \zeta_H$. By continuity, quantity (4) is never negative.

□

4 TOROIDS AND BLOWUPS

A *toroid*, in the sense that will be used here, is a circular arc, rotated in three-dimensional space, about the line through its endpoints. The only toroids that are of concern in the P3P problem are those generated by circular arcs whose endpoints are two of the three control points. Suppose for a moment that the endpoints are B and C. By the Inscribed Angle Theorem, all of the points p on the circular are such that the angle subtended at p by the rays to B and to C, are equal. That is, these points have the same view angle α. But clearly, this is, more generally, still the case for all of the points on the toroid generated by the arc; all of these points have the same view angle α. Moreover, a moment's reflection makes it clear that the points on this toroid are the only points in three-dimensional space whose first view angle equals this α.

Henceforth, this toroid is denoted $Toroid_\alpha$, where the label "α" lets us know that we are referring to the first view angle, *i.e.* the angle, at a given point, subtended by the rays to B and C. Similarly, $Toroid_\beta$ is the toroid consisting of points whose second view angle (the angle subtended by rays to A and C) have the same value β. Likewise, $Toroid_\gamma$ corresponds to the third view angle. Notice that for specified values of α, β and γ, the points that have these three view angles (if any) are the points that lie on the intersection of the three toroids $Toroid_\alpha$, $Toroid_\beta$ and $Toroid_\gamma$.

Given a value for α, besides $Toroid_\alpha$, there is another toroid of interest, namely the toroid corresponding to a first view angle of $\pi - \alpha$. This is obtained by taking the circular arc used to generate $Toroid_\alpha$, and replacing it with the circular arc with endpoints B and C, whose union with the original circular arc forms a circle (through B and C). This is also clear from the Inscribed Angle Theorem. Denote this new toroid, $Toroid_{\pi-\alpha}$. Similarly, we have $Toroid_{\pi-\beta}$ and $Toroid_{\pi-\gamma}$. The union of $Toroid_\alpha$ and $Toroid_{\pi-\alpha}$ is clearly obtained by rotating and entire circle through B and C, about the line through B and C. We will denote this $DoubleToroid_\alpha$, and refer to it as a *double toroid*. Similarly for $DoubleToroid_\beta$ and $DoubleToroid_\gamma$.

Very important special cases of toroids and double toroids occur when $\alpha = A$ or $\beta = B$ or $\gamma = C$, where A, B and C implicitly mean the interior angles of the control points triangle ABC. When $\alpha = A$, the toroid $Toroid_\alpha$ will also be denoted $Toroid_A$. It is one of the three *basic toroids*. The other two basic toroids are of course $Toroid_\beta$ when $\beta = B$, and $Toroid_\gamma$ when $\gamma = C$, and they are denoted $Toroid_B$ and $Toroid_C$, respectively. Similarly, we have the three *basic double toroids*: $DoubleToroid_A$, $DoubleToroid_B$ and $DoubleToroid_C$. There are also three other toroids of special interest, $Toroid_{\pi-A}$, $Toroid_{\pi-B}$ and $Toroid_{\pi-C}$, whose meaning should now be clear.

Lemma 7. *The following are equivalent formulas for the basic double toroid $DoubleToroid_A$:*

1. $$z^4 + 2[x^2+y^2-(y_2+y_3)xy-x_2-y_2-1]\,z^2 + (x^2+y^2-1)[x^2+y^2-2(x_2+x_3)x - 2(y_2+y_3)y+2x_2x_3+2y_2y_3-1] \;=\; 0;$$

2. $$z^4 + [\,2\zeta\overline{\zeta} - (\overline{\zeta_2}+\overline{\zeta_3})\,\zeta - (\zeta_2+\zeta_3)\,\overline{\zeta} - 2\,]\,z^2 + (\zeta\overline{\zeta}-1)\,[\,\zeta\overline{\zeta} - (\overline{\zeta_2}+\overline{\zeta_3})\,\zeta - (\zeta_2+\zeta_3)\,\overline{\zeta} + (1+\zeta_2\overline{\zeta_3}+\overline{\zeta_2}\zeta_2)\,] \;=\; 0.$$

Similar formulas are of course obtained for $DoubleToroid_B$ and $DoubleToroid_C$ by cycling the indices.

Proof. The double toroid $DoubleToroid_A$ can be described as the collection of points that satisfy $\cos^2\alpha = \cos^2 A$. By the Law of Cosines, this equation can be expressed as $4s_2^2s_3^2\cos^2 A = (s_2^2 + s_3^2 - d_1^2)^2$. But, $\cos^2 A = 1 - \sin^2 A = 1 - d_1^2/4$, $d_1^2 = (x_2-x_3)^2 + (y_2-y_3)^2$ and $s_1^2 = (x-x_1)^2+(y-y_1)^2+z^2$. The first formula stated in the theorem follows quickly from these facts, and the special properties of x_1, x_2, x_3, y_1, y_2 and y_3 that can be deduced from the restrictions placed on ζ_1, ζ_2 and ζ_3, and which are listed in (Rieck, 2018)[Lemma 1]. The second formula follows immediately from the first formula upon setting $x_j = (\zeta_j + \overline{\zeta_j})/2$ and $y_j = (\zeta_j - \overline{\zeta_j})/2i$, for $j = 1,2,3$.

□

The basic toroids, $Toroid_A$, $Toroid_B$ and $Toroid_C$, and their related toroids $Toroid_{\pi-A}$, $Toroid_{\pi-B}$ and $Toroid_{\pi-C}$, are particularly important for understanding the locations of weakly related points, including strongly related points. This will be carefully explored in the next section. However, the stage for this will be set here, by introducing a couple more concepts, and presenting a couple more lemmas.

Lemma 8. *When $A < \alpha < \pi - A$, $DoubleToroid_\alpha$ intersects the unit sphere in two circles that are reflections of each other about the xy-plane. Moreover, $Toroid_\alpha$ intersects the unit sphere in two circular arcs that are reflections of each other about the xy-plane. Similarly for the intersection of $Toroid_{\pi-\alpha}$ and the unit sphere.*

Additionally, when $\alpha < A$, $Toroid_\alpha$ does not extend inside the unit sphere, and when $\alpha > \pi - A$,

$Toroid_\alpha$ does not extend outside the unit sphere. In all of this, A and α can be replace with B and β, or with C and γ.

Proof. Let M be the midpoint of the segment BC. Let P be plane through M that is perpendicular to BC. This intersects the unit sphere in a great circle $\mathcal{G}$. Any point p on $\mathcal{G}$ determines a unique circle $\mathcal{C}$ containing B, C and p. By the Inscribed Angle Theorem, all the point along the arc $\mathcal{A}$ of $\mathcal{C}$ from B to C, containing p, have the same first view angle as p. Letting p move along $\mathcal{G}$ from one point in the xy-plane to the antipodal point in the xy plane, this first view angle clearly changes continuously from A to $\pi - A$, and so somewhere is equal to α (fixed value in the lemma).

The arc $\mathcal{A}$ for this p is clearly part of the intersection of $Toroid_\alpha$ and the unit sphere. But the same may be said about the reflection of $\mathcal{A}$ about the xy-axis. As p is allowed to vary, the entire unit sphere is swept out by the circular arc $\mathcal{A}$, so it is clear that no other points lie on the intersection of $Toroid_\alpha$ and the unit sphere. It is also immediately clear that if p is such that the points of $\mathcal{A}$ have first view angle α, then the points on the arc of $\mathcal{C}$ from B to C, *not* containing p, have first view angle $\pi - \alpha$, as do the reflections of this arc about the xy-plane. The first paragraph of the lemma now follows.

To prove the rest, suppose that $\alpha < A$. A portion of $Toroid_\alpha$ in the xy-plane is an arc between B and C that is outside the unit circle, apart from B and C. As this is rotated about the line BC, this arc cannot come into further contact with the unit sphere. This is because the above reasoning establishes that fact that at any point on the unit sphere, apart from B and C, the first view angle is between A and $\pi - A$. So $Toroid_\alpha$ is outside the unit sphere, apart from B and C. Similarly for the $\alpha > \pi - A$ case.

□

Henceforth, the control points triangle ABC will be assumed to be an acute triangle, and so all of the interior angles A, B and C are less that $\pi/2$. The pair of toroids $Toroid_A$ and $Toroid_{\pi-A}$, partitions three-dimensional space (with these toroids removed) into three parts: outside $Toroid_A$, between $Toroid_A$ and $Toroid_{\pi-A}$, and inside $Toroid_{\pi-A}$.

The digits "0", "1", "2" will be used (respectively) to distinguish between these three portions of space. Another digit (also "0", "1" or "2") will be similarly

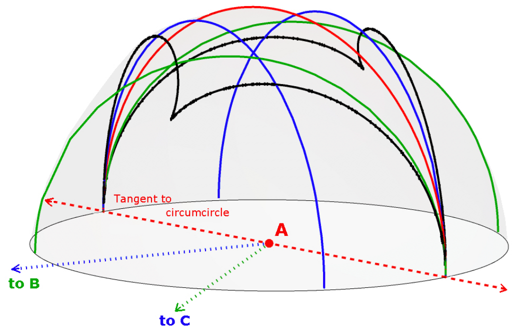

Figure 2: Upper half of $DoubleBlowup_A$

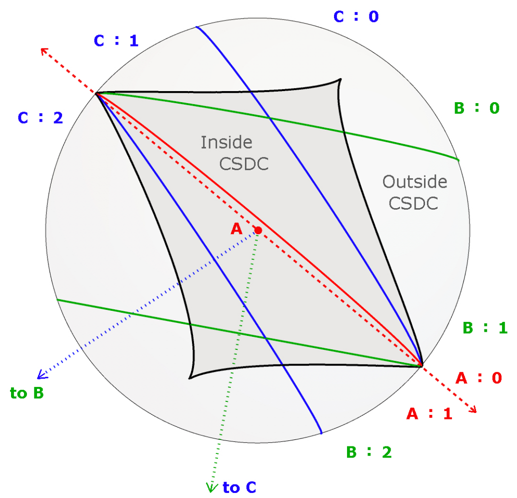

Figure 3: Top-down view of same $DoubleBlowup_A$

used to represent partitioning space based on $Toroid_B$ and $Toroid_{\pi-B}$. Again, one final digit will be used to represent partitioning space based on $Toroid_C$ and $Toroid_{\pi-C}$. In this way, we will speak of a *toroidal region abc* where each of a, b, c is 0, 1 or 2, and abc will be called a region *descriptor*. Here a indicates one of three portions of space based on $Toroid_A$ and $Toroid_{\pi-A}$, b similarly indicates a portion based on $Toroid_B$ and $Toroid_{\pi-B}$, and c similarly indicates a portion based on $Toroid_C$ and $Toroid_{\pi-C}$. For instance, toroidal region 010 consists of points outside $Toroid_A$ and $Toroid_C$, but between $Toroid_B$ and $Toroid_{\pi-B}$.

We will use the notation $Blowup_A$, and refer to this as the *blowup* of A, for the set of lines through A. Each line should be regarded as a possible tangent line of a differentiable curve passing through A. In the next section, such a curve will be the path of a smoothly moving *particle*. We will also use the notation $DoubleBlowup_A$, and refer to this as the *double cover* of the blowup of A, for the set of all rays that originate at A. Each such ray clearly corresponds to a line through A in a two-to-one way. In this sense, there are two elements of $DoubleBlowup_A$ for each element of $Blowup_A$.

