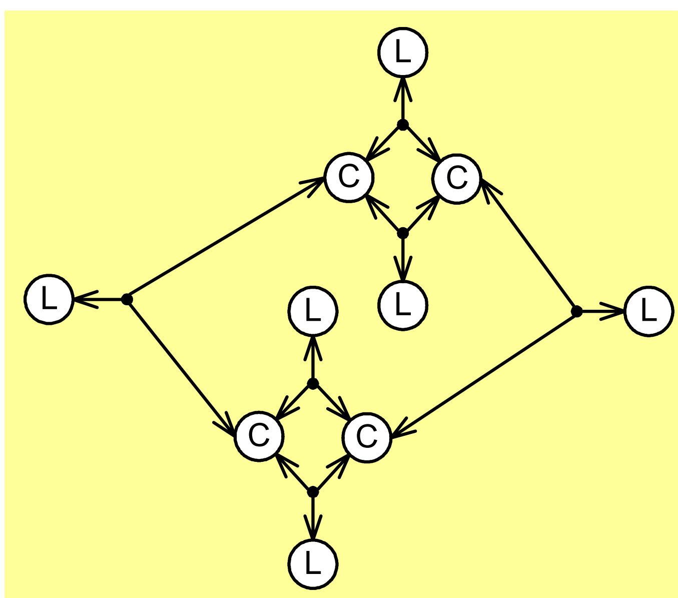
USING TENSOR DIAGRAMS TO REPRESENT AND SOLVE GEOMETRIC PROBLEMS
2002 Edition

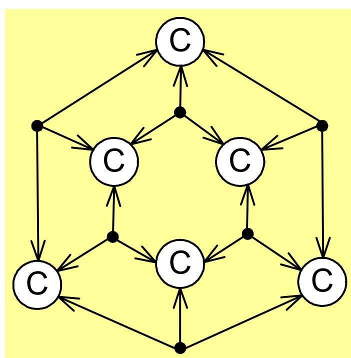
James F. Blinn
Microsoft Research
blinn@microsoft.com
http://www.research.microsoft.com/~blinn



The condition that a line is tangent to a cubic curve

Table of Contents

PART 0
A First Pass



An invariant of a cubic curve

Chapter 0-00

Introduction

Several years ago, Jim Kajiya loaned me a copy of a book called *Diagram Techniques in Group Theory* [G. E. Stedman, Cambridge England, Cambridge University Press, 1990]. This book described a graphical representation called a **Tensor Diagram** for the algebra used to solve various problems in mathematical physics. I was only able to understand the first chapter of the book, but even that was enough to excite me tremendously about adapting the technique to the algebra of homogeneous geometry that we are familiar with in computer graphics. Tensor diagrams actually show up in various guises in a lot of different fields. They are simply a way to represent generalizations of matrix and vector products, so anything that uses matrices and vectors can possibly benefit from the tensor diagram notion. Recently I have been playing more and more with these diagrammatic ways of doing algebra and have come up with a lot of interesting results. This Siggraph course and the accompanying course notes will present what I have figured out so far. Most of the geometric/algebraic problems I will apply this to are pretty well known, but some aren't and may even be new. Sometimes an application of TD will require a new way of looking at an old problem. To set this up, there will sometimes be long stretches of discussion that do not immediately mention TD's. The final punch line should make this worthwhile. I have been excited by how easily and prettily tensor diagrams can express the algebraic form of various geometric problems. And I hope to share that excitement with you.

How Mathematics Works

Here's how mathematics works. First you find a problem that you want to solve. After some study you find that solving that problem will require some preliminary tools and derivations. Working on those you find that they too require some precursors. You keep working backwards from the problem until you get to something you know and can prove. That is, you work backwards from the problem to the solution.

Now how do you present your results? You start with the simple constructions and work forwards. That is you start with something obvious and well known. Then you take off in a slightly new direction. Each derivation and new definition is not a great conceptual leap so your audience should be able to follow you easily. However, for someone seeing this for the first time many of the new definitions and derivations seem unmotivated.

Guess what? I'm going to do that too, but not completely. I will start out here by showing the general problem that I am interested in, to see where we ultimately are going. Then I will go back and start from scratch. I will, in fact, do this twice. The first time is a quick run through of the basic ideas. The second time I will go back and fill in some details. Along the way we will find some new ways of looking at various intermediate problems.

For example, in playing around with cubic polynomials I found out that I didn't understand quadratic polynomials as well as I thought. So I had to go back and perform the same operations on quadratics, which seemed trivial at first, but taught me something when I went back to cubics.

The Universe

There are two basic geometrical universes that I will discuss here: Euclidean geometry and Projective geometry. The primary difference between them is the allowable transformations that can be performed on geometric figures that will "make no difference" to their geometrical relationships. For Euclidean geometry the allowable transformations will be pure rotations and translations. For Projective geometry the allowable transformations will be a general perspective projection. In each case we will be interested in properties of a shape that remain the same (are invariant) when subjected to the transformation. These properties include:

Projective	Intersections Tangency Cross Ratios
Euclidean	All the above plus: Distance Angle Parallelism

Most of the algebraic tools we will develop will be useful in both situations.

Another axis of concern is dimensionality. I will discuss only 1, 2 and 3-dimensional versions of these three geometries. Higher dimensions are interesting too, but I won't go into them explicitly here. I will refer to the various situations using the following naming convention. The basic representation of a point in each of the geometries is as a vector. For homogeneous geometry we use an extra homogeneous coordinate. So we will have:

Euclidean	Projective
1D: $[X]$	1DH: $[x \quad w]$
2D: $[X \quad Y]$	2DH: $[x \quad y \quad w]$
3D: $[X \quad Y \quad Z]$	3DH: $[x \quad y \quad z \quad w]$

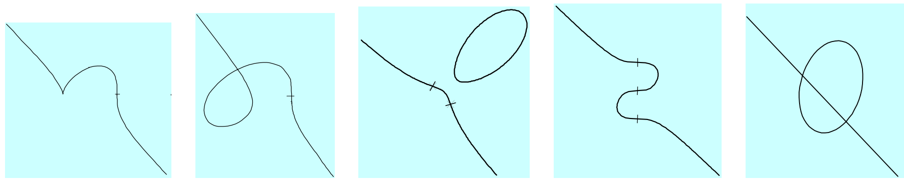
Note that, purely algebraically, a 2D Euclidean point will have much in common with a 1DH Projective point. Most of the manipulation machinery for 2D Euclidean points will be the same as that for 1DH Projective points. The same goes for (3D Euclidean and 2DH Projective) and (4D Euclidean and 3DH Projective). Where appropriate, then, I will label sections in the notes deal with one particular dimensionality as either **1DH(2D)**, or **2DH(3D)**, or **3DH(4D).**

The Problem

Now for the basic motivating problem: I want to understand cubic curves. In 2DH space, these are curves consisting of all points that satisfy the homogeneous equation

$$\begin{aligned} Ax^3 + 3Bx^2y + 3Cxy^2 + Dy^3 \\ +3Ex^2w + 6Fxyw + 3Gy^2w \\ +3Hxw^2 + 3Jyw^2 \\ +Kw^3 = 0 \end{aligned}$$

Given various choices for the coefficients *A* through *K* this curve can have various shapes. A selection of possible shapes appears below;

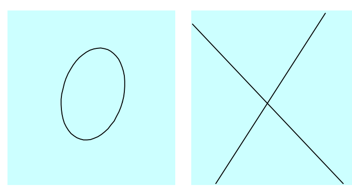


So, what sort of geometric questions do we want to ask? First off, given A…K which of the general shape categories above (as well as a few others) does the curve have? Next, note that a cubic curve can have up to three inflection points (marked by tick marks on the curves above). Note also that, if there are three, they are collinear. So, given A…K, how many inflection points are there, where are they, what is the equation of the line through them? The complexity of these questions is hinted at by examining the final diagram. It is different from the other three in that its algebraic equation can be factored into the product of a linear form and a quadratic form. Given A…K how can we tell if the cubic equation if factorable?

We can get a hint by going down a degree and ask the same question about quadratic curves. These satisfy the equation

$$\begin{aligned} &Ax^2 + 2Bxy + Cy^2 \\ &+2Dxw + 2Eyw \\ &+Fw^2 = 0 \end{aligned}$$

This expression can generate all conic sections or, if the equation is factorable, it can generate a pair of intersecting lines:



We will later show that the algebraic test for whether the quadratic is factorable involves evaluating the, so-called, discriminant of the quadratic function. If this discriminant is zero, the quadratic is factorable. The discriminant, moreover, is the determinant of a matrix

$$\det\begin{bmatrix} A & B & D \\ B & C & E \\ D & E & F \end{bmatrix} = ACF + 2BED - D^2C - E^2A - B^2F = 0$$

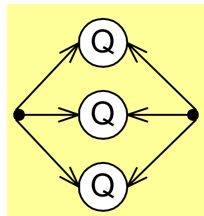
An equivalent discriminant for the cubic curve case is considerably more complicated. Paluszny and Patterson {M. Paluszny and R Patterson, *A Family of Tangent Continuous Cubic Algebraic Splines*, ACM Transactions on Graphics, Vol. 12, No. 3, July 1993, page 212} describe the cubic discriminant as a polynomial that is degree 12 in the coefficients *A…F* and that has over 10,000 terms. Manipulating this thing explicitly is… inconvenient. Actually, it's not that complicated. George Salmon, {G. Salmon, A Treatise on the Higher Plane Curves, Published by G. E. Stechert & Co., New York, 1934, A Photographic reprint of the Third Edition of 1879, pages 191, 192 and 199} showed that the discriminant is a function of two simpler quantities

$$\mathbf{D} = T^2 + 64S^3$$

Where S is degree 4 in *A…K* and has 25 terms, and T is degree 6 in *A…K* and has 103 terms. This is better, but still a bit complicated.

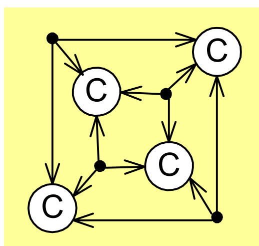
A Solution

The main interest in studying tensor diagrams is because they can enable us to write and manipulate such geometric tests more easily. As a foretaste of thing to come, I want to show you the Tensor Diagram forms of the above formulas. I'll explain how they translate into algebra later. The Tensor Diagram for the discriminant of a quadratic looks like:

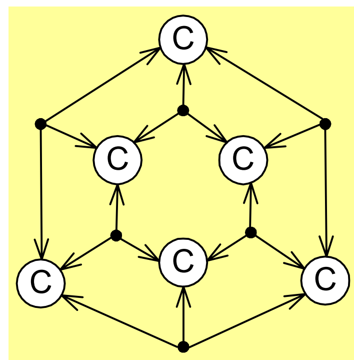


The diagrams for the T and S terms in the discriminant of the cubic look like:

$S = -\frac{1}{24}$



$T = -\frac{1}{6}$



These are not only pretty, but they are a whole lot simpler than a 10000-term polynomial.

The Matrix of knowledge

In wending ourselves through the various derivations here I will be constantly going back and forth between algebra (what is the geometric interpretation of this equation) and geometry (what is the equation that indicates this?)

In addition it is useful to visualize the various topics as a matrix with dimensionality along one axis and polynomial degree along another. We can traverse this matrix in various orders. Pure row order or pure column order is usually not the best route though. Sometimes a derivation within one of the cells will naturally lead us to generalizations across degree (up and down a column) and sometimes across dimensions (left and right across a row.)

	1DH	2DH	3DH
linear	Linear polynomial	Line in a plane	Line in space
quadratic	Quadratic polynomial	Quadratic curve	Quadratic surface
Pairs of quadratics	Resultants	Resultant	Resultant
Cubics	Polynomial	Curve	Surface
Pairs of cubics	Resultants	Resultants	Resultants

Quartic	Polynomial	Curve	Surface

What does it all mean

When the only tool you have is a hammer, you tend to see each problem as a nail. It is tempting, when investigating some new tool or technique to try to apply it to any and all situations that come along. I will try to avoid that here and, in the summary, try to probe the limitations of our tool and try to identify situations in which it is not useful. Not all problems may be well served by tensor diagrams. I will also attempt to address the limitations of this technique

The algebra represented by a tensor diagram can be written in conventional notation for many of the simpler applications, and it might seem like diagrams are not that remarkable an advance. But in more complex applications, when the diagram contains loops, the conversion to standard linear vector expression is not easy. That's where tensor diagrams shine.

About these notes

This is a "living document". That's a euphemism for something that is not yet finished. Some sections are a bit rough and incomplete both in exposition and in results. I am still playing around with these tools and will probably have some new results by the time the course itself is presented. I have attempted to pick all the "low hanging fruit" of this representation, but there are still some problems that I haven't nailed down that I have a nagging feeling have pretty simple solutions. I hope to have figured out some of the answers in the time between sending these notes to Siggraph and the actual presentation at the conference. Maybe you, the reader, will find those that I've missed. For updates to the notes, both from myself and any feedback from others, be sure to consult my web site at:

http://www.research.microsoft.com/~blinn

Some of these notes contain text cannibalized from several of my *Jim Blinn's Corner* articles in the *IEEE Computer Graphics and Applications* journal. I will identify these in the chapter introductions. Since each of these articles is meant to be relatively self contained, the beginning of the articles typically contains a review of relevant previous results. I have left several of these reviews intact so some chapters in these notes may seem a bit repetitious. My intention in doing this is that the alternative explanations may help readers new to the subject by showing the material in more than one way. The notes also have a lot of intentional redundancy. I often solve the same problem in several ways and the notes will present all of them. The notes are thus an archive of alternative ideas about the problems. This might help gain more insight into the machinery. Finally, some of the notes come from my original worksheets used in investigating the problems. I use a mathematical typesetting program and lots of cut-and-paste and drag-and-drop to do algebra. The resulting derivations are often painfully explicit and verbose. I've often left these in intact but you can skip quickly over them.

The illustrations of the actual tensor diagrams themselves are drawn on a yellow background. This looks nice, but also serves a pedagogical purpose. Each chunk within a yellow rectangle represents a term in an expression. Separate yellow boxed diagram chunks may be added together, perhaps will scalar multiples. Unfortunately due to time constraints I have not been able to put the yellow boxes in all the zillions of illustrations here. So when you look at a diagram without the yellow backgrounds, use it as an exercise to figure out where they should have gone.

Also, I will format the notes in a manner that I have always wanted in books I have read. I typically like to use diagrams as parts of speech. These will be inserted in the text where they are mentioned instead of referring to "figure xyz" that might be several pages away. I will also take advantage of one of the benefits of electronic publication in that there is no page limit, so I will *repeat* equations and figures wherever they might be re-used instead of, again, referring to equation numbers or figure numbers that may be an even larger number of pages away.

In many places I refer to a "symbolic algebra" program that explicitly evaluates the polynomial represented by a tensor diagram. This is a useful tool to make sure algebraic signs and multiplicative constants don't get dropped. This program is described more fully in chapter 21 of my new book *Jim Blinn's Corner: Notation, Notation, Notation,* Morgan Kauffman, 2002 (Hint, hint).

Chapter 0-01

Homogeneous Geometry

This chapter is cannibalized from
Uppers and Downers, Part 1
which is chapter 9 of *Jim Blinn's Corner: Dirty Pixels*

Mathematical research is largely a process of successive generalization. Generalizing the square root operation to negative values leads to complex number theory. Generalizing Euclidean transformations to perspective transformations leads to projective geometry and homogeneous coordinates. Often stuff we use in day-to-day calculations is just a special case of some more general theory that we aren't even aware of. In fact, sometimes this can lead to confusion—an example of "too little knowledge is a dangerous thing". This is the case for the standard vector/matrix formulation used to solve geometric problems in homogeneous coordinates. But once you see the generalization what was a problem before now becomes a thing of beauty. So, this time I'm going to discuss what's *really* going on with homogeneous coordinates.

This chapter starts with a review of the standard vector/matrix notation for homogeneous geometry and then does an expose of its deficiencies. In the next chapter I'll introduce a better notation (Einstein Index Notation) and show how it solves some complex problems easily. Finally, the rest of the document describes Tensor Diagrams, a representation of EIN based on graph theory.

First, some ground rules. We're dealing with pure projective geometry and homogeneous coordinates here. That means that such concepts as distance, angle measurement and parallelism are meaningless. We are only interested in the things that remain constant after a perspective transformation: intersections, tangency etc. Also, any nonzero scalar multiple of any of our entities still represents the same entity. Sometimes I will discard troublesome scales without much comment.

Naming Conventions

Since this is all about notation, let's review some typical conventions for naming things to give you a taste of the things we are trying to unify.

We'll start with the distinction between scalars, vectors and matrices. For the time being I will write scalars, vector components and matrix elements using italic letters; a, p_i, T_{ij}. I'll use roman letters for vectors: P. Different vectors of the same type will sometimes be distinguished with subscripts: P_1, P_2. I'll use bold face letters for matrices: **T**. We see right away that there is some potential for confusion with the meaning of subscripts; they can either identify vector elements or they can name different vectors. This is one of the things we want to avoid when we get to our improved technique.

Often we will need to represent vectors or matrices that are simple modifications of other vectors or matrices. I'll represent these by using the letter of the original with some diacritical mark appended. (The definitions of some of these operations are given below.) A transformed version of a vector P is P'. The adjoint of matrix **M** is written $\mathbf{M}^*$. The transpose of matrix **M** is $\mathbf{M}^{\mathrm{T}}$. The dual of matrix **M** is $\tilde{\mathbf{M}}$.

3D(2DH) Review

First, let's go through the standard litany of homogeneous representation in two dimensions. We represent two-dimensional homogeneous entities (2DH) in three dimensional space.

Points and Lines

Points are three element row vectors

$$\mathrm{P} = \begin{bmatrix} x & y & w \end{bmatrix}$$

Lines are three element column vectors

$$\mathrm{L} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

The matrix (or dot) product of the row and column vector is a scalar:

$$\mathrm{P} \cdot \mathrm{L} = ax + by + cw$$

If the value is zero, the point lies on the line.

Transformations

We transform points by multiplying on the right by a 3×3 matrix

$$\mathrm{P}' = \mathrm{P}\,\mathbf{T}$$

Transform lines by multiplying on the left by the adjoint of the matrix we used to transform points

$$\mathrm{L}' = \mathbf{T}^* \,\mathrm{L}$$

The adjoint is the transpose of the matrix of cofactors of the original matrix. An example element of the adjoint of **T** is

$$\left(\mathbf{T}^*\right)_{23} = T_{21}T_{13} - T_{11}T_{23}$$

The adjoint is the same as the inverse except for a scale factor. But we don't care about scale factors, so the adjoint and inverse are all the same to us.

Intersections

To find the line containing two given points, take their 3D cross product and write the result as a column vector.

$$\mathrm{P}_1 \times \mathrm{P}_2 = \begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix} \times \begin{bmatrix} x_2 & y_2 & z_2 \end{bmatrix} = \begin{bmatrix} y_1 w_2 - y_2 w_1 \\ w_1 x_2 - w_2 x_1 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}$$

To find the intersection of two lines, take their 3D cross product and write the result as a row vector.

$$\mathrm{L}_1 \times \mathrm{L}_2 = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \times \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} b_1 c_2 - b_2 c_1 & c_1 a_2 - c_2 a_1 & a_1 b_2 - a_2 b_1 \end{bmatrix}$$

Quadrics

A second order algebraic equation such as

$$Ax^2 + 2Bxy + 2Cxw + Dy^2 + 2Eyw + Fw^2 = 0$$

represents an arbitrary conic section (also called a quadric curve). This equation can be written in matrix form as

$$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} A & B & C \\ B & D & E \\ C & D & F \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} x & y & w \end{bmatrix} \mathbf{Q} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = 0$$

that is, a point P is on the conic if

$$\mathrm{P}\mathbf{Q}\mathrm{P}^{\mathrm{T}} = 0$$

Note that the matrix **Q** is symmetric.

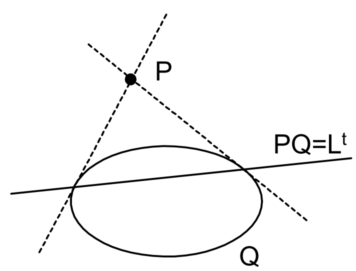
There is a variant of **Q** that is good for testing line tangency. We'll call this the "dual" of **Q** and write it as $\tilde{\mathbf{Q}}$. It so happens that the dual of **Q** is equal to the adjoint of **Q**. A line L is tangent to the conic if

$$\mathrm{L}^{\mathrm{T}}\tilde{\mathbf{Q}}\mathrm{L} = 0$$

Note that we had to put in a few transposes in the above two equations to make the row and column conformability rules of matrix multiplication work out.

Starting from an arbitrary point, we can draw two lines tangent to a given quadric. Connecting these two points of tangency gives a line called the "polar line". The vector for this line is just the product of the quadric matrix and the point vector.

$$\mathrm{P}\mathbf{Q} = \mathrm{L}^{\mathrm{T}}$$



What you get out, according to the rules of matrix multiplication, is a row vector. You have to transpose it to get the line into the column vector notation.

To transform a conic section, we transform the **Q** matrix by

$$\mathbf{Q}' = \mathbf{T}^{*}\mathbf{Q}\left(\mathbf{T}^{*}\right)^{\mathrm{T}}$$

that is, by pre- and post-multiplying by the adjoint of the point transformation matrix **T**. To transform the dual of **Q** we must multiply by the point transformation matrix

$$\left(\mathbf{Q}^{*}\right)' = \mathbf{T}^{\mathrm{T}}\left(\mathbf{Q}^{*}\right)\mathbf{T}$$

Oops

Now wait a minute. There's something fishy here. I've been preaching all along that points are row vectors and lines are column vectors. But the above equations have points showing up sometimes as columns and lines showing up sometimes as rows. It gets even worse when we go to three dimensions.

4D(3DH) Review

Now, let's generalize this to 3D homogeneous.

The Obvious Part

All the 2DH stuff generalizes pretty easily to three dimensions: just make the vectors four elements long and making the matrices 4×4. You then have to reinterpret the geometric meaning of things a bit. What was a line in 2DH (column vector) is now a plane in 3DH. I will typically use the letter E for 3DH planes.

$$\mathrm{E} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

What was a conic section in 2DH (symmetric matrix) is a family of 3DH surfaces consisting of ellipsoids, cones, cylinders, saddle points, and the like. Tangency of lines with a conic section becomes tangency of planes against the above surfaces.

The only slightly tricky part in going to 3DH involves the four dimensional generalization of the cross product. Geometrically this is the problem of finding a plane passing through three points. Now we can write the three dimensional cross product of P_1 and P_2 as

$$\mathrm{P}_1 \times \mathrm{P}_2 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

where

$$a = \det\begin{bmatrix} y_1 & w_1 \\ y_2 & w_2 \end{bmatrix}, \quad b = \det\begin{bmatrix} w_1 & x_1 \\ w_2 & x_2 \end{bmatrix}, \quad c = \det\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

By analogy the four dimensional cross product of P_1, P_2, and P_3, is a column vector with an a component of

$$a = \det\begin{bmatrix} y_1 & z_1 & w_1 \\ y_2 & z_2 & w_2 \\ y_3 & z_3 & w_3 \end{bmatrix}, \; b = -\det\begin{bmatrix} x_1 & z_1 & w_1 \\ x_2 & z_2 & w_2 \\ x_3 & z_3 & w_3 \end{bmatrix}$$

$$c = \det\begin{bmatrix} y_1 & y_1 & w_1 \\ y_2 & y_2 & w_2 \\ y_3 & y_3 & w_3 \end{bmatrix}, \; d = -\det\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix}$$

A symmetric formulation finds the coordinates of a point common to three planes.

Lines

Again, less well known is a homogeneous formulation for lines in three dimensions. I discuss this in more detail in the next chapter. Here are the highlights.

We represent a 3DH line as an anti-symmetric 4×4 matrix. Giving the six unique elements of the matrix the names p, q, r, s, t, u we can write the matrix as

$$\mathbf{L} = \begin{bmatrix} 0 & p & -q & r \\ -p & 0 & s & -t \\ q & -s & 0 & u \\ -r & t & -u & 0 \end{bmatrix}$$

Given two points P_1 and P_2, you can calculate the values of p, q, r, s, t, u for the line connecting them by

$$p = \det\begin{bmatrix} z_1 & w_1 \\ z_2 & w_2 \end{bmatrix},\ q = \det\begin{bmatrix} y_1 & w_1 \\ y_2 & w_2 \end{bmatrix},\ r = \det\begin{bmatrix} y_1 & z_1 \\ y_2 & z_2 \end{bmatrix}$$
$$s = \det\begin{bmatrix} x_1 & w_1 \\ x_2 & w_2 \end{bmatrix},\ t = \det\begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \end{bmatrix},\ u = \det\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

Given two planes E_1 and E_2, you can calculate p, q, r, s, t, u for the intersection line by using a similar set of expressions.

It turns out that these calculations will always generate a singular matrix for **L**. Blindly calculating the determinant of the matrix, though, gives the value $(pu - qt + sr)^2$. This means that the components of the line matrix will always satisfy the constraint

$$pu - qt + sr = 0$$

A given point $\begin{bmatrix} x & y & z & w \end{bmatrix}$ lies on the line **L** if their vector/matrix product gives four zeros

$$\begin{bmatrix} x & y & z & w \end{bmatrix}\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

If the point is not on the line, the two of them together determine a plane in space. The four numbers you get out of the product will be the components of the plane.

$$\begin{bmatrix} x & y & z & w \end{bmatrix}\mathbf{L} = \begin{bmatrix} a & b & c & d \end{bmatrix}$$

You just have to transpose the result to get it to be a column vector.

There is also a different form of the line matrix that is good for intersections with planes. We'll call this $\tilde{\mathbf{L}}$. This consists of the same six values as **L** but arranged differently

$$\tilde{\mathbf{L}} = \begin{bmatrix} 0 & -u & -t & -s \\ u & 0 & -r & -q \\ t & r & 0 & -p \\ s & q & p & 0 \end{bmatrix}$$

A given plane includes the line **L** if the vector/matrix product with the $\tilde{\mathbf{L}}$ form gives four zeros. If it doesn't, the four values give the point of intersection of the line and plane.

$$\tilde{\mathbf{L}}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

Again you must rewrite the result as a row vector.

It so happens that $\tilde{\mathbf{L}}$ is *almost* the adjoint of **L**. In fact, if you go through the adjoint calculation machinery you find that

$$\mathbf{L}^* = (pu - qt + sr)\tilde{\mathbf{L}}$$

If it weren't for the embarrassing fact that $pu - qt + sr = 0$ we'd be in business. As it is, we find the dual by just rearranging the elements. I will show a better way below.

You transform a line matrix by

$$\mathbf{L}' = \mathbf{T}^* \mathbf{L}\left(\mathbf{T}^*\right)^{\mathrm{T}}$$

You transform its dual form by

$$\tilde{\mathbf{L}}' = \mathbf{T}^{\mathrm{T}} \tilde{\mathbf{L}} \mathbf{T}$$

YIKES!

So what's going on here? Our whole concept of row and column vectors distinguishing between points and planes is crumbling! And well it should. It turns out that the somewhat pictorial matrix representation of all these geometric entities is simply not powerful enough to express all the things that can happen.

Chapter 0-02

A Homogeneous Formulation for Lines in 3 Space

This chapter is cannibalized from my 1977 Siggraph paper in *Computer Graphics* (Proc. Siggraph), Vol. 11, No. 2, 1977, page 237.

It repeats some of the previous chapter and then goes into more detail about 3D lines

Homogeneous coordinates have long been a standard tool of computer graphics. They afford a convenient representation for various geometric quantities in two and three dimensions. The representation of lines in three dimensions has, however, never been fully described. This paper presents a homogeneous formulation for lines in 3 dimensions as an anti-symmetric 4×4 matrix which transforms as a tensor. This tensor actually exists in both covariant and contravariant forms, both of which are useful in different situations. The derivation of these forms and their use in solving various geometric problems is described.

Introduction

We will assume the reader is somewhat familiar with the homogeneous representation of points and planes in 3-space. A good introduction may be found in [1]. Briefly, a point is represented as a four-component vector, usually written as

$$\begin{bmatrix} x & y & z & w \end{bmatrix}$$

Any non-zero multiple of this row vector represents the same point. The "real" components of the point may be discovered by dividing by the fourth component to obtain the three components:

$$\begin{bmatrix} x/w & y/w & z/w \end{bmatrix}$$

A plane is represented as a four-component column vector:

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Any non-zero multiple of this column vector represents the same plane. The first three components describe a vector normal to the plane and the fourth is related to its distance from the origin.

The dot product of a point (row) vector and a plane (column) vector is proportional to the distance from the point to the plane.

$$\begin{bmatrix} x & y & z & w \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = ax + by + cz + dw \propto D$$

A special case of this is the fact that, if the dot product is zero, the point lies in the plane. If the dot product is non-zero, we can find the actual distance by the following means. Construct a three dimensional vector of unit length perpendicular to the plane.

$$\begin{bmatrix} A & B & C \end{bmatrix} = \begin{bmatrix} a & b & c \end{bmatrix} / \sqrt{a^2 + b^2 + c^2}$$

Scale it up by D and add it to the position of the point. We should then have a point on the plane.

$$\begin{bmatrix} X & Y & Z \end{bmatrix} = \begin{bmatrix} \frac{x}{w} + DA & \frac{y}{w} + DB & \frac{z}{w} + DC \end{bmatrix}$$

Since this point is on the plane, its dot product with the plane vector will be zero. We now have an equation that can be solved for *D*.

$$0=\begin{bmatrix}X & Y & Z & 1\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}x+DAw & y+DBw & z+DCw & w\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}$$

$$D=-\frac{ax+by+cz+dw}{w\sqrt{a^2+b^2+c^2}}$$

The sign of *D* indicates which side of the plane the point was on. It can be ignored if only the distance is required.

An object defined in terms of homogeneous points may be transformed by multiplication of its points by a 4x4 matrix.

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{T}=\begin{bmatrix}x' & y' & z' & w'\end{bmatrix}$$

Any combination of scaling, translation, rotation, and perspective distortion may be represented by the matrix **T**. To determine the coordinates of a plane after it has undergone the same transformation we must pre-multiply by the inverse of **T**.

$$\mathbf{T}^{-1}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}a'\\b'\\c'\\d'\end{bmatrix}$$

Thus the dot product of the transformed point and plane is the same as the dot product of the original point and plane. The relationship of a point lying on a plane is preserved.

Suppose we are given three points and we wish to determine the components of the plane vector through them. That is, we wish to solve for *a*, *b*, *c*, and *d* in the equation:

$$\begin{bmatrix}x_1 & y_1 & z_1 & w_1\\x_2 & y_2 & z_2 & w_2\\x_3 & y_3 & z_3 & w_3\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

Consider a fourth point not in the plane of the other three. Its dot product with the desired plane vector will then be non-zero. We will call it *q*. The resulting equation is then:

$$\begin{bmatrix}x_1 & y_1 & z_1 & w_1\\x_2 & y_2 & z_2 & w_2\\x_3 & y_3 & z_3 & w_3\\x_4 & y_4 & z_4 & w_4\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\mathbf{M}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}0\\0\\0\\q\end{bmatrix}$$

This equation may be solved by multiplying both sides by the adjoint of **M**. The adjoint is the transpose of the matrix formed from the cofactors of the original matrix. The cofactor of an element of a matrix is found by erasing the row and column containing the element and computing the determinant of the remaining smaller matrix, finally flipping the sign if the sum of the row and column indices of the element is odd. Thus the cofactor of the x_4 term of **M** is:

$$\operatorname{cof}(x_4)=-\det\begin{bmatrix}y_1 & z_1 & w_1\\y_2 & z_2 & w_2\\y_3 & z_3 & w_3\end{bmatrix}$$

The product of a matrix and its adjoint is the identity matrix times the determinant of the original matrix. The product of the adjoint with the right side of the equation is just *q* times the right hand column. Our equation is now:

$$\det \mathbf{M} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = q \begin{bmatrix} \operatorname{cof} x_4 \\ \operatorname{cof} y_4 \\ \operatorname{cof} z_4 \\ \operatorname{cof} w_4 \end{bmatrix}$$

Now, since any non-zero multiple of a plane vector represents the same plane, we can neglect the q and $\det \mathbf{M}$ terms above. Finally, note that the cofactors do not contain any components of the arbitrarily chosen fourth point. This whole process can be represented in a shorthand notation:

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \det \begin{bmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} & \hat{\mathbf{l}} \\ x_1 & y_1 & z_1 & w_1 \\ x_2 & y_2 & z_2 & w_2 \\ x_3 & y_3 & z_3 & w_3 \end{bmatrix}$$

where

$$\hat{\mathbf{i}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \hat{\mathbf{j}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}; \quad \hat{\mathbf{k}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; \quad \hat{\mathbf{l}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

This is simply a generalization of the more familiar shorthand notation of the cross product of two vectors in ordinary three-space. The only problem that could arise is if the matrix **M** were singular. This only occurs if the three original points are collinear, whereupon there is no solution. In this case, the four cofactors are all zero. We can take the appearance of four zeros when looking for a plane through three points as an indication that the three points were collinear.

There is a similar mechanism for determining the point of intersection of three planes. That is, the homogeneous coordinates of the point of intersection is:

$$\begin{bmatrix} x & y & z & w \end{bmatrix} = \det \begin{bmatrix} a_1 & a_2 & a_3 & \hat{\mathbf{i}} \\ b_1 & b_2 & b_3 & \hat{\mathbf{j}} \\ c_1 & c_2 & c_3 & \hat{\mathbf{k}} \\ d_1 & d_2 & d_3 & \hat{\mathbf{l}} \end{bmatrix}$$

where, here

$$\hat{\mathbf{i}} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}; \quad \hat{\mathbf{j}} = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}; \quad \hat{\mathbf{k}} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}; \quad \hat{\mathbf{l}} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$$

Again, the appearance of four zeros when solving for the point of intersection indicates that the three planes do not have a single common point. They, in fact, intersect on a line.

The Homogeneous Line Representation

We shall now construct a homogeneous representation of lines in 3D taking the form of a 4x4 matrix we shall call **L**. It will have the property that any scalar multiple of it represents the same line. In addition, if a point vector is multiplied by **L**, a result of four zeros indicates that the point is on the line. The inspiration for this formulation comes from the Grassmann coordinate systems described in [2].

First re-consider the problem of finding the plane trough three points. If the four cofactors in the solution are all zero then the three points were collinear. We can re-interpret this as a condition upon a third point that will make it collinear with two others. Thus for two given points P_1 and P_2, a third point is collinear if:

$$\det\begin{bmatrix}\hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} & \hat{\mathbf{l}}\\ x_1 & y_1 & z_1 & w_1\\ x_2 & y_2 & z_2 & w_2\\ x & y & z & w\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

That is, we must have

$$-y\det\begin{bmatrix}z_1 & w_1\\ z_2 & w_2\end{bmatrix}+z\det\begin{bmatrix}y_1 & w_1\\ y_2 & w_2\end{bmatrix}-w\det\begin{bmatrix}y_1 & z_1\\ y_2 & w_2\end{bmatrix}=0$$

$$x\det\begin{bmatrix}z_1 & w_1\\ z_2 & w_2\end{bmatrix}-z\det\begin{bmatrix}x_1 & w_1\\ x_2 & w_2\end{bmatrix}+w\det\begin{bmatrix}x_1 & z_1\\ x_2 & z_2\end{bmatrix}=0$$

$$-x\det\begin{bmatrix}y_1 & w_1\\ y_2 & w_2\end{bmatrix}+y\det\begin{bmatrix}x_1 & w_1\\ x_2 & w_2\end{bmatrix}-w\det\begin{bmatrix}x_1 & y_1\\ x_2 & w_2\end{bmatrix}=0$$

$$x\det\begin{bmatrix}y_1 & z_1\\ y_2 & z_2\end{bmatrix}-y\det\begin{bmatrix}x_1 & z_1\\ x_2 & z_2\end{bmatrix}+z\det\begin{bmatrix}x_1 & y_1\\ x_2 & y_2\end{bmatrix}=0$$

Now, defining the six new coordinates:

$$p=\det\begin{bmatrix}z_1 & w_1\\ z_2 & w_2\end{bmatrix};\quad q=\det\begin{bmatrix}y_1 & w_1\\ y_2 & w_2\end{bmatrix};\quad r=\det\begin{bmatrix}y_1 & z_1\\ y_2 & z_2\end{bmatrix}$$

$$s=\det\begin{bmatrix}x_1 & w_1\\ x_2 & w_2\end{bmatrix};\quad t=\det\begin{bmatrix}x_1 & z_1\\ x_2 & z_2\end{bmatrix};\quad u=\det\begin{bmatrix}x_1 & y_1\\ x_2 & y_2\end{bmatrix}$$

We can write the four equations in matrix form:

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\begin{bmatrix}0 & p & -q & r\\ -p & 0 & s & -t\\ q & -s & 0 & u\\ -r & t & -u & 0\end{bmatrix}=\begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix}$$

The above anti-symmetric matrix is then our desired line representation **L**. Any non-zero multiple of **L** will still represent the same line. If a point is multiplied by **L** and four zeros result then the point is on the line.

Furthermore, if the point is not on the line, the four coordinates obtained will be the same values obtained if all three points were solved for their common plane. That is, they will be the components of the plane common to the point and the line:

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{L}=\begin{bmatrix}a & b & c & d\end{bmatrix}$$

We need only to transpose the row vector to get the plane vector in its more familiar column format.

There is an analogous process for generating the matrix representing the line formed by intersecting two planes. Given planes 1 and 2, the condition that a third plane contains their line of intersection is:

$$\det\begin{bmatrix}a_1 & a_2 & a & \hat{\mathbf{i}}\\ b_1 & b_2 & b & \hat{\mathbf{j}}\\ c_1 & c_2 & c & \hat{\mathbf{k}}\\ d_1 & d_2 & d & \hat{\mathbf{l}}\end{bmatrix}=\begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix}$$

That is, the four equations must be satisfied:

$$b\det\begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix} - c\det\begin{bmatrix} b_1 & b_2 \\ d_1 & d_2 \end{bmatrix} + d\det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} = 0$$

$$-a\det\begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix} + c\det\begin{bmatrix} a_1 & a_2 \\ d_1 & d_2 \end{bmatrix} - d\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix} = 0$$

$$a\det\begin{bmatrix} b_1 & b_2 \\ d_1 & d_2 \end{bmatrix} - b\det\begin{bmatrix} a_1 & a_2 \\ d_1 & d_2 \end{bmatrix} + d\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = 0$$

$$-a\det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} + b\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix} - c\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = 0$$

These can be written in matrix form:

$$\mathbf{K}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The matrix **K** is an anti-symmetric matrix that is a homogeneous representation of the line of intersection of the two planes. Any non-zero multiple of **K** represents the same line. The product of **K** and any other plane vector will yield four zeros if the line is contained in the plane. If the line is not contained in the plane then the product will yield the homogeneous coordinates of the point of intersection of the line with the plane:

$$\mathbf{K}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

We need only to transpose the point vector to get it in the more familiar row form. There is one somewhat surprising fact, however. For a given line, the matrix **L** formed by two points on the line is not the same as the matrix **K** formed by two planes intersecting on the line. We will now show this.

The Dual Line Representation

We first take note of another interpretation of the matrix **L**. Since each column yields a zero when multiplied by a point on the line we can think of it as a plane containing the line. Similarly each row of **K** can be thought of as a point on the line **K**. Thus **L** consists of four planes containing the line represented by **L** and **K** consists of four points on the line represented by **K**. Let us take any three planes of **L** and attempt to find the point common to them. Since we know that the planes intersect, not at a single point, but at a line we expect to get four zeros.

$$\det\begin{bmatrix} 0 & p & -q & \hat{\mathbf{i}} \\ -p & 0 & s & \hat{\mathbf{j}} \\ q & -s & 0 & \hat{\mathbf{k}} \\ -r & t & -u & \hat{\mathbf{l}} \end{bmatrix} =$$

$$\hat{\mathbf{i}}\times\det\begin{bmatrix} -p & 0 & s \\ q & -s & 0 \\ -r & t & -u \end{bmatrix} - \hat{\mathbf{j}}\times\det\begin{bmatrix} 0 & p & q \\ q & -s & 0 \\ -r & t & -u \end{bmatrix} + \hat{\mathbf{k}}\times\det\begin{bmatrix} 0 & p & -q \\ -p & 0 & s \\ -r & t & -u \end{bmatrix} - \hat{\mathbf{l}}\times\det\begin{bmatrix} 0 & p & -q \\ -p & 0 & s \\ q & -s & 0 \end{bmatrix} =$$

$$(pu - qt + sr)\begin{bmatrix} -s \\ -q \\ -p \\ 0 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

In order to make $x = y = z = w = 0$, as we know must be the case, we are forced to the conclusion that either $s = q = p = 0$ or $pu - qt + sr = 0$. By a similar operation on other choices of columns of **L** we find that the latter choice is correct. Thus, to reiterate, for any matrix **L** constructed from two point vectors to represent the line connecting them, the six coordinates will always satisfy the relation:

$$pu - qt + sr = 0 \qquad (*)$$

Given this relation we can construct the following matrix product:

$$\begin{bmatrix} 0 & -u & -t & -s \\ u & 0 & -r & -q \\ t & r & 0 & -p \\ s & q & p & 0 \end{bmatrix}\begin{bmatrix} 0 & p & -q & r \\ -p & 0 & s & -t \\ q & -s & 0 & u \\ -r & t & -u & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The middle matrix is just **L**. The product is all zeros either identically or by virtue of relation (*). How can we interpret the left-hand matrix? Since each row multiplied by **L** yields four zeros each row must be a point on the line. The left-hand matrix must be the same as **K**, that is, four points on the line stacked into a 4×4 matrix. The matrix **K** thus contains the same numbers as the matrix **L;** they are just arranged differently. We can now match the names of the coordinates with their values if calculated as the intersection of two planes:

$$u = -\det\begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix};\quad t = \det\begin{bmatrix} b_1 & b_2 \\ d_1 & d_2 \end{bmatrix};\quad s = -\det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}$$

$$p = -\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix};\quad q = \det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix};\quad r = -\det\begin{bmatrix} a_1 & a_2 \\ d_1 & d_2 \end{bmatrix}$$

Thus the homogeneous representation of a line exists in two dual forms generated by joining two points and by intersecting two planes. The six coordinate points generated in each case satisfy equation (*).

Distance Measurements

To further increase intuitive feel for the meaning of these six coordinates let us see where a given line intersects the plane at infinity. We multiply the **K** form of the line with the plane at infinity and get:

$$\begin{bmatrix} 0 & -u & -t & -s \\ u & 0 & -r & -q \\ t & r & 0 & -p \\ s & q & p & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -s \\ -q \\ -p \\ 0 \end{bmatrix}$$

The intersection is the point at infinity $\begin{bmatrix}-s & -q & -p & 0\end{bmatrix}$. That means that the 3D vector $\begin{bmatrix}s & q & p\end{bmatrix}$ points parallel to the line. Now let us determine the plane containing the line and the origin. We multiply the **L** form of the line with the origin and get:

$$\begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix}\mathbf{L} = \begin{bmatrix}-r & t & -u & 0\end{bmatrix}$$

This means that the 3D vector $\begin{bmatrix}-r & t & -u\end{bmatrix}$ points perpendicular to this plane. The dot product of these two vectors is zero; this is just relation (*). Thus $\begin{bmatrix}s & q & p\end{bmatrix}$ lies in the plane containing the line and the origin. If we compute the cross product of the two vectors we will get a third vector which is perpendicular to the line and pointing directly toward it.

$$\begin{bmatrix}tp+uq & rp-us & -rq-st\end{bmatrix} = \mathbf{T}$$

By making use of (*) it can be shown that the length of **T** is

$$|\mathbf{T}| = \sqrt{\left(r^2+t^2+u^2\right)\left(s^2+q^2+p^2\right)}$$

We can now compute the perpendicular distance, D, from the origin to the line. Place the normalized **T** at the origin and scale it up by the factor D. We should now be at the point on the line that is closest to the origin.

$$\begin{bmatrix}D\dfrac{(tp+uq)}{|\mathbf{T}|} & D\dfrac{(rp-us)}{|\mathbf{T}|} & D\dfrac{(-rq-st)}{|\mathbf{T}|} & 1\end{bmatrix}\mathbf{L} = \begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix}$$

Multiplying out and solving for D we get:

$$D = \sqrt{\frac{r^2+t^2+u^2}{s^2+q^2+p^2}}$$

This is the perpendicular distance from the origin to the line **L**.