While an element of $DoubleBlowup_A$ is technically a ray, we will want to think of it as representing a point that is infinitesimally close to A, in the direction of the ray. As such, it has three view angles α, β and γ associated with is, which naturally extends the definition of the view angles of ordinary points in space. It also makes sense to talk about the ordinary points in space that are weakly or strongly related to a "point" in $DoubleBlowup_A$. Now, $DoubleBlowup_A$ can naturally be visualized as a tiny (infinitesimally small, really) sphere, centered at A. Figures 2 and 3 show the upper half of $DoubleBlowup_A$, regarded as a sphere. Antipodal points on this sphere are weakly related, but not strongly related. This basically means that if a moving particle passes through A, at this exact moment, the particle instantaneously changes to a particle that is only weakly related to the original particle.

Lemma 9. *Given any (ordinary) point p on* $Toroid_A$*, with view angles* $\alpha_p = A$*,* β_p *and* γ_p*, there is exactly one point on* $DoubleBlowup_A$ *with view angles* α_p*,* β_p *and* γ_p*. Its antipodal point on this sphere has view angles* α_p*,* $\pi-\beta_p$ *and* $\pi-\gamma_p$*. Similarly for* $Toroid_B$ *and* $Toroid_C$*.*

Proof. Consider the three rays emanating from p, one pointing to A, one pointing to B, and one pointing to C. Translate and rotate this configuration of rays so that they now emanate from A, with the second ray still pointing to B, and the third ray still pointing to C. Regard the first ray as an element of $DoubleBlowup_A$. This point will have view angles A, $\pi-\beta_p$ and $\pi-\gamma_p$. Its antipodal point (on $DoubleBlowup_A$) will have view angles A, β_p and γ_p. This latter point is strongly related to p. The uniqueness of these points is immediately clear.

□

We will now explore the intersection of $DoubleBlowup_A$ with each of these: $Toroid_A$, $Toroid_B$, $Toroid_C$, $Toroid_{\pi-B}$, $Toroid_{\pi-C}$, and *CSDC*.

The intersection with $Toroid_A$ is simple. The tangent plane of $Toroid_A$ at A is simply the vertical extension of the tangent line for the circumcircle (unit circle) at A, which is also the tangent plane of the danger cylinder at A. This plane cuts the sphere $DoubleBlowup_A$ in a great circle. Based on this cut, one side of the sphere is outside $Toroid_A$ (and outside the danger cylinder), and the other side of the sphere is inside $Toroid_A$ (and inside the danger cylinder).

To understand the intersection of $DoubleBlowup_A$ with $DoubleToroid_B$, begin by noticing that $DoubleToroid_B$ has a singularity at A. While there is no tangent plane at A, there is instead a double cone that is tangent to $DoubleToroid_B$ at A. The intersection of $DoubleBlowup_A$ with $DoubleToroid_B$ is the same as the intersection of $DoubleBlowup_A$ with this double cone, which is clearly a pair of circles. On the sphere ($DoubleBlowup_A$), between these two circles, we have the points whose second view angle (β) is between B and $\pi-B$. On the other side of one of the circles are the points whose second view angle is less than B. On the other side of the other circles are the points whose second view angle is greater than $\pi-B$. Similarly for the intersection of $DoubleBlowup_A$ with $DoubleToroid_C$.

The intersection of $DoubleBlowup_A$ and *CSDC* involves a simple closed curve on one hemisphere and its reflection on the other hemisphere. Between these closed curves on the sphere lie the points that are outside *CSDC*. Inside each of the closed curves are the points that are inside *CSDC*. Figures 2 and 3 show how $DoubleBlowup_A$ is carved up based on the various intersection. Clearly the situation is similar for $DoubleBlowup_B$ and $DoubleBlowup_C$.

5 LOCATING RELATED POINTS

Throughout this section, it will always be assumed that the control points triangle is *acute*. Also, when counting points, we will restrict our attention to points in the upper-half space, recalling again that the reflection of any point about the xy-plane is strongly related to the first point. We will be concerned with moving particles, and their motion will be assumed to be continuous and smooth. Remember that by Corollary 1, points not on the danger cylinder or *CSDC* have exactly one weak relative, if they are outside *CSDC*, or

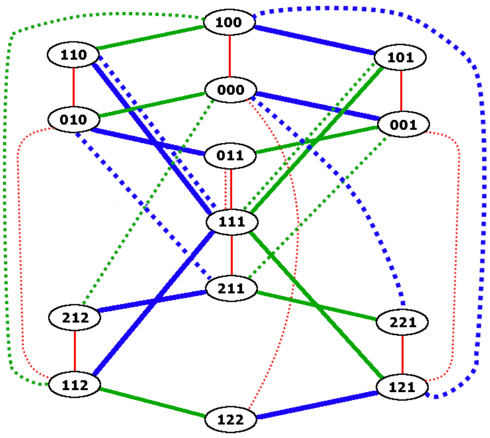

Figure 4: Transition graph of toroidal regions

exactly three weak relatives, if they are inside *CSDC*.

Lemma 9 will be repeatedly exploited in this section. It tell us that whenever a particle passes through a basic double toroid, in an ordinary way, exactly one of its weak relatives will pass through the corresponding control point. Notice too that when passing through a control point, a particle's corresponding digit in its region descriptor will change from 0 to 1, or vice versa, and each of the other two digits either remain 1, or else change from 0 to 2, or vice-versa.

Various regions, such as "in toroidal region 100, and inside *CSDC*" will be considered. While there is considerable evidence that such regions are connected, and arguments concerning this can be advanced in light of results in the next section, such connectivity will generally not be assumed here.

Lemma 10. *Figure 4 is a graph that has a node for each of the fourteen toroidal regions. A solid edge is drawn between two nodes of the graph if and only if the corresponding regions share a two-dimensional portion of a basic double toroid, as a common boundary. A dotted edge is drawn between two nodes if and only if it is possible to move smoothly from one region to the other by just passing through a control point. The edge thickness reflects which basic double toroid or which control point is used: thin for* $DoubleToroid_A$ *or A, medium for* $DoubleToroid_B$ *or B, and thick for* $DoubleToroid_C$ *or C.*

Proof. It must first be established that there are fourteen toroidal regions, and that these are as labeled in the graph. Because "0" and "2" cannot both occur in a region descriptor, because of Lemma 8 (see below), we are limited to these possibilities: 000, 100, 110, 111, 122, 112, and 222, and descriptors obtained from these by permuting the digits. However, region 222 does not exist. To see this, notice that $Toroid_{\pi-A} \cap Toroid_{\pi-B} \cap Toroid_{\pi-C}$ consists of a single point, namely, the orthocenter of the control points triangle. It is then easy to see that all other points are outside at least one of the three toroids involved here.

It is straightforward to check that each of the other descriptors describes a valid toroidal region, by simply identifying a point inside each region. Much of this can be seen by considering the double covers of the blowups of the control points (see Figure 3).

Next, the solid edges in the graph are easily checked. For instance, if a particle is in region 000, and close to $Toroid_A$, then it can pass through this surface, and will clearly arrive in region 100.

Checking the dotted edges is only slightly more tricky. For instance, if a particle in region 010, and near control point A, it can travel through A. In doing so, it will also pass through $Toroid_A$, moving inside it. Also, its second are third view angles will instantaneously (discontinuously) change to their supplementary angles. Thus, the particle will transition from region 010 to region 112. (Although the middle digit here remains "1", the second view angle will have changed to its supplementary angle.) The other dotted edges can be similarly checked.

□

Given a point p in some toroidal region, it is simple to determine up to four possible toroidal regions that could contain its weak relatives, since the view angles must either all be the same as those of p, or else one of the angles must be the same, and the other two must be supplementary to the corresponding angles of p. For instance, by this reasoning, if p is in region 122, then each of its weak relatives must be in region 122 or 100 or 120 or 102. However, regions 120 and 102 do not actually exist, because, by Lemma 8, a "0" implies being outside the unit sphere, while a "2" implies being inside the unit sphere. Therefore a toroidal region descriptor cannot contain both a "0" and a "2".

If p is instead in the region 111, then each of its weak relatives must also be in this region. However, when comparing p to a weak relative q, it will help to indicate when they have supplementary view angles instead of equal view angles. For this purpose, we might speak of p being in *state* 111, and q being in state $1\underline{1}\,\underline{1}$, as a way of saying that p and q are both located in region 111, have the same first view angle, but the other two view angles for q are supplementary to those for p. To indicate supplementary angles, it does not matter whether we underline a "1" in the state descriptor for p or for q, as long as we underline one or the other, but not both.

Lemma 11. *If a point is in the toroidal region 000, then so too are all of its weak relatives, and these are in fact, strong relatives.*

Proof. Approaching the limiting case discussed in (Rieck, 2015) (*i.e.* $z \to \infty$), all of the view angles tend to zero. Here we are dealing with points in the toroidal region 000, and so this region is nonempty. The results in (Rieck, 2015) indicate that in this limiting case, any point inside *CSDC* (which here is just the standard deltoid curve) has three strong relatives, and any point outside the *CSDC* has only one strong relatives.

Thus, in the limiting case, the weak relatives are all accounted for by these strong relatives. If a particle moves from the limiting case, and does not cross a basic toroid, it will remain in region 000, and all points of this region can be reached in this manner, since it is obvious that this region is connected. If the particle also avoids the CSDC, then its number of weak relatives will not change. Moreover, by continuity, these weak relatives will remain strong relatives.

□

Lemma 12. *Inside CSDC, four weakly related particles can be in any of the following quadruples of toroidal region states, and no others:*

$$(000,\ 000,\ 000,\ 000),\ (100,\ 100,\ 100,\ 122),$$
$$(110,\ 110,\ \underline{1}12,\ 1\underline{1}2),\ (111,\ \underline{1}\,\underline{1}1,\ \underline{1}1\underline{1},\ 1\underline{1}\,\underline{1}),$$

and similar state quadruples obtained from these by permuting the entries in the triples together, and/or by permuting the triples in a quadruple.

Proof. Consider a particle p, initially outside the three basic toroids, and inside *CSDC*. So the state quadruple for it and its weak relatives is initially (000, 000, 000, 000). Now, while remaining inside *CSDC*, it is possible for p to move to any point inside all the basic toroids, and inside *CSDC*, in such a way that it stays inside *CSDC*, and crosses each of the three basic toroids just one time. Without loss of generality, suppose that it crosses $Toroid_A$ first, then $Toroid_B$, and then $Toroid_C$. p could move in this way to any point in toroidal region 111, inside *CSDC*. Also, any point in region 100, inside *CSDC*, could be reached by stoping prior to crossing $Toroid_B$. Similarly for region 110, inside *CSDC*.

Regarding the particle ordering suggested by the quadruple notation, assume that p comes first. When p passes through $Toroid_A$, the other particles must do likewise, *except* precisely one of them must also pass through the control point A, by Lemma 9. If we assume that the last particle goes through A, then the state quadruple will become (100, 100, 100, 122).

When p next pass through $Toroid_B$, one of the particles must go through B, but this is not possible for the last particle. Assuming the third particle goes through B, the state quadruple (110, 110, $\underline{1}12$, $1\underline{1}2$) results. p will then pass through $Toroid_C$, but one of the first two particles, say the second one, would need to pass through C. The result would be (111, $\underline{1}\,\underline{1}1$, $\underline{1}1\underline{1}$, $1\underline{1}\,\underline{1}$). All of the state quadruples encountered so far are therefore possible. The state quadruples obtained symmetrically from these state quadruples are also achievable, of course.

Now, by continuity, while a particle remains in a particular toroidal region, and inside *CSDC*, the state quadruple associated with it and its weak relatives will not change. We have already accounted for each of the possible toroidal regions here, and thus we have determined all of the possible state quadruples.

□

Lemma 13. *Outside CSDC, for a pair of weakly related particles, the following pairs of toroidal region states are possible:*

$$(000,\ 000)\,,\ (100,\ 122)\,,\ (\underline{1}12,\ 1\underline{1}2)\,,$$

and similar state pairs obtained from these by permuting the entries in the triples together, and/or by swapping the triples in a pair.

Proof. To make the following argument for these state pairs, we will need to know that it is possible for a particle p to begin outside the three basic toroids, pass through one of them, say $Toroid_A$, and then pass through one of the apexes of this toroid, say B, all the while staying outside *CSDC*. Figure 3 suggests that this is possible, but producing a rigorous argument is a bit tedious. Such an argument should also ensure that any point outside *CSDC*, and in either region 100 or 112 is reachable in this way.

Let q be the weak relative of p. The state pair for (p,q), as p travels the suggested path, will change from (000, 000) to (100, 122), and then to ($\underline{1}$12, 1$\underline{1}$2), so the listed state pairs are achievable. So too are the state pairs symmetrically obtained from these state pairs.

□

Lemma 14. *If a particle is outside CSDC, and in the toroidal region 110 or 101 or 011, then its weakly related particle is in the same region, and these two particles are strongly related. Also, if a particle is outside CSDC, and in the toridal region 111, then it is not strongly related to its weakly related particle. Together with the previous lemma, this determines all possible state pairs when weakly related particles are outside CSDC.*

Proof. Suppose a particle p is in region 011, and outside *CSDC*. Let q be the weak relative of p. The only conceivable state pairs for (p,q) are these: (011, 011), (011, 0$\underline{1}\underline{1}$), (011, 2$\underline{1}$1) and (011, 21$\underline{1}$). However, (011, 2$\underline{1}$1) and (011, 21$\underline{1}$) are not possible because Lemma 13 shows that a point in region 211, and outside *CSDC*, is paired with another point in region 211, not region 011.

Next consider the state pair (011, 0$\underline{1}\underline{1}$). It can be checked that in this case, neither p nor q can pass through a control point. (See Figure 3 for the case of A, and for B and C, simply observe that both a "0" and a "2" would occur in a region descriptor.) So neither can pass through a basic double toroid, not without first going through *CSDC*. If p passes through *CSDC*, then q must also pass through it, and simultaneously a pair of (real) weak relatives will be created on the danger cylinder. After passing through *CSDC*, the resulting quadruple of toroidal region states for the resulting particles (p,q,r,s) would be have this form: (011, 0$\underline{1}\underline{1}$, ..., ...). However, this does not match the possibilities listed in Lemma 12, and hence is actually impossible. The only remaining possible state pair is (011, 011), and so p and q must be strongly related. Similarly if p in instead in region 101 or 110.

Next, suppose instead that p and its weak relative q have initial state pair (111, 111). Let p pass through $Toroid_A$ as q passes through A. The state pair changes to (011, 0$\underline{1}\underline{1}$)), which we know now is impossible. Thus p and q cannot be strongly related if they are in region 111. For each possible toroidal region, outside *CSDC*, the state pair for a particle in this region, and its weak relative, have now been identified.

□

We are now ready to state and prove the principle result of this section.

Theorem 3. *Assume the control points triangle is acute. Consider a general point p in the upper-half space of three dimensional (real) space. We will consider only the upper-half space in the following.*

1. *If p is outside CSDC, and outside all three basic toroids, then there is one other point with the same view angles as p.*
2. *If p is outside CSDC, and outside two basic toroids, but inside the other, then there are no other points with the same view angles as p.*
3. *If p is outside CSDC, and outside one basic toroid, but inside the other two, then there is one other points with the same view angles as p.*
4. *If p is outside CSDC, and inside all three basic toroids, then there are no other points with the same view angles as p.*
5. *If p is inside CSDC, and outside all three basic toroids, then there are three other points with the same view angles as p.*
6. *If p is inside CSDC, and outside two basic toroids, but inside the other, then there are two other points with the same view angles as p.*
7. *If p is inside CSDC, and outside one basic toroid, but inside the other two, then there is one other points with the same view angles as p.*
8. *If p is inside CSDC, and inside all three basic toroids, then there are no other points with the same view angles as p.*

Proof. Together, Lemmas 11 through 14 describe all of the possible state pairs (in the outside *CSDC* case) and possible state quadruples (in the inside *CSDC* case). If p is outside *CSDC*, and outside all three basic toroids, then the corresponding state pair for p and its weak relative is (000, 000), and so the two points

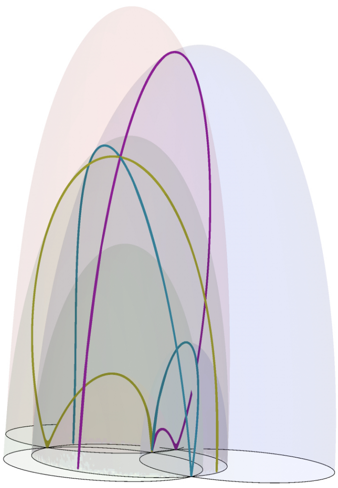

Figure 5: Pairwise intersections of the basic double toroids

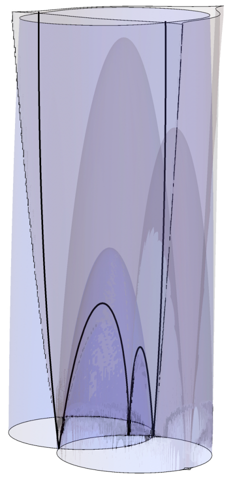

Figure 6: Intersection of *CSDC* and a basic double toroid

are strongly related. If instead, p is outside *CSDC*, outside two basic toroids, but inside the other basic toroid, then, without loss of generality, the state pair will be (100, 122), and so p does not have a strong relative. If instead, p is outside *CSDC*, outside one basic toroid, but inside the other two basic toroids, then, without loss of generality, the state pair will be (110, 110), and so the two points are strongly related. If instead, p is outside *CSDC*, and inside all three basic toroids, then, without loss of generality, it is in toroidal region 122, 112 or 111. In any of these cases, it is seen that p does not have a weak relative in the same state as p, and so p has no strong relatives.

If p is inside *CSDC*, and outside all three basic toroids, then Lemma 11 ensures that its three weak relatives are actually strongly related to p. If instead, p is inside *CSDC*, and outside two basic toroids, but inside the other basic toroid, then, without loss of generality, the state quadruple is (100, 100, 100, 122), and so p has two strong relatives. If instead, p is inside *CSDC*, and outside one basic toroid, but inside the other two basic toroids, then, without loss of generality, the state quadruple is (110, 110, $1\underline{1}2$, $\underline{1}12$), and so p has one strong relative. If instead, p is inside *CSDC*, and inside all three basic toroids, then, without loss of generality, it is in toroidal region 111, 122 or 112. In any of these cases, it is seen that p does not have a weak relative in the same state as p, and so p has no strong relatives.

□

6 INTERSECTIONS OF SURFACES

In order to gain a good sense of the various surfaces of relevance to the P3P problem, and in order to produce highly accurate images of how they intersect, parametrical formulas were obtained for the various intersection curves. In all cases, the curves are real rational curves when considered in "*xyZ*-space," meaning that the totality of possible triples (x, y, Z) in three-dimensional real space constitute a rational curve, in each case. It has proven to be quite useful here to regard the *xy*-plane as the complex plane, for easier manipulation of rather complicated formulas. However, this does *not* mean that we are here interested in "complexified" versions of the curves, that is, we are *not* treating x, y and Z as complex variables. They remain real variables, but as has been the practice throughout this paper, we are consolidating x and y into a single complex variable ζ by taking $\zeta = x + iy$.

The formulas given for each curve (below) provide explicit values for ζ, $\overline{\zeta}$ and Z as rational functions of ω, where ω is free to range over the unit complex numbers (*i.e.* $\omega \in \mathfrak{C}$). Clearly, formulas then become immediately available for x, y and z, by using $x = (\zeta + \overline{\zeta})/2$, $y = (\zeta - \overline{\zeta})/2i$, and $z = \pm\sqrt{Z}$.

Figure 5 shows the curves obtained, in the upper-half space (of *xyz*-space) when pairwise intersecting the three basic double toroids, $DoubleToroid_A$, $DoubleToroid_B$, $DoubleToroid_C$. Figure 6 shows the intersection of one of these double toroids with *CSDC* (and suffers a bit from a rendering issue).

Figure 7 shows the intersections of all the basic double toroids with each other, and with *CSDC*. This becomes rather messy inside the unit sphere. Notice too that the three basic double toroids intersect in a unique point in the upper-half space, which can be shown to be above the reflection of the orthocenter about the origin.

Theorem 4. *The (real) rational curve in xyZ-space, given by the following parameterization lies on the intersection of the double toroids $DoubleToroid_B$ and $DoubleToroid_C$ (when realized in xyZ-space):*

$$\zeta = \frac{(1+\zeta_2\zeta_3^2)}{(\zeta_3-\zeta_2)(1-\zeta_3+\zeta_2\zeta_3)}\rho\,\omega^{-1} + \frac{(1+\zeta_2^2\zeta_3)}{(\zeta_2-\zeta_3)(1-\zeta_2+\zeta_2\zeta_3)}\rho\,\omega$$

$$\overline{\zeta} = \frac{(1+\zeta_2^2\zeta_3)\zeta_3}{(\zeta_3-\zeta_2)(1-\zeta_3+\zeta_2\zeta_3)}\rho\,\omega^{-1} + \frac{(1+\zeta_2\zeta_3^2)\zeta_2}{(\zeta_2-\zeta_3)(1-\zeta_2+\zeta_2\zeta_3)}\rho\,\omega$$

$$\begin{aligned} Z = {} & (1+\zeta_2^2\zeta_3)(1+\zeta_2\zeta_3^2)\cdot \\ & [(\zeta_3-\zeta_2\zeta_3+\zeta_2\zeta_3^2)\omega^{-1}+(-\zeta_2+\zeta_2\zeta_3-\zeta_2^2\zeta_3)\omega]\cdot \\ & [(-1+\zeta_2-\zeta_2\zeta_3)\omega^{-1}+(1-\zeta_3+\zeta_2\zeta_3)\omega \\ & +(\zeta_3-\zeta_2)\rho]\ /\ [\zeta_2\zeta_3(\zeta_2-\zeta_3)^2\cdot \\ & (1-\zeta_2+\zeta_2\zeta_3)(1-\zeta_3+\zeta_2\zeta_3)] \end{aligned}$$

where $\rho = [(1-\zeta_2+\zeta_2\zeta_3)(1-\zeta_3+\zeta_2\zeta_3)/(\zeta_2\zeta_3)]^{1/2}$ is constant, and ω ranges over complex numbers such that $|\omega| = 1$.