Transforming Lines

A homogeneous point is transformed by post-multiplying by a 4x4 matrix. A homogeneous plane is transformed by pre-multiplying by the inverse of the point transformation matrix. We shall now derive the process whereby a homogeneous line is transformed. This procedure should preserve dot products just as the plane transformation does. That is, given the relationship:

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{L} = \begin{bmatrix}a & b & c & d\end{bmatrix}$$

We wish the transformed quantities to also satisfy the relationship:

$$\begin{bmatrix}x' & y' & z' & w'\end{bmatrix}\mathbf{L}' = \begin{bmatrix}a' & b' & c' & d'\end{bmatrix}$$

We can express the primed point and plane in terms of the unprimed by

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{T} = \begin{bmatrix}x' & y' & z' & w'\end{bmatrix}$$

$$\begin{bmatrix}a & b & c & d\end{bmatrix}\left(\mathbf{T}^{-1}\right)^t = \begin{bmatrix}a' & b' & c' & d'\end{bmatrix}$$

Combining these

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{T}\,\mathbf{L}' = \begin{bmatrix}a & b & c & d\end{bmatrix}\left(\mathbf{T}^{-1}\right)^t$$

$$\begin{bmatrix}x & y & z & w\end{bmatrix}\mathbf{T}\,\mathbf{L}'\mathbf{T}^t = \begin{bmatrix}a & b & c & d\end{bmatrix}$$

Comparing this with the original point, line, plane relation we can state that a solution is:

$$\mathbf{L} = \mathbf{T}\,\mathbf{L}'\mathbf{T}^t$$

or

$$\mathbf{T}^{-1}\,\mathbf{L}\left(\mathbf{T}^{-1}\right)^t = \mathbf{L}'$$

Matrices that represent quantities that transform in this way are called tensors. In addition, since the transformation matrix used is the inverse of the point transformation matrix, it is a contra-variant tensor.

By applying the analogous process to the **K** form of the line we get

$$\mathbf{T}\,\mathbf{K}\,\mathbf{T}^t = \mathbf{K}'$$

This is another tensor. This time the transformation matrix is the same as the point transformation matrix so it is a covariant tensor.

Intersecting Lines

We have so far examined the problem of whether a point is on a line and whether a line is in a plane. There remains the question of whether two lines intersect, and, if so, where. We can solve this by taking the point form of one line and multiplying it by the plane form of the other.

$$\mathbf{K}_1\,\mathbf{L}_2 = \mathbf{N}$$

Each row of $\mathbf{K}_1$, being a point of line 1, will generate a plane through that point and through line 2. If the two lines intersect, each of these will be the same plane: the plane containing the two lines. Likewise each column of $\mathbf{L}_2$, being a plane containing line 2, will generate a point at the intersection of that plane and line 1. If the two lines intersect, each of these will be the same point, the point of intersection of the lines. Thus each row of **N** is a plane vector for the plane common to the lines. Each column of **N** is a point vector for the intersection of the lines. **N** is the outer product of the point and the plane:

$$\mathbf{N} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \begin{bmatrix} a & b & c & d \end{bmatrix} = \begin{bmatrix} ax & bx & cx & dx \\ ay & by & cy & dy \\ az & bz & cz & dz \\ aw & bw & cw & dw \end{bmatrix}$$

Since the point of intersection always lies in the plane of intersection the inner product will be zero. This can be calculated as the trace of **N**. In terms of the components of $\mathbf{K}_1$ and $\mathbf{L}_2$ the trace of **N** has the value

$$\text{trace}\,\mathbf{N} = p_1u_2 - q_1t_2 + s_1r_2 + p_2u_1 - q_2t_1 + s_2r_1$$

Note the similarity to relation (*).

For lines that do not intersect (skew lines) the trace of **N** will be proportional to the perpendicular distance between them. This can be seen in the following manner. First consider the cross product of the direction vectors of the two lines.

$$\begin{bmatrix} s_1 & q_1 & p_1 \end{bmatrix} \times \begin{bmatrix} s_2 & p_2 & q_2 \end{bmatrix} = \begin{bmatrix} s_3 & p_3 & q_3 \end{bmatrix}$$

This vector will be perpendicular to both lines. A plane having $\begin{bmatrix} s_3 & q_3 & p_3 \end{bmatrix}$ as its $\begin{bmatrix} a & b & c \end{bmatrix}$ components will be parallel to both line 1 and line 2. We can find the particular such plane which contains line 1 by solving for d_1 in

$$\mathbf{K}_1 \begin{bmatrix} s_3 \\ q_3 \\ p_3 \\ d_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This yields four equations all of which can be shown to have the common solution

$$d_1 = -p_1u_2 + q_1t_2 - s_1r_2$$

Similarly, the plane parallel to line 1 that contains line 2 has

$$d_2 = p_2u_1 - q_2t_1 + s_2r_1$$

The perpendicular distance of each of these planes to the origin is

$$D_1 = \frac{d_1}{\sqrt{{s_3}^2 + {q_3}^2 + {p_3}^2}}; \qquad D_2 = \frac{d_2}{\sqrt{{s_3}^2 + {q_3}^2 + {p_3}^2}}$$

The perpendicular distance between the two planes and the perpendicular distance between the lines is

$$D_2 - D_1 = \frac{d_2 - d_1}{\sqrt{{s_3}^2 + {q_3}^2 + {p_3}^2}} = \frac{\operatorname{trace} \mathbf{N}}{\sqrt{{s_3}^2 + {q_3}^2 + {p_3}^2}}$$

If the trace is zero, the lines intersect. If the trace is non-zero, the perpendicular distance is as shown.

What, then, are the six homogeneous coordinates for the line along which this distance is measured? We already have the direction of the line as $[s_3 \quad q_3 \quad p_3]$. It remains to find r_3, t_3, and u_3. This can be accomplished by using the three facts that line 3 intersects line 1, line 3 intersects line 2, and the coordinates of line 3 must satisfy relation (*).

$$\operatorname{trace} \mathbf{K}_3 \mathbf{L}_2 = 0$$
$$\operatorname{trace} \mathbf{K}_1 \mathbf{L}_3 = 0$$
$$p_3 u_3 - q_3 t_3 + s_3 r_3 = 0$$

These three equations may then be solved for r_3, t_3, and u_3.

Conclusion

The line representation developed here can be used to solve many geometric problems in three dimensions. Its form, however, does lead to much redundant calculation for many problems of interest. Its main use may therefore be as a conceptual tool to generate formulas for desired geometrical quantities that are then simplified based on other knowledge of the problem.

References

[1] Newman, W. and Sproull, R. *Principles of Interactive Computer Graphics*, McGraw-Hill, 1973, pp. 467.

[2] Hodge, W.V.D. and Pedoe, D. *Methods of Algebraic Geometry*, Cambridge University Press, 1968, pp. 286.

(

Chapter 0-03

Einstein Index Notation

This chapter is cannibalized from
Uppers and Downers, Part 1
which is chapter 9 of *Jim Blinn's Corner: Dirty Pixels*

Tenser said the tensor.
Tenser said the tensor.
Tension, apprehension,
and dissension have begun.
—Alfred Bester, *The Demolished Man*

The Situation

In order to represent homogeneous geometry correctly, we have to borrow some notation from the world of tensor analysis. It turns out that physicists have been faced with this sort of thing for some time. They are concerned with somewhat different problems than we are here but their notation is readily adaptable. Let's build up to this gradually.

We recognize that there are two kinds of things: point-like and plane-like. Rather than cloud our minds with such concepts as rows and columns, we'd like to identify which kind of vector something is by some other sort of notational mechanism. The physicist's solution, translated into our terms, is to write the point-like things with superscript indices (the uppers of the title) and plane-like things with subscript indices (the downers). The point-like indices are called *contravariant* and plane-like ones are called *covariant*.

Putting the two different types of indices in two different locations keeps them distinguishable but also create an ambiguity: superscripts used to mean exponentiation, now they are contravariant coordinate indices. Subscripts used to be available to construct different names, now they are covariant coordinate indices, and we have to use entirely different letters for different names. This ambiguity is another example of a growing problem with mathematical notation:

There aren't enough squiggles to go around

Anyway, we can now write our point as

$$\mathrm{P} = \begin{bmatrix} \mathrm{P}^1 & \mathrm{P}^2 & \mathrm{P}^3 & \mathrm{P}^4 \end{bmatrix}$$

and a 3DH plane as

$$\mathrm{E} = \begin{bmatrix} \mathrm{E}_1 & \mathrm{E}_2 & \mathrm{E}_3 & \mathrm{E}_4 \end{bmatrix}$$

Matrices have two indices. Each one can be either covariant or contravariant. This makes for three possibilities: pure covariant (M_{ij}), pure contravariant (M^{ij}), and what are called "mixed" (M^i_j). It was the inability to distinguish between these that has been causing all our troubles.

One further note: The different type styles we previously needed to distinguish between scalars, vectors and matrices are no longer necessary. Since we can now easily determine the species of creature by how many indices it has, I'll just use italic letters when indices are used. Anything with one index is a vector; anything with two indices is a matrix. In fact, we can now have triply indexed critters (cubical matrices?) or quadruply indexed critters. These actually have practical use, as we will see below.

The Multiplication Machine

We will now represent vector and matrix multiplication in a different way. Remember, in the old style, the laws of matrix multiplication are just a shorthand notation for

$$\begin{bmatrix} x & y & z & w \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = ax + by + cz + dw$$

Now, instead of using single letter names we are using index numbers for the components.

$$\mathrm{P} \cdot \mathrm{E} = \begin{bmatrix} P^1 & P^2 & P^3 & P^4 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} = P^1 E_1 + P^2 E_2 + P^3 E_3 + P^4 E_4 = \sum_i P^i E_i$$

In fact, we can save valuable ink by noticing that the summation occurs so often that we can declare that it is implied by the fact that the point index and plane index are the same letter. This is often emphasized by the convention of using Greek letters for indices that are summed over. Thus the product of a point and a plane is

$$P^\alpha E_\alpha$$

This expression is a sort of prototype of the terms that are summed over. This might take a bit of getting used to, but it's worth it. This notation is credited to Albert Einstein, who invented it to shorten his calculations for general relativity; it is therefore called *Einstein Index Notation*.

We can make all of our row/column confusion go away by the following rule:

> Each index that is summed over must occur someplace in the prototype term exactly once as a covariant index and once as contravariant index. These indices “annihilate” each other, and the resultant product has one less of each kind of index.

Indices that are *not* summed over are called *free indices*. They must occur just once in the prototype term, and the same index must occur in the resultant term. For example, the product of two matrices in the old notation is

$$\mathbf{T} = \mathbf{MN}$$

In the new notation this would be

$$T_j^i = M_j^\alpha N_\alpha^i$$

Now for a shocker:

$$M_j^\alpha N_\alpha^i = N_\alpha^i M_j^\alpha$$

But, you say, matrix multiplication is not commutative! And you are right. The above expression is not the whole matrix product; it is a prototype for each term in the product. Within the term you just have numbers being multiplied; that *is* commutative. So, to mix metaphors somewhat, the two orderings of multiplication are

$$\mathbf{MN} = M_j^\alpha N_\alpha^i = N_\alpha^i M_j^\alpha$$

and

$$\mathbf{NM} = N_j^\alpha M_\alpha^i = M_\alpha^i N_j^\alpha$$

The New Order

Let's now reinterpret all of the above confusion in terms of this notation. First let's do the obvious stuff.

Points and Planes

A point has a single contravariant index,

$$P^i$$

A 2DH line or 3DH plane has a single covariant index,

$$E_i$$

A point times a line/plane is a scalar

$$P^\alpha E_\alpha = s$$

If it's zero the point lies on the line.

Quadrics

A quadric curve or surface has a pure covariant form for dealing with points. Note that both the points have their accustomed contravariant indices.

$$P^\alpha Q_{\alpha\beta} P^\beta = 0$$

The polar line to a quadric curve and the polar plane to a quadric surface is

$$E_i = P^\alpha Q_{\alpha i}$$

Note again the delightful consistency of indices.

The dual form of the quadric for testing line/plane tangency is pure contravariant

$$E_\alpha \tilde{Q}^{\alpha\beta} E_\beta = 0$$

Both line/planes have their accustomed covariant indices. The tilde over the Q is kind of redundant since the placement of the indices tells us everything, covariant for the normal form and contravariant for the dual form. We won't bother with the tildes any more; yet more clean ups.

The contravariant form of Q is the adjoint of the covariant form. This leads to another convenient rule that we will expand upon later. For now I'll just state:

Taking an adjoint flips the type of its indices

This means that the adjoint of a mixed tensor (which is a transformation matrix) is also a mixed tensor. If the only type of matrix you encountered was a transformation matrix, you wouldn't know there was such a thing as covariant or contravariant tensors.

3DH Lines

A 3DH line has a pure covariant tensor or a pure contravariant tensor. The plane containing the line and a point P uses the covariant form

$$P^{\alpha} L_{\alpha j} = E_j$$

The point of intersection of the line and a plane E uses the contravariant form

$$L^{\alpha i} E_{\alpha} = P^i$$

Transformations

A transformation matrix is a mixed tensor. It can transform a point

$$(P')^i = P^{\alpha} T_{\alpha}^i$$

or it can transform a line (2DH) or a plane (3DH):

$$(E')_i = T_i^{\alpha} E_{\alpha}$$

The covariant form of **Q** transforms like

$$(Q')_{ij} = \left(T^*\right)_i^{\alpha} Q_{\alpha\beta} \left(T^*\right)_j^{\beta}$$

Notice that we no longer need to use the superscript T to express transpose; the relevant indices are just swapped. Yet more economization. Explicit notation for adjoints *is* still necessary for now.

The contravariant form of **Q** transforms using **T**

$$(Q')^{ij} = T_{\alpha}^i Q^{\alpha\beta} T_{\beta}^j$$

Likewise, the covariant form of the 3DH line L transforms like

$$(L')_{ij} = \left(T^*\right)_i^{\alpha} L_{\alpha\beta} \left(T^*\right)_j^{\beta}$$

The contravariant form transforms like:

$$(L')^{ij} = T_{\alpha}^i L^{\alpha\beta} T_{\beta}^j$$

This is an example of the general transformation rule

> *To transform something, multiply in a* **T** *for each contravariant (point-like) index and a* $\mathbf{T}^*$ *for each covariant (plane-like) index.*

In fact, all these things are called tensors precisely because they transform according to this rule.

There is an interesting consequence of this. In order to transform a *transformation matrix* we must multiply both **T** and $\mathbf{T}^*$. Calling the transformation matrix **M**

$$(M')_j^i = \left(T^*\right)_j^{\beta} M_{\beta}^{\alpha} T_{\alpha}^i$$

This is a nice way of representing the standard trick of scaling about an arbitrary point by transforming the point to the origin ($\mathbf{T}^*$), scaling about the origin (**M**), and transforming back (**T**).

The Magic Epsilon

You might have noticed that we've encountered a lot of expressions of the form

$$\det\begin{bmatrix} p_1 & p_2 \\ r_1 & r_2 \end{bmatrix}$$

There's another gimmick that the physicists have come up with that's useful to abbreviate this type of thing: it's the Levi-Civa epsilon. In this chapter I can give just a hint of the wonders in store for us when using epsilon. To start out let's discuss this just in 2DH (3D) terms.

The 3D (2DH) epsilon

The three dimensional epsilon tensor has three indices, so it looks like

$$\varepsilon_{ijk}$$

Its elements are defined to be

$$\begin{aligned} \varepsilon_{123} &= \varepsilon_{231} = \varepsilon_{312} = 1 \\ \varepsilon_{321} &= \varepsilon_{213} = \varepsilon_{132} = -1 \\ \varepsilon_{ijk} &= 0 \quad \text{otherwise} \end{aligned}$$

You can visualize ε by thinking of it as a cube of numbers made by stacking up the matrices:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Multiplying two points P and S by epsilon gives

$$P^{\alpha} S^{\beta} \varepsilon_{\alpha\beta i} = L_i$$

a covariant vector. What is it? To find, say, the first element let's write down all the terms containing an epsilon that is nonzero.

$$\begin{aligned} L_1 &= P^{\alpha} S^{\beta} \varepsilon_{\alpha\beta 1} \\ &= P^2 S^3 \varepsilon_{231} + P^3 S^2 \varepsilon_{321} = P^2 S^3 - P^3 S^2 \end{aligned}$$

The other elements look similar. L is, in fact, the cross product of P and S, that is, the line connecting them.

There is also a contravariant epsilon that has the same numerical values. You can use it to take cross products of lines to find their point of intersection.

$$L_{\alpha} K_{\beta} \varepsilon^{\alpha\beta i} = P^i$$

Notice that I didn't even need to tell you that K was a line. You could tell by seeing that it has a single covariant index.

The epsilon can also be used to calculate adjoints. It turns out that

$$\left(M^*\right)^{ij} = \tfrac{1}{2} \varepsilon^{j\alpha\beta} \varepsilon^{i\gamma\delta} M_{\alpha\gamma} M_{\beta\delta}$$

Where does the 1/2 come from? Well, as an exercise, write down all the nonzero epsilon terms implied by the summation over α, β, γ, and δ. You will find that each product of the M terms will appear twice, requiring ½ to compensate. We can, however, ignore the ½ since we are being homogeneous.

The above is for pure covariant matrices; note how the result is contravariant. If you start with a contravariant matrix you multiply by two covariant epsilons and get a covariant result. For a mixed matrix you must multiply by one covariant and one contravariant epsilon and you will get a mixed result.

The 4D (3DH) epsilon

There is also a four dimensional epsilon that has four indices. It's defined by

$$\varepsilon_{ijkl} = \begin{cases} 1 & \text{if } ijkl \text{ is an even permutation of 1234} \\ -1 & \text{if } ijkl \text{ is an odd permutation of 1234} \\ 0 & \text{otherwise} \end{cases}$$

Given this, we can compactify a lot of the formulae for 3DH stuff.

The plane through three points is

$$E_i = P^\alpha S^\beta R^\gamma \varepsilon_{\alpha\beta\gamma i}$$

The point common to three planes uses the other form

$$P^i = E_\alpha F_\beta G_\gamma \varepsilon^{\alpha\beta\gamma i}$$

The adjoint of a 4×4 matrix is

$$\left(M^*\right)^{ij} = \tfrac{1}{6}\varepsilon^{j\alpha\beta\gamma}\varepsilon^{iabc} M_{\alpha a} M_{\beta b} M_{\gamma c}$$

Again, this is for a covariant matrix. Contravariant and mixed matrices require the same treatment as for 2DH.

The 3DH line through two points P and R is the covariant matrix

$$L_{ij} = P^\alpha R^\beta \varepsilon_{\alpha\beta ij}$$

The plane containing this line and another point S is

$$E_i = S^\alpha L_{\alpha i}$$

You find the contravariant version of a line by intersecting the two planes E and F:

$$L^{ij} = E_\alpha F_\beta \varepsilon^{\alpha\beta ij}$$

The point of intersection of this line with another plane G is

$$P^i = L^{i\alpha} G_\alpha$$

The covariant and contravariant line forms are related by

$$L_{ij} = \varepsilon_{ij\alpha\beta} L^{\alpha\beta}$$

Admittedly, implementing some of the above calculations by explicitly multiplying by epsilon is a bit idiotic. You wind up multiplying by a whole lot of zeroes and ones. The epsilon notation is good as a bookkeeping convenience.

What we've learned so far

All of geometry is reduced to tensor multiplication (well, almost all). And there are no embarrassing transposes. The concept of Row-ness or column-ness is superseded by the more general concept of covariant and contravariant indices. Plus we can feel really cool by sharing notation with General Relativity.

Everything is a tensor. A tensor is a multiply indexed array of numbers that transforms in a certain special way. Each index of a tensor can be one of two types: covariant or contravariant. We write covariant indices as subscripts and contravariant indices as superscripts.

A column vector in the old matrix notation becomes a tensor with a single covariant index and is called a covector. These typically represent lines (2DH) or planes (3DH). A row vector in matrix notation becomes a tensor with a single contravariant index. I'll simply call them vectors. These typically represent points in 2DH or 3DH.

Now, in standard 3D geometry you can form dot products of vectors and you can form cross products of vectors. There are a lot of calculations (notably lighting calculations) that take place in this Euclidean 3D space. In our homogeneous scheme, however, you can only form the dot product of a vector with a covector; you can't take the dot product of a vector with another vector. You use the cross product to combine two vectors, and the result is a covector. I'll stick with these rules throughout this chapter..

Tensor multiplication (a generalization of the dot product) involves summing over a pair of covariant and contravariant indices. A typical expression looks like

$$A^i_{\alpha j} B^\alpha = C^i_j$$

Summed indices (here, α) are called bound indices and disappear from the result. Unsummed (free) indices (the i and j) survive in the result. Since bound indices are purely local to the summation, we can arbitrarily change their name; we could just as well have written the above equation as

$$A^i_{\beta j} B^\beta = C^i_j$$

Free indices, however, can only be renamed globally; you have to do it consistently for each term on each side of an equation. Index name bookkeeping is one of the biggest pains of this method.

In the last chapter I also introduced the special tensor called ε. We can use ε to abbreviate the following tensor operations.

The cross product

$$A \times B = C$$

becomes

$$A^\alpha B^\beta \varepsilon_{\alpha\beta i} = C_i$$

The adjoint of a 3x3 matrix becomes

$$\left(M^*\right)^{ij} = \tfrac{1}{2}\varepsilon^{j\alpha\beta}\varepsilon^{i\gamma\delta} M_{\alpha\gamma} M_{\beta\delta} \qquad (1)$$

In homogeneous land we can drop the factor of 1/2.

The 3D (2DH) Epsilon-Delta Rule

If you multiply two epsilons together along one index, the rules of tensor multiplication give

$$\varepsilon_{\alpha jk}\varepsilon^{\alpha lm} = D^{lm}_{jk}$$

a 4-index mixed tensor. Since ε is filled with a fixed pattern of 0's, +1's, and −1's, you can work out the values of **D**. They turn out to be

$$D^{lm}_{jk} = \begin{cases} +1 & \text{if } l=j;\ m=k;\ l\neq m \\ -1 & \text{if } m=j;\ l=k;\ l\neq m \\ 0 & \text{otherwise} \end{cases}$$

You can write this in a more compact form by inventing yet another special tensor called δ.

$$\delta^i_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

In other words, δ is just an identity matrix. This might seem a bit pointless at first, but it lets us write the above as:

$$\varepsilon_{\alpha jk}\varepsilon^{\alpha lm} = \delta^l_j\delta^m_k - \delta^m_j\delta^l_k \qquad (2)$$

This fundamental identity is called the epsilon-delta rule and is useful as a way to simplify any expression you come across that contains the product of two epsilons.

A Simple Application

Take, for example, the following common 3D vector identity:

$$A\times(B\times C) = (A\cdot C)B - (A\cdot B)C \qquad (3)$$

First let's see if this makes sense in our homogeneous world of vectors and covectors. **B** and **C** can be vectors (that is,. 2DH points). Their cross product is a covector (a 2DH line). We cross this with **A**, which means that **A** must be a covector too. The result is a vector. Now look at the right side of the equation. It's also a vector, a weighted sum of the two vectors **B** and **C**. The weighting factors are dot products between vector-covector pairs. Everything fits in consistently with our scheme.

Writing the left side in Einstein index notation we get:

$$A_i\left(B^jC^k\varepsilon_{jkl}\right)\varepsilon^{ilm}$$

In this expression the letters *i*, *j*, *k* and *l* are all bound indices (since they each appear exactly twice) indicating that they are implicitly summed over. The letter *m* is a free index, indicating that the net result is a vector with the single superscript index *m*.

We are going to want to apply the epsilon-delta rule, but its definition (equation 2) has different letters for the indices on the epsilons. We can make our expression more similar to it by shuffling the indices of the epsilons to get the common summed index, *l*, as the first one. This is legal as long as we do an even permutation of the indices.

$$A_iB^jC^k\varepsilon_{ljk}\varepsilon^{lmi}$$

Then we need to do some pattern matching with equation 2. It's actually easiest to rewrite equation 2, renaming the indices to match up with the above expression. (This sort of thing is a nuisance we will be able to avoid with the graphical method.)

$$\varepsilon_{ljk}\varepsilon^{lmi} = \delta_j^m \delta_k^i - \delta_j^i \delta_k^m$$

Stuff in the deltas for the product of the epsilons and you get

$$A_i B^j C^k \left(\delta_j^m \delta_k^i - \delta_j^i \delta_k^m\right)$$

multiply out and apply several identities of the form

$$A_i \delta_k^i = A_k$$

You get

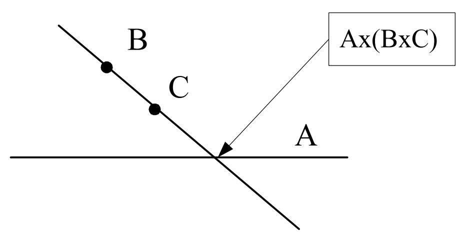
$$A_k B^m C^k - A_j B^j C^m$$

which, in old style vector notaton is

$$(\mathrm{A}\cdot\mathrm{C})\mathrm{B} - (\mathrm{A}\cdot\mathrm{B})\mathrm{C}$$

Ta daa...

A geometric interpretation of this algebra appears in the following figure.



It's a two dimensional version of the perspective shadow calculation I described in chapter six of *A Trip Down The Graphics Pipeline*. Think of B as a light source casting the shadow of point C onto the ground line A. The two sides of the equation are two ways of thinking about calculating the shadow location. You can think of it as finding the line through B and C, (B×C), and intersecting that line with line A, (A×(B×C)). Alternatively you can parametrize the line through B and C as a linear combination of the vectors B and C

$$\alpha\mathrm{B} + \beta\mathrm{C}$$

The α,β pair forms a one dimensional homogeneous coordinate for points on the line. Force this point to be on the line A by

$$\mathrm{A}\cdot\left(\alpha\mathrm{B} + \beta\mathrm{C}\right) = 0$$

or

$$\alpha(\mathrm{A}\cdot\mathrm{B}) + \beta(\mathrm{A}\cdot\mathrm{C}) = 0$$

a solution to this equation is

$$\alpha = \mathrm{A}\cdot\mathrm{C}$$
$$\beta = -\mathrm{A}\cdot\mathrm{B}$$

so the intersection is

$$(\mathrm{A}\cdot\mathrm{C})\mathrm{B}-(\mathrm{A}\cdot\mathrm{B})\mathrm{C}$$

Quadric/Line Tangency

Here's another use for the epsilon-delta rule.

Earlier I noted that a symmetric 3x3 covariant tensor, **Q**, represents a second order curve (conic sections and the like). For points on the curve

$$\mathrm{P}\mathbf{Q}\mathrm{P}^{\mathrm{T}}=0$$

Then I baldly stated that a line, L, is tangent to the curve if

$$\mathrm{L}^{\mathrm{T}}\mathbf{Q}^{*}\mathrm{L}=0$$

where $\mathbf{Q}^{*}$ is the adjoint of **Q**. Why should you believe me (other than the fact that I am extremely trustworthy)? We need to express the fact that L and **Q** have exactly one point in common. To do this we write an arbitrary point of L in the same way as mentioned above—as the weighted sum of two distinct points of L.

$$\mathrm{P}=\alpha\mathrm{R}+\beta\mathrm{S}$$

How, you might ask, do we decide what to use for R and S? Well, it doesn't matter. It's one of the funny things that often happen in mathematical proofs. We never actually have to know explicit coordinates for R or S, we just have to know that they exist and that

$$\mathrm{L}=\mathrm{R}\times\mathrm{S}$$

The points of L that intersect **Q** have

$$\mathrm{P}\mathbf{Q}\mathrm{P}^{\mathrm{T}}=0$$

or

$$(\alpha\mathrm{R}+\beta\mathrm{S})\mathbf{Q}(\alpha\mathrm{R}+\beta\mathrm{S})^{\mathrm{T}}=0$$

Multiplying this out and remembering that **Q** is symmetric gives

$$\alpha^2\left(\mathrm{R}\mathbf{Q}\mathrm{R}^{\mathrm{T}}\right)+2\alpha\beta\left(\mathrm{R}\mathbf{Q}\mathrm{S}^{\mathrm{T}}\right)+\beta^2\left(\mathrm{S}\mathbf{Q}\mathrm{S}^{\mathrm{T}}\right)=0$$

This is a homogeneous quadratic equation in (α,β) of the form

$$\alpha^2 a+\alpha\beta b+\beta^2 c=0$$

If L is tangent to **Q** this equation should have exactly one solution. Remembering our high school algebra this condition is

$$b^2-4ac=0$$

For our particular quadratic equation this translates to:

$$4\left(\mathrm{R}\mathbf{Q}\mathrm{S}^{\mathrm{T}}\right)^2 - 4\left(\mathrm{R}\mathbf{Q}\mathrm{R}^{\mathrm{T}}\right)\left(\mathrm{S}\mathbf{Q}\mathrm{S}^{\mathrm{T}}\right) = 0$$

We want to show that this condition is the same as

$$\mathrm{L}^{\mathrm{T}}\mathbf{Q}^{*}\mathrm{L} = (\mathrm{R}\times\mathrm{S})^{\mathrm{T}}\mathbf{Q}^{*}(\mathrm{R}\times\mathrm{S}) = 0$$

At this point, with conventional vector-matrix notation we're stuck. There's no convenient way to use the commonality between cross products and matrix adjoints.

Einstein to the rescue. Write the two expressions in Einstein index notation. The first is

$$R^iQ_{ij}S^jR^kQ_{kl}S^l - R^iQ_{ij}R^jS^kQ_{kl}S^l$$

and the second is

$$\left(R^iS^j\varepsilon_{ijm}\right)\left(\tfrac{1}{2}\varepsilon^{m\alpha\beta}\varepsilon^{n\gamma\delta}Q_{\alpha\gamma}Q_{\beta\delta}\right)\left(R^kS^l\varepsilon_{kln}\right)$$

You can then show that these two expressions are equal by doing a lot of renaming of indices and using the epsilon-delta rule twice. It works, but you have to be careful to keep all 10 index names straight. We want a still simpler scheme.

The 4D (3DH) Epsilon-Delta Rule

The somewhat imposing 4D version of the epsilon delta rule is

$$\begin{aligned}\varepsilon_{\alpha ijk}\varepsilon^{\alpha lmn} = {} & \delta_i^l\delta_j^m\delta_k^n + \delta_i^m\delta_j^n\delta_k^l + \delta_i^n\delta_j^l\delta_k^m \\ & -\delta_i^l\delta_j^n\delta_k^m - \delta_i^m\delta_j^l\delta_k^n - \delta_i^n\delta_j^m\delta_k^l\end{aligned}$$

This is what you must use if you are solving problems in 3D homogeneous coordinates. Note that all the delta terms have the same subscripts, but they just have even or odd permutations of the (*lmn*) letters in their superscripts.

From this we can evaluate another useful identity. Form the double summation by setting $i = l$ above. When you simplify, remember that a delta summed with itself is the trace of the identity—in 4D this would be the constant 4. That is

$$\delta_i^i = 4$$

When you work this out, it looks a lot like our old friend the 3D (2DH) rule

$$\varepsilon_{\alpha ijk}\varepsilon^{\alpha imn} = 2\left(\delta_j^m\delta_k^n - \delta_j^n\delta_k^m\right)$$

Addendum
A Programming Application

This vector/covector distinction can take on practical significance in the programming arena. One game that's currently played with modern programming languages is to define compound data types and to overload all the language's arithmetic operators to do arithmetic on the new data type. For example, one might define a data type for vectors (as a triple or a quadruple of numbers) and then define vector addition for the "+" operator, vector subtraction for the "–" operator, etc. in the obvious way. The problem comes when you get to multiplication. Of the two types of vector multiplication, the dot product and the cross product, which one should be meant by the "*" symbol? Now that we realize that there are two types of vectors we can remove the ambiguity. We need to define two different data types: vectors and covectors.

The storage structure, addition and subtraction operators are the same for both of these but multiplication depends on the operand type. Suppose the variable S is declared to be a scalar, V1 and V2 are vectors and C1 and C2 are covectors. Expressions containing the multiplication operator should be interpreted as follows

expression	product	result
S*V1	scalar	vector
S*C1	scalar	covector
V1*C1	dot	scalar
V1*V2	cross	covector
C1*C2	cross	vector

Chapter 0-04

A First Look At Tensor Diagrams

This chapter is cannibalized from
Uppers and Downers, Part 2
in Jim Blinn's Corner: Dirty Pixels

3D(2DH) Diagram Notation

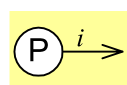
The basic technique is adapted from a book called *Diagram Techniques in Group Theory* (reference: Stedman, Geoffrey E., *Diagram Techniques in Group Theory*, Cambridge University Press, United Kingdom, 1990). We represent the product of a bunch of tensors in diagram form as a directed graph. The encoding is as follows

- Each tensor in the product is a node in the graph.
- Each index is an arc. A contravariant index is directed away from the node, a covariant index is directed towards the node.
- Bound indices (those that are summed over) are arcs connecting two nodes. The directedness of the arc represents the fact that you can sum only over a contravariant-covariant pair of indices.
- Free indices are "dangling" arcs.

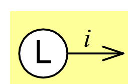
The nice thing about this is that bound indices don't need to be named (although I will sometimes do so here to more easily relate the Einstein index notation to the diagram notation). Only the topology of this diagram is important. Any rearranging that preserves topology is OK except for mirror reflections. These are not allowed since they imply transposition of some of the tensors. This is simply a sign flip for epsilon but might change the value of a nonsymmetric tensor.

Here are some examples.

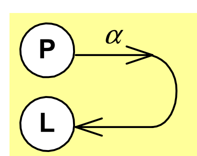
A point is P^i



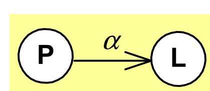
A line L_i



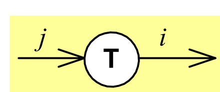
The dot product of a point and a line $P^\alpha L_\alpha$



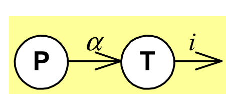
which is topologically equivalent to



A transformation matrix T^i_j

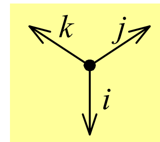


A transformed point $P^\alpha T^i_\alpha$

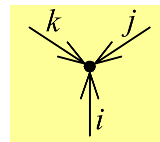


We write the special tensor epsilon (in either its covariant or contravariant forms) using a dot for the node and labeling index arcs counterclockwise around the node.

Covariant epsilon ε^{ijk}

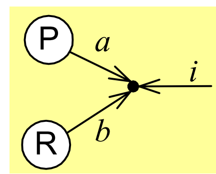


Contravariant epsilon ε_{ijk}



The cross product of two vectors is

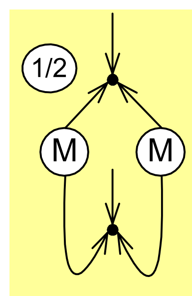
Cross product $P^{\alpha} R^{\beta} \varepsilon_{\alpha\beta i}$



Note that mirroring this flips the sign.

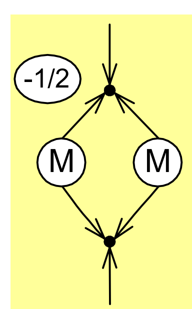
Equation 1 for the adjoint of a matrix appears in diagram form as

Adjoint of M $\frac{1}{2} \varepsilon^{j\alpha\beta} \varepsilon^{i\gamma\delta} M_{\alpha\gamma} M_{\beta\delta}$

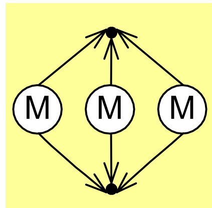


This is a bit clumsy. We can flip the lower epsilon over to make it a bit prettier but must compensate by prefixing with a minus sign.

Adjoint of M $\frac{1}{2} \varepsilon^{j\alpha\beta} \varepsilon^{i\gamma\delta} M_{\alpha\gamma} M_{\beta\delta}$

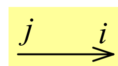


If we multiply the adjoint by the original matrix we get the identity matrix times the determinant of the original matrix. The trace (sum of the diagonal elements) of this 3×3 matrix is then three times the determinant. I'll leave it for you to verify that, graphically, the determinant of a matrix is proportional to



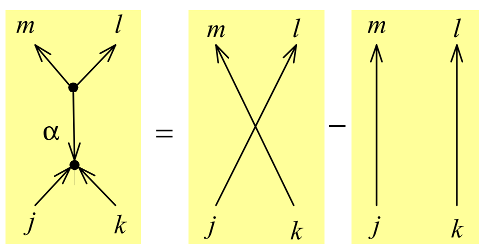
Since we are not concerned with scale factors, the only interesting feature of a determinant is whether it's zero or not. In other words, if the above diagram (which evaluates to a scalar) is zero, then the matrix is singular.

The other special symbol, delta, is just written as an arc (when necessary I will label it with two labels)

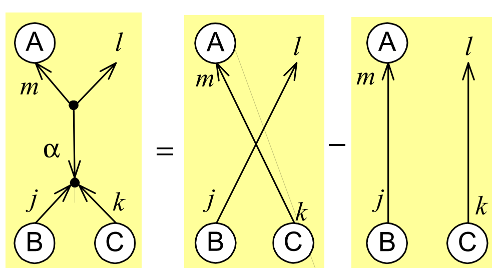
Kronecker delta	δ^i_j	

The Diagram for Epsilon-Delta

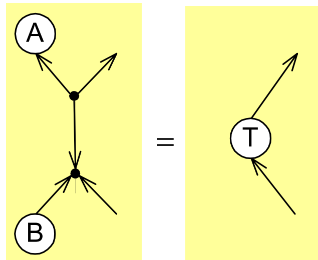
Now we get to the interesting part: the epsilon delta rule in diagram notation. Compare this with equation 2 to see how the index names work. Note especially that the arcs are carefully labeled counterclockwise around both epsilon nodes.



As an example, our proof of the vector identity of equation 3 in diagram notation looks like

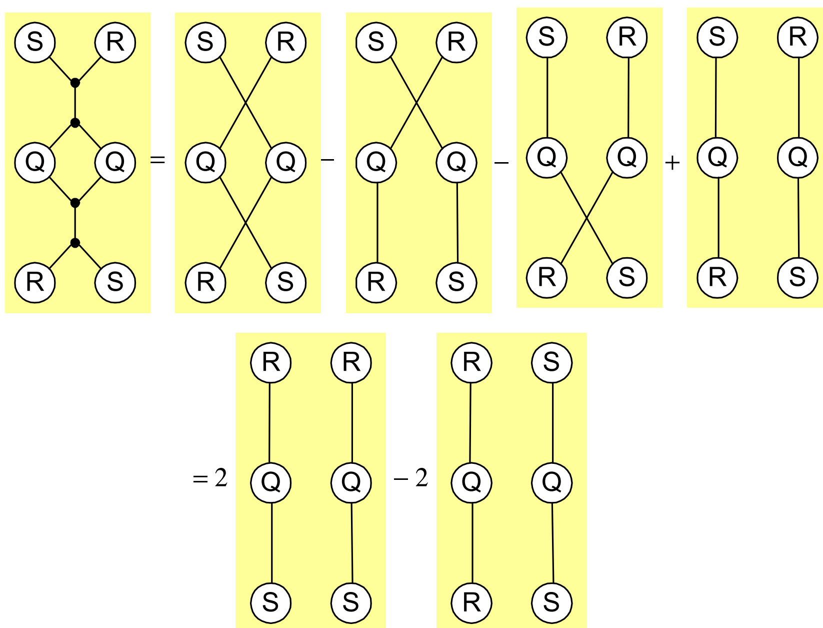


When I used this to calculate the perspective shadow I was primarily interested in a finding a transformation matrix to apply to arbitrary points C. In diagram notation this transformation is just



Note that this diagram has one covariant and one contravariant free index and thus qualifies as a transformation matrix.

Now we can do the line tangency proof diagrammatically. To show the essence of the proof I've left out the cluttering index names, a few global scale factors, and even the arrows. This diagram is



4D(3DH) Diagram Notation

The generalization of this to 4D(3DH) is generally straightforward. The only tricky part is the Levi-Civita epsilon symbol. This section is mostly a discussion about that.

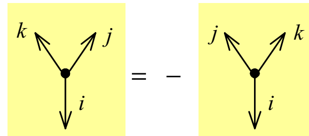
One of the nice things about diagram notation is that any geometric rearrangement of the diagram does not affect the algebraic result. We cannot use totally arbitrary diagram manipulations however. Anti-symmetric tensors, like epsilon, could have their sign changed if we are not careful. In three dimensions (and homogeneous two-dimensional geometry) this epsilon tensor has three indices and is written ε_{ijk}. The values of the components are

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$$
$$\varepsilon_{321} = \varepsilon_{213} = \varepsilon_{132} = -1$$
$$\varepsilon_{ijk} = 0 \quad \text{otherwise}$$

A cyclic permutation of the indices of ε doesn't change its value, that is

$$\varepsilon^{ijk} = \varepsilon^{jki} = \varepsilon^{kij}$$

In the diagram, we didn't need to explicitly identify which index was first. As long as you labeled them counterclockwise, you could start with any arc. As long as any diagram rearrangements do not permute the indices of the epsilons in such a manner that the sign changes we are OK. In three dimensions this means that mirror reflections are disallowed, as they are equivalent to an odd permutation of indices. (You can do mirror reflections, though, if you remember to introduce a minus sign into the diagram.)



ε_{ijk} $\qquad$ $-\varepsilon_{ijk}$

In four dimensions (used for homogeneous three-dimensional geometry) the epsilon tensor has four indices, ε_{ijkl} whose values are

$$\varepsilon_{ijkl} = 1 \qquad \text{if } ijkl \text{ is an even permutation of 1234}$$

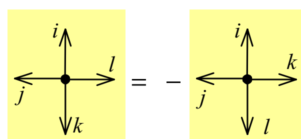
$$\varepsilon_{ijkl} = -1 \qquad \text{if } ijkl \text{ is an odd permutation of 1234}$$

$$\varepsilon_{ijkl} = 0 \qquad \text{otherwise}$$

In contrast to 3D, in 4D a cyclic permutation *does* change the sign, that is

$$\varepsilon^{ijkl} = -\varepsilon^{jkli}$$

In 4D diagram notation the 4D epsilon is simply be a four-pronged node. An odd permutation of indices for the 4D epsilon is not so geometrically obvious if the diagram form is simply a node with four lines, as in



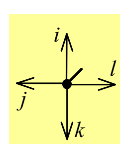
ε_{ijkl} $\qquad$ $\varepsilon_{ijlk} = -\varepsilon_{ijkl}$

We have to be careful, however. . In 4D(3DH), we would like a similar no-mirroring rule for the four-dimensional epsilon diagram. We need make sure the diagram notation allows us to keep track of this. There are two ways to do this. The sign of the four dimensional epsilon is not changed by mirror reflections (a mirror reflection generates an even permutation of the four arc/indices.) It is changed, however, by starting the counting of the indices at a different arc, as in the above figure.

Way 1

For this reason, the books I have read introduce an extra flag between the intended first and last arc in the four dimensional epsilon to mark the initial index. This flag makes the node look like it has five arcs and thus a mirror reflection will then change the sign. The diagram is

ε_{ijkl}



ε_{ijkl}

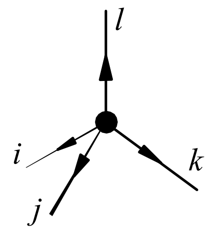
Exchange of any neighboring arcs (counting the tick mark as an honorary arc for now) flips the sign. In fact, in 4D, mirroring just the epsilon portion of a diagram does not change the sign.

These are the rules for allowable diagram rearrangements that I have seen in books. They always seemed klugy to me.

Way 2

Then I realized that the elegance of the diagram notation could be restored by simply expressing four-dimensional tensor multiplication as a three-dimensional diagram. The four-arc epsilon would then have arcs directed to the vertices of a tetrahedron. Even permutations of the indices become simple three-dimensional rotations of the epsilon tensor diagram. Odd permutations become mirror reflections, so mirroring is disallowed just as in two-dimensional diagrams.

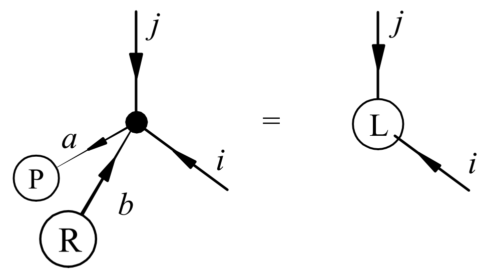
ε_{ijkl}



Use for 3DH Lines

Given this scheme, let's look at some more constructions

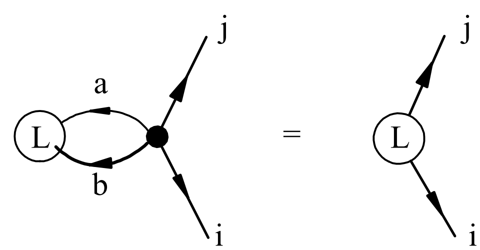
The 3DH homogeneous line between two points is $P^{\alpha}R^{\beta}\varepsilon_{\alpha\beta ij} = L_{ij}$



We should actually make the node for L be asymmetrical since it represents an anti-symmetric matrix.

Note that this tensor has two covariant indices. We can convert it to the contravariant form by multiplication by the 4D epsilon. This implements conversion between the "K form" and "L form" of the 4x4 line matrix as discussed in the previous chapter on homogeneous lines.