Proof. Using the equations discussed in the previous section for $DoubleToroid_B$ and $DoubleToroid_C$, eliminate Z $(= z^2)$, to obtain the equation $\zeta_2^2\zeta_3^2(1+\zeta_2^2\zeta_3)(1+\zeta_2\zeta_3^2)\zeta^2 + \zeta_2\zeta_3(1+\zeta_2^2\zeta_3)(1+\zeta_2\zeta_3^2)\overline{\zeta}^2 - \zeta_2\zeta_3(\zeta_2+\zeta_3+4\zeta_2^2\zeta_3^2+\zeta_2^4\zeta_3^3+\zeta_2^3\zeta_3^4)\zeta\overline{\zeta} - (\zeta_2^3\zeta_3^3-1)^2 = 0$. A direct check shows that the formulas given in the theorem for ζ and $\overline{\zeta}$ satisfy this equation. Next, set the formulas for $DoubleTorioid_B$ and $DoubleTorioid_C$ equal to each other, and solve the resulting equation for Z to find that $Z = (1-\zeta\overline{\zeta})[(\overline{\zeta_2}-\overline{\zeta_3})\zeta + (\zeta_2-\zeta_3)\overline{\zeta} + \zeta_1\overline{\zeta_3} - \zeta_1\overline{\zeta_2} + \overline{\zeta_1}\zeta_3 - \overline{\zeta_1}\zeta_3]\ /\ [(\overline{\zeta_2}-\overline{\zeta_3})\zeta + (\zeta_2-\zeta_3)\overline{\zeta}]$. This can then be manipulated using the formulas for ζ and $\overline{\zeta}$ to obtain the formula for Z in the theorem.

□

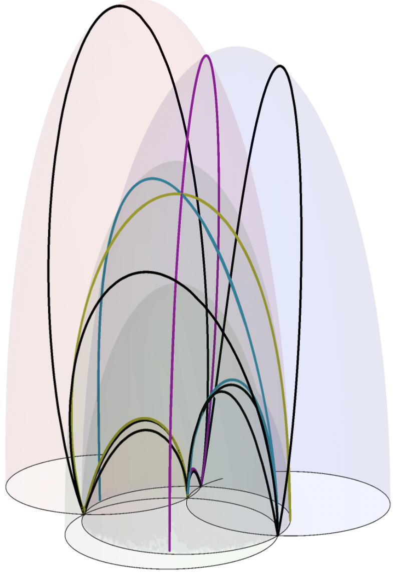

Figure 7: Several intersections

Theorem 5. *The (real) rational curve in xyZ-space, given by the following parameterization lies on the intersection of $DoubleToroid_A$ and CSDC (when realized in xyZ-space):*

$$\zeta = \frac{-1+2\zeta_2\zeta_3\omega+(\zeta_2+\zeta_3-\zeta_2^2\zeta_3^2)\omega^2-2\zeta_2\zeta_3(\zeta_2+\zeta_3)\omega^3-\zeta_2\zeta_3(\zeta_2+\zeta_3)\omega^6}{[1-\zeta_2\zeta_3(\zeta_2+\zeta_3)-2\zeta_2\zeta_3\omega-(\zeta_2+\zeta_3-\zeta_2^2\zeta_3^2)\omega^2]\,\omega^2}$$

$$\overline{\zeta} = \frac{-\zeta_2-\zeta_3-2(\zeta_2+\zeta_3)\omega^3+(\zeta_2\zeta_3(\zeta_2+\zeta_3)-1)\omega^4+2\zeta_2\zeta_3\omega^5-\zeta_2^2\zeta_3^2\omega^6}{[1-\zeta_2\zeta_3(\zeta_2+\zeta_3)-2\zeta_2\zeta_3\omega-(\zeta_2+\zeta_3-\zeta_2^2\zeta_3^2)\omega^2]\,\omega^2}$$

$$Z = \frac{-(\zeta_2+\zeta_3)(\zeta_2\zeta_3+\omega^2)(1-\zeta_2\zeta_3\omega)^2(1-\zeta_2\omega^2)^2(1-\zeta_3\omega^2)^2}{\zeta_2\zeta_3[1-\zeta_2\zeta_3(\zeta_2+\zeta_3)-2\zeta_2\zeta_3\omega-(\zeta_2+\zeta_3-\zeta_2^2\zeta_3^2)\omega^2]^2\,\omega^4}$$

where ω ranges over complex numbers such that $|\omega| = 1$.

Proof. If a point (ζ, z) is on *CSDC*, then $\zeta_L \in \mathfrak{D}$. In this case, there exists a unique $\omega \in \mathfrak{C}$ satisfying $\omega^2 - 2\overline{\omega} = \omega^2 - 2/\omega = \zeta_L$, as discussed in the proof of Theorem 2. Using the formula for ζ' in (2) in place of ζ here (because of Theorem 1), we obtain a polynomial equation connecting ω, ζ, $\overline{\zeta}$ and Z. If we assume further that (ζ, z) is on $DoubleToroid_A$, then Lemma 7 provides another polynomial equation relating ζ, $\overline{\zeta}$ and Z. By eliminating Z from these two equations, we obtain a single polynomial equation connecting the variables ω, ζ and $\overline{\zeta}$:

$$4\zeta_2\zeta_3(\zeta_2^2+\zeta_2\zeta_3+\zeta_3^2)-4\zeta_2\zeta_3(\zeta_2+\zeta_3)\zeta+(4\zeta\zeta_2^2\zeta_3^2-4\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3))\overline{\zeta}$$

$$+[4\zeta_2\zeta_3(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)-2(-\zeta_2^2-\zeta_3^2+\zeta_2^3\zeta_3^2+\zeta_2^2\zeta_3^3)\zeta+2(\zeta_2+\zeta_3)(-1+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)\zeta^2-2\zeta_2\zeta_3(\zeta_2+\zeta_3)\zeta^3+2\zeta_2\zeta_3(-3\zeta_2^2-4\zeta_2\zeta_3-3\zeta_3^2+\zeta_2^3\zeta_3^2+\zeta_2^2\zeta_3^3)\overline{\zeta}-2\zeta_2^2\zeta_3^2(\zeta_2^2+4\zeta_2\zeta_3+\zeta_3^2)\zeta\overline{\zeta}+2\zeta_2\zeta_3(2+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)\zeta^2\overline{\zeta}+6\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3)\overline{\zeta}^2-4\zeta_2^2\zeta_3^2\zeta\overline{\zeta}^2]\,\omega$$

$$+[-(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)^2+2(1+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)\zeta-(1+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)^2\zeta^2+(\zeta_2+\zeta_3)(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)\zeta^3-(\zeta_2+\zeta_3)\zeta^4-2\zeta_2\zeta_3(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)\overline{\zeta}+(-\zeta_2^2-\zeta_3^2+2\zeta_2^3\zeta_3^2+2\zeta_2^2\zeta_3^3+\zeta_2^4\zeta_3^4)\zeta\overline{\zeta}-\zeta_2\zeta_3(3\zeta_2^2+8\zeta_2\zeta_3+3\zeta_3^2+\zeta_2^3\zeta_3^2+\zeta_2^2\zeta_3^3)\zeta^2\overline{\zeta}+(1+2\zeta_2^2\zeta_3+2\zeta_2\zeta_3^2)\zeta^3\overline{\zeta}-\zeta_2\zeta_3(-2\zeta_2^2-3\zeta_2\zeta_3-2\zeta_3^2+2\zeta_2^3\zeta_3^2+2\zeta_2^2\zeta_3^3)\overline{\zeta}^2+2\zeta_2\zeta_3(\zeta_2+\zeta_3+\zeta_2^3\zeta_3+3\zeta_2^2\zeta_3^2+\zeta_2\zeta_3^3)\zeta\overline{\zeta}^2-\zeta_2\zeta_3(2+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)\zeta^2\overline{\zeta}^2-2\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3)\overline{\zeta}^3+\zeta_2^2\zeta_3^2\zeta\overline{\zeta}^3]\,\omega^2$$

$$+[-4\zeta_2\zeta_3(\zeta_2^2+\zeta_2\zeta_3+\zeta_3^2)+4\zeta_2\zeta_3(\zeta_2+\zeta_3)\zeta+4\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3)\overline{\zeta}-4\zeta_2^2\zeta_3^2\zeta\overline{\zeta}]\,\omega^3$$

$$+[-2\zeta_2\zeta_3(\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2)+(-\zeta_2^2-\zeta_3^2+\zeta_2^3\zeta_3^2+\zeta_2^2\zeta_3^3)\zeta-(\zeta_2+\zeta_3)(-1+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)\zeta^2+\zeta_2\zeta_3(\zeta_2+\zeta_3)\zeta^3-\zeta_2\zeta_3(-3\zeta_2^2-4\zeta_2\zeta_3-3\zeta_3^2+\zeta_2^3\zeta_3^2+\zeta_2^2\zeta_3^3)\overline{\zeta}+\zeta_2^2\zeta_3^2(\zeta_2^2+4\zeta_2\zeta_3+\zeta_3^2)\zeta\overline{\zeta}-\zeta_2\zeta_3(2+\zeta_2^2\zeta_3+\zeta_2\zeta_3^2)\zeta^2\overline{\zeta}-3\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3)\overline{\zeta}^2+2\zeta_2^2\zeta_3^2\zeta\overline{\zeta}^2]\,\omega^4$$

$$+[\zeta_2\zeta_3(\zeta_2^2+\zeta_2\zeta_3+\zeta_3^2)-\zeta_2\zeta_3(\zeta_2+\zeta_3)\zeta-\zeta_2^2\zeta_3^2(\zeta_2+\zeta_3)\overline{\zeta}+\zeta_2^2\zeta_3^2\zeta\overline{\zeta}]\,\omega^6=0.$$

A direct computation reveals that the equation is satisfied when the formulas for ζ and $\overline{\zeta}$ in the theorem are substituted. (This computation is quite tedious, and so best performed with software.) Moreover, the first of the two equations used to eliminate Z is linear in Z, and when solved for Z, reveals that

$$Z=\frac{(1-\zeta\overline{\zeta})\,[\zeta_2\zeta_3\zeta^2-(1+\zeta_2\zeta_3(\zeta_2+\zeta_3))\,\zeta-\zeta_2\zeta_3\overline{\zeta}+\zeta_2+\zeta_3+\zeta_2^2\zeta_3^2]}{\zeta_2\,\zeta_3\,[2+(\zeta^2-2\overline{\zeta})\,\omega+\omega^3]}.$$

When the substitutions for ζ and $\overline{\zeta}$ are made in this formula, it becomes straightforward to manipulate the resulting expression to obtain the formula given for Z in the theorem.

□

Theorem 6. *The (real) rational curve in xyZ-space, given by the following parameterization lies on the intersection of DoubleToroid$_A$ and the danger cylinder (when realized in xyZ-space):*

$$\zeta = \omega$$

$$\overline{\zeta} = 1/\omega$$

$$Z = \frac{(\zeta_2+\zeta_3)(\omega^2+\zeta_2\zeta_3)}{\zeta_2\zeta_3\,\omega}$$

where ω *ranges over complex numbers such that* $|\omega|=1$.

Proof. When $\overline{\zeta}$ is set equal to $1/\zeta$, meaning that (ζ,z) is on the danger cylinder, and this is substituted into the formula for *DoubleToroid$_A$*, we obtain a simple equation, as follows:

$$Z\,[(\zeta_2+\zeta_3)(\zeta^2+\zeta_2\zeta_3)-\zeta_2\zeta_3\,\zeta Z]\;=\;0.$$

Assuming that $Z\neq 0$, we find that

$$Z\;=\;\frac{(\zeta_2+\zeta_3)(\zeta^2+\zeta_2\zeta_3)}{\zeta_2\zeta_3\,\zeta}.$$

Of course, it is required that $\zeta\in\mathfrak{C}$. It is then clear that the stated parameterized curve is on the intersection of the danger cylinder and *DoubleToroid$_A$*.