The contravariant form is $L_{\alpha\beta}\varepsilon^{\alpha\beta ij} = L^{ij}$

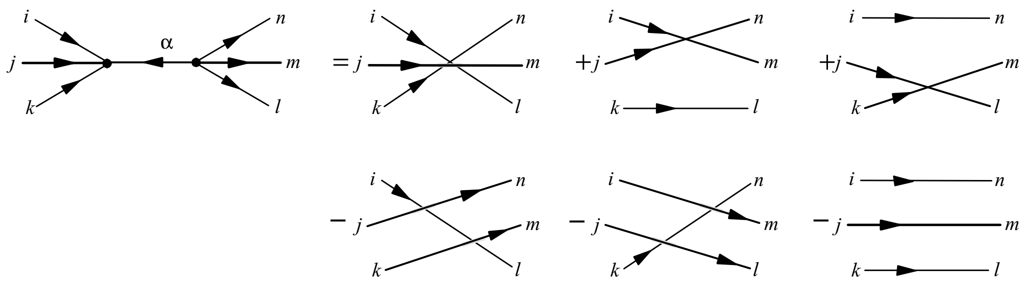


The 4D epsilon delta rule

This is

$$\varepsilon_{\alpha ijk}\varepsilon^{\alpha lmn} = \delta_i^l\delta_j^m\delta_k^n + \delta_i^m\delta_j^n\delta_k^l + \delta_i^n\delta_j^l\delta_k^m - \delta_i^l\delta_j^n\delta_k^m - \delta_i^m\delta_j^l\delta_k^n - \delta_i^n\delta_j^m\delta_k^l$$

The diagram looks like.



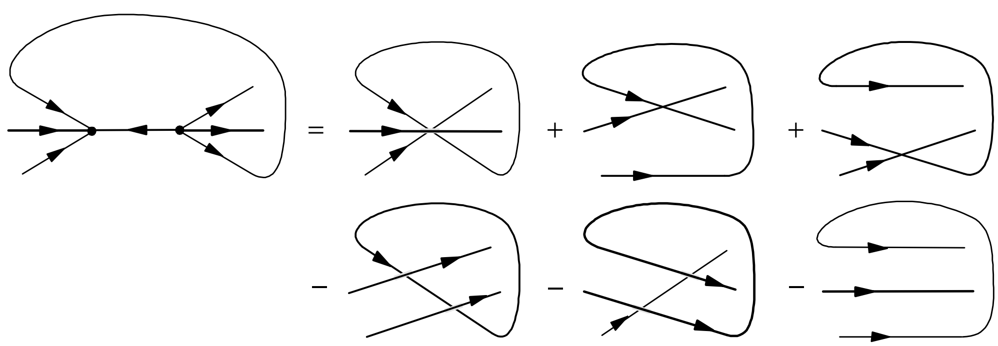
Keep in mind that this diagram is three dimensional; the two joined epsilons look sort of like an acetylene molecule. A mnemonic to help you remember it is to note that in the 2D projection of the diagram, the positive terms have an odd number of crossings (one or three) and the negative terms have an even number (zero or two). In three dimensions, the positive terms actually each have only one pair of arcs crossing. (The first of the terms looks lik it has three arcs crossing, but the horizontal one is ctually closer to tyou than the other two.) The negative terms have arcs that do not cross in three dimensions (The apparent crossings in the left tow negative terms do not actually cross in three dimensions; they are a dort of twisted version of the straight-through rightmost negative term.

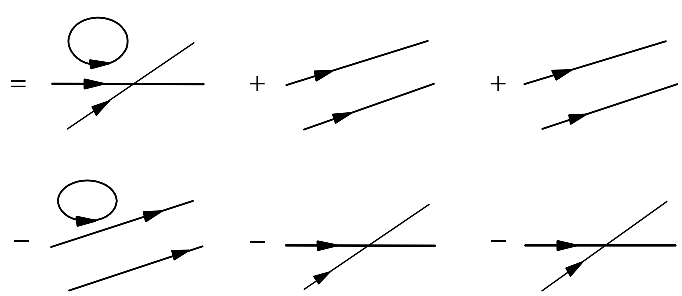
Let's now look at the double summation

$$\varepsilon_{\alpha\beta jk}\varepsilon^{\alpha\beta mn} = 2\left(\delta_j^m\delta_k^n - \delta_j^n\delta_k^m\right)$$

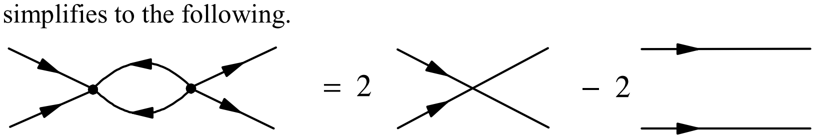
This can be found graphically from the previous diagram by connecting the *i* and *l* branch together and calling it β. This gives the following diagram

.



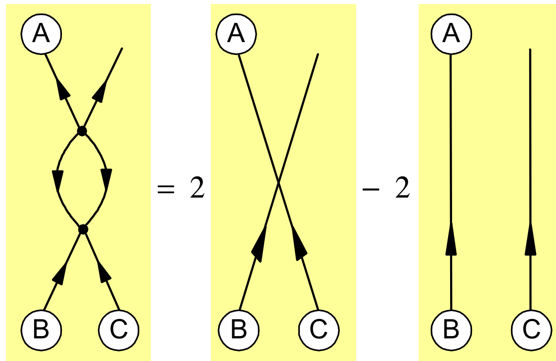


Note that a closed loop is the graphical representation of δ_i^i and is equal to 4, the trace of the 4x4 identity matrix. This all simplifies to the following.

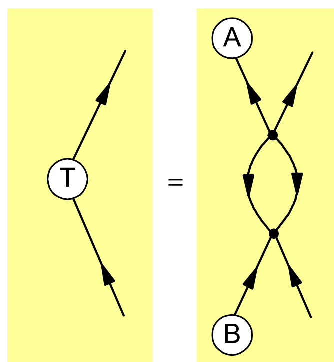


An Application

A sample application of this identity is as follows. Plug line A and points B and C into the diagram:



It shows the 4D (3DH) perspective shadow calculation; Point B casts the shadow of point C onto plane A. So, given that we will need to pass many point C's through the rendering process, we can pre-compute the transformation matrix **T** as:



(This is just the more realistic 3D generalization of the 2D problem we discussed earlier.)

References

Stedman, Geoffrey E., *Diagram Techniques in Group Theory*, Cambridge University Press, United Kingdom, 1990.

Chapter 0-05

Polynomial Discriminants and Curve Tangency

This chapter is cannibalized from my *Jim Blinn's Corner* articles
Polynomial Discriminants, Part 1 and Part 2
IEEE Computer Graphics and Applications
Nov/Dec 2000 and Jan/Feb 2001
It reiterates some of what is in the previous chapters.

I like beautiful equations. But beauty is sometimes subtle, or hidden by bad notation. In my next few columns I am going to reveal some of the hidden beauty in the explicit formulation of the discriminants of polynomials. Along the way I will drag in some clever algebra, promote some notational schemes from mathematical physics, and illustrate some ways of visualizing homogeneous space. This will ultimately lead us to some interesting ways to find roots of these polynomials, a task that will become more and more important as we computer graphicists struggle to break free of the tyranny of the polygon and move into rendering higher order surfaces.

So first, let's review discriminants.

Discriminants

I'll to soften you up a bit by starting with something already familiar: the quadratic equation.

Quadratics

A general quadratic equation is

$$ax^2 + bx + c = 0$$

We learn in high school that the solution of this equation is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

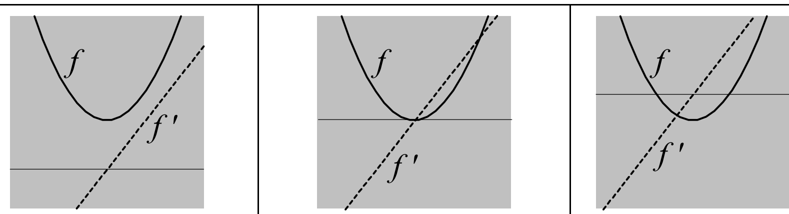
The discriminant is the value under the square root sign.

$$\boxed{\Delta = b^2 - 4ac}$$

The polynomial will have 0, 1, or 2 real roots depending on the sign of the discriminant. If it is negative, there are no real roots, if positive there are two distinct real roots, and if the discriminant is zero there is a double root (i.e. two coincident real roots). (See Figure 1). A more generalizable way to way to derive the discriminant is to note that it is zero if there is some parameter value where both the function and its derivative are zero. In other words we want to simultaneously solve:

$$ax^2 + bx + c = 0$$
$$2ax + b = 0$$

To find if this is possible for a given quadratic, just solve the derivative equation for x and plug it into the quadratic. The result is an expression in *abc* that simplifies to the above discriminant.



$\Delta_2 < 0$ | $\Delta_2 = 0$ | $\Delta_2 > 0$

Figure 1. Relation of Δ_2 to roots and derivative of quadratic polynomial *f*

Cubics

Stepping up to cubic equations, we have:

$$ax^3 + bx^2 + cx + d = 0$$

As with quadratics, we can find the discriminant by equating the function and its derivative to zero

$$\begin{aligned} ax^3 + bx^2 + cx + d &= 0 \\ 3ax^2 + 2bx + c &= 0 \end{aligned} \tag{0.1}$$

We can find the discriminant --- the condition on *abcd* that makes this possible--- by various methods. We could solve the quadratic for *x* and substitute that into the cubic equation. A more general technique is to form the, so-called, resultant of the two polynomials. This basically involves taking various linear combinations of them to form new polynomials of lower degree. I won't go into details here, but for cubics this process leads to something called the Sylvester determinant [ref 1,2,3]

$$\Delta = \frac{1}{a}\det\begin{bmatrix} a & 0 & 3a & 0 & 0 \\ b & a & 2b & 3a & 0 \\ c & b & c & 2b & 3a \\ d & c & 0 & c & 2b \\ 0 & d & 0 & 0 & c \end{bmatrix}$$

A similar technique is Bezout's method. [ref 3] This also takes various linear combinations to construct a smaller but more complicated matrix. We ultimately get

$$\Delta = \frac{1}{a}\det\begin{bmatrix} ab & 2ac & 3a \\ 2ac & 3ad + bc & 2b \\ 3ad & 2bd & c \end{bmatrix}$$

Of course we could take the easy way out and just look it up on the web [ref 1]. The result is

$$\begin{aligned} \Delta = c^2b^2 - 4db^3 - 4c^3a \\ +18abcd - 27d^2a^2 \end{aligned} \tag{0.2}$$

This ungainly mess is rather harder to remember than the quadratic discriminant. But it is useful. As with quadratics, it's value (or rather the square root of its value) figures prominently in the solution of the polynomial. If it's negative the cubic has exactly one real root; if positive there are three distinct real roots. And if it's zero the cubic has a double root and another single root, or possibly a triple root (see Figure 2)

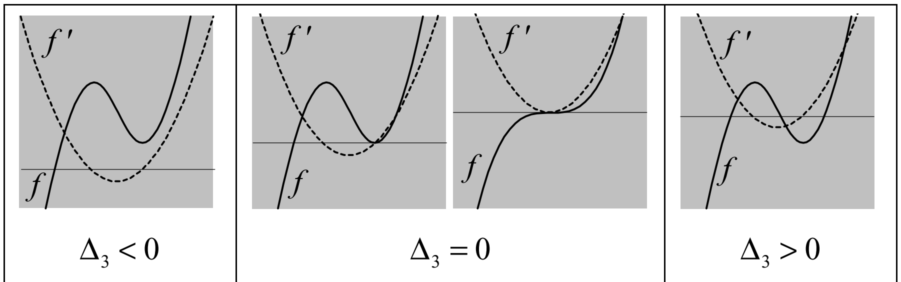


Figure 2. Relation of Δ_3 to roots and derivative of cubic polynomial *f*

Quartics

If you think that's bad, take a look at quartic polynomials. The equation is

$$ax^4 + bx^3 + cx^2 + dx + e = 0$$

Following the procedure outlined above we surf over to:

http://www.inwap.com/pdp10/hbaker/hakmemgeometry.html

http://mathworld.wolfram.com/DiscriminantPolynomial.html

we arrive at the truly stunning

$$\begin{aligned}\Delta_4 = &-27a^2d^4 + 18abcd^3 - 4b^3d^3 - 4ac^3d^2 \\ &+ b^2c^2d^2 + 144a^2cd^2e - 6ab^2d^2e - 80abc^2de \\ &+ 18b^3cde + 16ac^4e - 4b^2c^3e - 192a^2bde^2 \\ &- 128a^2c^2e^2 + 144ab^2ce^2 - 27b^4e^2 + 256a^3e^3\end{aligned} \tag{0.3}$$

Aesthetics

These discriminants look really ugly in their explicit form. But there is an interesting pattern embedded in them. Finding that pattern is our mathematical journey.

A Homogeneous Matrix Formulation

My first urge in any algebraic discussion is to write things in homogeneous form, in this case as homogeneous polynomials. This generalizes the parameter value from the simple quantity *x* to the homogeneous pair [*x w*]. The homogeneous quadratic equation is

$$ax^2 + bxw + cw^2 = 0$$

The main thing that homogeneity brings to the party is the addition of a new "parameter at infinity" at the value $\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}$. This means that if the parameter *a* is zero, the quadratic does not simply degenerate into a linear equation. Instead, it remains a quadratic, but it simply has one of its roots at infinity (w=0)

Next I want to indulge an even stronger algebraic urge: to write things in matrix form. To make this a bit neater I will first modify the notation for the coefficients to build in some constant factors. I'll write the quadratic equation as

$$Ax^2 + 2Bxw + Cw^2 = 0$$

This allows us to write the quadratic equation as a symmetric matrix product

$$\begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix} = 0 \tag{0.4}$$

(This way of representing a quadratic is related to a technique known as blossoming.) The solutions now become

$$\begin{aligned} x &= \frac{-2B \pm \sqrt{4B^2 - 4AC}}{2A} \\ &= \frac{-B \pm \sqrt{B^2 - AC}}{A} \end{aligned}$$

And the discriminant is $B^2 - AC$, which we can recognize as minus the determinant of the coefficient matrix

$$\Delta_2 = -\det\begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

Neat. We've expressed the formula for the discriminant in terms of a common matrix operation: the determinant.

Bumping up to cubics, I will again rename the coefficients

$$Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3 = 0 \tag{0.5}$$

To make a matrix representation of this analogous to equation (0.4) we want to arrange these coefficients into a 2x2x2 symmetric "cube" of numbers. There are various ways to show this but they are all a bit clunky. About the best you can do with conventional matrix notation is to think of the coefficients as a vector of 2x2 matrices. Equation (0.5) becomes

$$\begin{bmatrix} x & w \end{bmatrix}\left\{\begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}\right\}\begin{bmatrix} x \\ w \end{bmatrix} = 0 \tag{0.6}$$

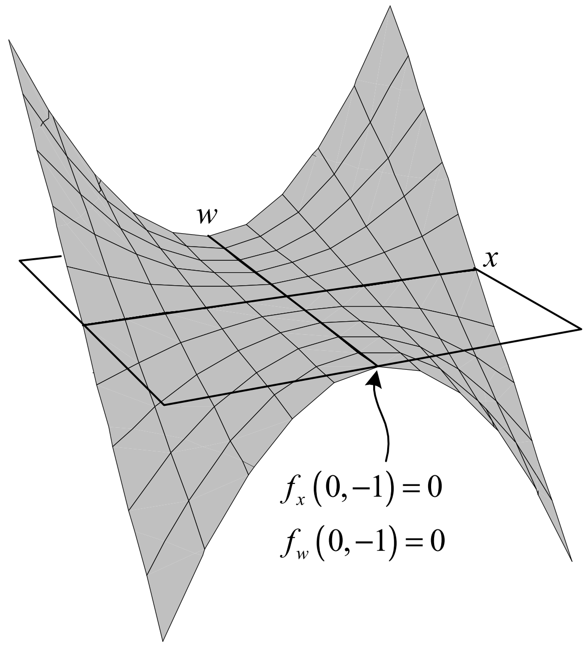
Once we have this triply indexed cube of numbers, we expect that the discriminant will be some sort of cubical generalization of the determinant. Let's find out what it is.

A Kinder, Gentler Cubic Discriminant

The defining property of the discriminant is that it is the condition that there is a parameter value where both the function and its derivative are zero. For a homogeneous cubic, we want the condition on (*ABCD*) that allows simultaneous solution of

$$\begin{aligned} f(x,w) &= Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3 = 0 \\ f_x(x,w) &= 3Ax^2 + 6Bxw + 3Cw^2 = 0 \\ f_w(x,w) &= 3Bx^2 + 6Cxw + 3Dw^2 = 0 \end{aligned} \tag{0.7}$$

To visualize what this means, note that having both the partial derivatives of *f* be zero means that the *f* function is tangent to the *f*=0 plane at that point. That is, there is a double root there See Figure 3,



Comparing this with the formulas in (0.1) it looks at first that going to homogeneous polynomials gives us an extra equation. But it really doesn't. That's because of the identity

$$3f = xf_x + wf_w$$

So the new derivative is just a linear combination of the function and the original x derivative. This means that we can pick any two equations out of (0.7) to work with further. I don't know about you, but the two I'm going to pick are the two lower order ones. Tossing out a constant factor of 3 we get the following two equations that we want to solve simultaneously

$$\begin{aligned} Ax^2 + 2Bxw + Cw^2 &= 0 \\ Bx^2 + 2Cxw + Dw^2 &= 0 \end{aligned} \tag{0.8}$$

Linear combinations helped us before; let's see what else they can do for us. If the two equations above are both zero, then any linear combination of them is zero too. We can get an equation without an x^2 term by:

$$\begin{aligned} &B\left(Ax^2 + 2Bxw + Cw^2\right) \\ -&A\left(Bx^2 + 2Cxw + Dw^2\right) = \\ &\quad 2\left(B^2 - AC\right)xw + \left(BC - AD\right)w^2 = 0 \end{aligned}$$

And we can symmetrically get an equation without the w^2 term by:

$$\begin{aligned} &D\left(Ax^2 + 2Bxw + Cw^2\right) \\ -&C\left(Bx^2 + 2Cxw + Dw^2\right) = \\ &\quad \left(AD - BC\right)x^2 + 2\left(BD - C^2\right)xw = 0 \end{aligned}$$

Tossing out common factors of x and w we see that we have knocked the simultaneous quadratics in equation (0.8) down to two simultaneous linear equations.

$$\begin{aligned}2\left(B^2-AC\right)x+\left(BC-AD\right)w=0\\\left(AD-BC\right)x+2\left(BD-C^2\right)w=0\end{aligned}\tag{0.9}$$

What we are saying is that if equations (0.9) can be satisfied, then equations (0.8) can be also. Each of the equations in (0.9) is easy to solve. The condition that the two solutions be equal leads us to

$$\frac{x}{w}=\frac{AD-BC}{2\left(B^2-AC\right)}=\frac{-2\left(BD-C^2\right)}{AD-BC}$$

The final expression for the discriminant is then

$$\Delta_3=4\left(C^2-BD\right)\left(B^2-AC\right)-\left(AD-BC\right)^2\tag{0.10}$$

If we multiplied this out and did the conversion from *ABCD* to *abcd* we would get equation (0.2), but equation (0.10) is certainly a lot prettier. But wait, it gets better.

With some imagination we can recognize that the various parenthesized quantities in equation (0.10) are (with a few sign flips that cancel each other out) just the determinants of various slices of the cube of coefficients. Let's give them names

$$\delta_1=AC-B^2=\det\begin{bmatrix}A&B\\B&C\end{bmatrix}$$

$$\delta_2=AD-BC=\det\begin{bmatrix}A&B\\C&D\end{bmatrix}$$

$$\delta_3=BD-C^2=\det\begin{bmatrix}B&C\\C&D\end{bmatrix}$$

We can now write the cubic resultant as

$$\Delta_3=4\delta_1\delta_3-\delta_2{}^2$$

I love symmetry in algebra. You will note that any time we come up with something that is the difference of two products I have an irresistible urge to write it as the determinant of a 2x2 matrix. And if it is the difference between the square of something and another product, I want to make it the determinant of a symmetric matrix. Satisfying this urge one more time gives

$$\Delta_3=-\det\begin{bmatrix}2\delta_1&\delta_2\\\delta_2&2\delta_3\end{bmatrix}\tag{0.11}$$

In other words, the cubic discriminant is a determinant of determinants. If this discriminant is zero we know that there is a double root. And we know what it is. In fact we have two formulations of it. Rewriting equation(0.9) in terms of the δ_i we have

$$\frac{x}{w}=\frac{\delta_2}{-2\delta_1}=\frac{-2\delta_3}{\delta_2}$$

Or, in more homogeneous terms the two equivalent formulations become

$$\begin{bmatrix}x&w\end{bmatrix}=\begin{bmatrix}\delta_2&-2\delta_1\end{bmatrix}$$

or

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} 2\delta_3 & -\delta_2 \end{bmatrix}$$

It's useful to have a choice here. If two of the δ_i's are zero, at least one of the two choices still generates a meaningful root.

Back To Our Roots

There is one more useful piece of information hidden here. To find it I'll expand the polynomial in terms of its roots. It will also be a little less cluttered if we go back to the non-homogeneous version.

$$\begin{aligned} f(x) &= (x - r_1)(x - r_2)^2 \\ &= x^3 + (-2r_2 - r_1)x^2 + \left(r_2^{\ 2} + 2r_1 r_2\right)x + \left(-r_2^{\ 2} r_1\right) \end{aligned}$$

The coefficients in our notation scheme are

$$\begin{aligned} A &= 1 \qquad & B &= \frac{-2r_2 - r_1}{3} \\ C &= \frac{r_2^{\ 2} + 2r_1 r_2}{3} \qquad & D &= -r_2^{\ 2} r_1 \end{aligned} \tag{0.12}$$

Plugging these into the definitions of the Δ_i and doing some simplifying we get:

$$\begin{aligned} \delta_1 &= -\frac{(r_2 - r_1)^2}{9} \\ \delta_2 &= \frac{2r_2 (r_2 - r_1)^2}{9} \\ \delta_3 &= \frac{-r_2^{\ 2} (r_2 - r_1)^2}{9} \end{aligned} \tag{0.13}$$

This means that if there is a *triple* root, where $r_1 = r_2$, then not only is $\Delta_3 = 0$ but all three components are zero $\delta_1 = \delta_2 = \delta_3 = 0$ From the definitions of the δ_i and equation (0.12) this means that:

$$\frac{B}{A} = \frac{C}{B} = \frac{D}{C} = -r_1$$

Or, in homogeneous terms, the triple root has the three equivalent formulations:

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} -B & A \end{bmatrix} \text{ or } \begin{bmatrix} -C & B \end{bmatrix} \text{ or } \begin{bmatrix} -D & C \end{bmatrix}$$

Again, the alternative formulations are useful. For example if $A = 0$ (implying $B = C = 0$) or if $D = 0$ (implying $B = C = 0$) at least one of the choices generates a meaningful root.

So, I'll summarize.

- For a quadratic, write the coefficients as a symmetric 2x2 matrix. The discriminant is minus the determinant of the matrix. If this is zero, the quadratic has a double root.
- For a cubic, write the coefficients as a symmetric 2x2x2 cube of numbers. Calculate the three sub-determinants $\delta_1, \delta_2, \delta_3$. If all three are zero, the polynomial has a triple root. Otherwise, the

discriminant is the determinant from equation (0.11). If this is zero, then the cubic polynomial has a double root (and an additional single root).

Quartics

Emboldened by this success let's go for broke and see what we can do with quartics. The homogeneous polynomial is

$$q(x,w) = Ax^4 + 4Bx^3w + 6Cx^2w^2 + 4Dxw^3 + Ew^4$$

We can think of this as a 2x2x2x2 hypercube of coefficients. The best we can do with matrix notation is as a 2x2 matrix of 2x2 matrices:

$$q(x,w) = \begin{bmatrix} x & w \end{bmatrix} \left\{ \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \\ \begin{bmatrix} B & C \\ C & D \end{bmatrix} & \begin{bmatrix} C & D \\ D & E \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \right\} \begin{bmatrix} x \\ w \end{bmatrix}$$

Transforming the big fat equation (0.3) to our new coefficient names gives

$$\begin{aligned} \Delta_4 = {} & A^3E^3 - 12A^2BDE^2 - 18A^2C^2E^2 + 54A^2CD^2E \\ & -27A^2D^4 + 54AB^2CE^2 - 6AB^2D^2E - 180ABC^2DE \\ & +108ABCD^3 - 54AC^3D^2 + 81AC^4E - 27B^4E^2 \\ & +108B^3CDE - 64B^3D^3 - 54B^2C^3E + 36B^2C^2D^2 \end{aligned} \tag{0.14}$$

We want to see if there is a prettier way to write this. Let's apply the same technique we used for the cubic; start with the desire to simultaneously solve for the two partial derivatives being zero.

$$\begin{aligned} q_x &= Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3 = 0 \\ q_w &= Bx^3 + 3Cx^2w + 3Dxw^2 + Ew^3 = 0 \end{aligned} \tag{0.15}$$

Now we go through our process of successively knocking down the degree by linear combinations. We form

$$\begin{aligned} & Aq_w - Bq_x = \\ & \quad 3\left(AC - B^2\right)x^2w + 3\left(AD - BC\right)xw^2 + \left(AE - BD\right)w^3 \end{aligned}$$

and

$$\begin{aligned} & Eq_x - Dq_w = \\ & \quad \left(AE - BD\right)x^3 + 3\left(BE - CD\right)x^2w + 3\left(CE - D^2\right)xw^2 \end{aligned}$$

Again, it's beneficial to give names to the parenthesized expressions above. By extension to the terminology for cubics I will define:

$$\begin{aligned} \delta_1 &= AC - B^2 & \delta_4 &= BE - CD \\ \delta_2 &= AD - BC & \delta_5 &= CE - D^2 \\ \delta_3 &= BD - C^2 & \delta_0 &= AE - BD \end{aligned} \tag{0.16}$$

So, in these terms, our linear combo trick has resulted in the two simultaneous equations

$$3\delta_1 x^2 + 3\delta_2 xw + \delta_0 w^2 = 0$$
$$\delta_0 x^2 + 3\delta_4 xw + 3\delta_5 w^2 = 0$$

Now we keep turning the crank. We reduce the above quadratics to linears by taking one linear combination to shave the x^2 term off one end and another combination to shave the w^2 term off the other end. This ultimately results in

$$\left(3\delta_0\delta_2 - 9\delta_1\delta_4\right)x + \left(\delta_0\delta_0 - 9\delta_1\delta_5\right)w = 0$$
$$\left(9\delta_5\delta_1 - \delta_0\delta_0\right)x + \left(9\delta_5\delta_2 - 3\delta_0\delta_4\right)w = 0$$

Again, mimicking our actions for the cubic, these two linear equations will have a common root if

$$\left(9\delta_1\delta_4 - 3\delta_0\delta_2\right)\left(9\delta_5\delta_2 - 3\delta_0\delta_4\right) - \left(9\delta_5\delta_1 - \delta_0\delta_0\right)^2 = 0 \tag{0.17}$$

The urge to write this as a matrix takes hold and we get:

$$\Delta_4 = \det\begin{bmatrix} \det\begin{bmatrix} 3\delta_1 & 3\delta_2 \\ \delta_0 & 3\delta_4 \end{bmatrix} & \det\begin{bmatrix} 3\delta_1 & \delta_0 \\ \delta_0 & 3\delta_5 \end{bmatrix} \\ \det\begin{bmatrix} 3\delta_1 & \delta_0 \\ \delta_0 & 3\delta_5 \end{bmatrix} & \det\begin{bmatrix} 3\delta_2 & \delta_0 \\ 3\delta_4 & 3\delta_5 \end{bmatrix} \end{bmatrix}$$

We might be tempted, then, to say that the quartic discriminant is a determinant of determinants of determinants. But that can't be right. Notice that equation (0.17) is eighth order in *ABCDE* while the correct version (0.14) is sixth order. What happened?

Watch me pull a rabbit out of my hat. Behold the identity:

$$\delta_1\delta_5 - \delta_2\delta_4 + \delta_0\delta_3 = 0 \tag{0.18}$$

If you like, you can convince yourself of this by plugging in the definitions of the δ_i's from equation (0.16), but how did I know to try this? Well, there's an interesting parallel between the arithmetic we did in defining the six δ_i values and the arithmetic involved in constructing the six components of a 3DH line from two 3DH points. I described this in some detail in {J. Blinn, *A Homogeneous Formulation for Lines in 3 space*, Siggraph 77, pp237-241} where I showed that the six values generated by equation (0.16) must always satisfy equation (0.18). Now lets use it.

If we multiply out equation (0.17) and do a little obvious factoring, we get

$$81\delta_1\delta_5\left(\delta_2\delta_4 - \delta_1\delta_5\right) + \delta_0(\text{buncha stuff})$$

Now apply the identity (0.18) and we get

$$81\delta_1\delta_5\left(\delta_0\delta_3\right) + \delta_0(\text{buncha stuff})$$

We now can factor δ_0 out of this to give the correct quartic discriminant

$$\Delta_4 = 81\delta_1\delta_3\delta_5 - 27\delta_1\delta_4{}^2 - 27\delta_2{}^2\delta_5$$
$$+9\delta_0\delta_2\delta_4 + 18\delta_0\delta_1\delta_5 - \delta_0{}^3$$

Nice but still not nice enough. Watch closely—my fingers never leave the keyboard. By applying the identity again we can turn this into

$$\Delta_4 = 81\delta_1\delta_3\delta_5 + 9\delta_0\delta_1\delta_5 + 18\delta_0\delta_2\delta_4$$
$$-27\delta_2{}^2\delta_5 - 27\delta_1\delta_4{}^2 - 9\delta_0{}^2\delta_3 - \delta_0{}^3$$

Which, as anyone can plainly see is

$$\Delta_4 = \det\begin{bmatrix} 3\delta_1 & 3\delta_2 & \delta_0 \\ 3\delta_2 & 9\delta_3 + \delta_0 & 3\delta_4 \\ \delta_0 & 3\delta_4 & 3\delta_5 \end{bmatrix}$$

This is just an application of the formula for the resultant of two cubics given in Kajiya [ref 4]. The two cubics in question are the two derivatives of equation (0.15)

This is pretty, but not pretty enough. There is another representation of the discriminant of a quartic that's even better. It's buried in some hundred-year-old lectures by Hilbert, reprinted recently in {D. Hilbert, *Theory of Algebraic Invariants,* Cambridge University Press, 1993, pages 72,74}. Hilbert defined two quantities that, translated into our terminology, are:

$$I_2 = AE - 4BD + 3C^2$$
$$I_3 = ACE - AD^2 - B^2E + 2BCD - C^3$$

Then the quartic discriminant happens to be

$$\Delta_4 = 27(I_3)^2 - (I_2)^3 \tag{0.19}$$

You can verify this for yourself by simple substitution. I won't wait…

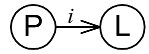
Behind The Curtain

The expression for the discriminant in equation (0.11) is a lot prettier than the one in equation (0.2). And the expression in Equation (0.19) is a lot prettier than Equation (0.3). But we will see that expressing them using Tensor Diagrams is even more beautiful.

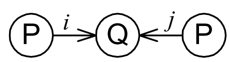
2D Homogeneous Geometry

Two-dimensional homogeneous geometry uses three element vectors, 3x3 matrices, 3x3x3 tensors etc, to represent various objects. I will denote such quantities in upper case boldface to distinguish them from polynomials discussed later. For example, a homogeneous point **P** is a three-element row vector, and a line **L** is a three-element column vector. The point lies on the line if the dot product $\mathbf{P}\cdot\mathbf{L}$ is zero. Table 1 shows different ways of expressing the dot product. Parts a, b, and c should be very familiar. Part d is the aforementioned Einstein Index Notation. Part e is the Tensor Diagram version. I will review their meaning below.

Table 1. Point on a Line	
a	$ax + by + cw = 0$
b	$\begin{bmatrix} x & y & w \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$
c	$\mathbf{P}\cdot\mathbf{L} = 0$

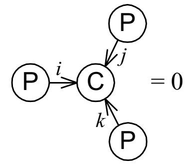
d	$P^i L_i = 0$
e	 $= 0$

Moving up to curves, the points on a second order (quadratic) curve satisfy the equation in Table 2(a). We can write this in matrix form by arranging the coefficients into the 3x3 symmetric matrix of Table 2(b).

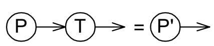
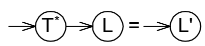
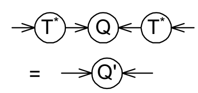
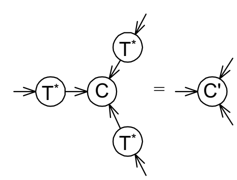
Table 2. Point on a Quadratic Curve	
a	$Ax^2 + 2Bxy + 2Cxw + Dy^2 + 2Eyw + Fw^2 = 0$
b	$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} A & B & C \\ B & D & E \\ C & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = 0$
c	$\mathbf{P} \cdot \mathbf{Q} \cdot \mathbf{P}^T = 0$
d	$P^i Q_{ij} P^j = 0$
e	 $= 0$

Next up, the points on a third order (cubic) curve satisfy Table 3(a). We can also write this by arranging the coefficients into a 3x3x3 symmetric generalization of a matrix. Doing this with conventional matrix notation is a bit weird. About the best we can do is to show it as a vector of matrices as in Table 3(b).

Table 3. Point on a Cubic curve	
a	$Ax^3 + 3Bx^2y + 3Cxy^2 + Dy^3 + 3Ex^2w + 6Fxyw + 3Gy^2w + 3Hxw^2 + 3Jyw^2 + Kw^3 = 0$
b	$\begin{bmatrix} x & y & w \end{bmatrix} \left\{ \begin{bmatrix} \begin{bmatrix} A & B & E \\ B & C & F \\ E & F & H \end{bmatrix} & \begin{bmatrix} B & C & F \\ C & D & G \\ F & G & J \end{bmatrix} & \begin{bmatrix} E & F & H \\ F & G & J \\ H & J & K \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \right\} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = 0$
c	???
d	$P^i P^j P^k C_{ijk} = 0$

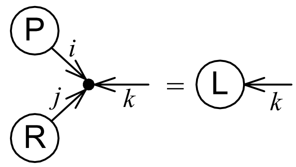
e	

Now let's talk about transformations. We geometrically transform points by post-multiplying by a 3x3 matrix: $\mathbf{P\,T} = \mathbf{P}'$, and we transform lines by pre-multiplying by the adjoint of the matrix: $\mathbf{T}^*\mathbf{L} = \mathbf{L}'$. Table 4 shows various ways to write these expressions, as well as those for transforming curves.

Table 4. Transformations				
	Point	Line	Quadratic	Cubic
b	$[x\ \ y\ \ w]\begin{bmatrix} i & j & k \\ l & m & n \\ o & p & q \end{bmatrix} = [x'\ \ y'\ \ w']$	$\begin{bmatrix} i^* & j^* & k^* \\ l^* & m^* & n^* \\ o^* & p^* & q^* \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} A' \\ B' \\ C' \end{bmatrix}$	$\begin{bmatrix} i^* & j^* & k^* \\ l^* & m^* & n^* \\ o^* & p^* & q^* \end{bmatrix}\begin{bmatrix} A & B & C \\ B & D & E \\ C & E & F \end{bmatrix}\begin{bmatrix} i^* & l^* & o^* \\ j^* & m^* & p^* \\ k^* & n^* & q^* \end{bmatrix} = \begin{bmatrix} A' & B' & C' \\ B' & D' & E' \\ C' & E' & F' \end{bmatrix}$	messy
c	$\mathbf{P\,T} = \mathbf{P}'$	$\left(\mathbf{T}^*\right)\mathbf{L} = \mathbf{L}'$	$\left(\mathbf{T}^*\right)\mathbf{Q}\left(\mathbf{T}^*\right)^T = \mathbf{Q}'$	messy
d	$P^i T_i^j = \left(P'\right)^j$	$\left(T^*\right)^i_j L_i = \left(L'\right)_j$	$\left(T^*\right)^i_k Q_{ij} \left(T^*\right)^j_l = \left(Q'\right)_{kl}$	$\left(T^*\right)^i_l \left(T^*\right)^j_m \left(T^*\right)^k_n C_{ijk} = \left(C'\right)_{lmn}$
e				

Finally, the cross product of two point-vectors **P** and **R** gives the line passing through them: $\mathbf{P}\times\mathbf{R} = \mathbf{L}$. See table 5. In a dual fashion, the cross product of two line-vectors **L** and **M** gives their point of intersection: $\mathbf{L}\times\mathbf{M} = \mathbf{P}$.

Table 5. The cross product	
a	$\begin{bmatrix} P^1R^2 - P^2R^1 \\ P^2R^0 - P^0R^2 \\ P^0R^1 - P^1R^0 \end{bmatrix} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \end{bmatrix}$
b	$\begin{bmatrix} P^0 & P^1 & P^2 \end{bmatrix}\times\begin{bmatrix} R^0 & R^1 & R^2 \end{bmatrix} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \end{bmatrix}$
c	$\mathbf{P}\times\mathbf{R} = \mathbf{L}$

e	$P^i R^j \varepsilon_{ijk} = L_k$
e	

The problem

In looking over these expressions we see that there are two hints that our notation has problems. The first is the need to take the transpose of **P** when multiplying by **Q.** This is very fishy; column matrices are supposed to represent lines, not points. In fact, there is something fundamentally different about matrices that represent transformations and matrices that represent quadratic curves. We cannot, however, distinguish between them with standard vector notation. The second problem is inability to conveniently represent entities with more than two indices. Our attempt to arrange the coefficients of a cubic polynomial into a triply indexed "cubical matrix" is an example of the problem.

Fortunately, there are two notational schemes that we can adapt from the world of theoretical physics to alleviate these shortcomings. They are "Einstein Index Notation" (EIN) and the diagram notation I referred to in my opening monologue. I originally called these "Feynman Diagrams" but there are enough differences to give them the more appropriate name "Tensor Diagrams". They are more like the diagrams in [Kuperberg, Greg, *Involutory Hopf algebras and 3-manifold invariants* International Journal of Mathematics, Vol 2, no. 1, (1991) pp 41-66, World Scientific Publishing Company.]

2DH Tensor Diagrams

Einstein index notation differentiates between two types of indices for vector/matrix elements: the point-like ones (which we will call contravariant and write as superscripts) and line-like ones (which we will call covariant and write as subscripts). Thus an element of a point-vector is P^i and an element of a line-vector is L_i. (Note that superscript indices are not the same as exponents. Mathematicians ran out of places to put indices and started overloading their notation. Live with it.). These things are easy to get mixed up and backwards so, for reference, I'll post the following reference diagram:

$$\mathbf{T}^{\text{CONTRA(point like)}}_{\text{CO(line like)}}$$

Dot products happen *only* between matching pairs of covariant and contravariant indices. Thus the dot of a point and a line is

$$\begin{aligned}\mathbf{P}\cdot\mathbf{L} &= \begin{bmatrix} P^1 & P^2 & P^3 \end{bmatrix}\cdot\begin{bmatrix} L_1 & L_2 & L_3 \end{bmatrix} \\ &= \sum_i P^i L_i\end{aligned}$$

We simplify further by omitting the sigma and stating that any super/subscript pair that has the same letter implicitly implies a summation over that letter. The EIN form of a dot product is then simply

$$P^i L_i \,.$$

A more complicated expression may have many tensors and superscripts and subscripts, and will implicitly be summed over all pairs of identical upper/lower indices. (These summations are also called tensor contractions). We can see this in the EIN for higher order curves in Table 2(d) and Table 3(c). Note that the expression for EIN is basically a model for the terms that are summed. Each individual factor in the notation is just a number, so the factors can be rearranged in any order, as was done in Table 3(c).

Tensor diagram notation is a translation of EIN into a graph. We represent a point as a node with an outward arrow indicating a contravariant index. A line, with its covariant index, is a node with an inward arrow. The dot product, i.e. the summation over the covariant/contravariant pair, is an arc connecting two nodes. See the bottom rows of Tables 1 through 3 for the diagram notation of the expressions we have seen so far. For many of the diagrams I will label the arcs with the index they correspond to in EIN. Some later, more complex, diagrams will not need this.

Transformations

A transformation matrix has one contravariant and one covariant index. Multiplying a point by such a matrix will "annihilate" its covariant index leaving a result that has a free contravariant index, making the result be a point. Table 4(d) shows the EIN form of the transformation of various quantities. Row (e) of the table shows how this translates into diagram notation. Now we can see the difference between the two types of matrices. A transformation matrix has one of each type of index (denoted with one arrow out and one arrow in); a quadratic matrix has two covariant indices (denoted with both arrows in). In Table 2(d) the two covariant/contravariant index pairs annihilate each other to produce a scalar.

Cross Products and Adjoints

We abbreviate the algebra for cross products and matrix adjoints by defining a three-index 3x3x3 element anti-symmetric tensor called the Levi-Civita epsilon. The elements of epsilon are defined to be

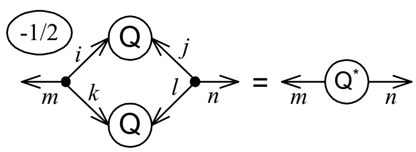
$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} &= +1 \\ \varepsilon_{321} = \varepsilon_{132} = \varepsilon_{213} &= -1 \\ \varepsilon_{ijk} &= 0 \quad \text{otherwise} \end{aligned} \tag{0.20}$$

Multiplying two vectors by epsilon forms their cross product. Since epsilon has three subscript indices, multiplying in two points with superscript indices will result in a vector with one remaining subscript index (a line). The diagram form of epsilon is a node with three inward pointing arcs. We will show this node as a small dot as in Table 5(d). You can imagine a similar table for the dual form, the cross product of two lines: $\mathbf{L} \times \mathbf{M} = \mathbf{P}$. Just use a contravariant form of epsilon, ε^{ijk}, so that $L_i M_j \varepsilon^{ijk} = P^k$ and flip the direction of all arrows in the diagram.

We must be careful about how the anti-symmetry of epsilon translates into a diagram. The convention is to label the arcs counterclockwise around the dot. A mirror reflection of an epsilon diagram will reverse the order of its indices, and therefore flip its algebraic sign.

Epsilon is also useful to form matrix adjoints. Table 6 shows various ways to denote the adjoint. The raw EIN expression $Q_{ij} Q_{kl} \varepsilon^{ikm} \varepsilon^{jln}$ gives twice the adjoint, so I had to insert a factor of ½ to get the correct answer. I've chosen to mirror the diagram for the first epsilon in the EIN (and introduce a minus sign) to make the whole diagram a bit prettier. These factors and signs clutter things up a bit but are necessary to get right.

Table 6. The adjoint of a 3x3 matrix	
a	$DF - E^2 = A^*$ $CE - BF = B^*$ $\vdots$

b	$\begin{bmatrix} \det\begin{bmatrix} D & E \\ E & F \end{bmatrix} & -\det\begin{bmatrix} B & E \\ C & F \end{bmatrix} & \det\begin{bmatrix} B & D \\ C & E \end{bmatrix} \\ -\det\begin{bmatrix} B & E \\ C & F \end{bmatrix} & \det\begin{bmatrix} A & C \\ C & F \end{bmatrix} & -\det\begin{bmatrix} A & B \\ C & E \end{bmatrix} \\ \det\begin{bmatrix} B & D \\ C & E \end{bmatrix} & -\det\begin{bmatrix} A & B \\ C & E \end{bmatrix} & \det\begin{bmatrix} A & B \\ B & D \end{bmatrix} \end{bmatrix} =$ $\begin{bmatrix} [B \quad D \quad E] & [C \quad E \quad F] & [A \quad B \quad C] \\ \times \quad , & \times \quad , & \times \\ [C \quad E \quad F] & [A \quad B \quad C] & [B \quad D \quad E] \end{bmatrix} = \begin{bmatrix} A^* & B^* & C^* \\ B^* & D^* & E^* \\ C^* & E^* & F^* \end{bmatrix}$
c	$\operatorname{adj}\mathbf{Q} = \mathbf{Q}^*$
d	$\frac{1}{2} Q_{ij} Q_{kl} \varepsilon^{ikm} \varepsilon^{jln} = \left(Q^*\right)^{mn}$
e	

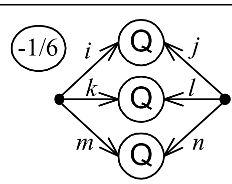
Now that we have the adjoint, the determinant is not far behind. We use the fact that

$$\mathbf{Q}^*\mathbf{Q} = \left(\det \mathbf{Q}\right)\mathbf{I}$$

We tie up the loose ends, literally, by taking the trace of this getting:

$$\operatorname{trace}\left(\mathbf{Q}^*\mathbf{Q}\right) = 3\det \mathbf{Q}$$

So in diagram terms, connect the adjoint from Table6(e) to another copy of **Q** and take the trace by connecting the two dangling arcs. Divide by 3 to get the determinant. The resulting diagram is in Table 7(e).

Table 7. Determinant of 3x3 matrix	
a	$ABD + 2BCE - C^2D - E^2A - B^2F = \det \mathbf{Q}$
b	…
c	$\det \mathbf{Q}$
d	$\frac{1}{6} Q_{ij} Q_{kl} Q_{mn} \varepsilon^{ikm} \varepsilon^{jln}$
e	

Homogeneous Polynomials

Now let's go down a dimension and take a look at one-dimensional homogeneous geometry. This is effectively the study of homogeneous polynomials. Basically we have the same thing as before, but everything is now composed of 2 element vectors, 2x2 matrices and 2x2x2 tensors, which I'll write as lower case boldface. A homogeneous linear equation is written in various notations in Table 8.

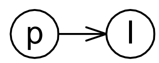
Table 8. Homogeneous Linear Equation	
a	$Ax + Bw = 0$
b	$\begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = 0$
c	$\mathbf{p} \cdot \mathbf{l} = 0$
d	$p^i l_i = 0$
e	 $= 0$

Table 9 shows a homogeneous quadratic equation

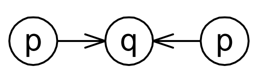
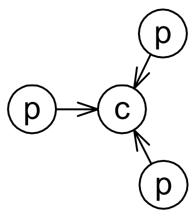
Table 9. Homogeneous Quadratic Equation	
a	$Ax^2 + 2Bxw + Cw^2 = 0$
b	$\begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = 0$
c	$\mathbf{pqp}^T = 0$
d	$p^i p^j q_{ij} = 0$
e	 $= 0$

Table 10 shows a homogeneous cubic equation. (Unfortunately I find that I have to use the letter C in two contexts, once as a coefficient and once as a tensor name. Live with it.)

Table 10. Homogeneous Cubic Equation	
a	$Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3 = 0$

b	$\begin{bmatrix} x & w \end{bmatrix} \left\{ \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \\ \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix} \right\} \begin{bmatrix} x \\ w \end{bmatrix} = 0$
c	messy
d	$p^i p^j p^k c_{ijk} = 0$
e	 $= 0$

The 2D Epsilon

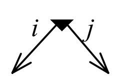
The only slightly subtle item is the form of the 2-element epsilon. Instead of having three indices, each with three values, the 2-element epsilon has two indices (making it a simple matrix) each with two values. By analogy to equation (0.20) the contravariant form of epsilon is

$$\varepsilon^{12} = +1$$
$$\varepsilon^{21} = -1$$
$$\varepsilon^{ij} = 0 \quad \text{otherwise}$$

In other words

$$\varepsilon = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

The Einstein notation is simply ε_{ij} or ε^{ij} and the diagram notation looks like

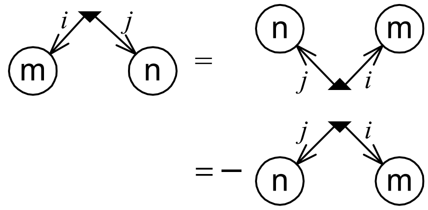


I have purposely constructed this icon to be asymmetrical. The convention is that when the diagram points down (as above) the first index is on the left. A mirror reflection of this diagram will perform a sign flip on the value of the diagram. If the diagram were not asymmetrical, a mirror flip would not be detectable.

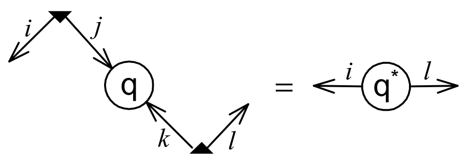
To drive this home, compare the EIN expressions

$$m^i \varepsilon_{ij} n^j = n^j \varepsilon_{ij} m^i = -n^j \varepsilon_{ji} m^i$$

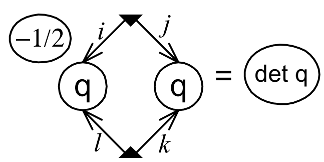
with their diagram counterparts. The first equality represents a rotation, second has a reflection.



Now let's use the epsilon. The adjoint of a 2x2 matrix, by analogy to Table 6, gives us Table 11

Table 11. The adjoint of a 2x2 matrix	
a	$\operatorname{adj}\mathbf{q} = \mathbf{q}^*$
b	$\varepsilon^{ij} q_{jk} \varepsilon^{lk} = \left(q^*\right)^{il}$
c	

We get the determinant by analogy to Table 7: multiply $\mathbf{q}^*$ by $\mathbf{q}$ and take the trace. This gives twice the determinant. Flip one of the epsilons to make the diagram neater. We get table 12:

Table 12. Determinant of 2x2 matrix	
a	$\det \mathbf{q}$
b	$\frac{1}{2} q_{jk} q_{li} \varepsilon^{ij} \varepsilon^{lk}$
c	

A 1DH Application: Discriminants

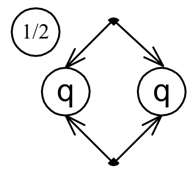
The discriminant of a polynomial is a condition on the coefficients that guarantees that the polynomial has double root. In our last installment we learned how to write this quantity in matrix terms. Now let's see how this looks in diagram form.

Quadratic

The discriminant of the quadratic polynomial from Table 5(a) is

$$B^2 - AC = -\det \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

In diagram form this is



(0.21)

Cubic

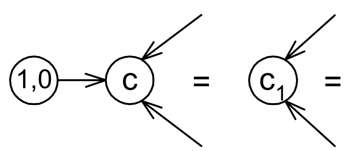
The discriminant of the cubic polynomial from table 10(a) is

$$\Delta_3 = \det\begin{bmatrix} 2\delta_1 & \delta_2 \\ \delta_2 & 2\delta_3 \end{bmatrix} \tag{0.22}$$

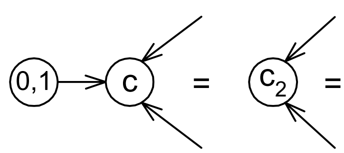
where the matrix elements are defined as

$$\delta_1 = AC - B^2 = \det\begin{bmatrix} A & B \\ B & C \end{bmatrix}$$
$$\delta_2 = AD - BC = \det\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$
$$\delta_3 = BD - C^2 = \det\begin{bmatrix} B & C \\ C & D \end{bmatrix}$$

What does this look like in diagram form? Let's look at the individual "slices" of the **c** tensor. We form these by multiplying one index by a "basis vector" like (1,0) or (0,1).

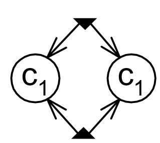


$= \begin{bmatrix} A & B \\ B & C \end{bmatrix}$

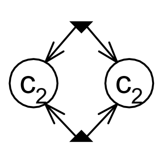


$= \begin{bmatrix} B & C \\ C & D \end{bmatrix}$

The determinants of these two matrices are

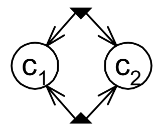


$= -2\det\begin{bmatrix} A & B \\ B & C \end{bmatrix} = -2\delta_1$



$= -2\det\begin{bmatrix} B & C \\ C & D \end{bmatrix} = -2\delta_3$

Now what happens if we mash together $\mathbf{c}_1$ and $\mathbf{c}_2$ as a sort of "cross determinant" with the diagram form:



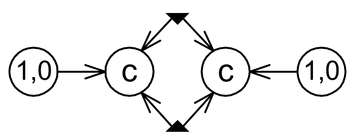
The value of this diagram is, in conventional matrix form

$$trace\left\{\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} B & C \\ C & D \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right\} =$$

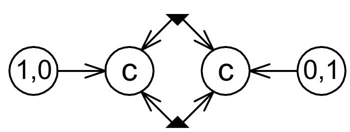
$$trace\left\{\begin{bmatrix} BC-AD & AC-B^2 \\ C^2-BD & 0 \end{bmatrix}\right\} =$$

$$BC-AD = -\delta_2$$

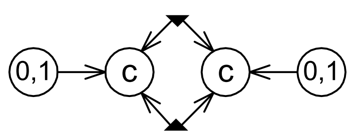
Now, remembering the definitions of $\mathbf{c}_1$ and $\mathbf{c}_2$, we have just shown that:



$$= -2\delta_1$$



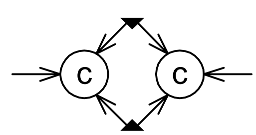
$$= -\delta_2$$



$$= -2\delta_3$$

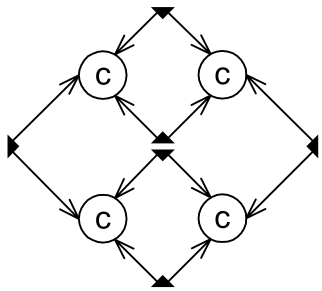
What we have just done is to find expressions for each of the elements of the matrix in equation (0.22) In other words:

$$-\begin{bmatrix} 2\delta_1 & \delta_2 \\ \delta_2 & 2\delta_3 \end{bmatrix} =$$



One interesting thing about this demonstration is that it shows why there are factors of 2 for the δ_1 and δ_3 entries, but not for the δ_2. Anyway, the final step is easy. The discriminant of the cubic $\mathbf{c}$ equals the determinant of this matrix (with the appropriate minus sign and scale factor)

$$\Delta_3 = -\tfrac{1}{2}$$



(0.23)

You can see this as a nice generalization of the discriminant diagram for the quadratic polynomial. Furthermore, a little scratch work will show that equation (0.23) is the simplest diagram that can be formed from epsilons and **c** nodes that is not identically zero.

Quartic

In the last chapter I showed a statement by Hilbert that the discriminant of the quartic polynomial

$$f(x,w) = Ax^4 + 4Bx^3w + 6Cx^2w^2 + 4Dxw^3 + Ew^4$$

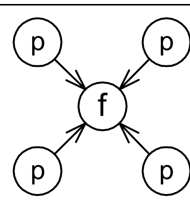
can be found by first calculating the two quantities:

$$\begin{aligned} I_2 &= AE - 4BD + 3C^2 \\ I_3 &= ACE - AD^2 - B^2E + 2BCD - C^3 \end{aligned} \qquad (0.24)$$

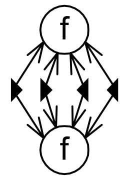
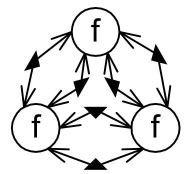
The discriminant is then:

$$\Delta_4 = 27(I_3)^2 - (I_2)^3$$

Now let's see if we can write this as a tensor diagram. Table 13 gives various ways to write the quartic equation.

Table 13. Homogeneous Quartic Equation	
a	$f(x,w) = Ax^4 + 4Bx^3w + 6Cx^2w^2 + 4Dxw^3 + Ew^4$
b	$f(x,w) = \begin{bmatrix} x & w \end{bmatrix} \left\{ \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \\ \begin{bmatrix} B & C \\ C & D \end{bmatrix} & \begin{bmatrix} C & D \\ D & E \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \right\} \begin{bmatrix} x \\ w \end{bmatrix}$
c	messy, messy
d	$f(p) = p^i p^j p^k p^l f_{ijkl}$
e	*f*(*p*) = 

The two simplest diagrams that you can form from 4-arc nodes and epsilons are:

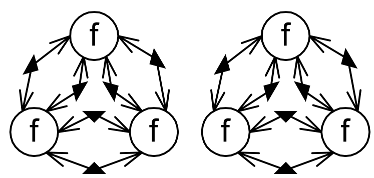
$F_2 \equiv$  and $F_3 \equiv$ 

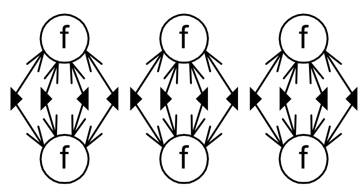
Using a symbolic algebra program, I've been able to evaluate these diagrams and verify that

$$
\begin{aligned}
F_2 &= 2AE - 8BD + 6C^2 \\
&= 2I_2 \\
F_3 &= -6ACE + 6AD^2 + 6B^2E - 12BCD + 6C^3 \\
&= -6I_3
\end{aligned}
$$

Hot damn… Hilbert's invariants match up with the two simplest possible diagrams. Some fiddling with constants gives us

$$
\begin{aligned}
\Delta_4 &= 27\left(\frac{F_3}{-6}\right)^2 - \left(\frac{F_2}{2}\right)^3 \\
&= \tfrac{1}{8}\left(6\left(F_3\right)^2 - F_2{}^3\right)
\end{aligned}
$$

$\Delta_4 = \frac{3}{4}$ 

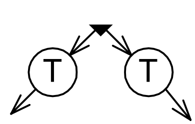
$-\frac{1}{8}$ 

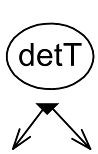
Invariants

The discriminant of a polynomial is an example of an "invariant" quantity. Invariant, in this case, means invariant under parameter transformations. When you calculate such a quantity for a polynomial its sign will remain unchanged if the polynomial is transformed parametrically. This makes sense since the number and multiplicity of roots of a polynomial do not change under parameter transformation.

1DH

Tensor diagrams are particularly useful to express invariant quantities because of the following identity:

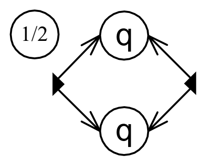


=


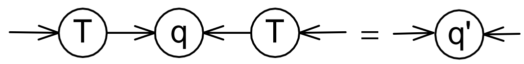
We can easily verify this by explicit calculation:

$$\mathbf{T}\varepsilon\mathbf{T}^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$
$$= \begin{bmatrix} 0 & ad-bc \\ bc-ad & 0 \end{bmatrix}$$
$$= (ad-bc) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

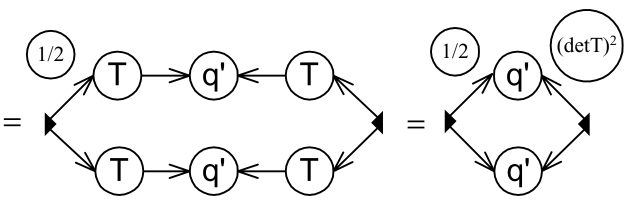
Now, let's apply this to the simplest of our discriminants, the quadratic. We start with

$\Delta_2 =$ 

We now do a parameter transformation on **q**. The 1DH analog to the equations in Table 4 is:



Putting this into our discriminant equation and applying our identity gives:

$\Delta_2 =$ 

In other words,

$$\operatorname{discr}\mathbf{q} = (\det\mathbf{T})^2 \operatorname{discr}\mathbf{q}'$$

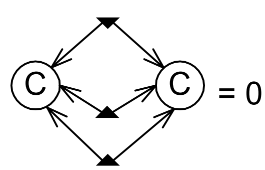
As long as we don't do anything silly, like transform by a singular matrix, the sign of the discriminant of a quadratic is invariant under coordinate transformation.

This seems pretty obvious, but there's a bigger idea lurking in it. It should be pretty simple to see that

> *Any diagram made up of a collection of nodes glued together with the appropriate number of epsilon nodes will represent a transformationally invariant quantity*

If there is an even number of epsilons, the sign is invariant; if there is an odd number of epsilons just the zeroness of the quantity is invariant.

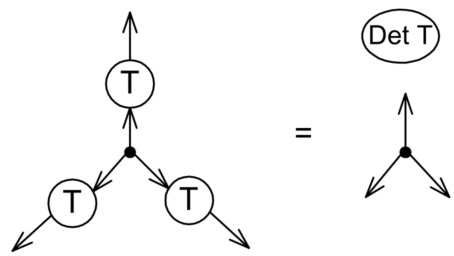
You can imagine any number of diagrams formed in this way; each of them represents some invariant quantity under parameter transformation. Many of these however will be uninteresting. For example you can show that the following diagram is identically zero



Hamilton's book (reference in the previous chapter) is all about algebraic rules for generating invariant quantities. We can do this much more simply with tensor diagrams. For example we know that the two expressions in equation (0.24) are invariant because tensor diagrams can generate them.

2DH

The 2DH epsilon has a similar identity involving transformation matrices that we had in 1DH:



This means that any 2DH tensor diagram made up of polynomial nodes and epsilons represents a transformational invariant.

A 2DH Application: Tangency

Now let's use this 1DH result to solve a 2DH geometry problem: tangency.

Quadratic with Line

Table 2 gave us the condition of a point being on a quadratic curve. How can we generate an expression that determines if a line **L** is tangent to curve **Q**? Let's start by assuming that we have two points on **L**. (We don't need to know how we found these two points. In fact, they will disappear shortly.) Then a general point on the line is

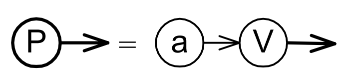
$$\mathbf{P}(\alpha,\beta)=\alpha\mathbf{R}+\beta\mathbf{S}$$

In matrix notation

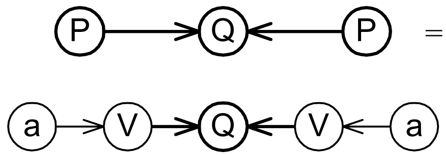
$$\mathbf{P}=\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}R^1 & R^2 & R^3\\ S^1 & S^2 & S^3\end{bmatrix}$$

$$\mathbf{P}=\mathbf{aV}$$

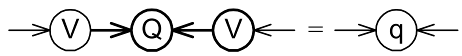
The 2x3 matrix is a sort of conversion from the world of 2D (1DH) vectors (homogeneous polynomials) to the world of 3D (2DH) vectors (homogeneous curves). If we plug this into the quadratic curve equation we get a homogeneous polynomial in (α,β) that evaluates the quadratic function at each point on the line. The condition of the line being tangent to the curve is the same as the condition that there is a double root to this polynomial. Let's write this in diagram form (For these mixed mode diagrams I'll make thicker arrows for the 3-element summations and thinner arrows for the 2-element summations)



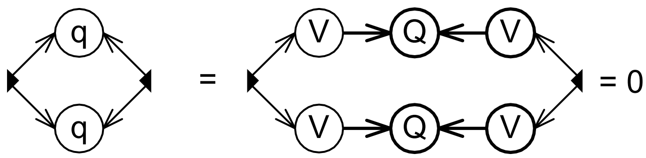
Plug this into the equation for the curve:



This turns the 3x3 symmetric quadratic curve matrix **Q** into a 2x2 symmetric quadratic polynomial matrix **q.** Just **q** by itself is

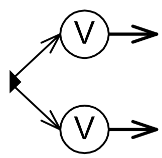


This polynomial has a double root iff its determinant is zero. Plugging this into the diagram form of the determinant gives us the condition that the polynomial has a double root, and thus that the line hits the curve at exactly one point.



(0.25)

Now look at the diagram fragment:

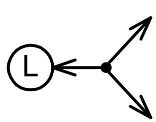


Write this as a matrix product:

$$\begin{bmatrix} R^1 & S^1 \\ R^2 & S^2 \\ R^3 & S^3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} R^1 & R^2 & R^3 \\ S^1 & S^2 & S^3 \end{bmatrix} =$$

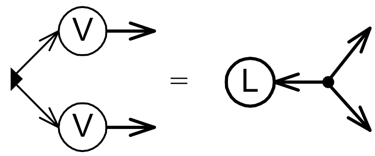
$$\begin{bmatrix} 0 & R^1S^2 - R^2S^1 & R^1S^3 - R^3S^1 \\ R^2S^1 - R^1S^2 & 0 & R^2S^3 - R^3S^2 \\ R^3S^1 - R^1S^3 & R^3S^2 - R^2S^3 & 0 \end{bmatrix}$$

You can recognize the elements of this matrix as the components of the cross product of the two points **R** and **S**. But these are just the elements of the line-vector **L** arranged into an anti-symmetric matrix. In diagram form we can show this as



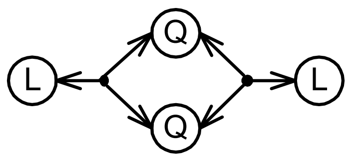
$$= \begin{bmatrix} 0 & L_3 & -L_2 \\ -L_3 & 0 & L_1 \\ L_2 & -L_1 & 0 \end{bmatrix}$$

We can therefore say that



(0.26)

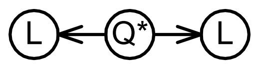
Note that the right hand side of this doesn't require explicit points on **L**, so if all you have are the **L** components you do not need to explicitly find points on **L**. Putting diagrams (0.25) and (0.26) together we get that the line **L** is tangent to curve **Q** if



= 0

(0.27)

This diagram, without the **L** nodes, is just the expression of the adjoint of the matrix **Q** from Table 9(c) (times minus two). In other words, while we use **Q** to test for point incidence, we use $\mathbf{Q}^*$ to test for line incidence (tangency):

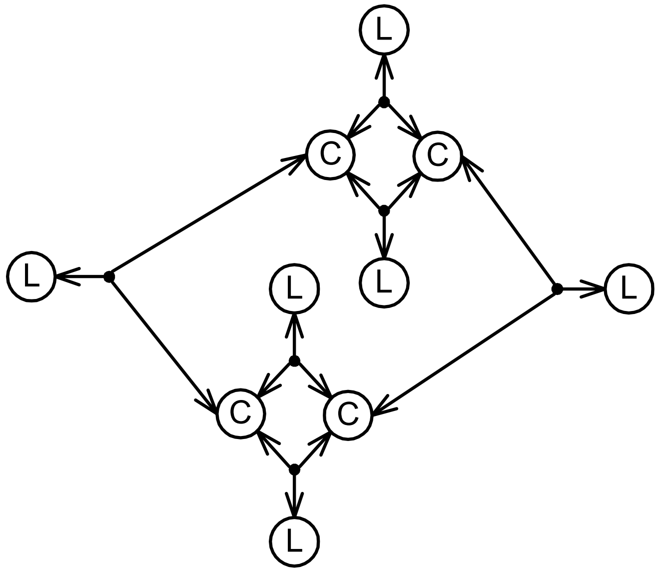


= 0

$$\mathbf{L}^T\left(\mathbf{Q}^*\right)\mathbf{L} = 0$$

Cubic with Line

So, going up an order, what is the condition of line **L** being tangent a cubic curve **C**? That is, we want an expression involving the vector **L** and the cubic coefficient tensor **C** that is zero if **L** is tangent to **C**. With the groundwork we've laid, this is easy. First, compare diagrams (0.21) and (0.27) to see how we converted the quadratic discriminant into a quadratic curve tangency equation. We just replaced each 2D epsilon with a 3D epsilon attached to a copy of **L,** and replace **q** with **Q**. Now do the same thing with the discriminant of a cubic polynomial (0.23). We get



This diagram represents a polynomial expression that is 4th order in **C** and 6th order in **L**. Since it has 18 arcs, the EIN version of this would require 18 index letters. All in all, it is something that would be rather difficult to arrive at in any other, non-diagram, way.

Since the tangency expression is 6th order in **L** it is reasonable to expect that it is possible to find a situation where there are six tangents to a cubic from a given point. This seems excessive, but it is possible as Figure 4 shows.

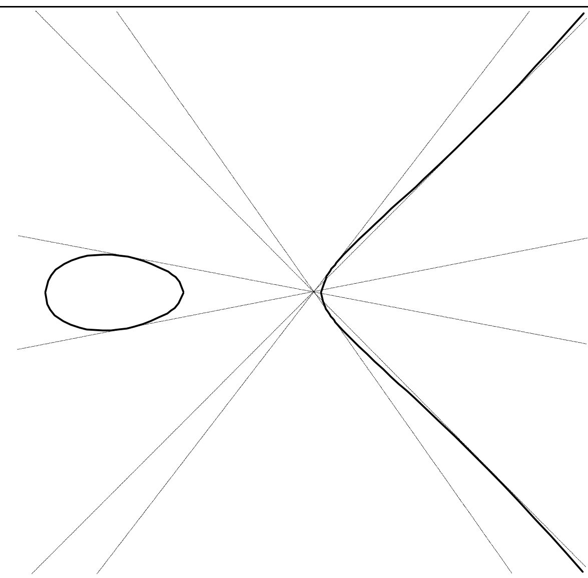


Figure 4. Six tangents from a point to a cubic curve.

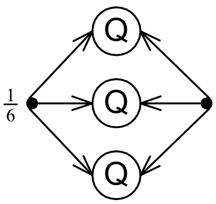
A 2DH Application: Discriminants

The concept of the discriminant also bumps up from 1DH-land to 2DH-land. Again, the discriminant being zero tells us that there are places where both the function and its derivatives are zero. Geometrically this means that there are places on the curve (function=0) where the tangent is not defined (derivative=0). This can happen if the curve is factorable into lower order curves; the points in question are the points of intersection of the lower order curves. Or it can mean that there are cusps or self-intersections in the curve. We'll see examples of all these below.

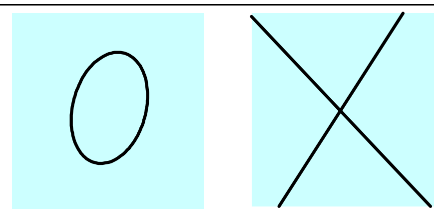
Quadratic

The discriminant of a quadratic curve is just the determinant of the matrix **Q**. In diagram notation this looks like

$$\Delta_2 = -\tfrac{1}{6}$$



If this discriminant is zero it means that the quadratic is factorable into two linear terms. Geometrically, it means that the curve is not a simple conic section, but a degenerate one consisting of two intersecting straight lines. See Figure 5.



$\Delta_2 \neq 0$ $\Delta_2 = 0$

Figure 5. Relation between quadratic discriminant and geometry

Cubic

An equivalent expression for the cubic curve case is considerably more complicated. As I mentioned in the introduction it is a function of two simpler quantities.

$$\Delta_3 = T^2 + 64S^3$$

The cubic discriminant relates to the geometry of the cubic curve as shown in Figure 6. Notice the varieties of cusp, self-intersection and lower-order-curve intersection that can happen when $\Delta_3 = 0$

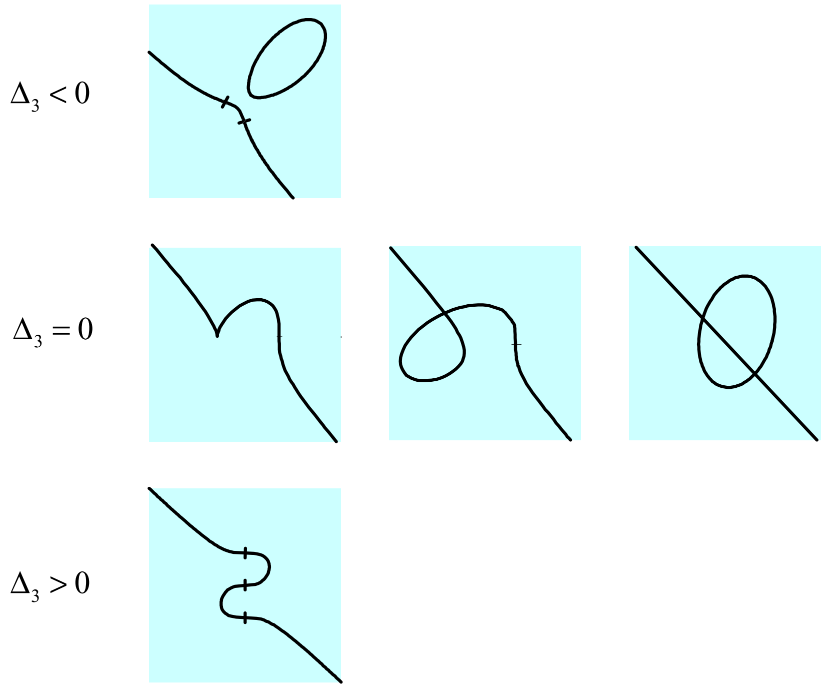
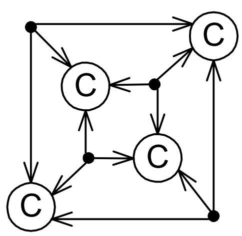


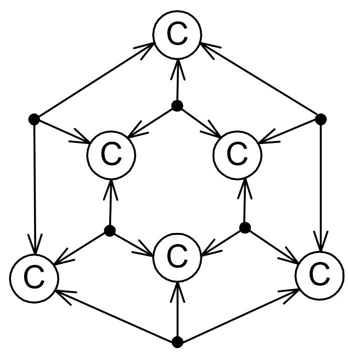
Figure 6. Relation between cubic discriminant and geometry

How can we express this as a tensor diagram? Lets work backwards and see what sort of simple diagrams we can make out of **C** nodes and epsilons. After some fooling around I came up with the following two:

$C_4 =$



$C_6 =$



Again, using a symbolic algebra program I have been able to verify that:

$$C_4 = -24S$$
$$C_6 = -6T$$

Salmon's invariants correspond to the two simplest tensor diagrams we can make for cubic curves! It's things like this that make be believe that I'm really on to something with all this tensor diagram nonsense. Some more fiddling with constants gives us:

$$\Delta_3 = \left(\frac{C_6}{-6}\right)^2 + 64\left(\frac{C_4}{-24}\right)^3$$
$$= \frac{1}{6^3}\left(6\left(C_6\right)^2 - \left(C_4\right)^3\right)$$

Relationships

There's something even more interesting going on here. Notice the similarity between the formula for the discriminants of a 1DH quartic polynomial and a 2DH cubic curve.

$$\text{1DH: } \Delta_4 = \tfrac{1}{8}\left(6\left(F_3\right)^2 - \left(F_2\right)^3\right)$$

$$\text{2DH: } \Delta_3 = \tfrac{1}{6^3}\left(6\left(C_6\right)^2 - \left(C_4\right)^3\right)$$

This means that there is a relationship between the possible root structures of a 4th order polynomial and the possible degeneracies of a 3rd order curve. That's one of the things I'm currently trying to understand.

Notation, Notation, Notation

A lot of the notational language of mathematical consists of the art of creative abbreviation. For example, a vector-matrix product $\mathbf{PT}$ is an abbreviation for a lot of similar looking algebraic expressions. However, clunky expressions like Table 3(b) showed that this notation is not powerful enough to allow us to easily manipulate the sort of expressions that we are encountering here. Einstein index notation has this power, but often gets buried under an avalanche of index letters. The tensor diagram method of drawing EIN is a better way to handle the index bookkeeping. What I've shown here is only the tip of the iceberg. There are a lot of other things that diagram notation can do well that I will cover in future columns.

Our languages help form how we think. I believe that this notation can help us think about these and similar problems and allow us to come up with solutions that we wouldn't find any other way.

References

[1] George Salmon, Modern Higher Algebra, Chelsea Publishing Company. (reprinted from the 1885 original)

[2] A. Kurosh, Higher Algebra, MIR publishers, Moscow, 1975.

[3] T. W. Sederberg, D. C. Anderson and R. N. Goldman, Implicit representation of parametric curves and surfaces, CVGIP 28, 72-84, 1984.

[4] J. Kajiya, "Ray Tracing Parametric Patches, *Proc. Siffraph 82*, ACM Press, New York, 1982, p. 248.

[5] [http://mathworld.wolfram.com/D/DiscriminantPolynomial.html]
[http://www.britannica.com/seo/d/discriminant/]

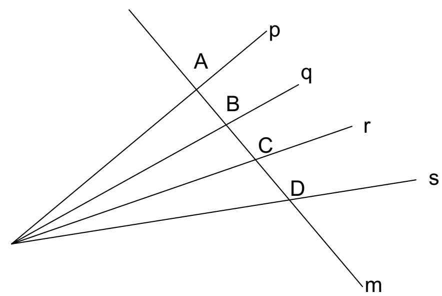
[6] Wee, C. and Goldman, R, *Elimination and Resultants Part 1*, IEEE CG&A January 1995

Chapter 0-06

2DH(3D)
The Cross Ratio

This chapter is cannibalized from my column
Cross Ratios
IEEE CG&A

Geometry is the study of what properties of a figure stay the same as the figure undergoes some transform. For example, in Euclidean geometry the allowable transformations are rotations and translations. Properties that stay constant include distances and angles. For Projective geometry, like we use with homogeneous coordinates, the transformations include perspective projections. In this case one thing that emphatically does not stay the same is geometric length. A property that does, however, is something called the cross ratio. This is defined by reference to the following figure where lines **p**, **q**, **r** and **s** all intersect at the same point.



In this figure the cross ratio is the ratio of the ratios of the following distances.

$$\chi = \frac{|\mathbf{AB}|/|\mathbf{BD}|}{|\mathbf{AC}|/|\mathbf{CD}|}$$

This value is constant no matter where line **m** is placed. It is also constant if the whole diagram undergoes a homogeneous transformation (possibly including perspective).

This has always puzzled me. If geometric distances aren't preserved, and ratios of geometric distances aren't preserved, then how come the ratio of ratios of geometric distances are preserved. Well, here's a quick demonstration.

The First Ratio

Let's begin by calculating the Euclidean distance between points **A** and **B**. We will start with the 2D homogeneous coordinates of each point, which we will name as follows:

$$\mathbf{A} = \begin{bmatrix} A_x & A_y & A_w \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} B_x & B_y & B_w \end{bmatrix}$$

To calculate a Euclidean distance we must turn these into the "real" coordinates by dividing out the homogeneous w coordinate giving the two 2D Euclidean points:

$$\begin{bmatrix} \frac{A_x}{A_w} & \frac{A_y}{A_w} \end{bmatrix} \; and \; \begin{bmatrix} \frac{B_x}{B_w} & \frac{B_y}{B_w} \end{bmatrix}$$

Then subtract giving

$$\begin{bmatrix} \frac{A_x}{A_w} - \frac{B_x}{B_w} & \frac{A_y}{A_w} - \frac{B_y}{B_w} \end{bmatrix}$$

The desired distance is the length of this 2D vector. The fact that this quantity doesn't treat all three of the *x*, *y*, and *w* components symmetrically is indicative of the fact the Euclidean distance is not a meaningful concept in projective geometry.

Anyway, if we bash ahead to get the Euclidean length of the above vector we would need to calculate the square root of the sum of the squares of these two components – wotta pain. Instead, let's pretend for a minute that the line **m** is horizontal. In that case, the length of the vector would equal the *x* component of the vector. In other words

$$\frac{A_x}{A_w} - \frac{B_x}{B_w} = |\mathbf{AB}|$$

In general, though, the line will be at some angle ϑ. In this case, the *x* component will equal the length times the cosine of the tilt angle, or

$$\frac{A_x}{A_w} - \frac{B_x}{B_w} = |\mathbf{AB}| \cos\vartheta$$

For the collinear segment **BD**, tipped by the same angle ϑ, the *x* component would be

$$\frac{B_x}{B_w} - \frac{D_x}{D_w} = |\mathbf{BD}| \cos\vartheta$$

Now here's where the first ratio will come in to simplify life. The top half of our desired cross ratio is the ratio between these segments **AB** and segment **BD**. And, by similarity, the ratio of the *x* components is the same as the ratio of the lengths. Taking the ratio of the above two equations we find

$$\frac{\frac{A_x}{A_w} - \frac{B_x}{B_w}}{\frac{B_x}{B_w} - \frac{D_x}{D_w}} = \frac{|\mathbf{AB}|\cos\vartheta}{|\mathbf{BD}|\cos\vartheta} = \frac{|\mathbf{AB}|}{|\mathbf{BD}|}$$

The angle dependency cancels. Yay! We can calculate the ratio of lengths without any squaring and square rooting.

The Second Ratio

Now let's look at the second ratio in the "ratio of ratios" game. Remember that what we are really interested in is

$$\chi = \frac{|\mathbf{AB}|/|\mathbf{BD}|}{|\mathbf{AC}|/|\mathbf{CD}|}$$

A little algebra on the **AB/BD** ratio turns it into

$$\frac{|\mathbf{AB}|}{|\mathbf{BD}|} = \frac{D_w}{A_w}\left(\frac{A_x B_w - A_w B_x}{B_x D_w - B_w D_x}\right)$$

By changing **B** to **C** everywhere we can see that the second of our two ratios is

$$\frac{|\mathbf{AC}|}{|\mathbf{CD}|} = \frac{D_w}{A_w}\left(\frac{A_x C_w - A_w C_x}{C_x D_w - C_w D_x}\right)$$

Now we can see why the ratio of ratios is something interesting homogeneously; it allows us to cancel the ugly *w* component ratios. The whole cross ratio then boils down to

$$\chi = \frac{|\mathbf{AB}|/|\mathbf{BD}|}{|\mathbf{AC}|/|\mathbf{CD}|} = \frac{A_x B_w - A_w B_x}{B_x D_w - B_w D_x}\;\frac{C_x D_w - C_w D_x}{A_x C_w - A_w C_x}$$

This is much more symmetric and closer to the sort of thing we expect we can make homogeneously constant.

Constancy with changing m

Now we have to ask why is this ratio independent of the position of the line **m**? The root cause must have something to do with the fact that the points **A**, **B**, **C** and **D** are generated from the lines **p**, **q**, **r** and **s** which have the special relationship of all intersecting at the same point.

First, then, here's a brief reminder of the relationship between homogeneous points and lines. You can calculate the point at the intersection of two lines by taking the cross product of the line vectors. This means that the intersection of the four lines is

$$c_1(\mathbf{p}\times\mathbf{q}) = c_2(\mathbf{q}\times\mathbf{r}) = c_3(\mathbf{r}\times\mathbf{s})$$

The inclusion of the constants c_i recognizes the fact that the cross products can represent the same intersection point even though there might be a homogeneous scale factor applied to all three of the x y and w components.

After some fooling around I have found that, for our purposes here, the neatest algebraic way to use this relationship is to express the lines **q** and **r** in terms of **p** and **s**. I'll write this as

$$\mathbf{q} = q_\alpha \mathbf{p} + q_\beta \mathbf{s}$$

$$\mathbf{r} = r_\alpha \mathbf{p} + r_\beta \mathbf{s}$$

You can think of the pairs $\begin{bmatrix} q_\alpha & q_\beta \end{bmatrix}$ and $\begin{bmatrix} r_\alpha & r_\beta \end{bmatrix}$ as 1-dimensional homogeneous coordinates for the collection of lines passing through the intersection of **p** and **s**. Any point passing through this intersection has an $[\alpha \quad \beta]$ pair that describes it. Any nonzero multiple of an $[\alpha \quad \beta]$ pair represents the same line though.

Now, let's relate this back to our points. Remember, the cross product is how we intersect lines so we have

$$\mathbf{A} = \mathbf{p}\times\mathbf{m}$$

$$\mathbf{B} = \mathbf{q}\times\mathbf{m}$$

$$\mathbf{C} = \mathbf{r}\times\mathbf{m}$$

$$\mathbf{D} = \mathbf{s}\times\mathbf{m}$$

Now write the interior point **B** in terms of the outside lines and points:

$$\begin{aligned}\mathbf{B} = \mathbf{q}\times\mathbf{m} &= \left(q_\alpha \mathbf{p} + q_\beta \mathbf{s}\right)\times\mathbf{m} \\ &= q_\alpha(\mathbf{p}\times\mathbf{m}) + q_\beta(\mathbf{s}\times\mathbf{m}) \\ &= q_\alpha \mathbf{A} + q_\beta \mathbf{D}\end{aligned}$$

We can then write the top half of the cross ratio as

$$\begin{aligned}|\mathbf{AB}|/|\mathbf{BD}| &= \frac{A_w}{D_w}\frac{A_xB_w - A_wB_x}{B_xD_w - B_wD_x} \\ &= \frac{A_w}{D_w}\frac{A_x\left(q_\alpha A_w + q_\beta D_w\right) - A_w\left(q_\alpha A_x + q_\beta D_x\right)}{\left(q_\alpha A_x + q_\beta D_x\right)D_w - \left(q_\alpha A_w + q_\beta D_w\right)D_x} \\ &= \frac{A_w}{D_w}\frac{q_\beta A_xD_w - q_\beta A_wD_x}{q_\alpha A_xD_w - q_\alpha A_wD_x} = \frac{A_w}{D_w}\frac{q_\beta}{q_\alpha}\end{aligned}$$

Similarly

$$|\mathbf{AC}|/|\mathbf{CD}| = \frac{A_w}{D_w}\frac{r_\beta}{r_\alpha}$$

The net cross ratio is now

$$\chi = \frac{|\mathbf{AB}|/|\mathbf{BD}|}{|\mathbf{AC}|/|\mathbf{CD}|} = \frac{q_\beta/q_\alpha}{r_\beta/r_\alpha}$$

The interesting thing about this is that the whole dependence on the location of **m** has disappeared, as well as the dependence on which coordinate (x) we chose to use as the measure of distance ratios. That is, it is only dependant on the orientations of the lines **q** and **r** relative to **p** and **s**. Any line **m'** intersecting these will generate the same cross ratio. The cross ratio is thus a property of the locations of the four original lines, rather than the extra line **m**. One key fact about this is that it contains only the ratios of the alpha/beta pairs, so it is also independent of arbitrary scaling of the homogeneous vector for **q** and **r**.

Constancy under perspective

Now we can also see why this is true when the four lines are transformed via an arbitrary transformation matrix. If we have a transformation **T** that changes our outer lines as

$$\mathbf{p}' = \mathbf{T}\mathbf{p}$$
$$\mathbf{s}' = \mathbf{T}\mathbf{s}$$

the inner lines will be

$$\begin{aligned}\mathbf{q}' = \mathbf{T}\mathbf{q} = \mathbf{T}\left(q_\alpha \mathbf{p} + q_\beta \mathbf{s}\right) &= q_\alpha \mathbf{T}\mathbf{p} + q_\beta \mathbf{T}\mathbf{s} \\ &= q_\alpha \mathbf{p}' + q_\beta \mathbf{s}'\end{aligned}$$

And similarly

$$\mathbf{r}' = r_\alpha \mathbf{p}' + r_\beta \mathbf{s}'$$

Even if we try to disguise the line $\mathbf{q}'$ by multiplying by some homogeneous scale we get the same value of cross ratio: the ratio of ratios:

$$\chi = \frac{q_\beta / q_\alpha}{r_\beta / r_\alpha}$$

A particularly interesting special case of this occurs when we take the outside lines **p** and **s** as the x and y-axes and transform by a simple non-uniform scale factor. Numerically this would be

$$\mathbf{p} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} q_\alpha \\ q_\beta \\ 0 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} r_\alpha \\ r_\beta \\ 0 \end{bmatrix}, \mathbf{s} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

From this we get:

$$\mathbf{p}' = \begin{bmatrix} f_x \\ 0 \\ 0 \end{bmatrix}, \mathbf{q}' = \begin{bmatrix} f_x q_\alpha \\ f_y q_\beta \\ 0 \end{bmatrix}, \mathbf{r}' = \begin{bmatrix} f_x r_\alpha \\ f_y r_\beta \\ 0 \end{bmatrix}, \mathbf{s}' = \begin{bmatrix} 0 \\ f_y \\ 0 \end{bmatrix}$$

The alpha and beta components of **q** and **r** have changed to

$$q'_\alpha = f_x q_\alpha, \quad q'_\beta = f_y q_\beta$$
$$r'_\alpha = f_x r_\alpha, \quad r'_\beta = f_y r_\beta$$

The lines **p** and **s** haven't moved. The lines **q** and **r** both have though. And the ratios of the lines **q** and **r** are different.

$$q'_\beta / q'_\alpha = \frac{f_y}{f_x}\left(q_\beta / q_\alpha\right)$$
$$r'_\beta / r'_\alpha = \frac{f_y}{f_x}\left(r_\beta / r_\alpha\right)$$

But the Ratio of Ratios remains the same:

$$\chi = \frac{q'_\beta / q'_\alpha}{r'_\beta / r'_\alpha} = \frac{q_\beta / q_\alpha}{r_\beta / r_\alpha}$$

In other words, even though a transformation might not move the outside lines **p** and **s**, it will move the inside lines **q** and **r** in such a manner that the change in the ratio $\left(q_\beta / q_\alpha\right)$ is matched by the same change in the ratio $\left(r_\beta / r_\alpha\right)$

Summary

The cross ratio is as much a property of the four mutually intersecting lines **p**, **q**, **r** and **s** as it is of the four collinear points **A**, **B**, **C** and **D**. Any line **m** that you throw across the lines will generate four points with the same cross ratio. You can, in turn, take any four collinear points **A**, **B**, **C** and **D,** and throw various collections of four mutually intersecting lines through them. Each of these line collections will have the same cross ratio. Finally, you can project any of these figures perspectively and also get an unchanged cross ratio.

The true homogeneous nature of the cross ratio can best be seen by writing it as:

$$\chi = \frac{q_\beta r_\alpha}{q_\alpha r_\beta}$$

and reviewing the effect of an arbitrary homogeneous scaling on each of the four lines. Remember that

$$\mathbf{q} = q_\alpha \mathbf{p} + q_\beta \mathbf{s}$$

$$\mathbf{r} = r_\alpha \mathbf{p} + r_\beta \mathbf{s}$$

Scaling only **q** will scale both q_α and q_β, but this cancels out in χ. Scaling only **r** will scale r_α and r_β, this will cancel. Scaling only **p** will scale q_α and r_α, cancel. Scaling only **s** will scale q_β and r_β, cancel. From this we can see the necessity of this arrangement of ratio of ratios in constructing a quantity that remains homogeneously meaningful. So, even though perspective transformations do not preserve distances, or ratios of distances, they do preserve ratios of ratios of distances.