□

Obviously, each of the preceding three theorems can be altered, by simply rotating the indices, to similarly handle the other basic double toroids. Moreover, for each of these theorems, the unit circle $\mathfrak{C}$ is also part of the intersection of the two surfaces being considered. In fact, the parameterized curve in the theorem, together with $\mathfrak{C}$, constitute the entire intersection of the two surfaces involved. This claim is not proved here, however.

7 THE DISCRIMINANT OF FINSTERWALDER'S CUBIC

The method of S. Finsterwalder, as descibed in (Haralick et al., 1994), is traditionally one of the better methods for solving the P3P problem, and is the basis for the most successful method currently used (known as "Lambda Twist"). The problem is reduced to finding a root of a cubic equation, which is equation

(14) in (Haralick et al., 1994) (but "λ" is here replaced with "Λ" for practical reasons):

$$G\Lambda^3 + H\Lambda^2 + I\Lambda + J \;=\; 0.$$

Using the notation of the present article, the coefficients of the cubic are as follows:

$$\begin{aligned}
G \;&=\; d_3^2\,[d_3^2(1-c_2^2) - d_2^2(1-c_3^2)\,] \\
H \;&=\; d_2^2(d_2^2-d_1^2)(1-c_3^2) + d_3^2(d_3^2+2d_1^2)(1-c_2^2) \\
&\quad + 2d_2^2 d_3^2\,(c_1c_2c_3 - 1) \\
I \;&=\; d_2^2(d_2^2-d_3^2)(1-c_1^2) + d_1^2(d_1^2+2d_3^2)(1-c_2^2) \\
&\quad + 2d_1^2 d_2^2\,(c_1c_2c_3 - 1) \\
J \;&=\; d_1^2\,[d_1^2(1-c_2^2) - d_2^2(1-c_1^2)\,].
\end{aligned}$$

Define $\mathcal{G} = G/\eta^2$, $\mathcal{H} = H/\eta^2$, $\mathcal{I} = I/\eta^2$, and $\mathcal{J} = J/\eta^2$. Then $\mathcal{G}/d_3^2$ and $\mathcal{J}/d_1^2$ are of the form described in (Rieck, 2014)[Theorem 1 and Lemma 10]. Moreover, each of $\mathcal{G}$, $\mathcal{H}$, $\mathcal{I}$ and $\mathcal{J}$ are of the form described in (Rieck, 2018)[Section 3], and satisfy conditions (5) found in Lemma 10 there. As such, each can be expressed as a linear combination of 1, $\mathcal{L}$ and $\mathcal{K}$, with coefficients that depend only on the control point positions, and where $\mathcal{K} = -1-\mathcal{R}$, and $\mathcal{L}$ and $\mathcal{R}$ are as in (Rieck, 2018). So,

$$\begin{aligned}
&\mathcal{K} = \big[\,(x_1^2-y_1^2+2x_1)\,(1-c_1^2) + (x_2^2-y_2^2+2x_2) \\
&\quad\cdot(1-c_2^2) + (x_3^2-y_3^2+2x_3)\,(1-c_3^2) \\
&\quad - 2x_H\,(1-c_1c_2c_3)\big]\,/\,\eta^2 \;\text{ and} \\
&\mathcal{L} = 2\,\big[\,(x_1-1)y_1\,(1-c_1^2) + (x_2-1)y_1\,(1-c_2^2) \\
&\quad + (x_3-1)y_1\,(1-c_3^2) + y_H\,(1-c_1c_2c_3)\,\big]\,/\,\eta^2.
\end{aligned} \tag{5}$$

Let us write $\mathcal{G} = \iota_G + \kappa_G\mathcal{K} + \lambda_G\mathcal{L}$, $\mathcal{H} = \iota_H + \kappa_H\mathcal{K} + \lambda_H\mathcal{L}$, $\mathcal{I} = \iota_I + \kappa_I\mathcal{K} + \lambda_I\mathcal{L}$, and $\mathcal{J} = \iota_J + \kappa_J\mathcal{K} + \lambda_J\mathcal{L}$, for constants ι_G, κ_G, λ_G, ι_H, κ_H, λ_H, ι_I, κ_I, λ_I, ι_J, κ_J and λ_J that depend only on the control point positions.

These coefficients can be found as follows. When $\mathcal{J}$ is described as a function of the position (x,y,z) of the optical center, if $x^2+y^2 = 1$, then (Rieck, 2014)[Theorem 1] guarantees that $\mathcal{J}$ does not depend on z. Moreover, (Rieck, 2014)[Lemma 10] provides a way to compute it. As there, let e_α be the distance from (x,y) to the control point (x_α, y_α) $(\alpha = 1,2,3)$. Let f_1 be the distance from (x,y) to the line through (x_2,y_2) and (x_3,y_3), and similarly for f_2 and f_3.

(Rieck, 2014)[Lemma 10] ensures that $\mathcal{J}/d_1^2$ equals $4(e_2^2-f_1^2-e_1^2+f_2^2)/d_3^2$ when $x^2+y^2=1$. Letting $t_\alpha = \tan(\phi_\alpha/2)$, it can be checked directly that $e_\alpha^2 = [(1+t_\alpha^2)(1+x^2+y^2) + 2(t_\alpha^2-1)x - 4t_\alpha y]/(1+t_\alpha^2)$. Also, $f_1^2 = [(t_2t_3+1) + (t_2t_3-1)x - (t_2+t_3)y]^2\,/\,[(1+t_2^2)(1+t_3^2)]$, and similarly for f_2^2 and f_3^2. Also, from (Rieck, 2014)[Lemma 1], $d_1^2 = 4(t_2-t_3)^2\,/\,(1+t_2^2)(1+t_3^2)$, and similarly for d_2^2 and d_3^2. The values of these various quantities now become of particular interest in the special cases where (x,y) is $(1,0)$ or $(-1,0)$ or $(0,1)$, which can now be immediately computed in terms of t_1, t_2 and t_3.

Next, consider $\mathcal{J} = \iota_J + \kappa_J\mathcal{K} + \lambda_J\mathcal{L}$. Assuming again, for the moment, that $x^2+y^2=1$, it is seen from (Rieck, 2018)[Theorem 1] that $\mathcal{K} = -1-\mathcal{R} = x^2 - 2x - y^2$ and $\mathcal{L} = 2(1+x)y$. So, if $(x,y) = (1,0)$, then we get $\mathcal{J} = \iota_J - \kappa_J$. If instead, $(x,y) = (-1,0)$, then we get $\mathcal{J} = \iota_J + 3\kappa_J$. If instead, $(x,y) = (0,1)$, then we get $\mathcal{J} = \iota_J - \kappa_J + 2\lambda_J$. Using the computed values from the previous paragraph, and using the fact that $t_1+t_2+t_3 = t_1t_2t_3$ (because $\phi_1+\phi_2+\phi_3 = 0$), it is then straightforward to compute the following values:

$$\begin{aligned}
\iota_J \;&=\; 4\,(t_2-t_3)^2\,(t_1t_2-1)\,(t_3^2-3)\,t_3 \\
&\quad\;/\,[\,(1+t_2^2)\,(1+t_3^2)^2\,(t_1-t_2)\,] \\
\kappa_J \;&=\; 4\,(t_2-t_3)^2\,(t_1t_2-1)\,t_3 \\
&\quad\;/\,[\,(1+t_2^2)\,(1+t_3^2)\,(t_1-t_2)\,] \\
\lambda_J \;&=\; 4\,(t_2-t_3)^2\,(t_1t_2-1) \\
&\quad\;/\,[\,(1+t_2^2)\,(1+t_3^2)\,(t_2-t_1)\,]
\end{aligned}$$

By symmetry, formulas for κ_G, λ_G and ρ_G can be obtained from the formulas for κ_J, λ_J and ρ_J, respectively, by simply making the transposition $1 \leftrightarrow 3$ on the indices.

An analysis of $\mathcal{I}$ begins by noticing that $I = d_1^2\,G\,/\,d_3^2 + (d_1^2-d_2^2+d_3^2)\,J\,/\,d_1^2 + d_1^2d_2^2$, and so $\mathcal{I} = d_1^2\,\mathcal{G}\,/\,d_3^2 + (d_1^2-d_2^2+d_3^2)\,\mathcal{J}\,/\,d_1^2 + d_1^2d_2^2$. From this we can deduce formulas for ι_I, κ_I and λ_I in terms of t_1, t_2 and t_3. Again, by symmetry, using the transposition $1 \leftrightarrow 3$ on the indices, we then obtain similar formulas for κ_H, λ_H and ρ_H.

Again using $t_1+t_2+t_3 = t_1t_2t_3$, it becomes possible to put all these coefficients into the following helpful form, which can be checked by hand or with the aid of algebraic manipulation software. This will then be used to compute the discriminant of Finsterwalder's cubic.

Lemma 15. *Set*

$$q \;=\; (1+t_1^2)\,(t_2-t_3)\,/\,[\,(1+t_3^2)\,(t_2-t_1)\,]$$

and

$$M = 4\,(t_1t_3-1)\,(t_2-t_1)^2\,/\,[\,(t_2-t_3)\,(1+t_1^2)^3\,].$$

Then $\mathcal{G} = \iota_G + \kappa_G \mathcal{K} + \lambda_G \mathcal{L}$*,* $\mathcal{H} = \iota_H + \kappa_H \mathcal{K} + \lambda_H \mathcal{L}$*,* $\mathcal{I} = \iota_I + \kappa_I \mathcal{K} + \lambda_I \mathcal{L}$*, and* $\mathcal{J} = \iota_J + \kappa_J \mathcal{K} + \lambda_J \mathcal{L}$*, with*

$$\begin{aligned}
\iota_G &= M(3-t_1^2)t_1 \\
\kappa_G &= -M(1+t_1^2)t_1 \\
\lambda_G &= M(1+t_1^2)
\end{aligned}$$

$$\begin{aligned}
\iota_H &= 3Mq(2t_1+t_3-t_1^2 t_3) \\
\kappa_H &= -Mq(2t_1+t_3+3t_1^2 t_3) \\
\lambda_H &= Mq(3+t_1^2+2t_1t_3)
\end{aligned}$$

$$\begin{aligned}
\iota_I &= 3Mq(t_1+2t_3-t_1t_3^2) \\
\kappa_I &= -Mq(t_1+2t_3+3t_1t_3^2) \\
\lambda_I &= Mq(3+t_3^2+2t_1t_3)
\end{aligned}$$

$$\begin{aligned}
\iota_J &= Mq^3(3-t_3^2)t_3 \\
\kappa_J &= -Mq^3(1+t_3^2)t_3 \\
\lambda_J &= Mq^3(1+t_3^2).
\end{aligned}$$

Lemma 16. *The discriminant of Finsterwalder's cubic equals a nonzero constant (depending on the control point positions) times*

$$(\mathcal{K}^2+\mathcal{L}^2+12\mathcal{K}+9)^2 - 4(2\mathcal{K}+3)^3. \qquad (6)$$

Proof of Lemma 16. The discriminant of Finsterwalder's cubic equals the discriminant of this cubic polynomial (in Λ):

$$\mathcal{G}\Lambda^3 + \mathcal{H}\Lambda^2 + \mathcal{I}\Lambda + \mathcal{J}.$$