Another Variant

Another way to see this puts the vectors for **p**, **q**, **r** and **s** more directly into the equation. We make use of our line **m** that does not pass through the common point of for **p**, **q**, **r** and **s** and solve explicitly for the quantities q_α q_β, r_α and r_β. We first take the cross products of the above equations for **q** and **r** with both **p** and **s**.

$$\mathbf{q}\times\mathbf{p} = q_\alpha \mathbf{p}\times\mathbf{p} + q_\beta \mathbf{s}\times\mathbf{p} = q_\beta \mathbf{s}\times\mathbf{p}$$

$$\mathbf{q}\times\mathbf{s} = q_\alpha \mathbf{p}\times\mathbf{s} + q_\beta \mathbf{s}\times\mathbf{s} = q_\alpha \mathbf{p}\times\mathbf{s}$$

$$\mathbf{r}\times\mathbf{p} = r_\alpha \mathbf{p}\times\mathbf{p} + r_\beta \mathbf{s}\times\mathbf{p} = r_\beta \mathbf{s}\times\mathbf{p}$$

$$\mathbf{r}\times\mathbf{s} = r_\alpha \mathbf{p}\times\mathbf{s} + r_\beta \mathbf{s}\times\mathbf{s} = r_\alpha \mathbf{p}\times\mathbf{s}$$

Each of these equations is a different homogeneous scaling of the common intersection point of **p,q,r,s**. We dot each equation with **m** (knowing we never get zero since **m** is not on this point.)

$$\mathbf{q}\times\mathbf{p}\cdot\mathbf{m} = q_\beta \left(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m}\right)$$

$$\mathbf{q}\times\mathbf{s}\cdot\mathbf{m} = q_\alpha \left(\mathbf{p}\times\mathbf{s}\cdot\mathbf{m}\right)$$

$$\mathbf{r}\times\mathbf{p}\cdot\mathbf{m} = r_\beta \left(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m}\right)$$

$$\mathbf{r}\times\mathbf{s}\cdot\mathbf{m} = r_\alpha \left(\mathbf{p}\times\mathbf{s}\cdot\mathbf{m}\right)$$

So we've just solved for the scalars

$$\frac{(\mathbf{q}\times\mathbf{p}\cdot\mathbf{m})}{(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m})}=q_\beta$$

$$-\frac{(\mathbf{q}\times\mathbf{s}\cdot\mathbf{m})}{(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m})}=q_\alpha$$

$$\frac{(\mathbf{r}\times\mathbf{p}\cdot\mathbf{m})}{(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m})}=r_\beta$$

$$-\frac{(\mathbf{r}\times\mathbf{s}\cdot\mathbf{m})}{(\mathbf{s}\times\mathbf{p}\cdot\mathbf{m})}=r_\alpha$$

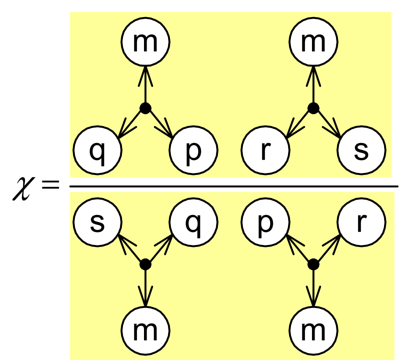
And the expression for the cross ratio becomes:

$$\chi=\frac{q_\beta r_\alpha}{q_\alpha r_\beta}=\frac{(\mathbf{q}\times\mathbf{p}\cdot\mathbf{m})(\mathbf{r}\times\mathbf{s}\cdot\mathbf{m})}{(\mathbf{q}\times\mathbf{s}\cdot\mathbf{m})(\mathbf{r}\times\mathbf{p}\cdot\mathbf{m})}$$

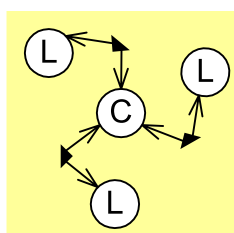
Now we can see more obviously why homogeneously scaling either **p, q, r** or **s** does not change the cross ratio.

Tensor Diagram Version

The tensor diagram version of the above ratio is simply:



PART 1
1DH (2D)



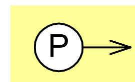
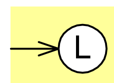
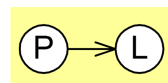
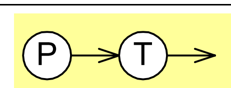
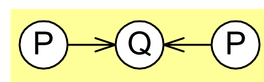
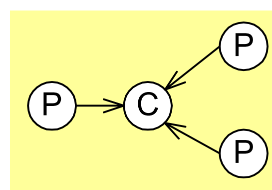
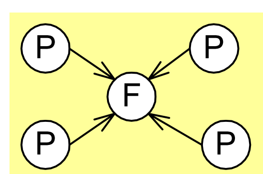
The resultant of a cubic and a linear polynomial

Chapter 1-00

1DH(2D)
Tensor Diagram Techniques

Now let's go back and fill in some details about 2D(1DH) diagrams. "Homogeneous" here means that we have 2D constructs (two element things) interpreted as 1D Homogeneous constructs (polynomials including parameter-at-infinity). In other words this chapter describes homogeneous polynomials.

The elements:

Item	*Matrix Notation*	*Diagram Notation*
Homogeneous parameter	$\begin{bmatrix} x & w \end{bmatrix}$	
Linear coefficients	$\begin{bmatrix} A \\ B \end{bmatrix}$	
Linear polynomial	$Ax + Bw = \begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix}$	
Transformation of parameter	$\begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} x' & w' \end{bmatrix}$	
Quadratic polynomial	$Ax^2 + 2Bxw + Cw^2 = \begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}$	
Cubic polynomial	$Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$ $= \begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}$	
Quartic polynomials	$Ax^4 + 4Bx^3w + 6Cx^2w^2 + 4Dxw^3 + Ew^4 =$ $\begin{bmatrix} x & w \end{bmatrix}\left\{\begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \\ \begin{bmatrix} B & C \\ C & D \end{bmatrix} & \begin{bmatrix} C & D \\ D & E \end{bmatrix} \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}\right\}\begin{bmatrix} x \\ w \end{bmatrix}$	

Derivatives

Since we are dealing with polynomials, let's do some polynomial stuff, like derivatives.

For a quadratic we have

$$f(x,w) = Ax^2 + 2Bxw + Cw^2 = \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

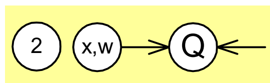
$$\frac{\partial}{\partial x} f(x,w) = 2Ax + 2Bw = \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

$$\frac{\partial}{\partial w} f(x,w) = 2Bx + 2Cw = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

The two components of the first derivative can best be written as a vector. From the above we can see that

$$\begin{bmatrix} f_x\big|_{[x,w]} \\ f_w\big|_{[x,w]} \end{bmatrix} = \begin{bmatrix} 2Ax + 2Bw \\ 2Bx + 2Cw \end{bmatrix} = 2 \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

In other words, a parametric value $[x,w]$ the vector containing the first derivatives of **Q** is simply



In a similar manner, the vector containing the first derivatives of the cubic polynomial represented by the coefficient matrix **C** above, evaluated at the parametric value $[x,w]$ will be:

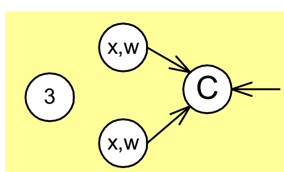
$$f(x,w) = Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$$

$$\frac{\partial}{dx} f(x,w) = 3Ax^2 + 6Bxw + 3Cw^2$$

$$\frac{\partial}{dw} f(x,w) = 3Bx^2 + 6Cxw + 3Dw^2$$

$$\begin{bmatrix} f_x\big|_{[x,w]} \\ f_w\big|_{[x,w]} \end{bmatrix} = \begin{bmatrix} 3Ax^2 + 6Bxw + 3Cw^2 \\ 3Bx^2 + 6Cxw + 3Dw^2 \end{bmatrix} = 3 \begin{bmatrix} \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \\ \hdashline \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} B & C \\ C & D \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \end{bmatrix}$$

We can write this more elegantly simply as the diagram:



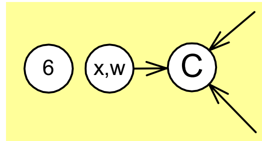
The second derivative of **C** form a 2x2 matrix:

$$\begin{bmatrix} f_{xx} & f_{xw} \\ f_{xw} & f_{ww} \end{bmatrix} = \begin{bmatrix} 6Ax + 6Bw & 6Bx + 6Cw \\ 6Bx + 6Cw & 6Cx + 6Dw \end{bmatrix}$$

$$= 6 \begin{bmatrix} Ax + Bw & Bx + Cw \\ Bx + Cw & Cx + Dw \end{bmatrix}$$

$$= 6 \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

Again, the diagram is simply



The epsilon

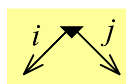
Definition

$$\varepsilon_{12} = 1$$
$$\varepsilon_{21} = -1$$
$$\varepsilon_{ij} = 0 \quad otherwise$$

As matrix

$$\varepsilon = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

As diagram



Symmetry

Be careful of signs and implied transposes when going back and forth to conventional vector/matrix products.

$$\begin{aligned} \mathbf{A}\varepsilon &= A^i \varepsilon_{ij} \\ &= \begin{bmatrix} A^1 & A^2 \end{bmatrix} \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{bmatrix} \\ &= \begin{bmatrix} A^1 & A^2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -A^2 & A^1 \end{bmatrix} \end{aligned}$$

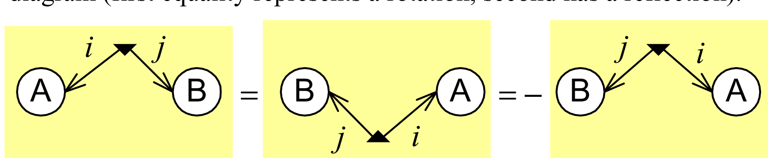
$$\begin{aligned} \varepsilon \mathbf{A}^T &= \varepsilon_{ij} A^j \\ &= \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{bmatrix} \begin{bmatrix} A^1 \\ A^2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} A^1 \\ A^2 \end{bmatrix} \\ &= \begin{bmatrix} A^2 \\ -A^1 \end{bmatrix} \end{aligned}$$

Note the asymmetry in the diagram. Geometric reflections flip the sign. We adopt the convention that when the diagram points down (as above) the first index is on the left. Then the EIN relations

$$A^i \varepsilon_{ij} B^j = B^j \varepsilon_{ij} A^i = -B^j \varepsilon_{ji} A^i$$

$$A^1 B^2 - A^2 B^1 = B^2 A^1 - B^1 A^2 = -\left(B^1 A^2 - B^2 A^1 \right)$$

Translate into the diagram (first equality represents a rotation, second has a reflection).



There exists both covariant and contravariant forms (arrows in / arrows out)

The Epsilon-Delta Identity

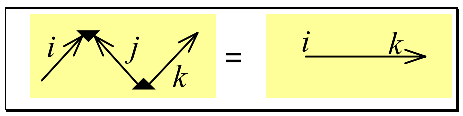
In 2DH the Epsilon-Delta identity has the following form:

$$\varepsilon^{ij}\varepsilon_{kj} = \delta^i_k$$

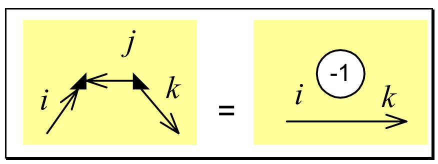
We can show this explicitly by enumerating all the combinations for i,j,k . Most of these are zero, those that are not appear in the table:

i	j	k	ε^{ij}	ε_{kj}	$\varepsilon^{ij}\varepsilon_{kj} = \delta^i_k$
1	2	1	+1	+1	1
2	1	2	-1	-1	1

In diagram form I sometimes call this the zig-zag identity:



A variant is the Cup identity:



Uses of epsilon

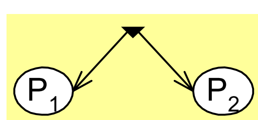
Homogeneous equality of two parameter values

In various notations we have:

$$\frac{x_1}{w_1} = \frac{x_2}{w_2}$$

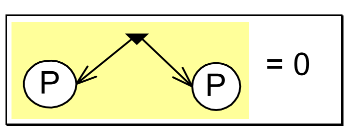
$$x_1 w_2 - w_1 x_2 = 0$$

$$\begin{bmatrix} x_1 & w_1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ w_2 \end{bmatrix} = 0$$



= 0

This gives the identity



Where P can be two identical copies of any complicated glop. This will be useful for complex diagram simplifications.

In conventional notation I will use the symbol $\hat{=}$ to represent “homogeneous equality”. That is

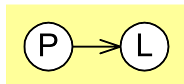
$$\mathbf{P}_1 \triangleq \mathbf{P}_2 \quad \Leftrightarrow \quad \mathbf{P}_1 = w\mathbf{P}_2, \text{for some } w \neq 0$$

Solution to Linear Equation

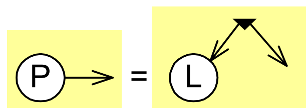
The simplest polynomial, the linear, has the equation

$$Ax + Bw = \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = 0$$

with the diagram:



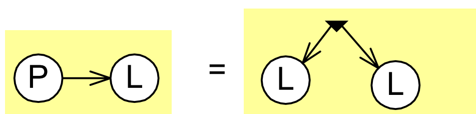
Given the column vector **L,** the solution **P** can be written:



This just means that

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} -B & A \end{bmatrix}$$

solves the equation. The proof is simple; plugging **P** into **L** gives us



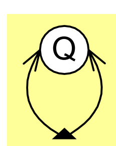
= 0

This seems simple, but we will actually use it a lot.

Another identity

For a symmetric matrix Q

$$\begin{aligned} trace(\mathbf{Q}\varepsilon) &= trace\left(\begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) \\ &= trace\left(\begin{bmatrix} -B & A \\ -C & B \end{bmatrix} \right) \\ &= 0 \end{aligned}$$

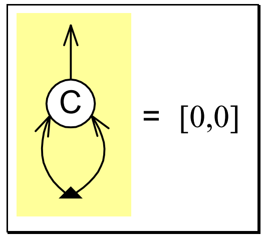


= 0

Any vector times a cubic polynomial tensor yields a symmetric matrix

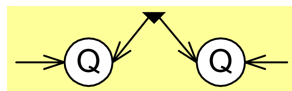
$$\begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} Ax + Bw & Bx + Cw \\ Bx + Cw & Cx + Dw \end{bmatrix}$$

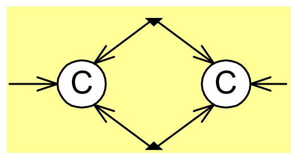
so we have



What is symmetrical?

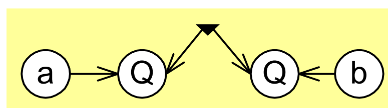
NO





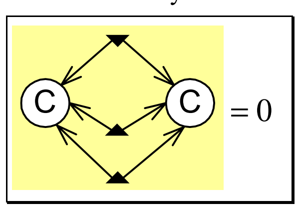
YES

Why? Mirroring the first diagram flips one epsilon, sign change. Mirroring the second diagram flips two epsilons, two sign changes. It's best, when thinking about symmetry to think of each dangling arc as having a unique dummy vector plugged in to it.



Another point about symmetry. In 3D diagrams the anti-symmetry of the (three input) epsilon is accurately represented by mirror operations in the 2D space of the diagram. In 4D diagrams we have mirror operations in the 3D space of the diagrams. In 2D diagrams we are discussing here, we could theoretically draw them all in a 1D space. Then the (1D) mirroring would mean a sign flip. We chose to us the more general 2D space for diagrams that requires us to add an asymmetry to the diagram so that 2D mirror operations work properly to represent sign flips. We might also chose to draw 3D diagrams in 3D space, but we would need to make the (three input) epsilon asymmetrical in 3D space by, for example, dishing it in like an ammonia molecule.

Anyway, since the diagram of the two joined cubics is symmetrical we have another identity:



Adjoint of matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^* = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

We can write this as

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -c & -d \\ a & b \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

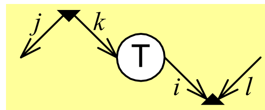
Einstein index notation

$$\varepsilon^{jk}\mathbf{T}_k^i\varepsilon_{li} = \left(\mathbf{T}^*\right)_l^j$$

To make sure the indices are arranged correctly, we look at a specific element:

$$\varepsilon^{j=1,k=2}\mathbf{T}_{k=2}^{i=2}\varepsilon_{l=1,i=2} = \left(\mathbf{T}^*\right)_{l=1}^{j=1}$$

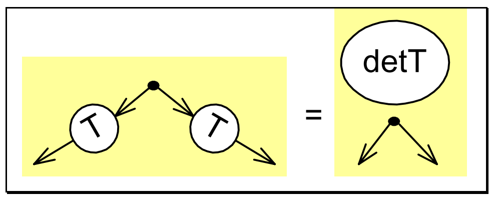
Diagram notation



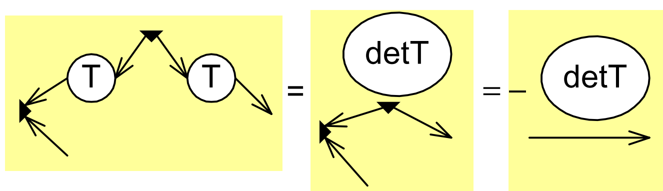
Determinant of matrix

Here are various ways of writing the same identities. These all apply to both transformation matrices (mixed tensors) as well as to symmetric matrices (second order polynomials)

Two transforms applied to an epsilon gives epsilon times determinant. This is the fundamental identity that we use to show transformation invariant properties of various expressions.

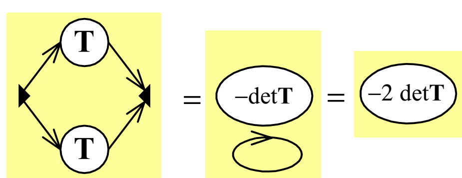


Now multiply both sides on the left an epsilon.



The leftmost diagram above is just **T** times minus the adjoint of **T** (since one of the epsilons is flipped with respect to the adjoint diagram). The above equation is just the diagram way of saying that "Adjoint times matrix = identity times determinant"

Now connect the dangling ends to form the trace of this. This gives the scalar (-2* determinant) (because trace of identity=2)



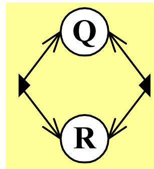
Cross Determinant

Given two different symmetric matrices

$$\mathbf{Q} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} D & E \\ E & F \end{bmatrix}$$

what is



$$trace\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} D & E \\ E & F \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}=$$

$$trace\begin{bmatrix} -B & A \\ -C & B \end{bmatrix}\begin{bmatrix} -E & D \\ -F & E \end{bmatrix}=$$

$$trace\begin{bmatrix} BE-AF & AE-BD \\ CE-BF & BE-CD \end{bmatrix}=$$

$$-AF+2BE-CD$$

The 1D(2DH) Epsilon-Epsilon Rule

This is the most useful identity I have found to manipulate 1DH diagrams. There are several variants which I catalog in the next chapter.

Basic Version

Here it is:

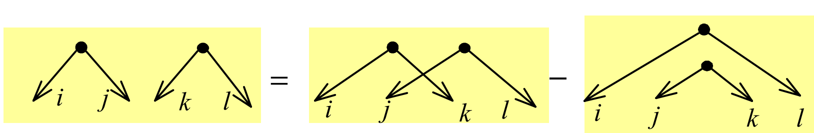
$$\varepsilon_{ij}\varepsilon_{kl}=\varepsilon_{ik}\varepsilon_{jl}-\varepsilon_{il}\varepsilon_{jk}$$

To see this most explicitly, look at table of values. Each index has two possible values (1 and 2) and there are four indices, so there are 16 possible table entries. Most of these will be zero. The following table shows only the nonzero values.

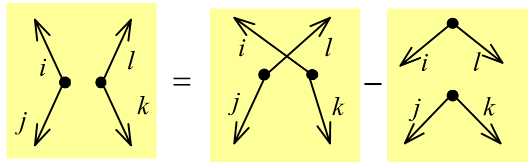
i	j	k	l	$\varepsilon_{ij}\varepsilon_{kl}$	$\varepsilon_{ik}\varepsilon_{jl}$	$\varepsilon_{il}\varepsilon_{jk}$
1	2	1	2	+1*+1=**+1**	0* 0	+1*-1=**-1**
1	2	2	1	+1*-1=**-1**	+1*-1=**-1**	0* 0
2	1	1	2	-1*+1=**-1**	-1*+1=**-1**	0* 0
2	1	2	1	-1*-1=**+1**	0* 0	-1*+1=**-1**
1	1	2	2	0* 0	+1*+1=**+1**	+1*+1=**+1**
2	2	1	1	0* 0	-1*-1=**-1**	-1*-1=**-1**

You can see that column 5 equals column 6 minus column 7 of the table.

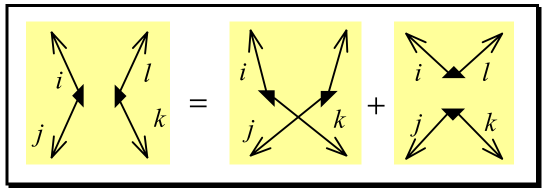
Expressing this in various diagram forms we get



Rearranging

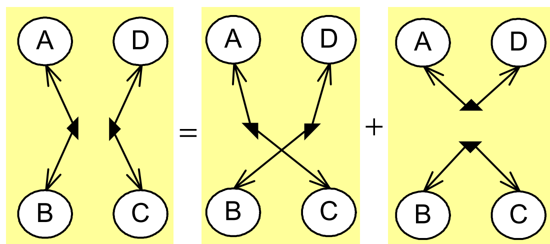


Sign flip to make prettier:



An interpretation

We can see this as an expression of a standard algebra trick, adding and subtracting a fortuitous term to enable factoring. To see this we apply the identity to the four nodes:



Recall that an epsilon between nodes in the homogeneous equivalent of the difference between them. In other words, this is the homogeneous equivalent of the equation

$$(A-B)(C-D)=(C-A)(D-B)+(D-A)(B-C)$$

This is an identity. A little algebra shows that both the right and left sides of this equation equal

$$(A-B)(C-D)+(AB+CD)-(AB+CD)$$

Another Interpretation

The Epsilon-Epsilon identity operates on four 2-vectors. We can also see it as revisiting something we have already seen that operates on two 4-vectors: generating a 3DH line from two 3DH planes. In chapter 0-2 we started with two column vectors representing planes:

$$\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{bmatrix}, \begin{bmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{bmatrix}$$

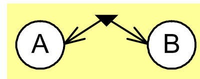
We calculated the six values

$$u=-\det\begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix}; \quad t=\det\begin{bmatrix} b_1 & b_2 \\ d_1 & d_2 \end{bmatrix}; \quad s=-\det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}$$
$$p=-\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}; \quad q=\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix}; \quad r=-\det\begin{bmatrix} a_1 & a_2 \\ d_1 & d_2 \end{bmatrix}$$

And we showed that any such 6 numbers calculated in this fashion must satisfy the identity

$$pu - qt + sr = 0$$

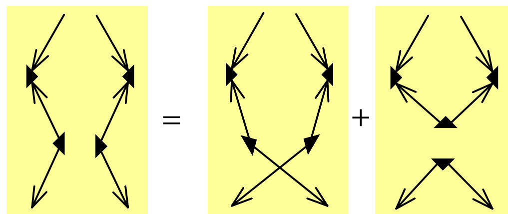
You can see the connection by, for example, matching p with

$p = -$ 

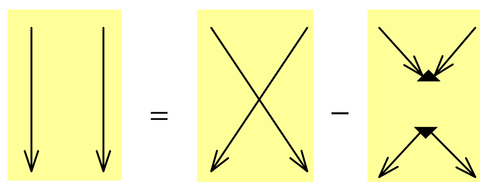
and so forth for *q, r, s, t*

2DH Epsilon-Delta version

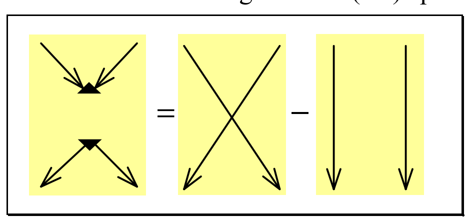
Another variant is derived from multiplying the above form on the bottom by two epsilons.



Apply some zigzag and cup identities;



Rearrange to get something that looks a lot like the original 2DH(3D) epsilon delta identity.



The Game Plan

Application of the epsilon-delta identity is useful for factoring polynomials.

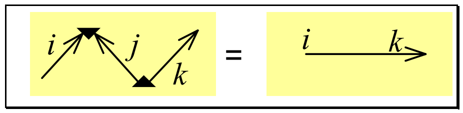
Given some complicated diagram (representing a polynomial), we can often factor it by picking two epsilons and applying the epsilon*epsilon identity. This gives the difference of two modified diagrams. We will typically find that one of several things happens:

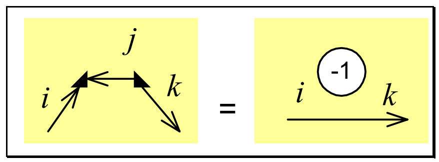
1. One of the diagrams on the right hand side is identically zero. (This is the motivation for amassing a catalog of diagram identities that equal zero). The remaining diagram has fallen into two disjoint pieces, the factors. $D_1 = 0 + d_2 d_3$
2. One of the diagrams is minus the original diagram $D_1 = -D_1 + D_2$. In other words $D_1 = \frac{1}{2} D_2$. The second D_2 might be also in two disjoint pieces.

Chapter 1-01

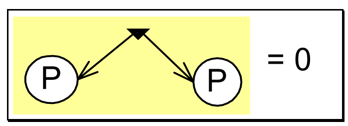
1DH(2D)
Diagram Identity Catalog

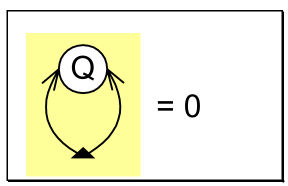
Epsilon Delta Identity

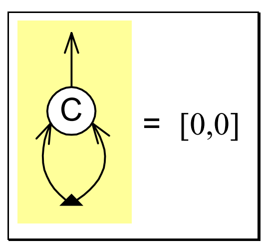


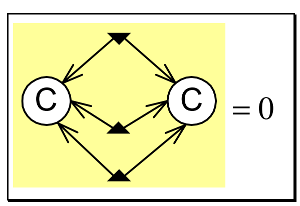


Epsilon applied to symmetric tensors

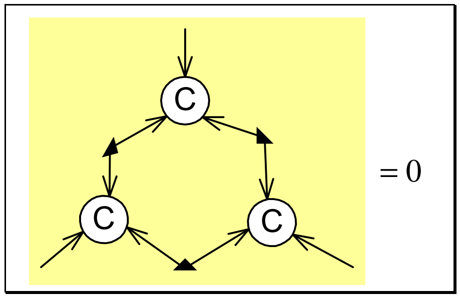






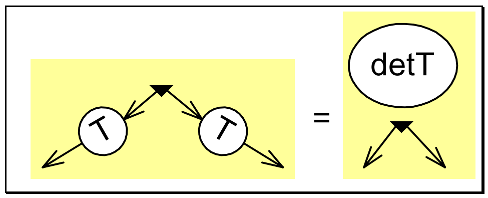


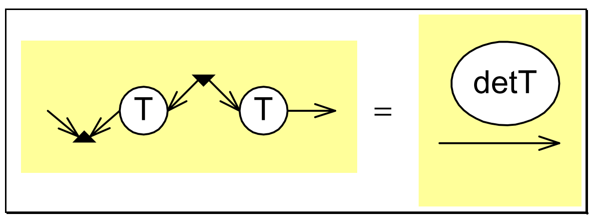
See last section for a proof of the following

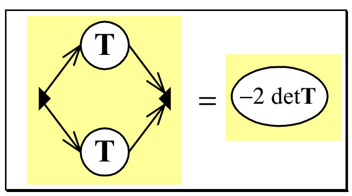


Determinants

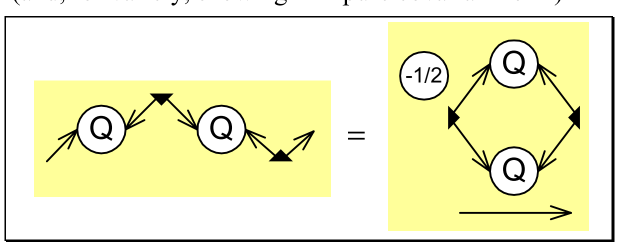
There are all permutations of covariant and contravariant forms of the following. I will just show representative examples:



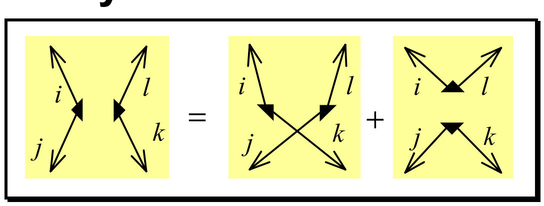


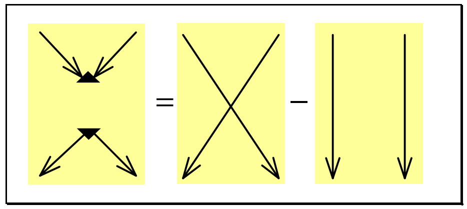


Combining these last two (and, for variety, showing it in pure covariant form)

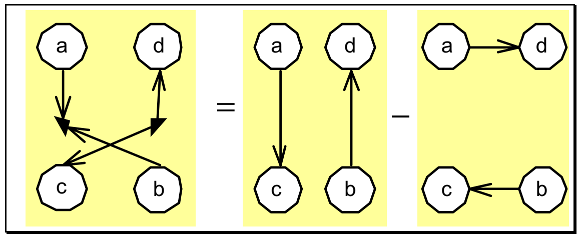


Epsilon-Epsilon Identity

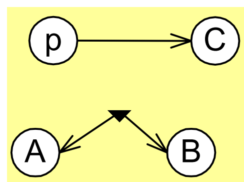




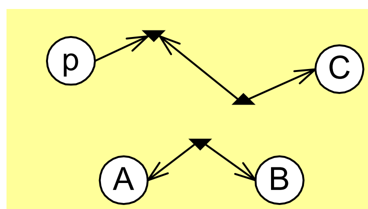
another variant



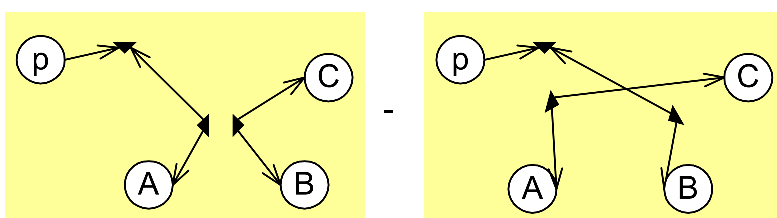
We now derive another variation. Start with



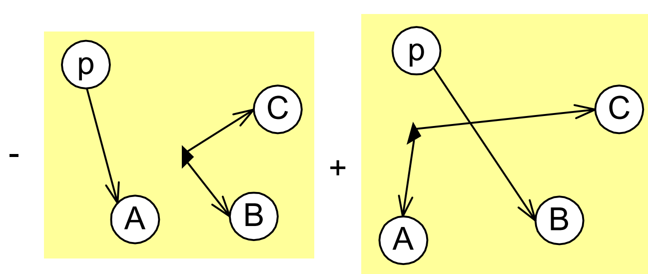
Insert a zig,zag



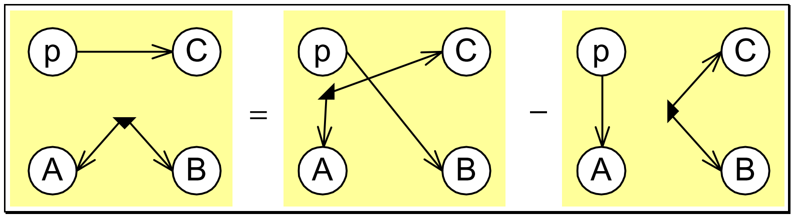
Apply epsilon/delta



Straighten out the dish arcs



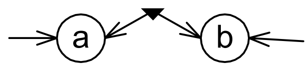
Net identity



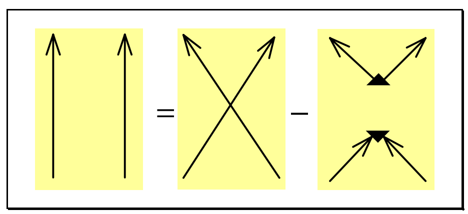
Chains of Quadratics

Two

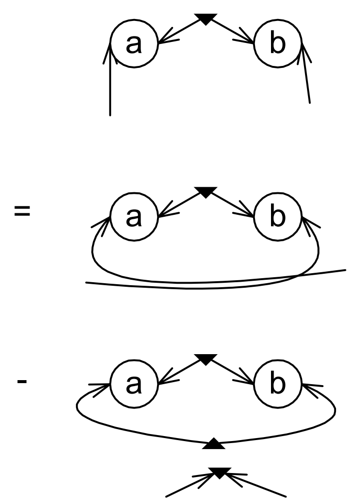
Identites for



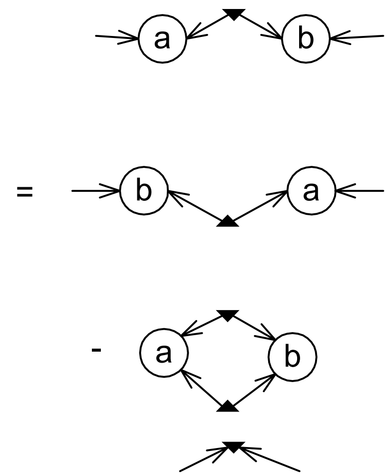
using identity



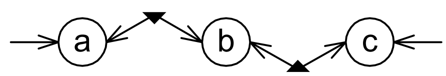
gives



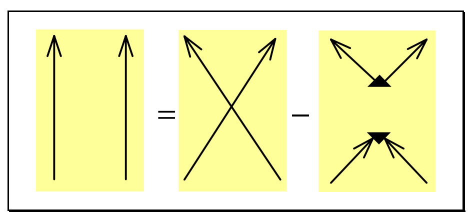
rearrange



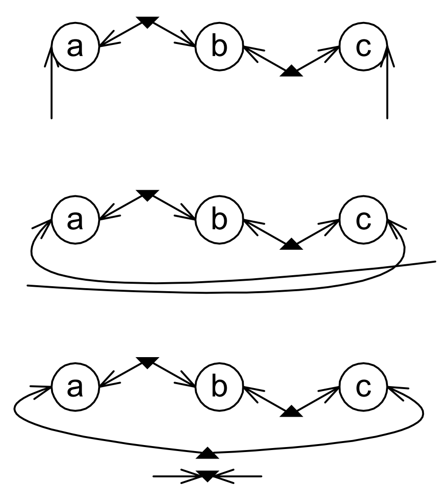
Three



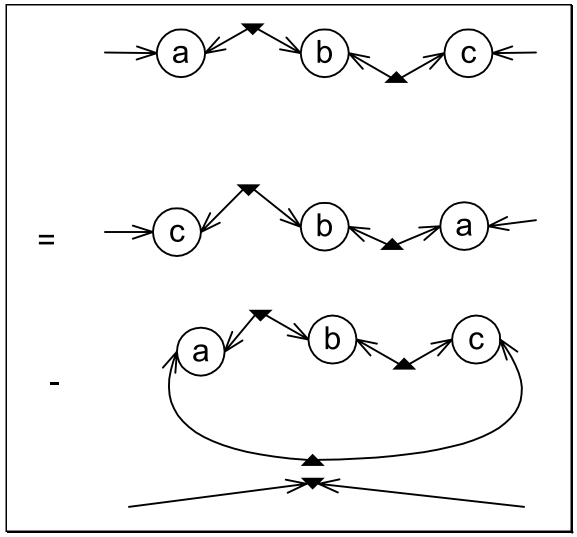
using identity



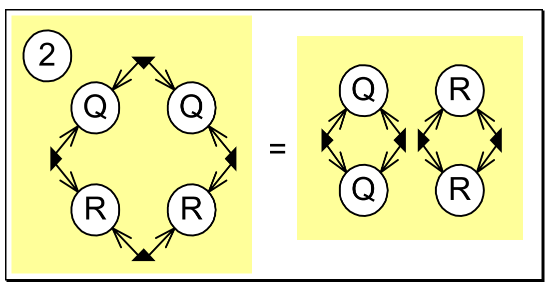
gives

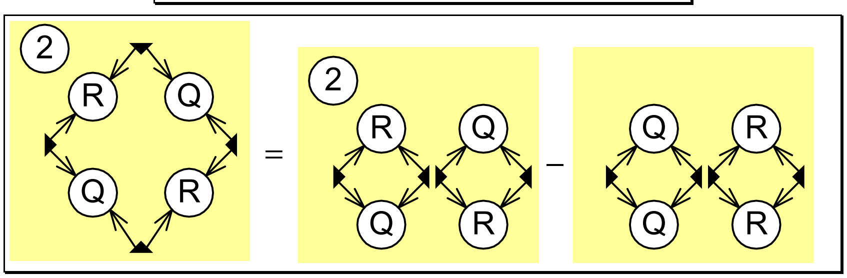


Neaten



Rings



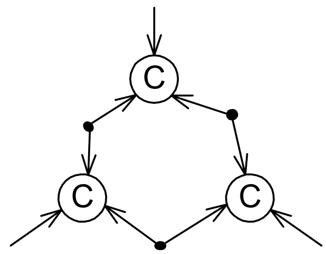


Proof of Ring of 3 Cubics Identity

Given the symmetric tensor **C** defining a homogeneous cubic polynomial

$$Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$$
$$= \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$
$$= \begin{bmatrix} x & w \end{bmatrix} \mathbf{C} \begin{bmatrix} x \\ w \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

The proposed identity is

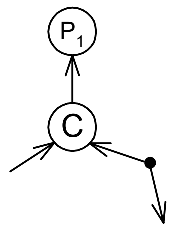


= 0

We have three identical copies of **C,** but independent of whatever glop is attached to the three free arcs.

Brute Force

Divide the diagram into three identical sections. Plug a placeholder point into each section. Then each section of this looks like a symmetric matrix times an epsilon



Name the elements of the 2x2 matrix formed from just the product P_1C as $A_1 \ldots C_1$. Then algebraically the diagram above evaluates to:

$$\begin{bmatrix} A_1 & B_1 \\ B_1 & C_1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -B_1 & A_1 \\ -C_1 & B_1 \end{bmatrix}$$

The whole ring is the trace of the three products

$$trace \begin{bmatrix} -B_1 & A_1 \\ -C_1 & B_1 \end{bmatrix} \begin{bmatrix} -B_2 & A_2 \\ -C_2 & B_2 \end{bmatrix} \begin{bmatrix} -B_3 & A_3 \\ -C_3 & B_3 \end{bmatrix} =$$

$$trace \begin{bmatrix} B_1B_2 - A_1C_2 & -B_1A_2 + A_1B_2 \\ C_1B_2 - B_1C_2 & -C_1A_2 + B_1B_2 \end{bmatrix} \begin{bmatrix} -B_3 & A_3 \\ -C_3 & B_3 \end{bmatrix} =$$

$$trace \begin{bmatrix} -B_1B_2B_3 + A_1C_2B_3 + B_1A_2C_3 - A_1B_2C_3 & B_1B_2A_3 - A_1C_2A_3 - B_1A_2B_3 + A_1B_2B_3 \\ -C_1B_2B_3 + B_1C_2B_3 - B_1B_2C_3 + C_1A_2C_3 & C_1B_2A_3 - B_1C_2A_3 - C_1A_2B_3 + B_1B_2B_3 \end{bmatrix}$$

$$= -B_1B_2B_3 + A_1C_2B_3 + B_1A_2C_3 - A_1B_2C_3 + C_1B_2A_3 - B_1C_2A_3 - C_1A_2B_3 + B_1B_2B_3$$

$$= +A_1C_2B_3 + B_1A_2C_3 - A_1B_2C_3 + C_1B_2A_3 - B_1C_2A_3 - C_1A_2B_3$$

$$= +A_1C_2B_3 + B_1A_2C_3 + C_1B_2A_3 - A_1B_2C_3 - B_1C_2A_3 - C_1A_2B_3$$

$$= -\det \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix}$$

Now for the definitions of the matrix elements

$$A_1 = Ax_1 + Bw_1$$
$$B_1 = Bx_1 + Cw_1$$
$$C_1 = Cx_1 + Dw_1$$
$$\begin{bmatrix} A_1 & B_1 & C_1 \end{bmatrix} = x_1 \begin{bmatrix} A & B & C \end{bmatrix} + w_1 \begin{bmatrix} B & C & D \end{bmatrix}$$

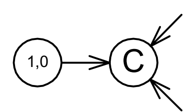
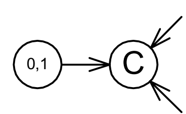
$$\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} A & B & C \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} B & C & D \end{bmatrix}$$

The determinant of this is manifestly zero (show this using general calculation for determinant of sum of matrices…)

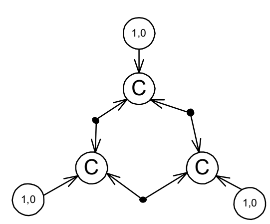
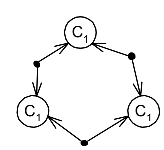
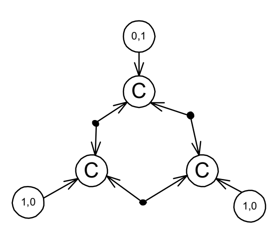
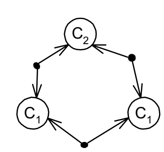
This shows an interesting relationship between the 2x2 matrix and the 3 element vector of ABC.

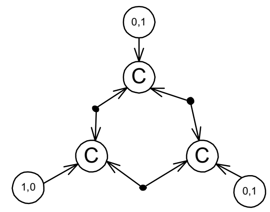
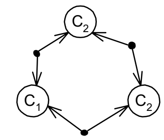
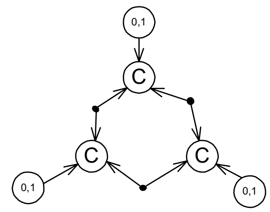
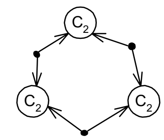
More Elegant

Name the two slices (watch out for overloading of the letter C, italic C is one of the elements of tensor **C**)

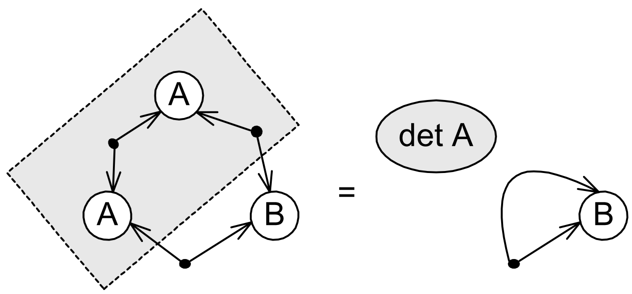
$\mathbf{C}_1 = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$	
$\mathbf{C}_2 = \begin{bmatrix} B & C \\ C & D \end{bmatrix}$	

The big diagram is another cubic with elements extracted by multiplying various input arcs by various basis vectors. I'll put hats over the names of the elements.

$\hat{A}$		
$\hat{B}$		

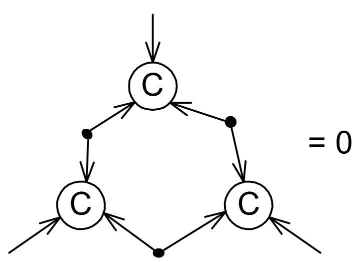
$\hat{C}$		
$\hat{D}$		

We can see that each of these components is identically zero because each diagram is of the form



Summary

So we have the identity:



This is significant because it doesn't matter what kind of glop is attached to the free arcs (indices). In any complicated diagram, if there is a cycle of three C's connected with epsilons in it somewhere, the whole diagram is identically zero.

Proposition: A similar ring of any odd number of identical nodes is zero. (I believe this is true, but don't have a complete proof.)

Chapter 1-02

1DH(2D) Transformations

This chapter introduces some important notational conventions for multiplying and substituting tensor diagrams. The basic motivation is to study transformation matrices, but the techniques introduced here are important for many other situations

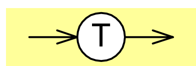
A 2x2 transformation matrix changes 1DH points (i.e. homogeneous parameter values) to other 1DH points.

$$\begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix}$$

Let's investigate some ways of looking at this. Many of these results may seem trivial, but they will turn out to be useful, and their generalization to higher dimensionality is aided by thoroughly understanding these simple cases. This investigation will also serve to introduce some important general techniques in diagram manipulation.

SubAtomic Physics

The diagram notation for a transformation matrix is simply



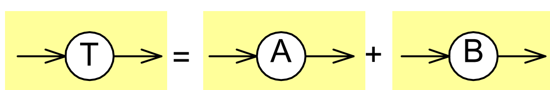
Let's take a look inside the circle and see what its "internal structure" can be.

Sum of Matrices

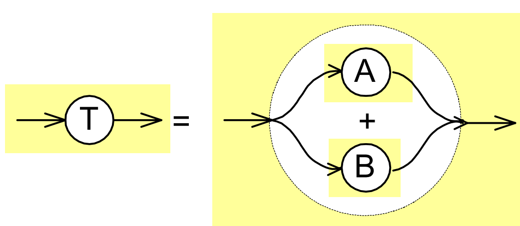
One possible internal structure is if **T** is the sum of two other matrices **A** and **B**

$$\mathbf{T} = \mathbf{A} + \mathbf{B}$$

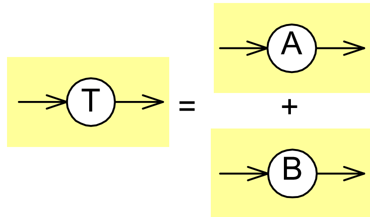
The straightforward diagram for this would be



Now whenever this is used, whatever plugs into the left of T must plug into the left of both A and B. Likewise, whatever plugs into the right of T must plug into the right of both A and B. To emphasize this it might also be instructive to think of the structure as:

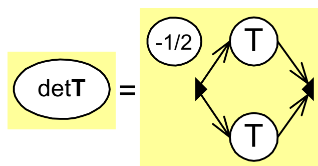


But for the sake of visual simplicity I will often compromise on

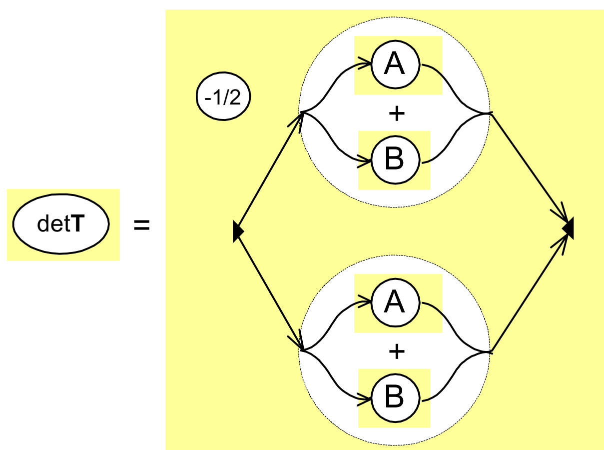


And we will rely on the geometric positioning of the input and output arcs to keep track so that the input to the left side of **T** is also applied to the left sides of each of the terms of **T**. This is easy for this case partly because **T** is a mixed tensor so that its inputs and outputs are distinguished by inward and outward pointing arcs. We have to be a bit more careful in more complicated cases involving pure covariant or pure contravariant tensors.

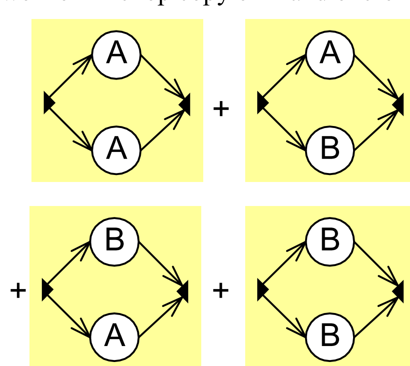
As an illustration of the use of this, let's see what the determinant of **T** is. We have from our definition of epsilon:



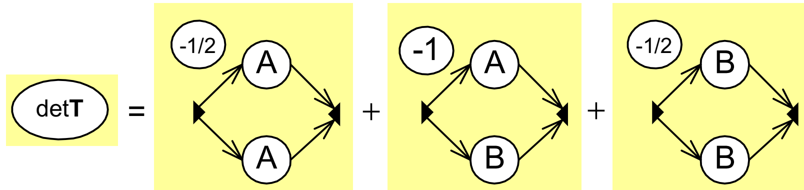
Notice that I've put scalar factors into a node with no outgoing arcs. This is a reasonable metaphor for a scalar. Taking this diagram and plugging in the internal structure of **T** gives us:



This is a configuration we will see very often in later sections. What we are going to do next is to distribute the multiplication by the epsilons over the addition of the two internal terms of **T**. Leaving aside the common factor of -½ for a moment we can see that the result will be the sum of the four terms found by taking each combination of one-of-two from the top copy of **T** and one-of-two from the bottom copy of **T.**



The second and third terms of this are equal (a mirror reflection can take one to the other, and there are an even number of epsilons, thus an even number of sign flips) so we have the final formula:



In more conventional notation

$$\det\left(\mathbf{A}+\mathbf{B}\right)=\det\mathbf{A}+fcn\left(\mathbf{A},\mathbf{B}\right)+\det\mathbf{B}$$

Where *fcn* is the "cross-determinant" of **A** and **B** indicated by the middle diagram term above.

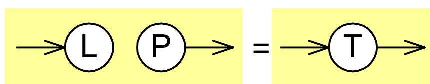
Outer product

Another way to generate a matrix is as the "outer product" of two vectors, here a covariant and a contravariant vector. Matrix notation is

$$\begin{bmatrix} a \\ b \end{bmatrix}\begin{bmatrix} x & w \end{bmatrix}=\begin{bmatrix} ax & aw \\ bx & bw \end{bmatrix}$$

$$\mathbf{L}\ \mathbf{P}=\mathbf{T}$$

The diagram notation for this would be:

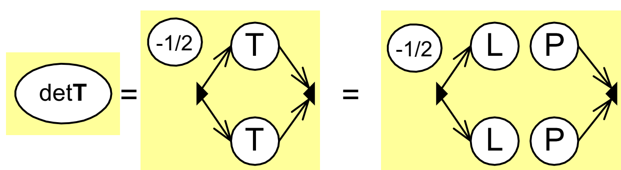


This shows that all the nodes in a diagram need not be connected into one mesh; two disjoint diagram fragments represents the product of the two fragments.

A matrix formed in this way will be singular. We can see this by taking the determinant. Doing this with vector/matrix notation basically requires us to explicitly evaluate the value

$$\det\begin{bmatrix} ax & aw \\ bx & bw \end{bmatrix}=axbw-bxaw=0$$

Doing it with diagrams is better. Plugging in the internal structure of **T**



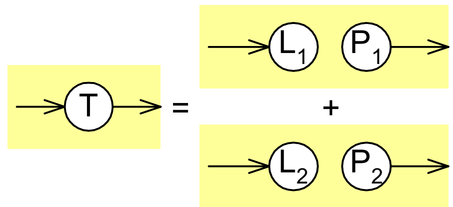
So the determinant of **T** is the product of two diagrams, each of which is zero since they each have identical things on both sides of an epsilon. It's not only zero, its zero squared.

Sum of Outer Products

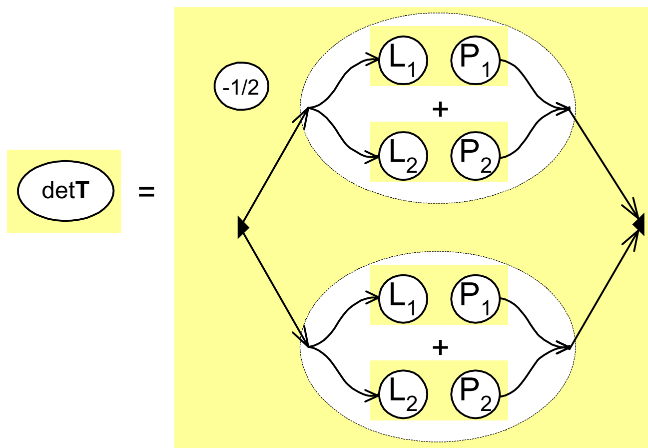
In order to make a nonsingular matrix we need the sum of more than one LP outer product. Let's try

$$\mathbf{T}=\mathbf{L}_1\mathbf{P}_1+\mathbf{L}_2\mathbf{P}_2$$

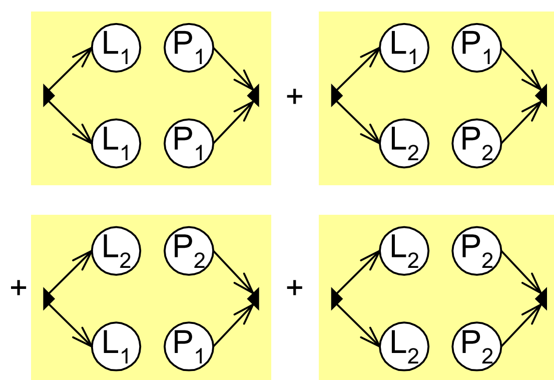
Representing this internal structure as a diagram gives us



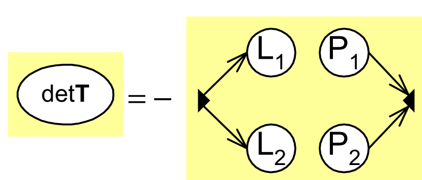
Now what is the determinant of our new matrix **T**? I will do this explicitly one last time. Plug the new diagram into the 2-epsilon representation of the determinant and get



Expanding out the terms we get -1/2 times these four terms:



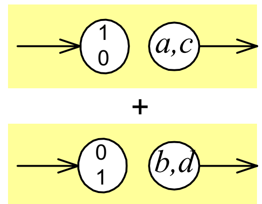
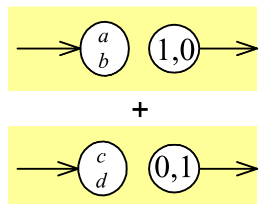
Two of these terms are zero (squared) and the other two are equal (with two sign flips to get the epsilons into the same configuration). Using this "2" to cancel the factor of ½ we get:



So, in order for the matrix **T** to be nonsingular we just require that L_1 be distinct from L_2 and that P_1 be distinct from P_2.

Constructing a desired matrix

Given a desired matrix there are any number of ways of decomposing it into the sum of two outer products. The simplest along with their diagram form are:

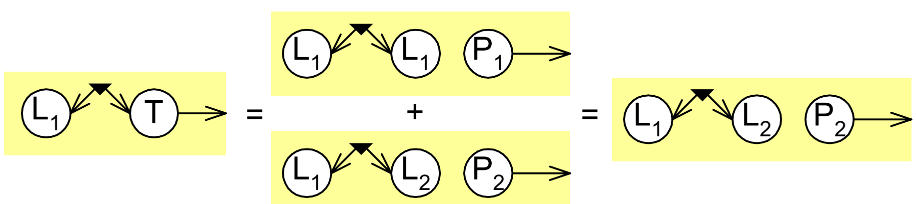
$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} a & c \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} b & d \end{bmatrix}$ Equation(asdf)	
$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix}$	

This may seem particularly trivial, but it does show that we only need to sum two outer products to construct any matrix. The formulas will actually be useful in some situations where we have calculated some complicated diagram form for the two points (rows) or the two (1DH) "lines" (columns). That is, it allows us to "glue" two rows or two columns together into a matrix.

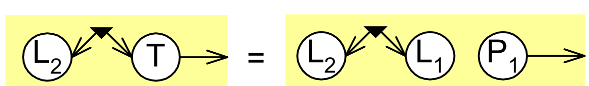
In fact, for any given matrix **T** and for any two given distinct lines $\mathbf{L_1}$ and $\mathbf{L_2}$ it is possible to find $\mathbf{P_1}$ and $\mathbf{P_2}$ that satisfy $\mathbf{T} = \mathbf{L}_1\mathbf{P}_1 + \mathbf{L}_2\mathbf{P}_2$. First multiply **T** by the single "point" that is the solution to $\mathbf{L_1}$. In conventional matrix notation this looks like:

$$\begin{aligned} \mathbf{L}_1\varepsilon\mathbf{T} &= (\mathbf{L}_1\varepsilon\mathbf{L}_1)\mathbf{P}_1 + (\mathbf{L}_1\varepsilon\mathbf{L}_2)\mathbf{P}_2 \\ &= (\mathbf{L}_1\varepsilon\mathbf{L}_2)\mathbf{P}_2 \end{aligned}$$

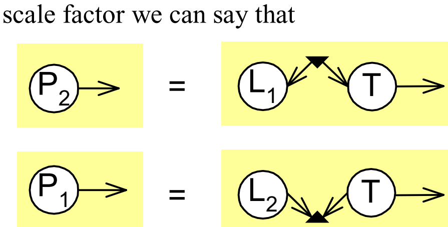
In diagram notation this looks like:



In other words **T** transforms the root of $\mathbf{L_1}$ into (a homogeneous factor times) $\mathbf{P_2}$. In a similar manner we have:



Notice that the homogeneous scale factor of $\mathbf{L}_1\varepsilon\mathbf{L}_2$ appears reversed (minus) in this equation. If we toss out this (nonzero) homogeneous scale factor we can say that



Notice that I've expressed the minus sign in the definition of $\mathbf{P_1}$ as a flip of the epsilon.

We can now address several matrix construction problems in these terms:

Eigenvectors/values

The eigenvectors of matrix **T** are the vectors that remain unchanged homogeneously (except for a nonzero homogeneous scale λ.

$$\mathbf{TL} = \lambda \mathbf{L}$$

An alternative form is to look at the "point" type vectors that remain unchanged

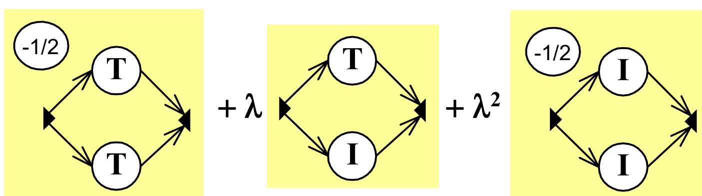
$$\mathbf{PT} = \lambda \mathbf{P}$$

Eigenvalue calculation

We calculate the eigenvalues of a matrix by solving the characteristic equation

$$\det\left(\mathbf{T} - \lambda \mathbf{I}\right) = 0$$

In the 2x2 case this will give a quadratic equation in λ. Our diagram notation for the determinant of a sum of matrices and a little imagination on distribution of scalar factors gives the characteristic equation as: (note two minuses making a plus in the middle term.)



The three diagrams represent, respectively, the determinant of **T**, minus the trace of **T**, and the constant 1. These are all invariants under Euclidean transformations of **T** and **I** so the characteristic equation and its roots are Euclidean invariants of the matrix **T.** (expand on this). The solution can give two distinct real eigenvalues, one real double eigenvalue, or zero real eigenvalues.

Construction from Eigenvectors

Suppose we want to construct a matrix from two given eigenvectors $\mathbf{L_1}$ and $\mathbf{L_2}$ and two given eigenvalues λ_1, λ_2. This is basically the problem we had above where we want to find $\mathbf{P_1}$ and $\mathbf{P_2}$ in

$$\mathbf{T} = \mathbf{L}_1\mathbf{P}_1 + \mathbf{L}_2\mathbf{P}_2$$

Such that

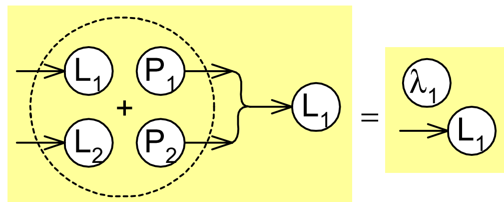
$$\mathbf{TL}_1 = \left(\mathbf{L}_1\mathbf{P}_1 + \mathbf{L}_2\mathbf{P}_2\right)\mathbf{L}_1 = \lambda_1\mathbf{L}_1$$

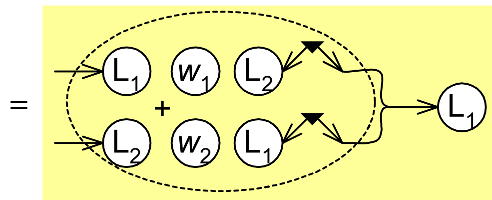
$$\mathbf{TL}_2 = \left(\mathbf{L}_1\mathbf{P}_1 + \mathbf{L}_2\mathbf{P}_2\right)\mathbf{L}_2 = \lambda_2\mathbf{L}_2$$

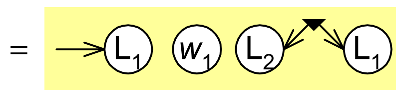
In a manner similar to the previous matrix construction problem, we make $\mathbf{P_2}$ be some homogeneous scalar times the root of $\mathbf{L_1}$, and make $\mathbf{P_1}$ be a homogeneous scalar times the root of $\mathbf{L_2}$. This means that $\mathbf{P}_2 \cdot \mathbf{L}_1 = 0$ and the first equation above turns into:

$$\mathbf{L}_1\mathbf{P}_1\mathbf{L}_1 = \mathbf{L}_1\left(w_1\mathbf{L}_2\varepsilon\right)\mathbf{L}_1 = \lambda_1\mathbf{L}_1$$

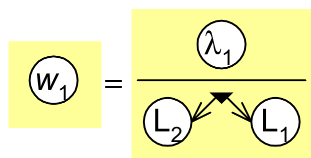
Which, in diagram form looks like

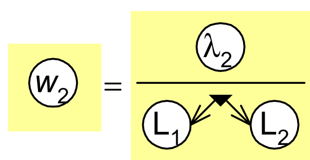




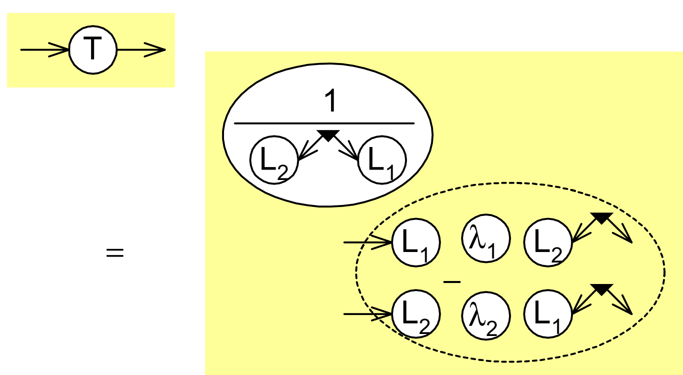


so we simply pick w_1 and (with a similar derivation) w_2 to be:

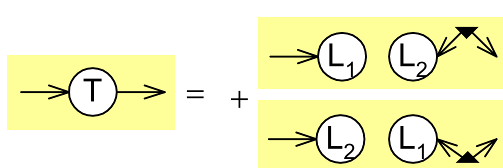




The final **T** in all its glory is:



Sometimes we don't care about the exact value of the eigenvalues. We just want a matrix with two given eigenvectors. Here we can just use:



Exemplary Transformations

In non-homogeneous terms the 1DH transformation is

$$\hat{X} = \frac{\hat{x}}{\hat{w}} = \frac{ax + bw}{cx + dw} = \frac{aX + b}{cX + d}$$

Suppose we want to specify the transformation so that matches a collection of given input/output pairs. That is, we are given a bunch of $X_i, \hat{X}_i$ and we want to find *a,b,c,d* such that

$$\hat{X}_i = \frac{aX_i + b}{cX_i + d}$$

way 1 Brute Force

Each input/output pair generates a linear equation in *a,b,c,d*

$$aX_i + b - cX_i\hat{X}_i - d = 0$$

Each such equation gives us a row in the matrix version:

$$\begin{bmatrix} X_0 & 1 & -X_0\hat{X}_0 & -\hat{X}_0 \\ X_1 & 1 & -X_1\hat{X}_1 & -\hat{X}_1 \\ X_2 & 1 & -X_2\hat{X}_2 & -\hat{X}_2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0$$

The solution for a,b,c,d looks algebraically like the problem of finding the 3DH plane on three 3DH points. The answer is

$$a = \det \begin{bmatrix} 1 & -X_0\hat{X}_0 & -\hat{X}_0 \\ 1 & -X_1\hat{X}_1 & -\hat{X}_1 \\ 1 & -X_2\hat{X}_2 & -\hat{X}_2 \end{bmatrix}, c = \det \begin{bmatrix} X_0 & 1 & -\hat{X}_0 \\ X_1 & 1 & -\hat{X}_1 \\ X_2 & 1 & -\hat{X}_2 \end{bmatrix}$$

$$b = -\det \begin{bmatrix} X_0 & -X_0\hat{X}_0 & -\hat{X}_0 \\ X_1 & -X_1\hat{X}_1 & -\hat{X}_1 \\ X_2 & -X_2\hat{X}_2 & -\hat{X}_2 \end{bmatrix}, d = \det \begin{bmatrix} X_0 & 1 & -X_0\hat{X}_0 \\ X_1 & 1 & -X_1\hat{X}_1 \\ X_2 & 1 & -X_2\hat{X}_2 \end{bmatrix}$$

Way 2: Intermediate matrix

Before pursuing this, let’s first change our outlook to do the whole thing homogeneously. We would like to find the matrix that transforms three given input points $\begin{bmatrix} x_i & w_i \end{bmatrix}$ to three given output points $\begin{bmatrix} \hat{x}_i & \hat{w}_i \end{bmatrix}$

$$\begin{bmatrix} x_i & w_i \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} \hat{x}_i & \hat{w}_i \end{bmatrix}$$

It turns out that we cannot do exactly that. What we can do is to find three output points that are homogeneously equivalent to the desired points. That is we will get

$$\begin{bmatrix} x_i & w_i \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \hat{s}_i \begin{bmatrix} \hat{x}_i & \hat{w}_i \end{bmatrix}$$

The exact values of the $\hat{s}_i$ will be dictated by the input/output values and their evaluation will be an automatic side effect of finding the transformation matrix. Now let's proceed to find the matrix in a new way.

We will solve for the transformation in two stages. The first transforms the inputs to the three canonical points $[1 \quad 0],[0 \quad 1],[1 \quad 1]$ and the second transforms these points to the three desired outputs. We'll do the second one first. We want to find the matrix that transforms:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} a_2 & c_2 \\ b_2 & d_2 \end{bmatrix}=\begin{bmatrix} \hat{s}_0\hat{x}_0 & \hat{s}_0\hat{w}_0 \\ \hat{s}_1\hat{x}_1 & \hat{s}_1\hat{w}_1 \\ \hat{s}_2\hat{x}_2 & \hat{s}_2\hat{w}_2 \end{bmatrix}$$

Taking the first two rows gets:

$$\begin{bmatrix} a_2 & c_2 \\ b_2 & d_2 \end{bmatrix}=\begin{bmatrix} \hat{s}_0\hat{x}_0 & \hat{s}_0\hat{w}_0 \\ \hat{s}_1\hat{x}_1 & \hat{s}_1\hat{w}_1 \end{bmatrix}=\begin{bmatrix} \hat{s}_0 & 0 \\ 0 & \hat{s}_1 \end{bmatrix}\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}$$

So we just need to find $\hat{s}_0, \hat{s}_1$. We do this by looking at the third row of above:

$$[1 \quad 1]\begin{bmatrix} a_2 & c_2 \\ b_2 & d_2 \end{bmatrix}=[\hat{s}_2\hat{x}_2 \quad \hat{s}_2\hat{w}_2]=\hat{s}_2[\hat{x}_2 \quad \hat{w}_2]$$

Plug in our solution for $a_2 \ldots d_2$

$$[1 \quad 1]\begin{bmatrix} \hat{s}_0 & 0 \\ 0 & \hat{s}_1 \end{bmatrix}\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}=\hat{s}_2[\hat{x}_2 \quad \hat{w}_2]$$

Solve this homogeneously by multiplying on the right by the adjoint the xw matrix

$$[\hat{s}_0 \quad \hat{s}_1]\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}^*=\hat{s}_2[\hat{x}_2 \quad \hat{w}_2]\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}^*$$

This evaluates to

$$\begin{aligned}[\hat{s}_0 \quad \hat{s}_1](\hat{x}_0\hat{w}_1-\hat{w}_0\hat{x}_1) &= \hat{s}_2[\hat{x}_2 \quad \hat{w}_2]\begin{bmatrix} \hat{w}_1 & -\hat{w}_0 \\ -\hat{x}_1 & \hat{x}_0 \end{bmatrix} \\ &= \hat{s}_2[\hat{x}_2\hat{w}_1-\hat{w}_2\hat{x}_1 \quad \hat{x}_0\hat{w}_2-\hat{w}_0\hat{x}_2]\end{aligned}$$

This all homogeneously makes sense if we pick:

$$\begin{aligned}\hat{s}_0 &= \hat{x}_2\hat{w}_1-\hat{w}_2\hat{x}_1 \\ \hat{s}_1 &= \hat{x}_0\hat{w}_2-\hat{w}_0\hat{x}_2 \\ \hat{s}_2 &= \hat{x}_0\hat{w}_1-\hat{w}_0\hat{x}_1\end{aligned}$$

The final second/half transformation is then

$$\begin{aligned}\mathbf{T}_2 = \begin{bmatrix} a_2 & c_2 \\ b_2 & d_2 \end{bmatrix} &= \begin{bmatrix} \hat{x}_2\hat{w}_1-\hat{w}_2\hat{x}_1 & 0 \\ 0 & \hat{x}_0\hat{w}_2-\hat{w}_0\hat{x}_2 \end{bmatrix}\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix} \\ &= \begin{bmatrix} \hat{x}_0\hat{w}_1\hat{x}_2-\hat{x}_0\hat{x}_1\hat{w}_2 & \hat{w}_0\hat{w}_1\hat{x}_2-\hat{w}_0\hat{x}_1\hat{w}_2 \\ \hat{x}_0\hat{x}_1\hat{w}_2-\hat{w}_0\hat{x}_1\hat{x}_2 & \hat{x}_0\hat{w}_1\hat{w}_2-\hat{w}_0\hat{w}_1\hat{x}_2 \end{bmatrix}\end{aligned}$$

Now we need the other half, we want:

$$\begin{bmatrix} x_0 & w_0 \\ x_1 & w_1 \\ x_2 & w_2 \end{bmatrix}\begin{bmatrix} a_1 & c_1 \\ b_1 & d_1 \end{bmatrix} = \begin{bmatrix} x_0 & w_0 \\ x_1 & w_1 \\ x_2 & w_2 \end{bmatrix}\mathbf{T}_1 = \begin{bmatrix} s_0 & 0 \\ 0 & s_1 \\ s_2 & s_2 \end{bmatrix}$$

Again we can generate the transform only up to the nonzero scalars s_0, s_1, s_2. It should come as no surprise to find that the desired matrix is just the adjoint of the form we got for T_2 if we were to replace the output points with the input points. So

$$\mathbf{T}_1 = \left\{ \begin{bmatrix} x_2 w_1 - w_2 x_1 & 0 \\ 0 & x_0 w_2 - w_0 x_2 \end{bmatrix}\begin{bmatrix} x_0 & w_0 \\ x_1 & w_1 \end{bmatrix} \right\}^*$$
$$= \begin{bmatrix} x_0 & w_0 \\ x_1 & w_1 \end{bmatrix}^* \begin{bmatrix} x_2 w_1 - w_2 x_1 & 0 \\ 0 & x_0 w_2 - w_0 x_2 \end{bmatrix}^*$$
$$= \begin{bmatrix} w_1 & -w_0 \\ -x_1 & x_0 \end{bmatrix}\begin{bmatrix} x_0 w_2 - w_0 x_2 & 0 \\ 0 & x_2 w_1 - w_2 x_1 \end{bmatrix}$$
$$= \begin{bmatrix} x_0 w_1 w_2 - w_0 w_1 x_2 & w_0 x_1 w_2 - w_0 w_1 x_2 \\ w_0 x_1 x_2 - x_0 x_1 w_2 & x_0 w_1 x_2 - x_0 x_1 w_2 \end{bmatrix}$$

The final matrix is the product of $\mathbf{T}_1$ and $\mathbf{T}_2$. I will explicitly calculate at least one of the entries of this to show something important. The 1,1 element of $\mathbf{T}$ will be:

$$\begin{aligned}
\mathbf{T}_{[1,1]} &= \left(x_0 w_1 w_2 - w_0 w_1 x_2\right)\left(\hat{x}_0 \hat{w}_1 \hat{x}_2 - \hat{x}_0 \hat{x}_1 \hat{w}_2\right) \\
&\quad + \left(w_0 x_1 w_2 - w_0 w_1 x_2\right)\left(\hat{x}_0 \hat{x}_1 \hat{w}_2 - \hat{w}_0 \hat{x}_1 \hat{x}_2\right) \\
&= x_0 w_1 w_2 \hat{x}_0 \hat{w}_1 \hat{x}_2 - w_0 w_1 x_2 \hat{x}_0 \hat{w}_1 \hat{x}_2 - x_0 w_1 w_2 \hat{x}_0 \hat{x}_1 \hat{w}_2 + w_0 w_1 x_2 \hat{x}_0 \hat{x}_1 \hat{w}_2 \\
&\quad + w_0 x_1 w_2 \hat{x}_0 \hat{x}_1 \hat{w}_2 - w_0 w_1 x_2 \hat{x}_0 \hat{x}_1 \hat{w}_2 - w_0 x_1 w_2 \hat{w}_0 \hat{x}_1 \hat{x}_2 + w_0 w_1 x_2 \hat{w}_0 \hat{x}_1 \hat{x}_2 \\
&= x_0 \hat{x}_0 w_1 \hat{w}_1 w_2 \hat{x}_2 \\
&\quad + w_0 \hat{w}_0 w_1 \hat{x}_1 x_2 \hat{x}_2 \\
&\quad + w_0 \hat{x}_0 w_1 \hat{x}_1 x_2 \hat{w}_2 - w_0 \hat{x}_0 w_1 \hat{x}_1 x_2 \hat{w}_2 \\
&\quad + w_0 \hat{x}_0 x_1 \hat{x}_1 w_2 \hat{w}_2 \\
&\quad - w_0 \hat{w}_0 x_1 \hat{x}_1 w_2 \hat{x}_2 \\
&\quad - w_0 \hat{x}_0 w_1 \hat{w}_1 x_2 \hat{x}_2 \\
&\quad - x_0 \hat{x}_0 w_1 \hat{x}_1 w_2 \hat{w}_2 \\
&= \left(x_0 \hat{x}_0\right)\left(w_2 \hat{x}_2\right)\left(w_1 \hat{w}_1\right) - \left(x_0 \hat{x}_0\right)\left(w_1 \hat{x}_1\right)\left(w_2 \hat{w}_2\right) \\
&\quad + \left(x_2 \hat{x}_2\right)\left(w_1 \hat{x}_1\right)\left(w_0 \hat{w}_0\right) - \left(x_1 \hat{x}_1\right)\left(w_2 \hat{x}_2\right)\left(w_0 \hat{w}_0\right) \\
&\quad + \left(x_1 \hat{x}_1\right)\left(w_0 \hat{x}_0\right)\left(w_2 \hat{w}_2\right) - \left(x_2 \hat{x}_2\right)\left(w_0 \hat{x}_0\right)\left(w_1 \hat{w}_1\right)
\end{aligned}$$

In other words, the value of T_{11} (and the other elements as well) is unchanged if we permute the points P_0 P_1 P_2. We can also see that this is just the homogeneous version of the equation for the *a* element above (with a change of sign that doesn't matter homogeneously)

$$a=\det\begin{bmatrix} w_0\hat{w}_0 & -x_0\hat{x}_0 & -w_0\hat{x}_0 \\ w_1\hat{w}_1 & -x_1\hat{x}_1 & -w_1\hat{x}_1 \\ w_2\hat{w}_2 & -x_2\hat{x}_2 & -w_2\hat{x}_2 \end{bmatrix}$$

Calculationally, its probably best to form the matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ and multiply them. But the permutation of points can give us some tools to improve things numerically. By comparing the values for the $s_i, \hat{s}_i$ a choice can be made of which two of the three input/output pairs to pick for P_0 and P_1 for best numerical stability.

Way3: Diagrams

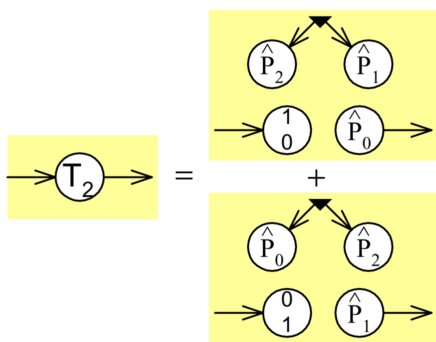
Let's see how this last technique looks when translated into tensor diagrams. We start with the second transformation

$$\mathbf{T}_2=\begin{bmatrix} \hat{x}_2\hat{w}_1-\hat{w}_2\hat{x}_1 & 0 \\ 0 & \hat{x}_0\hat{w}_2-\hat{w}_0\hat{x}_2 \end{bmatrix}\begin{bmatrix} \hat{x}_0 & \hat{w}_0 \\ \hat{x}_1 & \hat{w}_1 \end{bmatrix}$$

We first write it in the format of equation (asdf):

$$\mathbf{T}_2=\left(\hat{x}_2\hat{w}_1-\hat{w}_2\hat{x}_1\right)\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}\hat{x}_0 & \hat{w}_0\end{bmatrix}+\left(\hat{x}_0\hat{w}_2-\hat{w}_0\hat{x}_2\right)\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}\hat{x}_1 & \hat{w}_1\end{bmatrix}$$

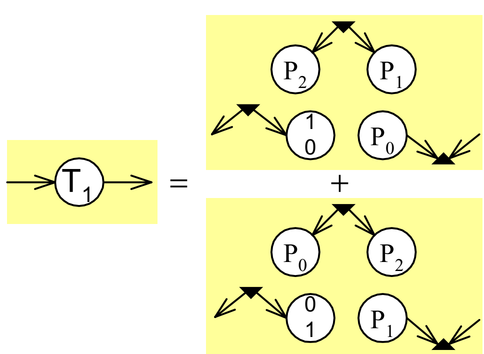
the translateion into diagrams is then straightforward:



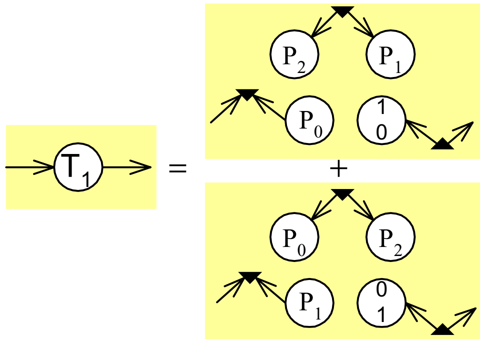
Next, the first transformation $\mathbf{T}_1$ is

$$\mathbf{T}_1=\left\{\begin{bmatrix} x_2w_1-w_2x_1 & 0 \\ 0 & x_0w_2-w_0x_2 \end{bmatrix}\begin{bmatrix} x_0 & w_0 \\ x_1 & w_1 \end{bmatrix}\right\}^*$$

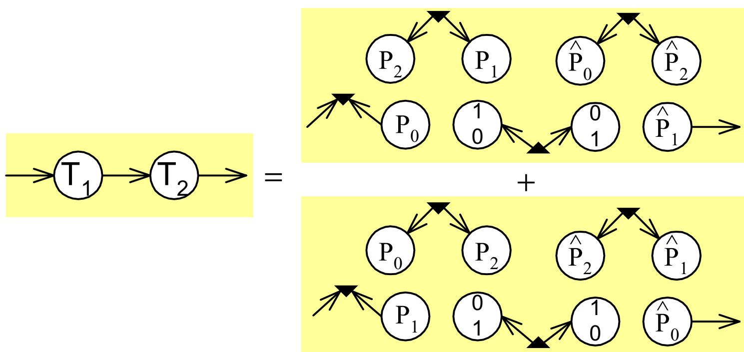
This is just the adjoint of the $\mathbf{T}_2$ expression but without the hats over the **P**s.



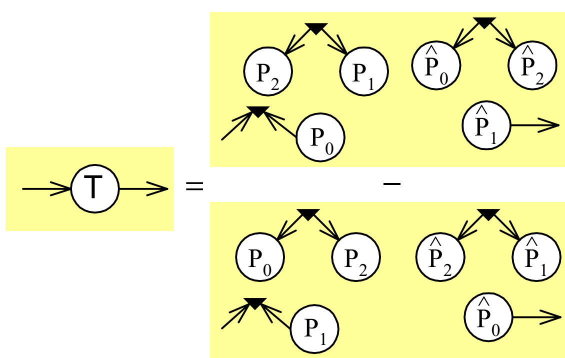
Notice that the adjoint operation has reversed the left/right order of inputs/outputs to $\mathbf{T}_1$. In order to plug $\mathbf{T}_1$ into $\mathbf{T}_2$ it will be easier to rearrange it like:



Now the product of $\mathbf{T_1}$ and $\mathbf{T_2}$ will have four terms, the permutations of two from $\mathbf{T_1}$ and two from $\mathbf{T_2}$. But if you look at the "interface" between the two, we see the basis vectors (1,0) and (0,1) in combinations that will make two of the four terms zero. The remaining two terms then gives us:



The basis vector expression in the top terms evaluates to +1 and the one in the bottom term evaluates to -1 so we have the final exemplary transformation as:



(Check this)

Chapter 1-03

1DH(2D)
Roots of A Quadratic

This chapter deals with solving quadratic equations.

The Conventional Problem

Given *A,B,C* we want to find the value of *X* that satisfies

$$ax^2 + bx + c = 0$$

Conventional solution

The solution we all learned in high school is the quadratic equation.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

There will be two roots if the discriminant $b^2 - 4ac$ is positive, one root (actually a double root) if it's zero, and no (real) roots if it's negative.

There is a numerical unpleasantness with this however. If $ac \ll b^2$ the value of the square root is nearly equal to *B* so calculating one of the two roots requires subtraction of two nearly equal values. We'll solve this shortly.

The Homogeneous Problem

What we really want to do is to solve homogeneous quadratic equations. We want to find the (two) $\begin{bmatrix} x & w \end{bmatrix}$ pairs that satisfy

$$Ax^2 + 2Bxw + Cw^2 = \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = 0$$

Note that I have built a constant of 2 into the linear (*B*) and changed them all to upper case to avoid confusion. This will make some later calculations easier.

The Homogeneous solution

The conventional solution now has two difficulties

1) The numerical problem.
2) The conventional formulation is not homogeneous. That is, its form is not nicely symmetrical with respect to *A* and *C* like the original polynomial expression was. In particular it won't work when $A = 0$, a perfectly reasonable occurrence with homogeneous quadratic polynomials.

We need to reformulate the solution to eliminated these difficulteis. Fortunately, the same approach solves both problems.

We start out by applying the quadratic equation to the homogeneous form. There are two ways to do this.

First solve for *x* in terms of *w*

$$x = \frac{-2Bw \pm \sqrt{4B^2w^2 - 4ACw^2}}{2A}$$
$$= w\frac{-B \pm \sqrt{B^2 - AC}}{A}$$

We can write the two roots homogeneously as:

$$\begin{bmatrix} x_{1a} & w_{1a} \end{bmatrix} = \begin{bmatrix} -B + \sqrt{B^2 - AC} & A \end{bmatrix}$$
$$\begin{bmatrix} x_{2a} & w_{2a} \end{bmatrix} = \begin{bmatrix} -B - \sqrt{B^2 - AC} & A \end{bmatrix}$$

Next solve for *w* in terms of *x*

$$w = \frac{-2Bx \pm \sqrt{4B^2x^2 - 4ACx^2}}{2C}$$
$$= x\frac{-B \pm \sqrt{B^2 - AC}}{C}$$

write these two solutions homogeneously as:

$$\begin{bmatrix} x_{1b} & w_{1b} \end{bmatrix} = \begin{bmatrix} C & -B + \sqrt{B^2 - AC} \end{bmatrix}$$
$$\begin{bmatrix} x_{2b} & w_{2b} \end{bmatrix} = \begin{bmatrix} C & -B - \sqrt{B^2 - AC} \end{bmatrix}$$

How do these two formulations relate?

They both (homogeneously) give the same two answers. For example we can see that $\begin{bmatrix} x_{1a} & w_{1a} \end{bmatrix}$ is homogeneously equal to $\begin{bmatrix} x_{2b} & w_{2b} \end{bmatrix}$ by multiplying them through an epsilon:

$$\begin{bmatrix} x_{1a} & w_{1a} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_{2b} \\ w_{2b} \end{bmatrix} =$$
$$\begin{bmatrix} -B + \sqrt{B^2 - AC} & A \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} C \\ -B - \sqrt{B^2 - AC} \end{bmatrix} =$$
$$\begin{bmatrix} -A & -B + \sqrt{B^2 - AC} \end{bmatrix} \begin{bmatrix} C \\ -B - \sqrt{B^2 - AC} \end{bmatrix} = -AC + \left(B^2 - \left(B^2 - AC\right)\right)$$
$$= 0$$

Just to beat this to death, we can multiply $\begin{bmatrix} x_{1a} & w_{1a} \end{bmatrix}$ by the constant $\left(-B - \sqrt{B^2 - AC}\right)$ and get $\begin{bmatrix} x_{2b} & w_{2b} \end{bmatrix}$. Recall that I am using $\triangleq$ to mean “is homogeneously equal to”:

$$\begin{aligned}
&\begin{bmatrix} -B+\sqrt{B^2-AC} & A \end{bmatrix} \triangleq \\
\left(-B-\sqrt{B^2-AC}\right)&\begin{bmatrix} -B+\sqrt{B^2-AC} & A \end{bmatrix} = \\
&\begin{bmatrix} AC & A\left(-B-\sqrt{B^2-AC}\right) \end{bmatrix} = \\
&A\begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix} \triangleq \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix}
\end{aligned}$$

And we can multiply $\begin{bmatrix} x_{2a} & w_{2a} \end{bmatrix}$ by a constant and get (homogeneously) $\begin{bmatrix} x_{1b} & w_{1b} \end{bmatrix}$

$$\begin{aligned}
&\begin{bmatrix} -B-\sqrt{B^2-AC} & A \end{bmatrix} \triangleq \\
\left(-B+\sqrt{B^2-AC}\right)&\begin{bmatrix} -B-\sqrt{B^2-AC} & A \end{bmatrix} = \\
&\begin{bmatrix} AC & A\left(-B+\sqrt{B^2-AC}\right) \end{bmatrix} = \\
&A\begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix} \triangleq \begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix}
\end{aligned}$$

I went to all this effort to match up the roots properly because our final solution will take one root from each of the pairs above. In order to make sure we get the proper two roots we needed to realize which paired with which. We have

$$\text{Root1} \qquad \begin{bmatrix} -B+\sqrt{B^2-AC} & A \end{bmatrix} \triangleq \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix}$$

$$\text{Root2} \qquad \begin{bmatrix} -B-\sqrt{B^2-AC} & A \end{bmatrix} \triangleq \begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix}$$

This means that we can avoid subtracting something nearly equal to *B* from *B* by taking the choice where the sign of the square root matches the sign of –*B*.

An algorithm for picking which root formulation

$B \geq 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B-\sqrt{B^2-AC} & A \end{bmatrix}$
$B \leq 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} -B+\sqrt{B^2-AC} & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix}$

Note that in each case we only have to evaluate a quantity like $-B+\sqrt{B^2-AC}$ once. That same value appears in both of the two solutions. This algorithm only has us adding two positive numbers or adding two negative numbers. It is numerically sound, but it's also aesthetically pleasing since *A* and *C* appears symmetrically in the solution pairs. Note also that I was careful to make the first root for $B \geq 0$

correspond with the first root for $B \le 0$. This might be important in cases where you were continually evaluating the roots while *B* varied.

This formulation is also more stable to odd inputs. For example, if A=0 the conventional solution screws up. It gives us

$$\begin{bmatrix} x_{1a} & w_{1a} \end{bmatrix} = \begin{bmatrix} -B + \sqrt{B^2 - AC} & A \end{bmatrix} = \begin{bmatrix} -B + |B| & 0 \end{bmatrix}$$
$$\begin{bmatrix} x_{2a} & w_{2a} \end{bmatrix} = \begin{bmatrix} -B - \sqrt{B^2 - AC} & A \end{bmatrix} = \begin{bmatrix} -B - |B| & 0 \end{bmatrix}$$

Depending on the sign of B one of these is the nonsensical root [0 0]. What should happen? In the non-homogeneous case, *A*=0 means that we don't really have a quadratic. In the homogeneous case, though, this is a perfectly respectable quadratic polynomial; it simply has one of its roots at $\begin{bmatrix} 1 & 0 \end{bmatrix}$, that is, where $X = x/w = \infty$. When A=0 our homogeneous solution gives

$B \ge 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -B - \|B\| \end{bmatrix} = \begin{bmatrix} C & -2B \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B - \|B\| & 0 \end{bmatrix} = \begin{bmatrix} -2B & 0 \end{bmatrix}$
$B \le 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} -B + \|B\| & 0 \end{bmatrix} = \begin{bmatrix} -2B & 0 \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & -B + \|B\| \end{bmatrix} = \begin{bmatrix} C & -2B \end{bmatrix}$

This still has problems, though, if A=B=0. This, again, is a perfectly reasonable occurrence. Let's look at all such problem cases

Various potential problem cases:

B=A=0

The equation is:

$$Cw^2 = 0$$

The correct solution is a double root at $w = 0$, that is $\begin{bmatrix} x & w \end{bmatrix} \triangleq \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix}$. Instead, the above machinery gives us one of:

$B \ge 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & 0 \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$
$B \le 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & 0 \end{bmatrix}$

We would be better off taking root 1 from the first choice and root 2 from the second choice.

B=C=0

Equation is:

$$Ax^2 = 0$$

The correct solution is a double root at $x = 0$, that is $\begin{bmatrix} x & w \end{bmatrix} \triangleq \begin{bmatrix} 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \end{bmatrix}$. Instead we get

$B \ge 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} 0 & A \end{bmatrix}$
$B \le 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} 0 & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$

We would rather take root 2 from first choice and root 1 from second choice. The situation is the reverse of the previous one.

A=C=0

No real problem here. The equation is

$$2Bxw = 0$$

with the desired solutions $\begin{bmatrix} x & w \end{bmatrix} \triangleq \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \end{bmatrix}$. The standard machinery gives us:

$B \ge 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} 0 & -B - \sqrt{B^2} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B - \sqrt{B^2} & 0 \end{bmatrix}$
$B \le 0$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} -B + \sqrt{B^2} & 0 \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} 0 & -B + \sqrt{B^2} \end{bmatrix}$

Final "standard" algorithm

Generating the appropriate choices for B=0 we get

$B > 0$		$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -B - \sqrt{B^2 - AC} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B - \sqrt{B^2 - AC} & A \end{bmatrix}$
$B < 0$		$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} -B + \sqrt{B^2 - AC} & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & -B + \sqrt{B^2 - AC} \end{bmatrix}$
$B = 0$	$\lvert A \rvert \ge \lvert C \rvert$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} \sqrt{-AC} & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -\sqrt{-AC} & A \end{bmatrix}$

	$\lvert A\rvert \le \lvert C\rvert$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -\sqrt{-AC} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & \sqrt{-AC} \end{bmatrix}$

Recall that we were careful to make the $\begin{bmatrix} x_1 & w_1 \end{bmatrix}$ solution vary continuously (homogeneously) for *B* going from positive to zero to negative, although there is a jump via a homogeneous scale when *B* passes zero. (That is, we made sure that the $\begin{bmatrix} x_1 & w_1 \end{bmatrix}$ solution doesn't abruptly jump over the other solution during this transition).