This in turn is a nonzero constant times the discriminant of the cubic polynomial (in $\Psi = q\Lambda$)

$$\mathfrak{g}\Psi^3 + \mathfrak{h}\Psi^2 + \mathfrak{i}\Psi + \mathfrak{j}$$

where

$$\begin{aligned}
\mathfrak{g} &= (3-t_1^2)t_1 - (1+t_1^2)t_1\mathcal{K} + (1+t_1^2)\mathcal{L} \\
\mathfrak{h} &= 3(2t_1+t_3-t_1^2t_3) - (2t_1+t_3+3t_1^2t_3)\mathcal{K} \\
&\quad + (3+t_1^2+2t_1t_3)\mathcal{L} \\
\mathfrak{i} &= 3(t_1+2t_3-t_1t_3^2) - (t_1+2t_3+3t_1t_3^2)\mathcal{K} \\
&\quad + (3+t_3^2+2t_1t_3)\mathcal{L} \\
\mathfrak{j} &= (3-t_3^2)t_3 - (1+t_3^2)t_3\mathcal{K} + (1+t_3^2)\mathcal{L}
\end{aligned}$$

To manage this computation, let us introduce new (atomic) indeterminates K and L, and then locally (just inside this proof) define

$$\begin{aligned}
g &= (3-t_1^2)t_1 - (1+t_1^2)t_1K + (1+t_1^2)L \\
h &= 3(2t_1+t_3-t_1^2t_3) - (2t_1+t_3+3t_1^2t_3)K \\
&\quad + (3+t_1^2+2t_1t_3)L \\
i &= 3(t_1+2t_3-t_1t_3^2) - (t_1+2t_3+3t_1t_3^2)K \\
&\quad + (3+t_3^2+2t_1t_3)L \\
j &= (3-t_3^2)t_3 - (1+t_3^2)t_3K + (1+t_3^2)L
\end{aligned}$$

The discriminant of the cubic polynomial (in w)

$$gw^3 + hw^2 + iw + j.$$

is rather simple, even when the coefficients are expanded as above. It turns out that it equals the following, which is straightforward to check:

$$4(t_1-t_3)^6\left[4(2K+3)^3 - (K^2+L^2+12K+9)^2\right].$$

The claim in the lemma now follows by substituting $\mathcal{K}$ and $\mathcal{L}$ for K and L, respectively.

□

The main result of this section is the following.

Theorem 7. *Grunert's system of three equations, i.e. the first three equations in (1), has a repeated solution, if and only if the corresponding cubic polynomial of Finsterwalder has a repeated root, if and only if* $(\mathcal{K}^2+\mathcal{L}^2+12\mathcal{K}+9)^2 = 4(2\mathcal{K}+3)^3$*. In this case, the quartic polynomial of Grunert also has a repeated root.*

Proof. The second "if and only if" is automatic since a polynomial has a repeated root if and only if its discriminant vanishes. Clearly any repeated solution to Grunert's system corresponds to a repeated root of the cubic polynomial. The converse must be established.

It is straightforward to check that ζ_L, defined in (2), equals $\mathcal{K} + \mathcal{L}i$. The discriminant of Finsterwalder's cubic (6) can then be seen to be a nonzero constant times $\mathcal{D}$, defined in (3). By Lemma 6, it can only be zero when a solution point is on the danger cylinder or *CSDC*, and in the latter case, some weak relative of the point is on the danger cylinder. Either way, the Grunert system, whose unknowns are s_1, s_2 and s_3, has a repeated solution, by (Rieck, 2014)[Proposition 1].

This establishes the first sentence of the theorem being proved here. Now, if Grunert's system has a repeated solution, then clearly Grunert's quartic polynomial has a repeated root, which is the second sentence.

□

Note: It is known that it is possible to have a repeated root of Grunert's quartic polynomial without having a repeated solution of Grunert's system. However, as we now see, a repeated root of Finsterwalder's cubic polynomial always ensures a repeated solution to Grunert's system, and hence a repeated solution to the extended Grunert system. In this sense, Finsterwalder's cubic is arguably more fundamental than Grunert's quartic.

8 CONCLUSION

Our initial motivation in this research was a desire to understand, as completely as possible, when points in space have the same three "view angles" to the three given "control points" of the P3P problem. More precisely, we wanted to know, given a reference point, where are the other points with the same view angles, and how many such points are there? We feel that we succeeded in this objective, especially in light of Theorem 3.

To truly understand this theorem, and why it is true, it is necessary to appreciate the nature of the various surfaces involved, and how they intersect. Quite a lot of detail in this regard has been presented. We argue that it is also helpful to understand when two points are related in the weaker sense that the lines connecting them to the control points constitute isomorphic triples of lines, in the sense that a rigid motion can be applied to move one of these triples of lines to the other.

While much has been accomplished in the present article, there certainly remains an opportunity to come to an even better understanding of how points are related, *viz.* the P3P problem. We believe that the dynamic approach that we used here could also benefit future researchers in this subject.

REFERENCES

Faugère, J.-C., Moroz, G., Rouillier, F., and El-Din, M. S. (2008). Classification of the perspective-three-point problem, discriminant variety and real solving polynomial systems of inequalities. In *ISSAC'08, 21st Int. Symp. Symbolic and Algebraic Computation*, 79–86.

Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F. (2003). Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(8): 930–943.

Grunert, J. A. (1841). Das pothenotische problem in erweiterter gestalt nebst über seine anwendungen in der geodäsie. In *Grunerts Archiv für Mathematik und Physik*, 1: 238–248.

Haralick, R. M., Lee, C.-N., Ottenberg, K., and Nölle, N. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. *J. Computer Vision*, 13(3): 331–356.

Rieck, M. Q. (2014). A fundamentally new view of the perspective three-point pose problem. *J. Mathematical Imaging and Vision*, 48(3):499-516.

Rieck, M. Q. (2015). Related solutions to the perspective three-point pose problem. *J. Mathematical Imaging and Vision*, 53(2):225-232.

Rieck, M. Q. (2018). On the discriminant of Grunert's system of algebraic equations and related topics *J. Mathematical Imaging and Vision*, 60(5): 737-762.

Sun, F.-M. and Wang, B. (2010). The solution distribution analysis of the P3P problem. In *SMC '10, Int. Conf. Systems, Man and Cybernetics*, 2033–2036.

Wang, B., Hu, H., and Zhang, C. (2019) Companion surface of danger cylinder and its role in solving variation of P3P problem tt arXiv:1906.08598 [cs.CV],

Wang, B., Hu, H., and Zhang, C. (2019) New insights on multi-solution distribution of the P3P problem tt arXiv:1901.11464 [cs.CV],

Zhang, C.-X. and Hu, Z.-Y. (2006). Why is the danger cylinder dangerous in the P3P problem? *Acta Automatica Sinica*, 32(4): 504–511.

APPENDIX

The following series of lemmas constitute a proof of the irreducibility claim made about the polynomial $\mathcal{P}$ in Lemma 5.

Lemma 17. *The following three polynomials in indeterminates* ζ *and* ω *(for any* $\zeta_H \in \mathbb{C}$*) are absolutely irreducible:* $\zeta\omega - 1$, $\zeta^2\omega^2 - 4(\zeta^3 + \omega^3) + 18\zeta\omega - 27$, $\zeta^2 - \zeta_H\zeta - \omega + \overline{\zeta_H}$.

Proof. Without loss of generality, we will regard the complex number field $\mathbb{C}$ as the field of coefficients, and show these polynomials are irreducible. The first polynomial is trivially irreducible. Next, suppose there is a non-trivial factorization $\zeta^2\omega^2 - 4(\zeta^3 +$

$\omega^3) + 18\zeta\omega - 27 \;=\; q_1 q_2$. The sum of the degrees of q_1 and q_2 must equal four. So each has degree 1, 2 or 3.

Without loss of generality, q_1 has degree three and q_2 has degree one, or else both have degree two. The first case then is $q_1 = a\zeta^3 + b\zeta^2\omega + c\zeta\omega^2 + d\omega^3 + e\zeta^2 + f\zeta\omega + g\omega^2 + h\zeta + j\omega + k$ and $q_2 = l\zeta + m\omega + n$. The second case is $q_1 = a\zeta^2 + b\zeta\omega + c\omega^2 + d\zeta + e\omega + f$ and $q_2 = g\zeta^2 + h\zeta\omega + j\omega^2 + k\zeta + l\omega + m$. Both cases are easily checked and found to be impossible. We thus conclude that $\zeta^2\omega^2 - 4(\zeta^3 + \omega^3) + 18\zeta\omega - 27$ is absolutely irreducible.

Next, consider a possible factorization of $\zeta^2 - \zeta_H\zeta - \omega + \overline{\zeta_H}$. This would need to be a product of linear polynomials, say $a\zeta + b\omega + c$ and $d\zeta + e\omega + f$. Clearly we need $ad = 1$, $be = 0$, and *w.l.o.g.*, $b = 0$. The coefficient of $\zeta\omega$ is then ae, and so we need either $a = 0$ or $e = 0$. But these are both clearly impossible. Therefore, $\zeta^2 - \zeta_H\zeta - \omega + \overline{\zeta_H}$ is absolutely irreducible.

□

Lemma 18. *Write*

$$\mathcal{P} \;=\; \mathcal{A}_4 Z^4 + \mathcal{A}_3 Z^3 + \mathcal{A}_2 Z^2 + \mathcal{A}_1 Z + \mathcal{A}_0\,,$$

where $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$ are polynomials in ζ and $\overline{\zeta}$. Then

$$\begin{aligned}
\mathcal{A}_0 \;&=\; (\zeta^2 - \zeta_H\zeta - \overline{\zeta} + \overline{\zeta_H})^2\,(\overline{\zeta^2} - \overline{\zeta_H}\,\overline{\zeta} - \zeta + \zeta_H)^2 \\
&\quad \cdot (\zeta\overline{\zeta} - 1)^2\,, \\
\mathcal{A}_1 \;&=\; (\zeta\overline{\zeta} - 1)\,\mathcal{A}_1' \;\; \textit{and} \\
\mathcal{A}_4 \;&=\; \zeta^2\,\overline{\zeta^2} - 4(\zeta^3 + \overline{\zeta^3}) + 18\zeta\overline{\zeta} - 27\,,
\end{aligned}$$

where $\mathcal{A}_1'$ is a polynomial of degree eight in ζ and $\overline{\zeta}$. Moreover, the polynomials $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ and $\mathcal{A}_4$ have respective degrees twelve, ten, eight, six and four.

Proof. Upon setting $Z = 0$, Z_L becomes $(\zeta^2 - \zeta_H\zeta - \overline{\zeta} + \overline{\zeta_H})\;(\zeta\overline{\zeta} - 1)$, $\overline{Z_L}$ becomes $(\overline{\zeta}^2 - \overline{\zeta_H}\,\overline{\zeta} - \zeta + \zeta_H)$ $(\zeta\overline{\zeta} - 1)$, $\widehat{\mathcal{D}}$ becomes $Z_L^2\overline{Z_L^2}$, and the claim concerning $\mathcal{A}_0$ is immediate. Regarding $\mathcal{D}$ as a function of Z_L, $\overline{Z_L}$ and Z, it can also be checked that upon setting $Z = 0$, the quantities $\partial\widehat{\mathcal{D}}/\partial Z_L$, $\partial\widehat{\mathcal{D}}/\partial\overline{Z_L}$ and $\partial\widehat{\mathcal{D}}/\partial Z$, become $2Z_L\overline{Z_L^2}$, $2Z_L^2\overline{Z_L}$ and $-4(Z_L^3 + \overline{Z_L}^3)$, respectively. Expanding these in terms of ζ_H, $\overline{\zeta_H}$, ζ and $\overline{\zeta}$, it can be observed that each of the resulting three expressions has $(\zeta\overline{\zeta} - 1)^3$ as a factor.