When $B = 0, \lvert A\rvert = \lvert C\rvert$ there doesn't seem to be a good reason for picking one solution set over another. This is only relevant for $A = -C$ since $A = C$ gives us no real roots.

The discriminant and ABC space

Of course there are only real roots if the discriminant (under the square root) is positive.

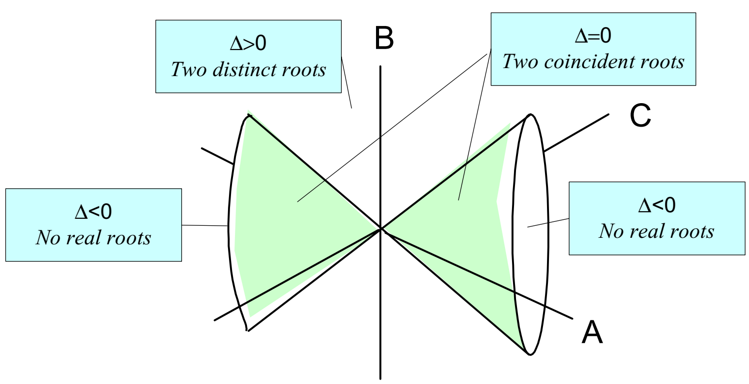
$$\Delta = B^2 - AC$$

Note that this is minus the determinant of the coefficient matrix

$$\Delta = -\det\begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

The sign of this (plus minus or zero) tells how many real roots we get.

It is useful to visualize the space of possible quadratics in ABC space. The surface where the discriminant is zero is a cone with its axis along the A=C line, and with the *A* and *C* axes embedded in it. Note that it is not a circular cone; the cross section is elliptical.



Looking at the relative volumes involved, it is comforting to know that, statistically at least, you are much more likely to stumble upon a quadratic with two real roots.

General solution schema

We are going to come up with some other ways of solving quadratics. Let's first see how our current solution was arrived at. Solving polynomials is typically done by a transformation of the parameter space chosen to make a new polynomial that has no linear term. In homogeneous terms a general transformation of the parameter will be a 2x2 matrix:

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Transformed polynomial is

$$[x \quad w]\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix} =$$
$$[\hat{x} \quad \hat{w}]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix}$$
$$= [\hat{x} \quad \hat{w}]\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B} & \hat{C} \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix}$$

Pick the transformation to get rid of the $\hat{B}$ term so that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} \hat{A} & 0 \\ 0 & \hat{C} \end{bmatrix}$$

Then we just solve the equation

$$\hat{A}\hat{x}^2 + \hat{C}\hat{w}^2 = 0$$

the solutions are:

$$\frac{\hat{x}^2}{\hat{w}^2} = -\frac{\hat{C}}{\hat{A}}$$
$$\frac{\hat{x}}{\hat{w}} = \pm\sqrt{-\frac{\hat{C}}{\hat{A}}}$$

This has real solutions only if $\hat{A}$ and $\hat{C}$ have different signs. Depending on which one is negative the solutions can also be written;

$$[\hat{x} \quad \hat{w}] = \begin{bmatrix} \pm\sqrt{-\hat{C}} & \sqrt{\hat{A}} \end{bmatrix} \text{ or } \begin{bmatrix} \pm\sqrt{\hat{C}} & \sqrt{-\hat{A}} \end{bmatrix}$$

Then we simply transform these roots back to the original parameter space via

$$[x \quad w] = [\hat{x} \quad \hat{w}]\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Choices of transformation

Now let's look at various ways to select the transformation matrix

Conventional choice

Use a "translation" of the form

$$[x \quad w] = [\hat{x} \quad \hat{w}]\begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix}$$

so that

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B} & \hat{C} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} A & B \\ A\delta + B & B\delta + C \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} A & A\delta + B \\ A\delta + B & A\delta^2 + 2B\delta + C \end{bmatrix}$$

Pick $\delta = -\frac{B}{A}$ to make $\hat{B} = 0$ and we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{B}{A} & 1 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B} & \hat{C} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \frac{-B^2 + AC}{A} \end{bmatrix}$$

Root is

$$\frac{\hat{x}}{\hat{w}} = \pm\sqrt{-\frac{\hat{C}}{\hat{A}}} = \pm\sqrt{\frac{B^2 - AC}{A^2}}$$

$$\begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{B^2 - AC} & A \end{bmatrix}$$

Going back to original parameter space:

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} \pm\sqrt{B^2 - AC} & A \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{B}{A} & 1 \end{bmatrix} = \begin{bmatrix} -B \pm \sqrt{B^2 - AC} & A \end{bmatrix}$$

Homogeneous choice

Our alternative of solving for *w* in terms of *x* effectively uses a “perspective” transformation that looks like the transpose of the previous “translation” transformation.

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}$$

Here we have

$$\begin{bmatrix}\hat{A} & \hat{B}\\ \hat{B} & \hat{C}\end{bmatrix}=\begin{bmatrix}1 & \delta\\ 0 & 1\end{bmatrix}\begin{bmatrix}A & B\\ B & C\end{bmatrix}\begin{bmatrix}1 & 0\\ \delta & 1\end{bmatrix}$$
$$=\begin{bmatrix}A+\delta B & B+\delta C\\ B & C\end{bmatrix}\begin{bmatrix}1 & 0\\ \delta & 1\end{bmatrix}$$
$$=\begin{bmatrix}A+2B\delta+C\delta^2 & B+\delta C\\ B+\delta C & C\end{bmatrix}$$

Pick $\delta=-\dfrac{B}{C}$ to make $\hat{B}=0$ and we have

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}1 & -\dfrac{B}{C}\\ 0 & 1\end{bmatrix}$$
$$\begin{bmatrix}\hat{A} & \hat{B}\\ \hat{B} & \hat{C}\end{bmatrix}=\begin{bmatrix}\dfrac{AC-B^2}{C} & 0\\ 0 & C\end{bmatrix}$$

Root is

$$\frac{\hat{x}}{\hat{w}}=\pm\sqrt{-\frac{\hat{C}}{\hat{A}}}=\pm\sqrt{\frac{C^2}{B^2-AC}}$$
$$\begin{bmatrix}\hat{x} & \hat{w}\end{bmatrix}=\begin{bmatrix}C & \pm\sqrt{B^2-AC}\end{bmatrix}$$

Going back to original parameter space:

$$\begin{bmatrix}x & w\end{bmatrix}=\begin{bmatrix}C & \pm\sqrt{B^2-AC}\end{bmatrix}\begin{bmatrix}1 & -\dfrac{B}{C}\\ 0 & 1\end{bmatrix}=\begin{bmatrix}C & -B\pm\sqrt{B^2-AC}\end{bmatrix}$$

Comparison

Let's look at this whole transformation process in *ABC* space. We have

$$\begin{bmatrix}\hat{A} & \hat{B}\\ \hat{B} & \hat{C}\end{bmatrix}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}A & B\\ B & C\end{bmatrix}\begin{bmatrix}a & c\\ b & d\end{bmatrix}$$
$$=\begin{bmatrix}aA+bB & aB+bC\\ cA+dB & cB+dC\end{bmatrix}\begin{bmatrix}a & c\\ b & d\end{bmatrix}$$
$$=\begin{bmatrix}a^2A+2abB+b^2C & acA+(bc+ad)B+bdC\\ acA+(bc+ad)B+bdC & c^2A+2cdB+d^2C\end{bmatrix}$$

Write this in the form

$$\begin{bmatrix} a^2 & 2ab & b^2 \\ ac & bc+ad & bd \\ c^2 & 2cd & d^2 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix}=\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix}$$

We are using either of two transformation matrices

$$\begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}$$

So our two choices for *a,b,c,d* give the following transformations in *ABC* space.

$$\begin{bmatrix} 1 & 0 & 0 \\ \delta & 1 & 0 \\ \delta^2 & 2\delta & 1 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix}=\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 2\delta & \delta^2 \\ 0 & 1 & \delta \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} A \\ B \\ C \end{bmatrix}=\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix}$$

What this means is that the "path" an [*A,B,C*] point takes under this transformation is a parabolic arc from its original position to the *B*=0 plane. The path is perpendicular to the *A* axis in the first case, and perpendicular to the *C* axis in the second case

. DIAGRAM.

Eigenvector choice

In order to avoid ill conditioned transformations, let's see what happens if we pick the transformation to be the two (orthogonal, unit length) eigenvectors of the matrix. Then $\hat{A}$ and $\hat{C}$ are the eigenvalues.

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}=\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

We are dipping into Euclidean geometry when we talk about unit length of the eigenvectors. In fact, eigenvectors are not precisely defined here since the product of a vector and the matrix is not a vector, it's a covector. I.e the expression

$$\mathbf{PQ}=\lambda\mathbf{P}$$

is meaningless given our rules of co/contravaniance. We can patch this up later. I derive in another chapter the explicit solution for eigenvectors/values here. Punchline is, first define

$$\Delta=\left(\frac{A-C}{2}\right)^2+B^2$$

Get the eigenvalues from

$A+C<0$	$\lambda_2=\frac{A+C}{2}-\sqrt{\Delta}$ $\lambda_1=\frac{AC-B^2}{\lambda_2}$
$A+C>0$	$\lambda_1=\frac{A+C}{2}+\sqrt{\Delta}$ $\lambda_2=\frac{AC-B^2}{\lambda_1}$

Get the unit length eigenvectors from normalizing the vectors from the choices:

	$\mathbf{v}_1$	$\mathbf{v}_2$
$C-A>0$	$\begin{bmatrix} B & \frac{C-A}{2}+\sqrt{\Delta} \end{bmatrix}$	$\begin{bmatrix} -\left(\frac{C-A}{2}+\sqrt{\Delta}\right) & B \end{bmatrix}$
$C-A<0$	$\begin{bmatrix} -\left(\frac{C-A}{2}-\sqrt{\Delta}\right) & B \end{bmatrix}$	$\begin{bmatrix} B & \frac{C-A}{2}-\sqrt{\Delta} \end{bmatrix}$

So transformed roots are from

$$\frac{\hat{x}}{\hat{w}} = \pm\sqrt{\frac{\lambda_1}{\lambda_2}}$$

And original roots look something like:

$$\begin{bmatrix} x & w \end{bmatrix} = \hat{x}\mathbf{v}_1 + \hat{w}\mathbf{v}_2$$

Geometric interpretation; the transformation *rotates* the vector $\begin{bmatrix} A & B & C \end{bmatrix}$ about the line $A+C=0$ until it lies in the B=0 plane.
Not sure if this buys us anything.

Most general choice

Now here's the reason for this chapter. We will attempt to catalog *all* transformations that result in $\hat{B}=0$. We can write the formulas for $\hat{A},\hat{B},\hat{C}$ as

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B} & \hat{C} \end{bmatrix} = \begin{bmatrix} a^2A+2abB+b^2C & c(aA+bB)+d(aB+bC) \\ c(aA+bB)+d(aB+bC) & c^2A+2cdB+d^2C \end{bmatrix}$$

First, pick any values for *a,b* (this selection will effectively parameterize our class of solutions). Then find the appropriate values for *c,d*. to make $\hat{B}=0$. These will be

$$c = -(aB+bC)$$
$$d = (aA+bB)$$

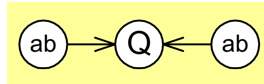
Then

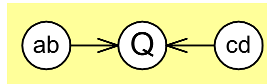
$$\begin{aligned}\hat{C} &= c^2A+2cdB+d^2C \\ &= (aB+bC)^2A - 2(aB+bC)(aA+bB)B + (aA+bB)^2C\end{aligned}$$

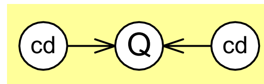
This mess can be factored. It happens that it equals

$$\hat{C} = \hat{A}\left(AC-B^2\right)$$

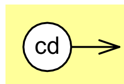
How do we know this? Here's where diagram techniques shine. In diagram terms we have

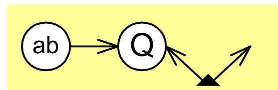
$\hat{A} =$ 

$\hat{B} =$ 

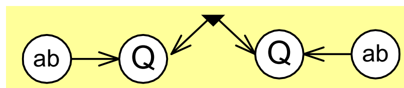
$\hat{C} =$ 

We have defined *cd* so that

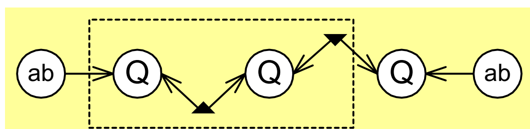


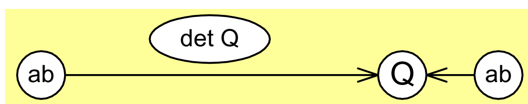
=


This makes $\hat{B} = 0$ since, in diagram form it gives the same glop on both sides of an epsilon.

$\hat{B} =$ 

Then plugging the definition of *cd* into the definition of $\hat{C}$ gives us

$\hat{C} =$ 

= 

The factorization of $\hat{C}$ has become almost trivial.
This means we have reduced the transformed matrix to:

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B} & \hat{C} \end{bmatrix} = \hat{A}\begin{bmatrix} 1 & 0 \\ 0 & AC - B^2 \end{bmatrix}$$

The transformed equation is:

$$\frac{\hat{x}}{\hat{w}} = \pm\sqrt{-\frac{\hat{C}}{\hat{A}}} = \pm\sqrt{B^2 - AC}$$

$$\begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{B^2 - AC} & 1 \end{bmatrix}$$

and going back to the original coordinate system gives us

$$\begin{aligned}[x \quad w] &= \left[\pm\sqrt{B^2-AC} \quad 1\right]\begin{bmatrix} a & b \\ -aB-bC & aA+bB \end{bmatrix} \\ &= \left[\pm a\sqrt{B^2-AC}-aB-bC \quad \pm b\sqrt{B^2-AC}+aA+bB\right] \\ &= a\left[-B\pm\sqrt{B^2-AC} \quad A\right]-b\left[C \quad -B\mp\sqrt{B^2-AC}\right]\end{aligned}$$

We can see that these are the two solution schema we chose between in the earlier section, blended by a and $-b$. We found that, effectively, we needed to get one root from one choice of (a,b) and the other root from a different choice of (a,b).

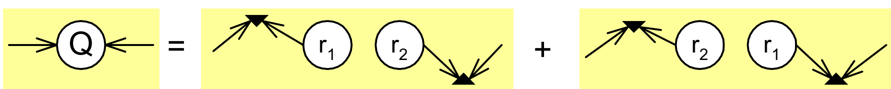
Watch the determinant

Note that the determinant of the transformation matrix is

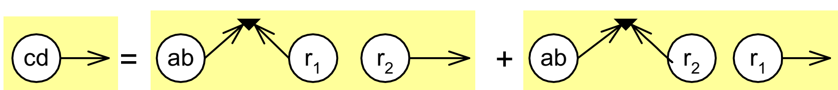
$$\det\begin{bmatrix} a & b \\ -aB-bC & aA+bB \end{bmatrix} = a^2A+2abB+b^2C=\hat{A}$$

So the only time this won't work is if the (a,b) we picked was already a root of **Q**.

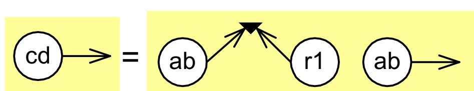
Another way to look at this is by saying that **Q** is composed of the outer product of its two roots, symmetrized.



We have constructed cd to be **Q** multiplied in on the left side by ab, and by an epsilon in on the right. The new epsilons merge with the existing ones and we have:



SO if ab equals one of the roots, say r_2, then the second term of this is zero and



That is, cd (the second row of the matrix) is a scalar times ab (the first row of the matrix) and the matrix is singular

Optimizing a,b

We can pick any two random values for a and b above as long as they themselves are not roots of the quadratic. Typically we pick either (1,0) or (0,1) depending on the sign of B. The various values we pick will all generate a roots that are homogeneously identical, that being:

We can also pick them via a criterion that optimizes the length of the x,w vector.

$$[x \quad w]=\left[a\left(\pm\sqrt{B^2-AC}-B\right)+b(-C) \quad a(A)+b\left(\pm\sqrt{B^2-AC}+B\right)\right]$$

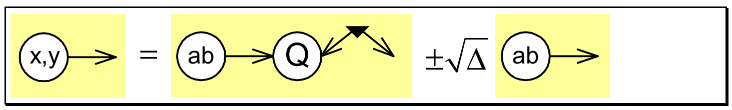
etc…

Roots in Diagram Notation

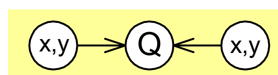
Let's play around with our general result some more. Moving the minus from b to inside the vector it multiplies we have

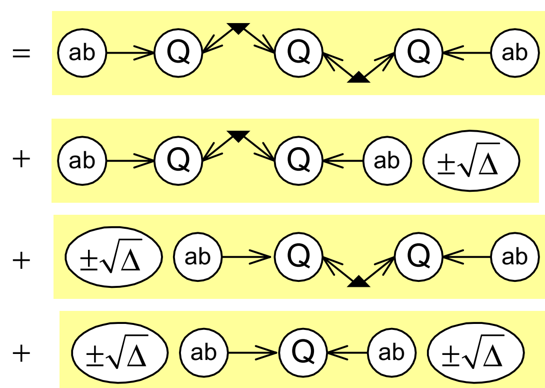
$$\begin{bmatrix} x & w \end{bmatrix} = a\begin{bmatrix} -B \pm\sqrt{B^2 - AC} & A \end{bmatrix} + b\begin{bmatrix} -C & B \pm\sqrt{B^2 - AC} \end{bmatrix}$$
$$= \begin{bmatrix} a & b \end{bmatrix}\left\{\begin{bmatrix} -B & A \\ -C & B \end{bmatrix} \pm\sqrt{B^2 - AC}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\}$$

Where $[a,b]$ can by *any* vector that doesn't make this zero. This is important because the matrix in curly brackets is singular. Now note the following interesting fact: the first matrix in curly brackets is just **Q** times epsilon. We can therefore write the solution in diagram form (using Δ for the discriminant (which is minus the determinant of **Q**)) as:



We can see that this solves the quadratic by plugging it in:



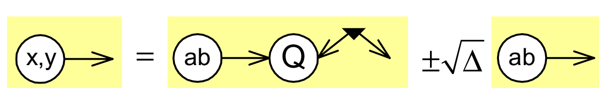


The second and third terms are identically zero since they are identical glop on either side of an epsilon. And the first and last terms since $\Delta = -\det \mathbf{Q}$.

This is amusing to be sure, but it becomes much more relevant when we generalize to higher dimensions.

Interpretation

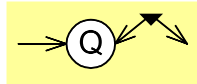
Let's get some more insight into what this means. The diagram form of the root



is the product of an arbitrary parameter "point" $\begin{bmatrix} a & b \end{bmatrix}$ with a matrix that we can write as:

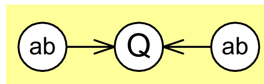
$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix}\left\{\mathbf{T} \pm \sqrt{\Delta}\mathbf{I}\right\}$$

Where

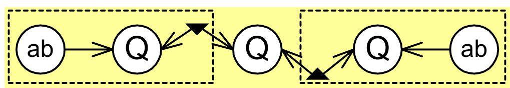
T = 

$$\mathbf{T} = \begin{bmatrix} -B & A \\ -C & B \end{bmatrix}$$

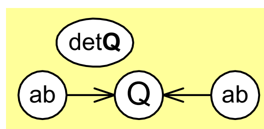
Let's seek some intuition into what **T** does. We first see that it is a mixed tensor (one arrow in and one arrow out) so it is a parameter transformation matrix; it maps an $\begin{bmatrix} a & b \end{bmatrix}$ value into another value $\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}$. How does this new point relate to **Q**? Let's compare the "before" quantity



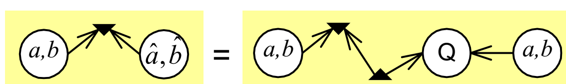
with the "after" quantity (dotted lines represent the transformed [a,b]



An identity transforms this to



In order for there to be real roots, the discriminant must be positive, meaning the determinant must be negative. If we consider **Q** as dividing the 1DHomogeneous parameter space into two regions (where $\mathbf{PQP}^{\mathrm{T}}$ is positive or negative), then the transformation **T** "inverts" the two regions; if a point **P** is in the positive region, the transformation **PT** is in the negative region (and vice versa). Points *on* **Q** will remain unmoved by **T** (except for a homogeneous scale). We can see this immediately by looking at the epsilon product of $\begin{bmatrix} a & b \end{bmatrix}$ and $\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}$



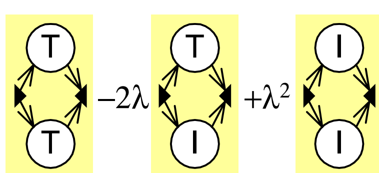
So if $\begin{bmatrix} a & b \end{bmatrix}$ is on **Q** then $\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}$ is homogeneously equal to it. Points on **Q** are just the roots of the polynomial. In other words

*The roots of **Q** are the eigenvectors of **T***

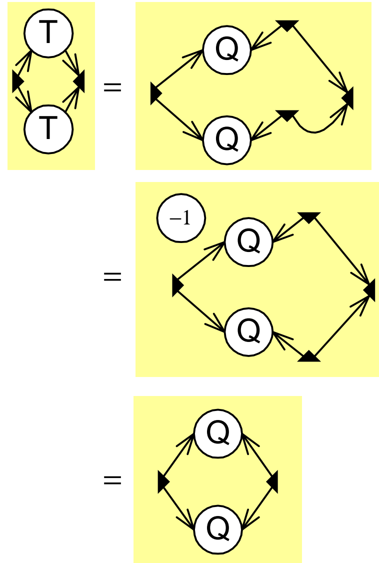
The characteristic equation for the eigenvalues of **T** is

$$\det\{\mathbf{T} - \lambda \mathbf{I}\} = 0$$

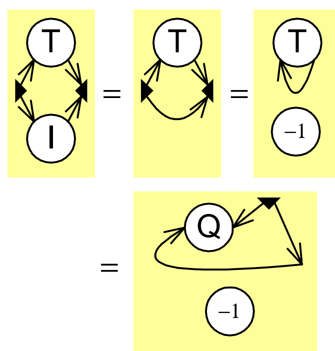
In diagram form this expands to:



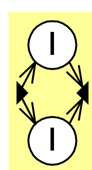
The three individual diagram fragments evaluate to:



$= 2\Delta$



$= \; 0$



$= -2$

The characteristic equation is then

$$2\Delta - 2\lambda^2 = 0$$

So we have

*The eigenvalues of **T** are* $\lambda = \pm\sqrt{\Delta}$

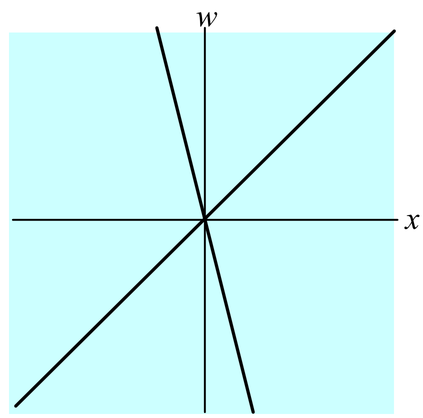
Compare this with the expression for the roots of **Q**

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \left\{ \mathbf{T} \pm \sqrt{\Delta}\mathbf{I} \right\}$$

The matrix in curly braces is just the matrix we took the determinant of to get the characteristic equation for **T**. It is singular; it maps any [a,b] to a root of **Q**, except for the (two?) choices of [a,b] that are perpendicular to a root of **Q**.

Geometric Interpretation

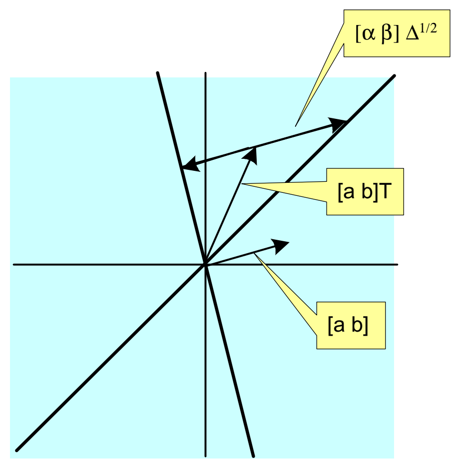
Suppose the two roots of a quadratic are represented by the two dark lines in the following [x w] space plot



Given that roots of **Q** satisfy

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \mathbf{T} \pm \sqrt{\Delta} \begin{bmatrix} a & b \end{bmatrix}$$

We can illustrate this relationship for a particular [a b] as:



For a different [ab] we get…

But if we pick an [a b] that is itself a root of **Q** we are unable to generate the other root since the diagram will look like…

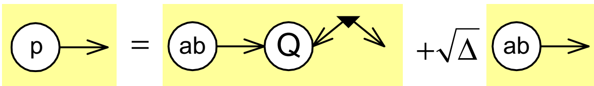
The numerical problem solution is where b^2>>ac and this is where… Interestingly it is not the choice that is +- that screws up, but the one that is small (???).

Dividing Out A Root

The following problem (and its generalization to higher order polynomials) will come up in later discussions: We want to solve it in diagram terms:

Given a quadratic **Q** and one of its roots **p**, find the other root **q**:

We can, of course, solve this by simply finding the roots as above and matching one with **p**. But it happens that there is a better way. We first note that:



Signs of roots

See section on Quartic polynomials

Chapter 1-04

1DH(2D)
Resultant of 2 Quadratics

The resultant of two polynomials is a function of their coefficients that indicates whether they have a common zero. We will start by constructing this function for some simpler polynomial pairs where one of the polynomials is linear. These calculations are somewhat trivial, but they will be of use in the higher order problem with two quadratics. We then generate the resultant of two quadratics by a procedure that we will later generalize to find resultants of higher order polynomials.

Two linear equations

Is there a common root of the two linear equations?

$$K_1x + K_0w = 0$$
$$L_1x + L_0w = 0$$

If these are both zero, then any linear combination is also zero:

$$L_1\left(K_1x + K_0w\right) - K_1\left(L_1x + L_0w\right) = 0$$
$$K_1L_1x + K_0L_1w - K_1L_1x - K_1L_0w = 0$$
$$\left(K_0L_1 - K_1L_0\right)w = 0$$

So if both of the original equations are zero for some x,w then the parenthesized expression above has to be zero. This, then, is our formula for the resultant of two linear polynomials.

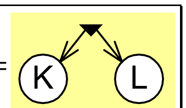
$$\boxed{\rho\left(K,L\right) = K_0L_1 - K_1L_0}$$

Since a linear polynomial has only one root, it is not surprising that the resultant is the formula for homogeneous equality of the two polynomials.

We can write the resultant in matrix form:

$$\begin{bmatrix} K_1 & K_0 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} L_1 \\ L_0 \end{bmatrix} = 0$$

Diagram notation the resultant looks like

$\rho\left(K,L\right) =$ 

Linear and Quadratic

Find conditions for a common root to a linear and a quadratic polynomial

$$Q_2x^2 + 2Q_1xw + Q_0w^2 = 0$$
$$L_1x + L_0w = 0$$

Again take linear a combo:

$$L_1\left(Q_2x^2+2Q_1xw+Q_0w^2\right)-Q_2x\left(L_1x+L_0w\right)=0$$
$$\left(Q_2L_1x^2+2Q_1L_1xw+Q_0L_1w^2\right)-\left(Q_2L_1x^2+Q_2L_0xw\right)=0$$
$$\left(2Q_1L_1-Q_2L_0\right)x+Q_0L_1w=0$$

We now have two linear equations

$$\left(2Q_1L_1-Q_2L_0\right)x+Q_0L_1w=0$$
$$L_1x+L_0w=0$$

Plug this into the formula we got in the previous section for two linear polynomials:

$$\rho\left(L,Q\right)=\left(Q_0L_1\right)L_1-\left(2Q_1L_1-Q_2L_0\right)L_0$$

Simplify a bit and the final resultant formulation is

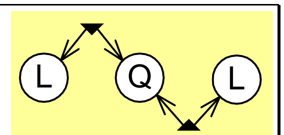
$$\boxed{\rho\left(L,Q\right)=Q_0{L_1}^2-2Q_1L_1L_0+Q_2{L_0}^2}$$

Write as matrix:

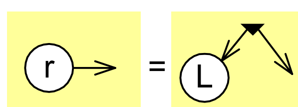
$$\begin{bmatrix}L_1 & L_0\end{bmatrix}\begin{bmatrix}Q_0 & -Q_1\\ -Q_1 & Q_2\end{bmatrix}\begin{bmatrix}L_1\\ L_0\end{bmatrix}=$$
$$\begin{bmatrix}L_1 & L_0\end{bmatrix}\begin{bmatrix}Q_2 & Q_1\\ Q_1 & Q_0\end{bmatrix}^*\begin{bmatrix}L_1\\ L_0\end{bmatrix}=0$$

That is, L times the adjoint of the Q matrix. The diagram looks like

$\rho\left(L,Q\right)=$

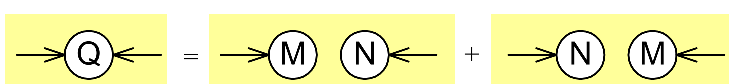


Of course this is perfectly obvious if we recall that the diagram form of the root, *r*, of linear equation **L** is:



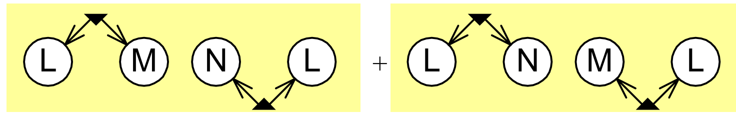
The resultant is, therefore, just seeing if this root also satisfies **Q**.

Another way to look at this is to visualize the "internal structure" of Q as the symmetrized outer product of the two roots, call them M and N:

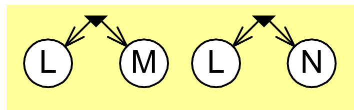


So the resultant actually calculates:

$\rho(Q,L) =$



$= 2$



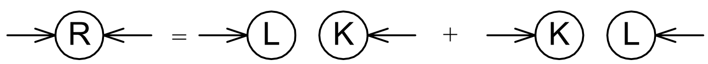
We can interpret this as follows: It is the product of the (homogeneous) differences of the factors of Q (these are M and N) with the factors of L (this is just L).

Quadratic and Quadratic

Now lets find the resultant of two quadratics

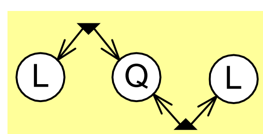
$$Q(x,w) = Q_2 x^2 + 2Q_1 xw + Q_0 w^2$$
$$R(x,w) = R_2 x^2 + 2R_1 xw + R_0 w^2$$

We will do this by visualizing the internal structure of Q as we did above (with roots M and N) and R as being the symmetrized outer product of its two roots, call them L and K:

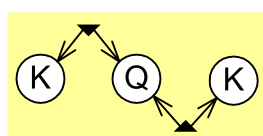


The desired resultant is the products of all the root pairs between the two (K,M) (K,N) (L,M) (L,N). Of course we want this in terms of the quadratics Q and R rather than their roots. We already know how to bunch some of these together. We know that:

$\rho(L,Q) =$

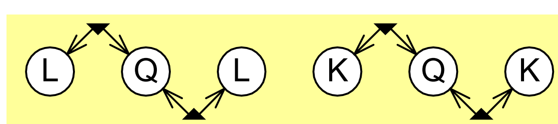


$\rho(K,Q) =$



The first of these represents the root-products (L,M) and (L,N). The second represents the root products (K,M) and (K,N). So we can write the desired resultant without needing to know the roots of Q

$\rho(Q,R) =$

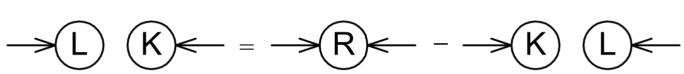


(0.1)

Now we want to fiddle with this to make it not explicitly contain the roots L and K but instead only the quadratic R.

General strategy

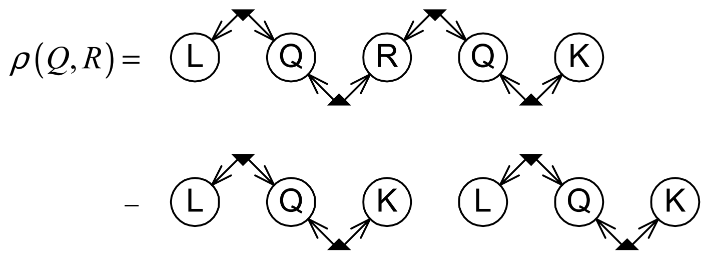
The basic technique is to use the following variant of the internal structure of R:



We look for a L,K pair in the diagram and replace it with the right hand side

Step 1

Apply this to the middle LK pair in equation of Equation (0.1):

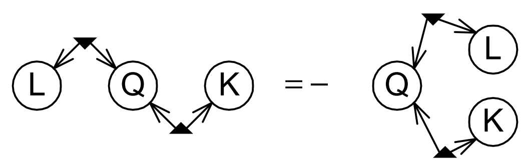


(0.2)

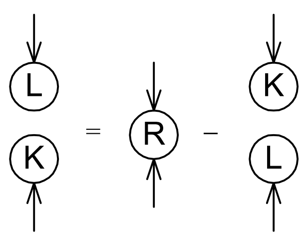
We have now removed one of the LK pairs from the first term and replaced it with an R. We then continue with this process hoping that we can replace all such LK pairs with R's. The technique requires some intuition on the substitutions we can do so it does not yet qualify as an algorithm. In fact, many of the substituions you will see below requires a lot of fiddling and dead ends. I am just showing you the paths that worked. I have the hope that port-mortem examination of these manipulations will ultimately lead to an algorithmic recipe.

Step 2

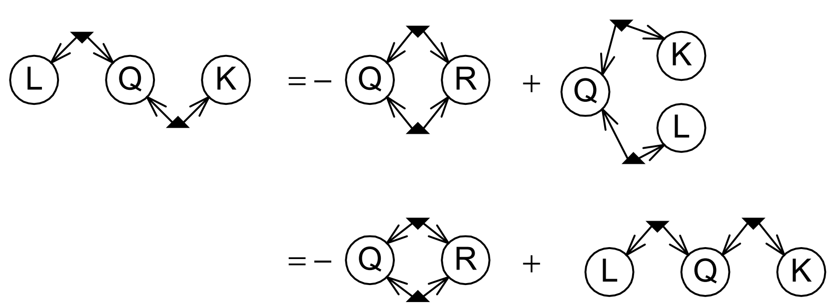
Look at the diagram fragment from the second term of (0.2)



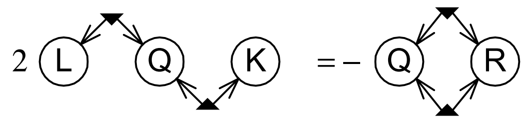
Apply our R equation in a slightly different orientation:



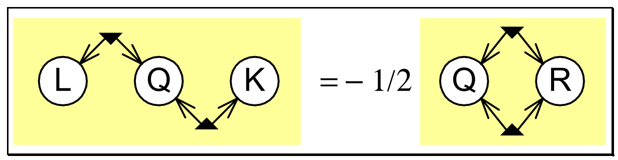
This gives



The last term of this (with an epsilon mirror and a sign flip) is equal to the lhs of the equation. We can then move it over the equal sign and get

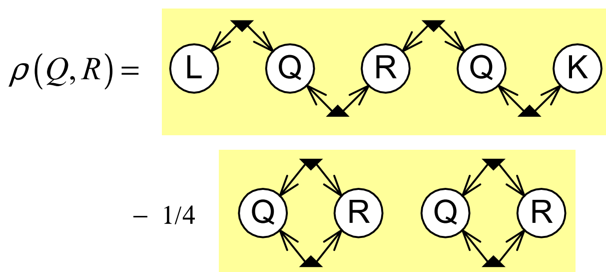


I will immortalize this as the identity



(0.3)

Our Equation (0.2) now becomes:

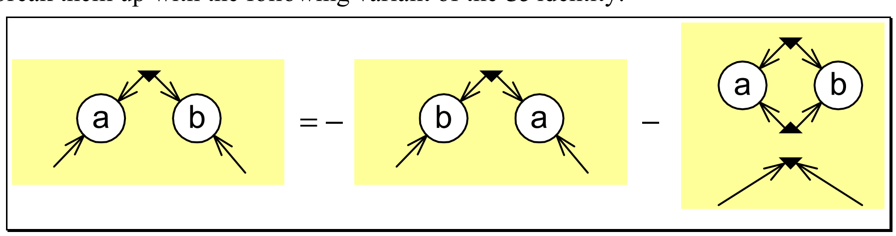


(0.4)

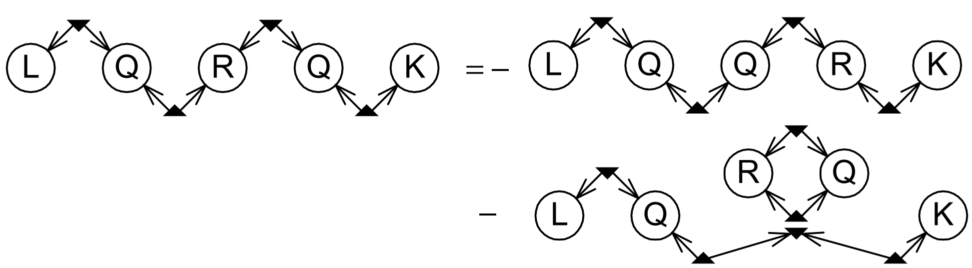
Only one LK pair left…

Step 3

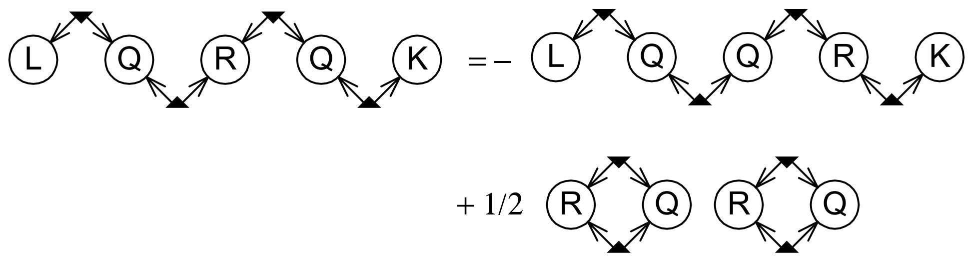
Looking at the first term of (0.4) we see the sequence QRQ. Whenever we see such chains we should have the urge to break them up with the following variant of the $\varepsilon\delta$ identity:



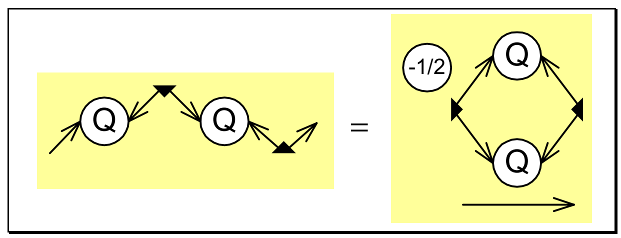
In this case this gives us:



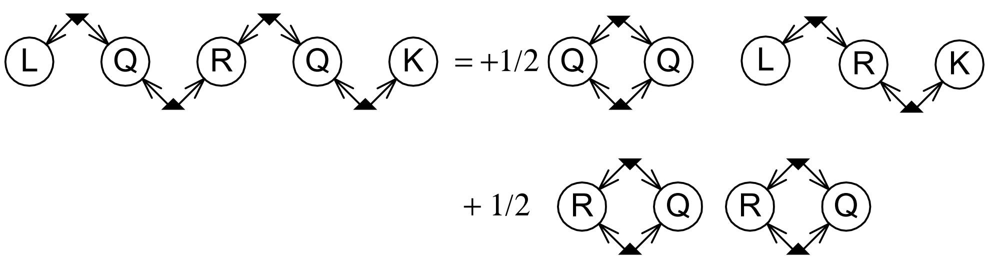
The last term of this simplifies via our new identity (0.3) to give



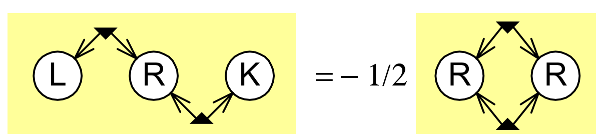
Note that in the first term of the rhs the chain has changed from QRQ to QQR. We then apply the identity



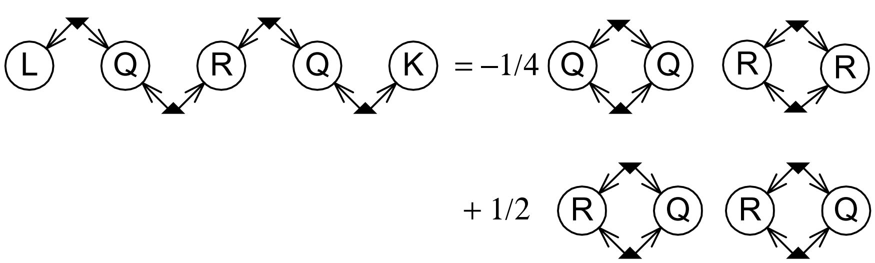
and get



Finally, we note that in the derivation of identity (0.3) the Q node is just a placeholder. We never used its contents; it can contain anything, even R. In other words we have:



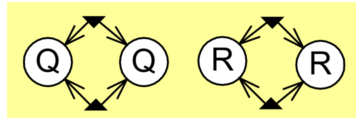
So we have boiled down our chain LQRQK to

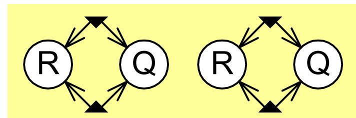


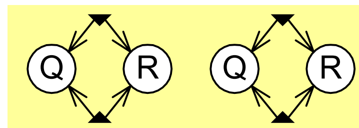
We have successfully gotten rid of all L and K nodes.

Step 4

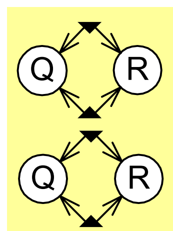
Put this together with equation (0.4) and you get:

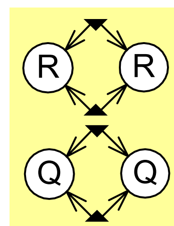
$\rho(Q,R) = \quad -1/4$ 

$+\ 1/2$ 

$-\ 1/4$ 

Which cleans up to our final answer:

$\rho(Q,R) = \ 1/4 \Big\{$ 

$-$ 

$\Big\}$

Another Way

There is another way to look at this problem that might generalize more easily to higher order functions. If the two quadratics have a common root then any linear combination of them will also have that root. Let us, then, look at the root structure of the quadratic

$$\alpha\mathbf{Q} + \beta\mathbf{R}$$

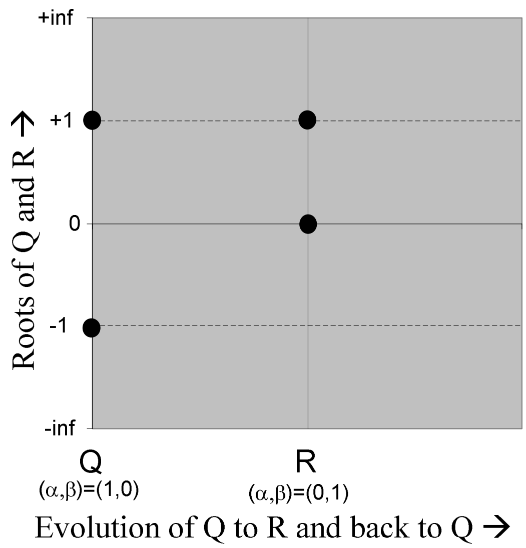
In order to see all the roots, including infinite roots (with w=0) I will map them into a finite range by plotting

$$\tan^{-1}(x/w)$$

Likewise we can think of (alpha,beta) as a sort of homogeneous coordinate of the blend so I will plot that axis as

$$\tan^{-1}(\alpha/\beta)$$

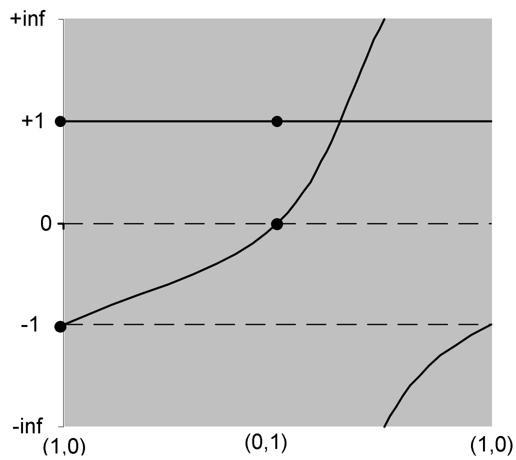
For example if **Q** has roots (+1) and (-1) and **R** has the roots (+1) and (0) the plot would look like



We then have three situations

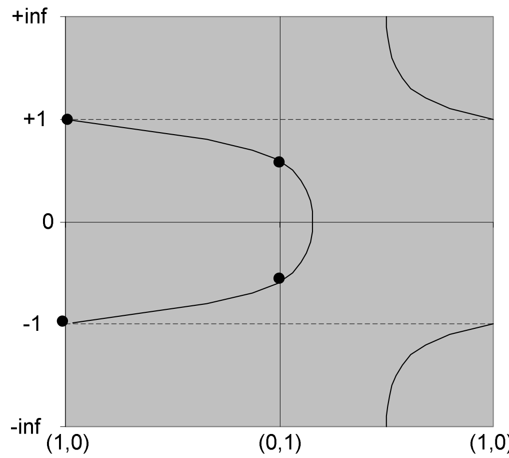
Q and R have a common root

The common root in this example is +1. The other root interpolates from (-1) at Q up to (0) at R, wraps around infinity and interpolates back to (-1) when we get back to Q.