Regarding $\widehat{\mathcal{D}}$ as a polynomial function of ζ, $\overline{\zeta}$ and Z instead, it follows that $\partial\widehat{\mathcal{D}}/\partial Z$ has $(\zeta\overline{\zeta} - 1)^3$ as a factor, when Z is set to zero. Therefore $\partial\mathcal{P}/\partial Z$ has $\zeta\overline{\zeta} - 1$ as a factor, upon setting $Z = 0$. Therefore, $\zeta\overline{\zeta} - 1$ is a factor of $\mathcal{A}_1$. Now, As $Z \to \infty$, ζ_L approached $\zeta^2 - 2\overline{\zeta}$. When $\zeta^2 - 2\overline{\zeta}$ is substituted for ζ_L into the formula (3) for $\mathcal{D}$, the result is $(\zeta\overline{\zeta} - 1)^2\,[\zeta^2\overline{\zeta^2} - 4(\zeta^3 + \overline{\zeta^3}) + 18\zeta\overline{\zeta} - 27]$. (There is a nice geometric reason for this that will not be discussed here.) The claim concerning $\mathcal{A}_4$ follows immediately from this.

For $j = 0, 1, 2, 3, 4$, the claim is that the polynomial $\mathcal{A}_j$ has degree $12 - 2j$, as a polynomial in ζ, $\overline{\zeta}$ and z. This is clearly an upper bound since $\mathcal{P}$ has degree twelve. The fact that this is the actual degree of $\mathcal{A}_j$ follows from the fact that $\mathcal{A}_j$ has a non-vanishing $(\zeta\overline{\zeta})^{6-j}$ term. To see this, expand $\widehat{\mathcal{D}}$ but only focus on terms of the form $(\zeta\,\overline{\zeta})^{8-\alpha} Z^\alpha$, which are real terms of maximal degree. Only the $\zeta^2 Z$ and $\zeta^3\,\overline{\zeta}$ parts of the expression for Z_L can contribute here. Similarly for $\overline{Z_L}$. The $Z_L^2\overline{Z_L^2}$ term in the formula for $\widehat{\mathcal{D}}$ contributes $(\zeta^2 Z + \zeta^3\overline{\zeta})^2\,(\overline{\zeta^2}Z + \overline{\zeta^3}\,\zeta)^2 \;=\; \zeta^4\,\overline{\zeta^4}\,(Z + \zeta\overline{\zeta})^4$. This has a non-vanishing term of the form $(\zeta\,\overline{\zeta})^{8-\alpha}\,Z^\alpha$ for $\alpha = 0, 1, 2, 3, 4$. No other terms in the formula for $\widehat{\mathcal{D}}$ contribute terms of this sort; they only contribute terms of lesser degrees. This establishes the claims concerning the degrees of $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ and $\mathcal{A}_4$.

□

Lemma 19. $\zeta\overline{\zeta} - 1$, $\zeta^2 - \zeta_H\,\zeta - \overline{\zeta} + \overline{\zeta_H}$ *and* $\overline{\zeta^2} - \overline{\zeta_H}\,\overline{\zeta} - \zeta + \zeta_H$ *are not factors of* $\mathcal{A}_1'$.

Proof. As in the proof of Lemma 18, observe that $(\partial^2 Z_L/\partial\overline{\zeta}^2, \partial^2\overline{Z_L}/\partial\overline{\zeta}^2)$ and $(\partial^3 Z_L/\partial\overline{\zeta}^3, \partial^3\overline{Z_L}/\partial\overline{\zeta}^3)$ become respectively $(-2\zeta, 4 + 2Z - 2\overline{\zeta_H}\zeta)$ and $(0, 6\zeta)$, upon setting $\overline{\zeta}$ to $1/\zeta$. Also, $(\partial^2\widehat{\mathcal{D}}/\partial Z_L^2, \partial^2\widehat{\mathcal{D}}/\partial Z_L\partial\overline{Z_L}, \partial^2\widehat{\mathcal{D}}/\partial\overline{Z_L}^2)$ becomes $\big(2(1 + 2\zeta - 2\zeta^2)\;(1 - 2\zeta + 6\zeta^2 + 4\zeta^3 + 4\zeta^4)Z^2/\zeta^4$, $-2(4 - 19\zeta^3 + 4\zeta^6)Z^2/\zeta^3$, $-2(2 - 2\zeta - \zeta^2)(4 + 4\zeta + 6\zeta^2 - 2\zeta^3 + \zeta^4)Z^2 \,/\, \zeta^2\big)$. Also, $(\partial^3\widehat{\mathcal{D}}/\partial Z_L^3$, $\partial^3\widehat{\mathcal{D}}/\partial Z_L^2\partial\overline{Z_L}$, $\partial^3\widehat{\mathcal{D}}/\partial Z_L\partial\overline{Z_L}^2, \partial^3\widehat{\mathcal{D}}/\partial\overline{Z_L}^3)$ becomes $(-24Z\;,\; 4(1 - 2\zeta^3)Z/\zeta^2, 4(\zeta^3 - 2)Z/\zeta\;,\; -24Z)$. Applying the rules of partial differential calculus now, it is discovered when $\overline{\zeta}$ is set to $1/\zeta$, that $\partial^2\widehat{\mathcal{D}}/\partial\overline{\zeta}^2$ becomes a polynomial in ζ and Z, times Z^2/ζ^4, and that $\partial^3\widehat{\mathcal{D}}/\partial\overline{\zeta}^3$ becomes a polynomial in ζ and Z, times Z/ζ^3. The coefficient of Z^0 in this latter polynomial is *not* identically zero. It follows that $(\zeta\overline{\zeta} - 1)^4$ divides the Z^0 term of $\widehat{\mathcal{D}}$, and that $(\zeta\overline{\zeta} - 1)^3$ divides the Z^1 term, but $(\zeta\overline{\zeta} - 1)^4$ does not divide the Z^1 term. So, $\zeta\overline{\zeta} - 1$ does not divide $\mathcal{A}_1'$.

Next, when $\zeta^2 - \zeta_H\zeta + \overline{\zeta_H}$ is substituted for $\overline{\zeta}$ in the expansion of $\widehat{\mathcal{D}}$, the coefficient of Z is $4(\zeta_H -$

$\zeta)^3(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1)^6$. Since this is not identically zero, it follows that $\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H}$ is not a factor of the coefficient of Z in the expansion of $\widehat{\mathcal{D}}$, and hence not a factor of $\mathcal{A}_1'$. Similarly for $\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H$.

□

The next lemma is straightforward (if a bit tedious) to check:

Lemma 20. *When $\overline{\zeta}$ is set to $1/\zeta$, the following formulas result:*

$$\mathcal{A}_0'=(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1)^4\,/\,\zeta^6$$
$$\mathcal{A}_1'=2(\zeta-1)(\zeta^2+\zeta+1)(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1)^3/\zeta^6$$
$$\mathcal{A}_2=(\zeta+1)^2(\zeta^2-\zeta+1)^2(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1)^2/\zeta^6$$
$$\mathcal{A}_3=4(\zeta+1)^2(\zeta^2-\zeta+1)^2(\overline{\zeta_H}\zeta^2-2\zeta+\zeta_H)/\zeta^4$$
$$\mathcal{A}_4=-4(\zeta+1)^2(\zeta^2-\zeta+1)^2/\zeta^3$$
$$\mathcal{A}_2-\mathcal{A}_1'+\mathcal{A}_0'\;=\;(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1)^2\cdot(\zeta_H^2\zeta^2-2\overline{\zeta_H}\zeta_H\zeta+4\zeta+\overline{\zeta_H})^2\,/\,\zeta^4$$

where $\mathcal{A}_0'=\mathcal{A}_0/(\zeta\overline{\zeta}-1)^2=(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})^2(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)^2$ *and* $\mathcal{A}_1'=\mathcal{A}_1/(\zeta\overline{\zeta}-1)$.

Another lemma that can be checked directly, though algebraic manipulation software certainly helps, is the following.

Lemma 21. *When $\zeta_H=0$, the following formulas result:*

$$\mathcal{A}_0=(\zeta^2-\overline{\zeta})^2(\overline{\zeta}^2-\zeta)^2(\zeta\overline{\zeta}-1)^2$$
$$\mathcal{A}_1=(\zeta\overline{\zeta}-1)(2\zeta^4\overline{\zeta}^4-5\zeta^5\overline{\zeta}^2-5\zeta^2\overline{\zeta}^5+\zeta^6+\overline{\zeta}^6+12\zeta^3\overline{\zeta}^3-\zeta^4\overline{\zeta}-\zeta\overline{\zeta}^4-8\zeta^2\overline{\zeta}^2+2\zeta^3+2\overline{\zeta}^3)$$
$$\mathcal{A}_2=6\zeta^4\overline{\zeta}^4-18\zeta^5\overline{\zeta}^2-18\zeta^2\overline{\zeta}^5+\zeta^6+\overline{\zeta}^6+52\zeta^3\overline{\zeta}^3+12\zeta^4\overline{\zeta}+12\zeta\overline{\zeta}^4-78\zeta^2\overline{\zeta}^2+6\zeta^3+6\overline{\zeta}^2+18\zeta\overline{\zeta}$$
$$\mathcal{A}_3=4\zeta^3\overline{\zeta}^3-14\zeta^4\overline{\zeta}-14\zeta^4\overline{\zeta}+52\zeta^2\overline{\zeta}^2+6\zeta^3+6\overline{\zeta}^3-72\zeta\overline{\zeta}$$

We are ready now to consider various ways that the polynomial $\mathcal{P}$ might possibly be factored, but will subsequently show that each of these is actually not possible.

Lemma 22. *Suppose that $\mathcal{P}$, treated as a polynomial in x, y and Z, can be factored nontrivially into real polynomials of x, y and Z, as $\mathcal{P}=p_1p_2$. Then, up to interchanging p_1 and p_2, there are just three possibilities:*

Case 1.

$$p_1\;=\;\mathcal{A}_4Z+\mathcal{B}\;,\;\;p_2\;=\;Z^3+\mathcal{C}Z^2+\mathcal{D}Z+\mathcal{E};$$

Case 2.

$$p_1\;=\;\mathcal{A}_4Z^2+\mathcal{B}Z+\mathcal{C}\;,\;\;p_2\;=\;Z^2+\mathcal{D}Z+\mathcal{E};$$

Case 3.

$$p_1\;=\;\mathcal{A}_4Z^3+\mathcal{B}Z^2+\mathcal{C}Z+\mathcal{D}\;,\;\;p_2\;=\;Z+\mathcal{E};$$

where $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ are real polynomials in x and y (but also be regarded as polynomials in ζ and $\overline{\zeta}$).

Proof. This is an immediate consequence of the fact that $\mathcal{A}_4$ is absolutely irreducible as a polynomial in the variables ζ and $\overline{\zeta}$, and the fact that $\mathcal{P}$ cannot have a non-constant factor that does not involve Z. Such a factor would necessarily divide both $\mathcal{A}_0$ and $\mathcal{A}_4$, which is impossible due to the Lemma 17. This leaves us with only the above possibilities.

□

Lemma 23. *Case 1 in Lemma 22 is not possible.*

Proof. Assume there is such a factorization. Clearly, dim $\mathcal{B}$ + dim $\mathcal{E}$ = dim $\mathcal{A}_0$ = 12. Let dim $\mathcal{B}=6+\varepsilon_1$, dim $\mathcal{C}=2+\varepsilon_2$, dim $\mathcal{D}=4+\varepsilon_3$, and dim $\mathcal{E}=6-\varepsilon_1$. Since, $\mathcal{A}_4\mathcal{C}+\mathcal{B}=\mathcal{A}_3$, we see that if $\varepsilon_1>0$ or $\varepsilon_2>0$, then we must have $\varepsilon_1=\varepsilon_2>0$, in order to cancel terms. Since, $\mathcal{A}_4\mathcal{D}+\mathcal{B}\mathcal{C}=\mathcal{A}_2$, we see that if $\varepsilon_1+\varepsilon_2>0$ or $\varepsilon_3>0$, then we must have $\varepsilon_1+\varepsilon_2=\varepsilon_3>0$. Since, $\mathcal{A}_4\mathcal{E}+\mathcal{B}\mathcal{D}=\mathcal{A}_1$, we see that if $-\varepsilon_1>0$ or $\varepsilon_1+\varepsilon_3>0$, then we must have $-\varepsilon_1=\varepsilon_1+\varepsilon_3>0$.

Suppose for a moment that $\varepsilon_1>0$. Then, $\varepsilon_2>0$, so $\varepsilon_1+\varepsilon_2>0$, so $\varepsilon_3>0$, so $\varepsilon_1+\varepsilon_3>0$, so $-\varepsilon_1>0$, a contradiction. Suppose instead that $\varepsilon_1<0$. Then, $\varepsilon_1+\varepsilon_3>0$, so $\varepsilon_3>0$, so $\varepsilon_1+\varepsilon_2>0$, so $\varepsilon_2>0$, so $\varepsilon_1>0$, a contradiction. Therefore, $\varepsilon_1=0$, and we can also see that $\varepsilon_2\le 0$ and $\varepsilon_3\le 0$.

Recall that $\mathcal{A}_0=(\zeta\overline{\zeta}-1)^2(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})^2(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)^2$, and that the factors here are absolutely irreducible. Observe that $\mathcal{B}$ and $\mathcal{E}$ are essentially real polynomials of x and y (since p_1 and p_2 are presumed to be essentially real polynomials of x, y and Z), of degree six. Thus, we are forced to conclude that $\mathcal{B}$ and $\mathcal{E}$ are both constant multiples of the real polynomial $(\zeta\overline{\zeta}-1)(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)$. (It is real as a polynomial in x, y and Z.) In fact, we may assume that both $\mathcal{B}$ and $\mathcal{E}$ equal this polynomial. Since, $\mathcal{A}_4\mathcal{E}+\mathcal{B}\mathcal{D}=\mathcal{A}_1$, $(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)$ must be a divisor of $\mathcal{A}_1$. But this is not true, by Lemma 19.

□

Lemma 24. *Case 2 in Lemma 22 is not possible.*

Proof. Assume there is such a factorization. Clearly, dim $\mathcal{C}$ + dim $\mathcal{E}$ = dim $\mathcal{A}_0$ = 12. Let dim $\mathcal{B} = 6+\varepsilon_1$, dim $\mathcal{C} = 8+\varepsilon_2$, dim $\mathcal{D} = 2+\varepsilon_3$, and dim $\mathcal{E} = 4-\varepsilon_2$. Since, $\mathcal{A}_4\mathcal{D}+\mathcal{B} = \mathcal{A}_3$, we see that if $\varepsilon_1 > 0$ or $\varepsilon_3 > 0$, then we must have $\varepsilon_1 = \varepsilon_3 > 0$. Since, $\mathcal{A}_4\mathcal{E}+\mathcal{B}\mathcal{D}+\mathcal{C} = \mathcal{A}_2$, we see that if any of ε_2, $-\varepsilon_2$ and $\varepsilon_1+\varepsilon_3$ are positive, then it must equal another one of these. Since, $\mathcal{B}\mathcal{E}+\mathcal{C}\mathcal{D} = \mathcal{A}_1$, we see that if $\varepsilon_1-\varepsilon_2 > 0$ or $\varepsilon_2+\varepsilon_3 > 0$, then we must have $\varepsilon_1-\varepsilon_2 = \varepsilon_2+\varepsilon_3 > 0$.

Suppose for a moment that $\varepsilon_2 > 0$. Then, $\varepsilon_1+\varepsilon_3 > 0$, so either $\varepsilon_1 > 0$ or $\varepsilon_3 > 0$, so $\varepsilon_1 = \varepsilon_3 > 0$, so $\varepsilon_2+\varepsilon_3 > 0$, so $\varepsilon_1-\varepsilon_2 = \varepsilon_2+\varepsilon_3 > 0$, so $-\varepsilon_2 = \varepsilon_2$, a contradiction. Next suppose instead that $\varepsilon_2 < 0$. Then, $\varepsilon_1+\varepsilon_3 > 0$, so either $\varepsilon_1 > 0$ or $\varepsilon_3 > 0$, so $\varepsilon_1 = \varepsilon_3 > 0$, so $\varepsilon_1-\varepsilon_2 > 0$, so $\varepsilon_1-\varepsilon_2 = \varepsilon_2+\varepsilon_3 > 0$, so $-\varepsilon_2 = \varepsilon_2$, a contradiction. Therefore, $\varepsilon_2 = 0$, and we can also see that $\varepsilon_1 \le 0$ and $\varepsilon_3 \le 0$.

$\mathcal{C}$ and $\mathcal{E}$ are essentially real polynomials of x and y of degrees eight and four, respectively. We may now easily reduce to one of two cases:

Case 2A.

$$\mathcal{C} = (\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})^2\,(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)^2$$
and $\mathcal{E} = (\zeta\overline{\zeta}-1)^2$;

Case 2B.

$$\mathcal{C} = (\zeta\overline{\zeta}-1)^2\cdot (\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})\,(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H) \text{ and}$$
$$\mathcal{E} = (\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})\,(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H).$$

Case 2B is quickly eliminated because it implies that $(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})\,(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)$ divides $\mathcal{B}\mathcal{E}+\mathcal{C}\mathcal{D} = \mathcal{A}_1$, which Lemma 19 indicates is not true.

Assume Case 2A now. Since $\zeta\overline{\zeta}-1$ divides $\mathcal{B}\mathcal{E}$ and $\mathcal{A}_1$, but not $\mathcal{C}$, it must divide $\mathcal{D}$. Since dim $\mathcal{D} \le 2$, we must have $\mathcal{D} = \lambda(\zeta\overline{\zeta}-1)$ for a constant λ. From $\mathcal{B}\mathcal{E}+\mathcal{C}\mathcal{D} = \mathcal{A}_1$ and $\mathcal{C} = \mathcal{A}'_0$, we get $(\zeta\overline{\zeta}-1)\mathcal{B}+\lambda\mathcal{A}'_0 = \mathcal{A}'_1$. Upon setting $\overline{\zeta} = 1/\zeta$, we must have $\lambda\mathcal{A}'_0 = \mathcal{A}'_1$ (identically). But from Lemma 20, we see this means that we need $\lambda(\zeta^3-\zeta_H\zeta^2+\overline{\zeta_H}\zeta-1) = 2(\zeta-1)(\zeta^2+\zeta+1)$ (identically). This is only possible when $\zeta_H = 0$, in which case, we must set $\lambda = 2$. Assuming this, it can then be seen that $\mathcal{B} = (\mathcal{A}_1-\mathcal{C}\mathcal{D})/\mathcal{E} = 2(\zeta^3\overline{\zeta}^3-3\zeta^4\overline{\zeta}-3\zeta\overline{\zeta}^4+9\zeta^2\overline{\zeta}^2-2\zeta^3-2\overline{\zeta}^3)$. It follows that $\mathcal{A}_4\mathcal{D}+\mathcal{B}-\mathcal{A}_3$ = $2(3-\zeta-\overline{\zeta})(9+3\zeta+3\overline{\zeta}+\zeta^2+\overline{\zeta}^2-\zeta\overline{\zeta})$. Since this is not identically zero, Case 2A is impossible, even when $\zeta_H = 0$.

□

Lemma 25. *Case 3 in Lemma 22 is not possible.*

Proof. Assume there is such a factorization. Clearly, dim $\mathcal{D}$ + dim $\mathcal{E}$ = dim $\mathcal{A}_0$ = 12. Let dim $\mathcal{B} = 6+\varepsilon_1$, dim $\mathcal{C} = 8+\varepsilon_2$, dim $\mathcal{D} = 10+\varepsilon_3$, and dim $\mathcal{E} = 2-\varepsilon_3$. Since, $\mathcal{A}_4\mathcal{E}+\mathcal{B} = \mathcal{A}_3$, we see that if $-\varepsilon_3 > 0$ or $\varepsilon_1 > 0$, then we must have $-\varepsilon_3 = \varepsilon_1 > 0$. Since, $\mathcal{B}\mathcal{E}+\mathcal{C} = \mathcal{A}_2$, we see that if $\varepsilon_1-\varepsilon_3 > 0$ or $\varepsilon_2 > 0$, then we must have $\varepsilon_1-\varepsilon_3 = \varepsilon_2 > 0$. Since, $\mathcal{C}\mathcal{E}+\mathcal{D} = \mathcal{A}_1$, we see that if $\varepsilon_2-\varepsilon_3 > 0$ or $\varepsilon_3 > 0$, then we must have $\varepsilon_2-\varepsilon_3 = \varepsilon_3 > 0$.

Suppose for a moment that $\varepsilon_3 > 0$. Then, $\varepsilon_2 = 2\varepsilon_3 > 0$, so $\varepsilon_1-\varepsilon_3 = \varepsilon_2$, so $\varepsilon_1 = 3\varepsilon_3 > 0$, so $-\varepsilon_3 = \varepsilon_3$, a contradiction. Suppose instead that $\varepsilon_3 < 0$. Then $\varepsilon_1 = -\varepsilon_3 > 0$, so $\varepsilon_1-\varepsilon_3 > 0$, so $\varepsilon_2 = \varepsilon_1-\varepsilon_3 > 0$, so $\varepsilon_2-\varepsilon_3 > 0$, so $\varepsilon_3 = \varepsilon_2-\varepsilon_3 > 0$, a contradiction. Therefore, $\varepsilon_3 = 0$, and we can also see that $\varepsilon_1 \le 0$ and $\varepsilon_2 \le 0$.

Similar to previous reasoning, in the present case, we must have $\mathcal{D} = (\zeta\overline{\zeta}-1)\,(\zeta^2-\zeta_H\zeta-\overline{\zeta}+\overline{\zeta_H})^2\,(\overline{\zeta^2}-\overline{\zeta_H}\,\overline{\zeta}-\zeta+\zeta_H)^2$, and $\mathcal{E} = \zeta\overline{\zeta}-1$. Now, $\mathcal{C} = (\mathcal{A}_1-\mathcal{D})/\mathcal{E} = \mathcal{A}'_1-\mathcal{A}'_0$. Then, $\mathcal{B} = (\mathcal{A}_2-\mathcal{C})/\mathcal{E} = (\mathcal{A}_2-\mathcal{A}'_1+\mathcal{A}'_0)/\mathcal{E}$. However, by considering the $\overline{\zeta} = 1/\zeta$ situation again, it is clear that $\mathcal{E}$ is not a divisor of $\mathcal{A}_2-\mathcal{A}'_1+\mathcal{A}'_0$. Therefore, Case 3 is not possible.

□