Both roots of R lie between the roots of Q

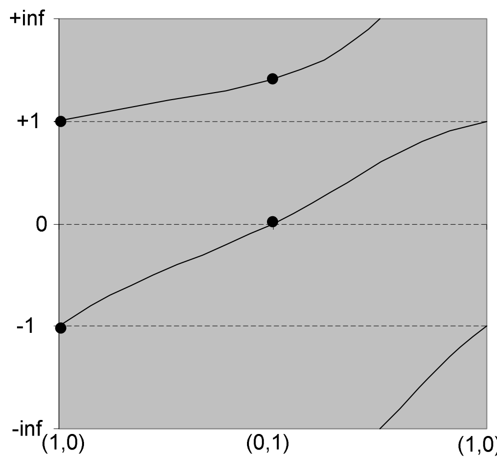
The evolution looks as follows. Note that there are quadratics on the interpolation that have no real roots.



The roots of Q and R are interleaved

This looks as below. All points along the interpolation have two distinct real roots.

$$\det(\alpha Q + \beta R) =$$

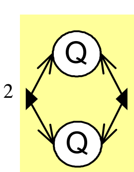


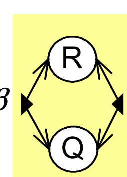
Detection of which case

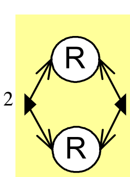
We can tell if we are in the first of these situations by seeing if the interpolation can have a double root for exactly one value of (alpha, beta). Double roots happen at a particular (alpha, beta) if

$$\det(\alpha\mathbf{Q} + \beta\mathbf{R}) = 0$$

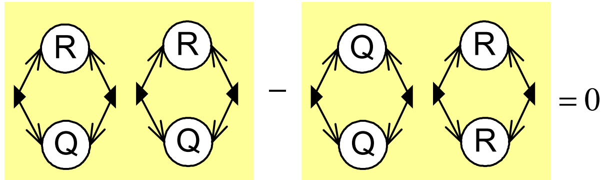
This is a quadratic function of (alpha, beta). Our diagram techniques show us that this is:

$\det(\alpha\mathbf{Q} + \beta\mathbf{R}) = \alpha^2$ 

$+2\alpha\beta$ 

$+\beta^2$ 

So the condition that this has a double root as a function of (alpha, beta) is that



This is just our resultant diagram again. This also shows that the sign of the resultant can tell us which of the other two cases we have: interleaved, vs. enclosed roots.

Discriminant of Cubic Polynomial

Once we have the resultant of two quadratics, the discriminant of a homogeneous cubic is not far behind. The cubic is:

$$\mathbf{C}(x,w) = Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$$

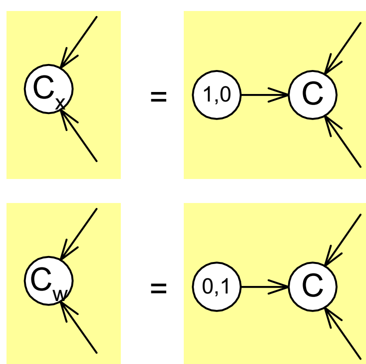
In general, the discriminant of a polynomial is the resultant of the polynomial and its derivative. An alternative in the homogeneous case is to just take the resultant of the two partial derivatives wrt *x* and *w*.

$$\mathbf{C}_x = 3Ax^2 + 6Bxw + 3Cw^2$$
$$\mathbf{C}_w = 3Bx^2 + 6Cxw + 3Dw^2$$

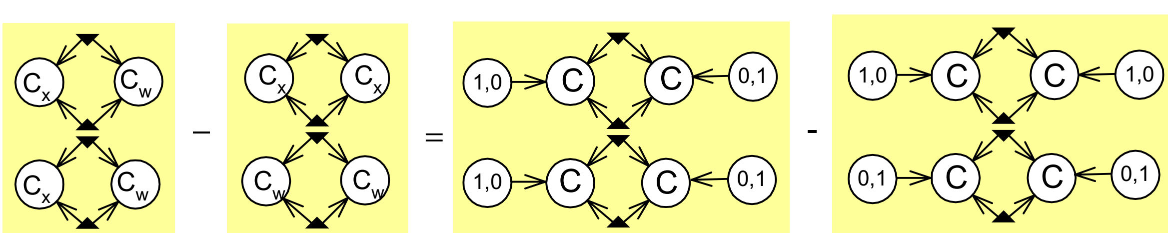
Any common roots between $\mathbf{C}_x$ and $\mathbf{C}_w$ are also shared by **C** because of the identity:

$$x\mathbf{C}_x + w\mathbf{C}_w = 3\mathbf{C}$$

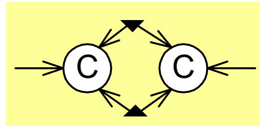
In diagram terms, the two partial derivatives (as quadratic polynomials) look like:



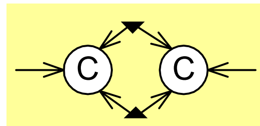
Then plugging this in to our diagram for the resultant of $\mathbf{C}_x$ and $\mathbf{C}_w$ gives the following (I will drop the factor of ¼ for now):



The rhs of this has various instances of the 2x2 matrix

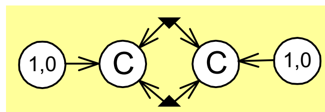


Giving names to these elements we have

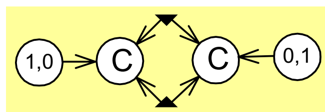


$$= \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

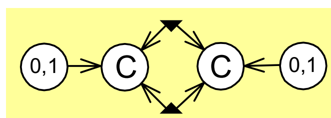
Then we can see that



$= A$



$= B$



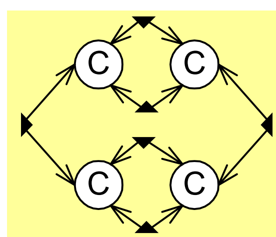
$= C$

In other words, the discriminant is

$$discr(\mathbf{C}) = B^2 - AC \tag{0.5}$$

This is (minus) the determinant of the 2x2 matrix. Using our diagram form for the determinant of a 2x2 matrix gives us:

$discr(\mathbf{C}) =$

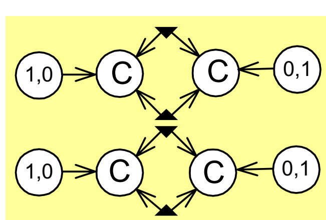


which we have seen in part one of these notes.

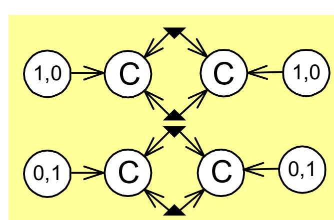
Another way

The recognition of the formula for the determinant in (0.5) was easy because we had a 2x2 matrix. I want to re-derive this in a more general manner. We start with

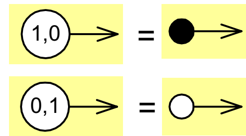
$discr(\mathbf{C}) =$



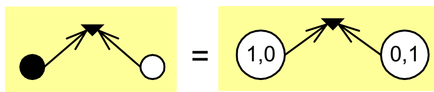
-



To make things a bit more visible I will use the abbreviation:



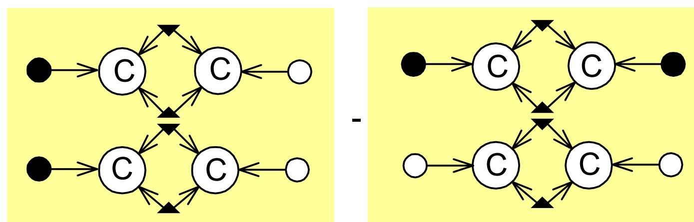
and note that:



$$= \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}0\\ 1\end{bmatrix} = 1$$

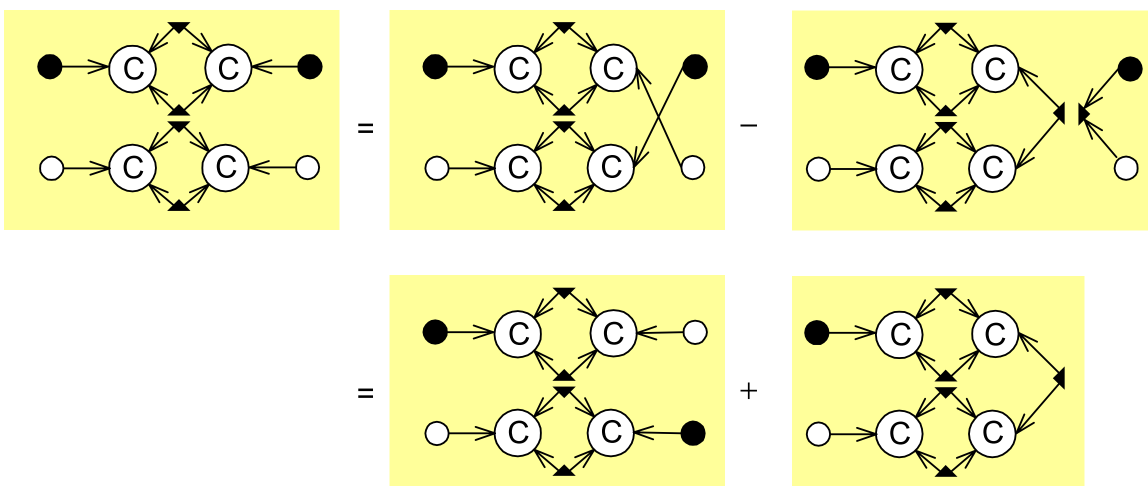
We then have

$discr(\mathbf{C}) =$



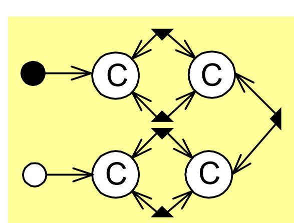
(0.6)

Apply $\varepsilon\delta$ to the rightmost term of this:

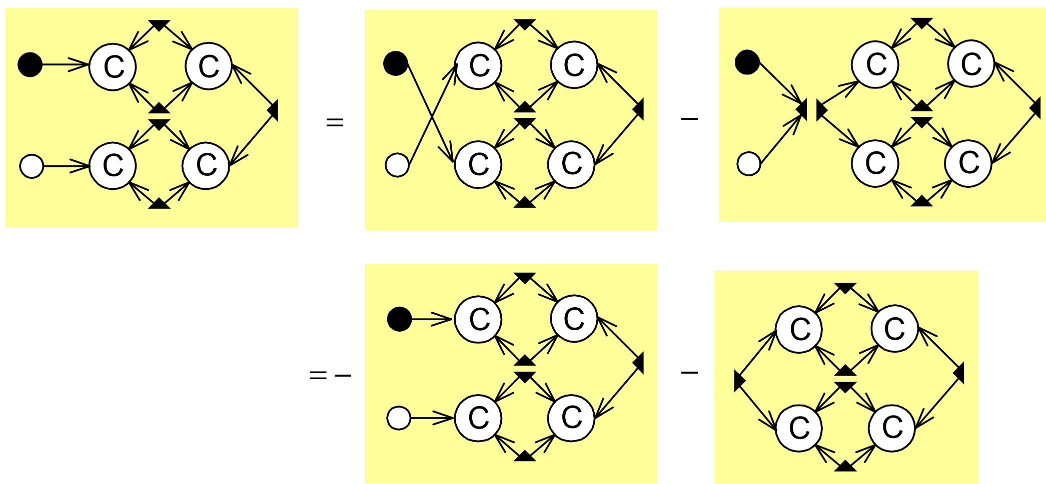


The first term of this nicely cancels out the first term of (0.6) and we have

$discr(\mathbf{C}) =$ -

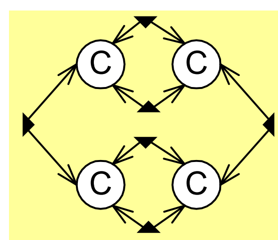


Now do another $\varepsilon\delta$ on this term:



Notice that the first term on the rhs was mirrored (with the introduction of a minus sign since it has an odd number of epsilons) to get the version on the second line. Move this term over the equals getting twice the lhs. Move the 2 to the rhs and we finally net:

$$discr(\mathbf{C}) = 1/2$$



Aside from a constant factor and perhaps a global minus sign (which I still have to clean up in these notes) we get our answer again.

Chapter 1-05

1DH(2D)
Relationship Between Two Quadratic Polynomials

There are some important principles in the study of homogeneously represented geometry that first become apparent when we consider the relationship between pairs of Quadratic Polynomials. Some of these might seem fairly trivial in this context, but they lay the groundwork for more complex situations.

To recap:

We are given two quadratics (slight change of notation)

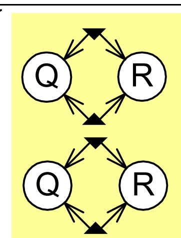
$$Q(x,w) = Ax^2 + 2Bxw + Cw^2$$
$$R(x,w) = Dx^2 + 2Exw + Fw^2$$

Or in matrix form

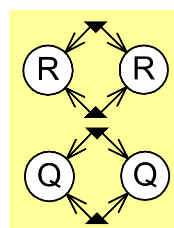
$$Q(x,w) = \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$
$$R(x,w) = \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} D & E \\ E & F \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

We found that the two polynomials share a common root if the resultant is zero. In diagram form, the resultant is equal to

$\rho(Q,R) = 1/4 \Big\{$



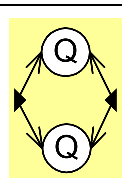
$-$

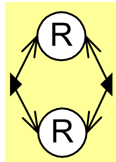
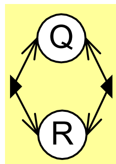


$\Big\}$

The Invariance of Invariants

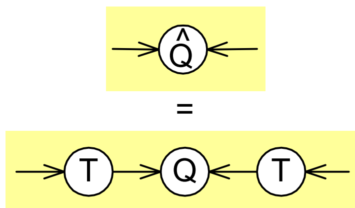
Let's take a closer look at the three diagram fragments from the calculation of the resultant. I will tabulate them along with their explicit value

Diagram	Value
	$-2\left(AC - B^2\right)$

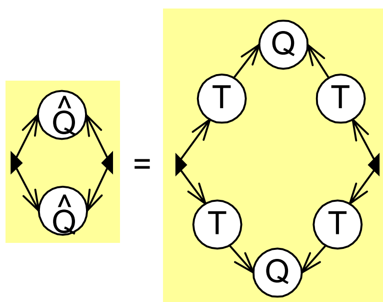
	$-2\left(DF-E^2\right)$
	$-AF+2BE-CD$

Transformation Invariance

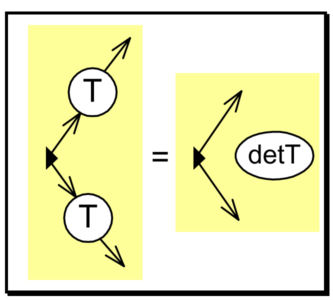
There is an interesting property that each of these three quantities shares: They are invariant (or at least their sign is invariant) under coordinate transformations. To see why, let's see what happens if we take the pair **Q,R** and transform them both by some transformation **T.** We have:



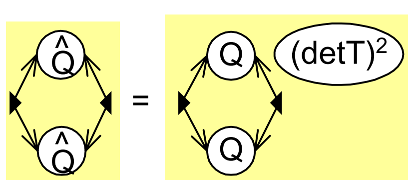
Plug this into the first diagram of the table and we get:



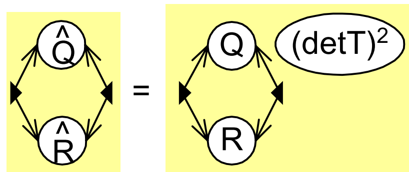
Now apply the identity

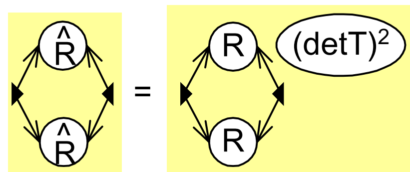


So we have:



and similarly:





What does this mean? Since the determinant of **T** appears only in squared form, it means that the signs of the three small diagrams don't change under the (same) coordinate transformation of (both) Q and R. The sign (or zeroness) of these three quantities is an inherent property of the relationship between Q and R and is not changed by simply reparametrizing them (both).

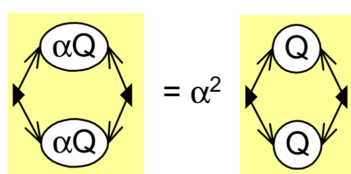
In fact, in Euclidean geometry we deal only with transformations that have unit determinant (they are pure rotations). Using this machinery in Euclidean geometry we can say even more. Not just the sign, but also the numerical value of the diagrams remains constant under geometric transformations.

This is a fundamental principle that we will bring with us both to higher order curves and to higher dimensionality:

> *Any diagram constructed of coefficient nodes glued together with epsilon nodes represents a transformationally invariant quantity. One of my main interests is to do two things: Given such a diagram, find out what geometric invariant it represents. And given a geometric invariant, construct a diagram that "signals" it.*

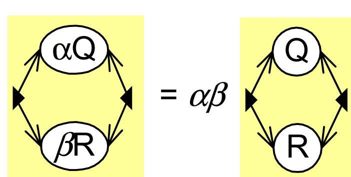
Homogeneous Scale Invariance

We are looking at arithmetic tests that purport to tell us something about the relationship between the roots of the two quadratics **Q** and **R**. Any such test must satisfy the homogeneous scaling condition: plugging in any nonzero scalar multiple of **Q** or any nonzero scalar multiple of **R** should give the same answer. Applying this to our transformational invariants gives the following:



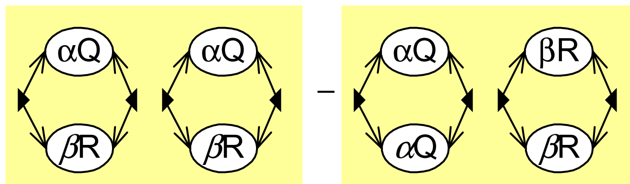
What this means is that, no matter how we scale **Q** the sign (or the zeroness) of this diagram will not change. In other words there are three states of this quantity that are preserved both by geometric transformations and by homogeneous scaling; it can be (`+,0,-`). A similar statement holds for the RR diagram.

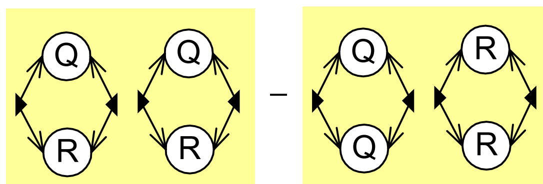
On the other hand, if we scale **Q** and **R** by different nonzero scales, the following diagram equation holds



By proper choice of α or β we could flip the sign of this diagram fragment. So the only thing that remains invariant under both transformation and homogeneous scaling is its zeroness; it has two states (zero, and nonzero)

Finally, the resultant itself contains this last term, but squared. This means that

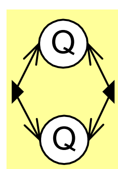
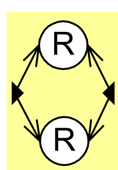
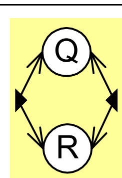
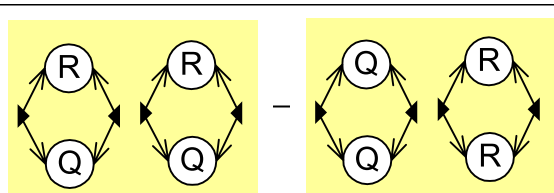


$= \alpha^2 \beta^2$ 

and we cannot flip the sign by any combinations of values for α β. The resultant has three states (+,0,-)

Net invariants

We therefore have four algebra-to-geometry indicators for the relation between Q and R.

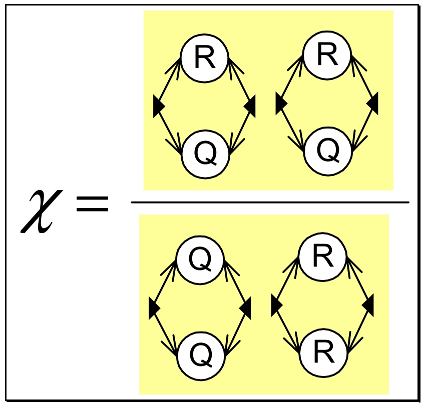
Invariant	Invariant values
	Negative, Zero, Positive
	Negative, Zero, Positive
	Zero, Nonzero
	Negative, Zero, Positive

There are potentially 3x3x2x3=54 different configurations of Q and R that are distinguishable by this technique. However, since the resultant is a function of the other three quantities, its sign is not independent of them, and there are fewer than 54.

But there's another calculation we can do that gives a continuous invariant.

A new type of cross ratio

Now lets take another look at our resultant of quadratics Q and R. Let us construct the diagram expression

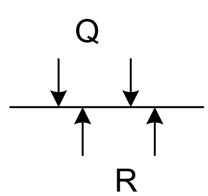


Now consider what happens to χ as Q and R are transformed geometrically by a matrix **T**. The factors of det**T** cancel out and χ retains its same numerical value. Now consider what happens if either Q or R (or both) are homogeneously scaled. The scale cancels out and χ retains its same numerical value. In other words χ is a new type of cross ratio. It's not obviously the ratio of four distances as with the 4-line case. But its quantitative numerical value, not just its sign or zeroness, is preserved under geometric transformations (in this case polynomial parameter transformations) and under homogeneous scaling of the coefficients. In particular, if $\chi = 1$, the resultant is zero and the polynomials have a common root.

These are the types of algebraic-geometric indicators that we are looking for in higher degree and dimensionality situations.

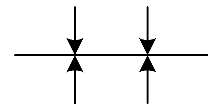
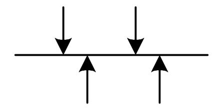
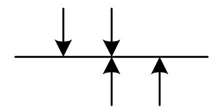
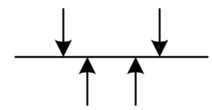
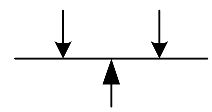
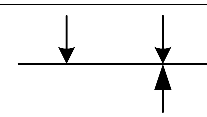
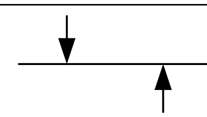
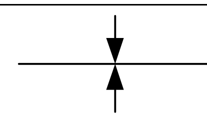
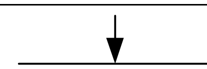
The possibilities

Let's take a look at the various possibilities for the complete relationship between the roots of these two equations. I will diagram the roots as follows



The possibilities are

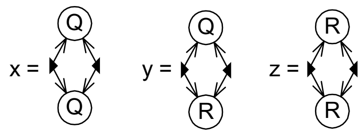
Table 1 – Possible Root Interactions				
Name	Nbr roots of **Q**	Nbr roots of **R**	Interaction	Example

22S			Both same	
22I		2	Interleaved	
22C			One common	
22E	2		Enclosed	
21D		1 double	Disjoint	
21C			Common	
20		0		
11D		1 double	disjoint	
11C	1 double		Common	
10		0		
00D	0	0	Same imaginary	
00C			Different imaginary	

(I have not included "dual" categories that are just swapping **Q** and **R**) Which of these 12 categories the pair **Q R** resides in is independent of any (common) homogeneous transform applied to them and so represents a homogeneously invariant property. The resultant being zero only distinguishes between the categories that are shaded above from those that are unshaded. This forces us to the conclusion that the resultant does not give the whole story about the relationship between the quadratics. Let us now try to relate these categories to invariant algebraic quantities.

Quadratic Phase Space

To see what's happening, let use the three-dimensional space defined by the three diagram invariants:

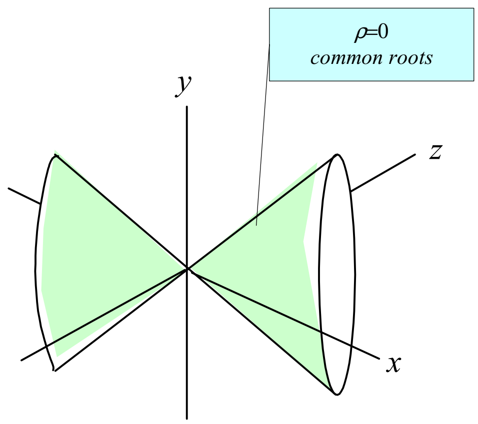


In terms of these coordinates the resultant and cross ratios are (neglecting a constant scale factor for the resultant).

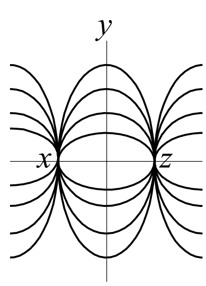
$$\rho(Q,R) = xz - y^2$$

$$\chi = \frac{xz}{y^2}$$

The $\rho(Q,R)=0$ surface will be a (non circular) cone with axis along the $x=z$ line.



This surface is where $xz = \chi y^2$ for $\chi = 1$. Other surfaces of constant χ are other cones along this same axis (for positive chi) or along the $x=-z$ line (for negative chi). Looking down the $x=z$ line these would look like



The x and z axes are embedded in all such cones since $0 = xz - \chi y^2$ is satisfied for x=y=0 or for y=z=0 for any value of chi.

Homogeneous Scaling

We can homogeneously scale the quadratic matrices without changing the location (or relationship between) the roots

$$\hat{\mathbf{Q}} = \alpha\mathbf{Q}$$
$$\hat{\mathbf{R}} = \beta\mathbf{R}$$

so that we get:

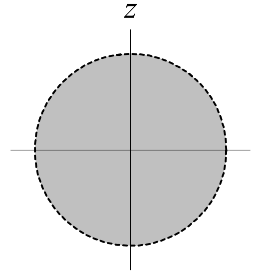
$$\hat{x} = \alpha^2 x$$
$$\hat{y} = \alpha\beta y$$
$$\hat{z} = \beta^2 z$$
$$\hat{\rho}(Q,R) = \alpha^2\beta^2\rho(Q,R)$$
$$\hat{\chi} = \chi$$

The sign (and zeroness) of the resultant doesn't change.
The value of the cross ratio doesn't change.

Map to unit half ball

We can pick α and β to map *xyz* space to a simpler $\hat{x}, \hat{y}, \hat{z}$ space. In particular we can attempt scale them so the $\hat{x}, \hat{y}, \hat{z}$ coordinates are on the unit ball. We can also pick signs of α and β to make $\hat{y}$ positive. We have collapsed *x,y,z* space down to the unit half ball. A top view looks like the unit disk with

$$\hat{y} = +\sqrt{1 - \hat{x}^2 - \hat{z}^2}$$



The curve of zero resultant

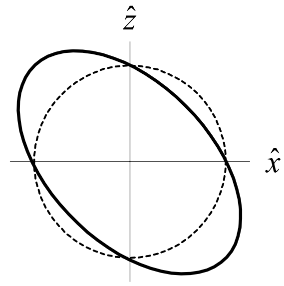
Let us project the intersection line of the zero-resultant cone and the unit sphere onto the $\hat{x}, \hat{z}$ diagram:

Projected onto this $\hat{x}, \hat{z}$ space the cone

$$\begin{aligned}\rho(Q,R) &= \hat{x}\hat{z} - \hat{y}^2 = 0 \\ &= \hat{x}\hat{z} - \left(1 - \hat{x}^2 - \hat{z}^2\right) \\ &= \hat{x}^2 + \hat{x}\hat{z} + \hat{z}^2 - 1\end{aligned}$$

This happens to be a tilted ellipse that passes thru the points

$$\begin{bmatrix}0 & \pm 1\end{bmatrix}, \begin{bmatrix}\pm 1 & 0\end{bmatrix}, \begin{bmatrix}\pm\sqrt{1/3} & \pm\sqrt{1/3}\end{bmatrix}, \begin{bmatrix}\pm 1 & \mp 1\end{bmatrix}$$



only the part inside the dotted line corresponds to the rho=0 surface.

Other curves of constant cross ratio

These curves satisfy

$$\chi\left(1-\hat{x}^2-\hat{z}^2\right)=\hat{x}\hat{z}$$

$$0=\hat{x}^2+\frac{1}{\chi}\hat{x}\hat{z}+\hat{z}^2-1$$

The following points are always on the curve independent of chi.

$$\left[\hat{x},\hat{z}\right]=\left[0,\pm 1\right] \text{ and } \left[\hat{x},\hat{z}\right]=\left[\pm 1,0\right]$$

A more general set can be recognized by substituting x=p+q, z=p-q.

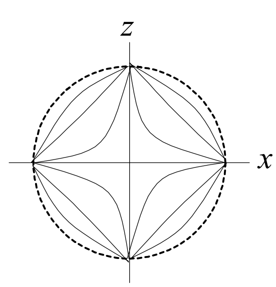
$$1=\left(p+q\right)^2+\frac{1}{\chi}\left(p+q\right)\left(p-q\right)+\left(p-q\right)^2$$

$$1=p^2+2pq+q^2+\frac{1}{\chi}\left(p^2-q^2\right)+p^2-2pq+q^2$$

$$1=p^2+q^2+\frac{1}{\chi}\left(p^2-q^2\right)+p^2+q^2$$

$$p^2\left(2+\frac{1}{\chi}\right)+q^2\left(2-\frac{1}{\chi}\right)=1$$

We can recognize this as an ellipse with eccentricity determined by chi. All such ellipses pass through the points mentioned above. Also have hyperbolas if chi is big enough. In the extreme when we have chi=0 we have the x and z axes. So lines of constant chi are:



χ=0 gives x and z axes

χ>0 gives curve in first and third quadrants

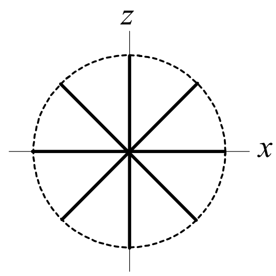
χ=infinity gives circle

$\chi<0$ gives curve in second and fourth quadrants

points (1,0,0) (-1,0,0) (0,0,1) (0,0,-1) are on all curves

More collapsing

By making α and β different we can stay on the unit ball but move points around on the constant χ surfaces (since for any α β the value of χ is constant). This means that we can transform any point that is not on the x or z axis onto one of the diagonals.



Now instead of a disk we have four intersecting lines. But we're not done yet. Lets look at the various special cases. We can catalog these points according to table 1.

Case (11C) x=y=z=0

Here, xyzhat cannot map to the unit ball. But this is a perfectly reasonable nondegenerate situation: it means that both Q and R have double roots and that they are the same. In other words Q and R are (homogeneously) the same.

Case (2..) x=0

Double root for Q

example

$$\mathbf{Q}=\left(\left(\frac{x}{w}\right)-q_1\right)\left(\left(\frac{x}{w}\right)-q_1\right)$$
$$=\left(\frac{x}{w}\right)^2-2q_1\left(\frac{x}{w}\right)^2+q_1^{\,2}$$
$$\triangleq x^2+w^2=\begin{bmatrix}1 & -q_1\\ -q_1 & q_1^{\,2}\end{bmatrix}$$

$$x=0$$
$$y=-aq_1^{\,2}-2bq_1-c$$
$$z=2\left(b^2-ac\right)$$

rescale Q and R

$$\hat{x}=0$$
$$\hat{y}=\alpha\beta y$$
$$\hat{z}=\beta^2 z$$

Subcase 2a Z=0

Double root for R

$$\hat{x} = 0$$
$$\hat{y} = \alpha\beta y$$
$$\hat{z} = 0$$

pick alpha beta to make yhat=1

$$\boxed{(\hat{x}, \hat{y}, \hat{z}) = (0,1,0)}$$

this is the case if Q and R each have double roots but they are different.

Subcase 2b Z>0

Two real roots for R

$$\hat{x} = 0$$
$$\hat{y} = \alpha\beta y$$
$$\hat{z} = \beta^2 z$$

Pick alpha beta to map y,z onto unit circle and to make them equal in magnitude. That equal magnitude then must be root(1/2)

$$\beta^2 = \frac{\sqrt{1/2}}{z}$$

$$y^2 + z^2 = 1 = \alpha^2\beta^2 y^2 + \beta^4 z^2$$

$$1 = \alpha^2 \frac{\sqrt{1/2}}{z} y^2 + \frac{1/2}{z^2} z^2$$

$$1/2 = \alpha^2 \frac{\sqrt{1/2}}{z} y^2$$

$$\frac{z}{y^2}\sqrt{1/2} = \alpha^2$$

So net cords are

$$\boxed{(\hat{x}, \hat{y}, \hat{z}) = \left(0, \sqrt{1/2}, \sqrt{1/2}\right)}$$

An appropriate alpha exists as long as y!=0

SubSubCase 2ba y=0

In this case we have **R**(Q,R)=0 so R shares a root with Q.

$$\mathbf{R}=\left(\left(\frac{x}{w}\right)-r_1\right)\left(\left(\frac{x}{w}\right)-q_1\right)$$
$$=\left(\frac{x}{w}\right)^2-\left(q_1+r_1\right)\left(\frac{x}{w}\right)^2+q_1r_1$$
$$\triangleq x^2+w^2=\begin{bmatrix}1 & -\frac{r_1+q_1}{2}\\ -\frac{r_1+q_1}{2} & q_1r_1\end{bmatrix}$$

$$x=0$$
$$y=-q_1{}^2-2\left(-\frac{r_1+q_1}{2}\right)q_1-q_1r_1=0$$
$$z=2\left(\left(\frac{r_1+q_1}{2}\right)^2-q_1r_1\right)=\frac{1}{2}\left(r_1-q_1\right)^2$$

then pick a,b to scale z^ to equal 1

$$\boxed{\left(\hat{x},\hat{y},\hat{z}\right)=\left(0,0,1\right)}$$

Subcase 2c Z<0

No real roots for R

$$\mathbf{R}=\left(\left(\frac{x}{w}\right)+r_R-ir_I\right)\left(\left(\frac{x}{w}\right)+r_R+ir_I\right)$$
$$=\left(\frac{x}{w}\right)^2+\left(\frac{x}{w}\right)2r_R+\left(r_R{}^2+r_I{}^2\right)$$
$$=\begin{bmatrix}1 & r_R\\ r_R & r_R{}^2+r_I{}^2\end{bmatrix}$$

then

$$x=0$$
$$y=-q_1{}^2-2r_Rq_1-\left(r_R{}^2+r_I{}^2\right)$$
$$=-q_1{}^2-2r_Rq_1-r_R{}^2-r_I{}^2$$
$$=-\left(q_1{}^2+r_R{}^2\right)-r_I{}^2$$
$$z=2\left(r_R{}^2-\left(r_R{}^2+r_I{}^2\right)\right)$$
$$=-2r_I{}^2$$

SubSubcase 2c1 Y=0

This can never happen unless (from above equation for y)

$$0=-\left(q_1{}^2+r_R{}^2\right)-r_I{}^2$$

this can only be the case if the root of R is double and equals the double root of Q. So the point (0,0,-1) cannot happen.

This is an interesting point, that there are some regions in x,y,z space that cannot be generated by any Q, R pair.

BUT can have this if Q=0 identically. Then x=y=0 and no restrictions on R and z

SubSubcase 2c2 Y!=0

Here we can scale to make:

$$\beta^2=-\frac{\sqrt{1/2}}{z}$$

$$y^2+z^2=1=\alpha^2\beta^2y^2+\beta^4z^2$$

$$1=-\alpha^2\frac{\sqrt{1/2}}{z}y^2+\frac{1/2}{z^2}z^2$$

$$1/2=-\alpha^2\frac{\sqrt{1/2}}{z}y^2$$

$$\frac{-z}{y^2}\sqrt{1/2}=\alpha^2$$

So net cords are

$$\boxed{\left(\hat{x},\hat{y},\hat{z}\right)=\left(0,\sqrt{1/2},-\sqrt{1/2}\right)}$$

Case 3 z=0

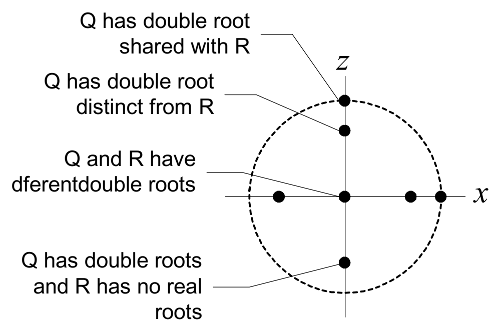
Same as above but swapping x and z.

$$\boxed{\left(\hat{x},\hat{y},\hat{z}\right)=\left(\sqrt{1/2},\sqrt{1/2},0\right)}$$

$$\boxed{\left(\hat{x},\hat{y},\hat{z}\right)=\left(1,0,0\right)}$$

$$\boxed{\left(\hat{x},\hat{y},\hat{z}\right)=\left(-\sqrt{1/2},\sqrt{1/2},0\right)}$$

The x and z axes have then collapsed to



Case 4a x,z are + + or --

Both Q and R have two distinct roots.

Pick alpha beta to make $x^2 = z^2$

$$\alpha^2 x = \beta^2 z$$

This means that for some gamma

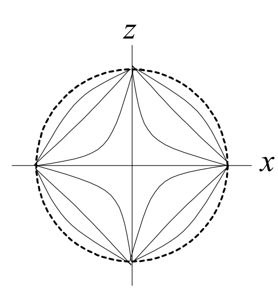
$$\alpha^2 = \gamma z$$
$$\beta^2 = \gamma x$$

And the unit ball requirement gives

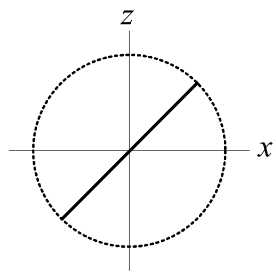
$$\left(\alpha^2 x\right)^2 + \left(\alpha\beta y\right)^2 + \left(\beta^2 z\right)^2 = 1$$
$$\left(\gamma z x\right)^2 + \left(\gamma\sqrt{zx}\,y\right)^2 + \left(\gamma x z\right)^2 = 1$$
$$\gamma^2 z^2 x^2 + \gamma^2 z x y^2 + \gamma^2 z^2 x^2 = 1$$
$$\gamma^2\left(2z^2x^2 + zxy^2\right) = 1$$
$$\gamma = \pm\frac{1}{\sqrt{2z^2x^2 + zxy^2}}$$

if z,x have same sign the square root is guaranteed to be real. The sign of gamma is equal to the sign of x and z.

What we are doing is sliding the points along the curve of constant chi



To the diagonal



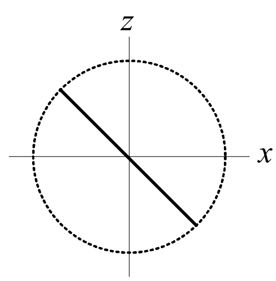
Point on resultant=0 surface if

$$\mathbf{R}(Q,R) = \hat{x}^2 - \hat{y}^2 = 0$$
$$\hat{x} = \hat{y} = \hat{z} = \sqrt{1/3}$$

what about points, though that are x=z=-root(1/3) ? This lies on R=0 surface so have common roots. The common roots are a complex conjugate pair though. This can only happen if Q and R are homogeneously equal.

Case 4b x,z are -+ or +-

Here one has no roots and the other has two. Can scale with alpha, beta to lie on back diagonal



Chapter 1-06

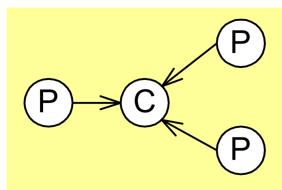
1DH(2D)
Roots of a Homogeneous Cubic Polynomial

The Problem

Find roots of a homogeneous cubic polynomial. That is, find $[x,w]$ such that

$$\begin{aligned}&Ax^3+3Bx^2w+3Cxw^2+Dw^3\\&=\begin{bmatrix}x & w\end{bmatrix}\begin{bmatrix}\begin{bmatrix}A & B\\B & C\end{bmatrix} & \begin{bmatrix}B & C\\C & D\end{bmatrix}\end{bmatrix}\begin{bmatrix}x\\w\end{bmatrix}\begin{bmatrix}x\\w\end{bmatrix}\\&=0\end{aligned}$$

In diagram form, find P



$= 0$

We would like this formulation to be symmetric in the coefficients A,D and B,C in the same manner as the solution we got for homogeneous quadratic polynomials.

Define Intermediate quadric

We have shown earlier that if we define

$$\Delta_1 = AC - B^2 = \det\begin{bmatrix}A & B\\B & C\end{bmatrix}$$

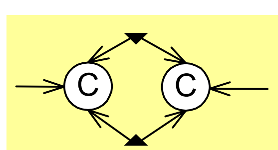
$$\Delta_2 = AD - BC = \det\begin{bmatrix}A & B\\C & D\end{bmatrix}$$

$$\Delta_3 = BD - C^2 = \det\begin{bmatrix}B & C\\C & D\end{bmatrix}$$

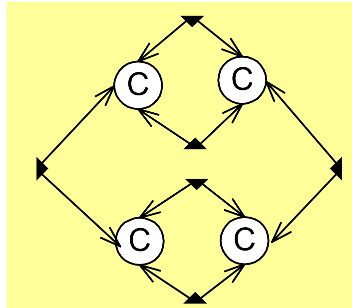
Then the cubic discriminant is

$$\Delta = -\det\begin{bmatrix}2\Delta_1 & \Delta_2\\\Delta_2 & 2\Delta_3\end{bmatrix}$$

The diagram form of this matrix is



So, the discriminant, the determinant of this (with homogeneous scale of 2) is

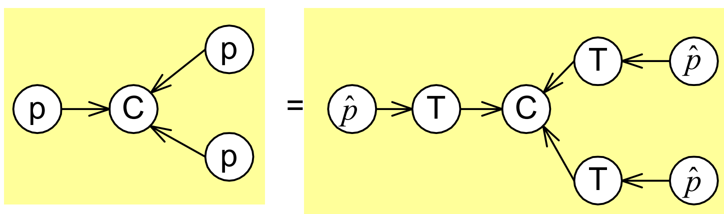


Define transformation of parameter space

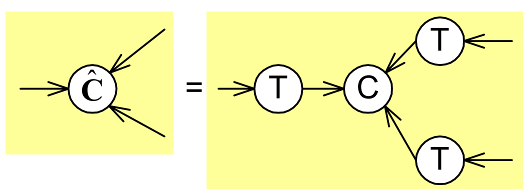
$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\mathbf{p} = \hat{\mathbf{p}}\mathbf{T}$$

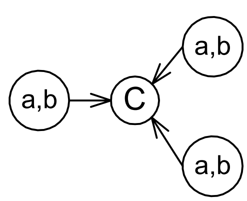
This will transform cubic $\mathbf{C}$ into $\hat{\mathbf{C}}$. In diagram form we have

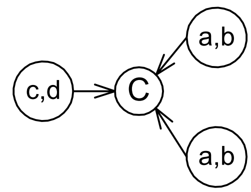
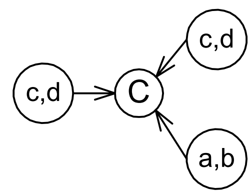
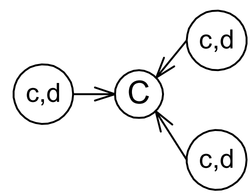


So



Name each coefficient of transformed cubic with hats too. We get the diagram form of each of these four values by plugging in the appropriate basis vectors to the free indices, which extract the appropriate rows from the transformation matrix. This gives us the four diagrams:

$\hat{A}$	

<table>
<tr><td>$\hat{B}$</td><td>
</td></tr>
<tr><td>$\hat{C}$</td><td>
</td></tr>
<tr><td>$\hat{D}$</td><td>
</td></tr>
</table>

Or in boringly explicit algebra

$$\hat{A} = a^3A + 3a^2bB + 3ab^2C + b^3D$$
$$\hat{B} = c\left(a^2A + 2abB + b^2C\right) + d\left(a^2B + 2abC + b^2D\right)$$
$$\hat{C} = a\left(c^2A + 2dcB + d^2C\right) + b\left(c^2B + 2cdC + d^2D\right)$$
$$\hat{D} = c^3A + 3c^2dB + 3cd^2C + d^3D$$

Solution Strategy

Pick **T** to make $\hat{B} = 0$. There are lots of choices of **T** that will make this happen

Conventional choice

(I have adjusted this for the current notation and for homogeneity of parameters)

Make **T** a 'parameter translation'

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ \delta & 1 \end{bmatrix}$$

Plug in the rows of **T** to the definition to give a formula for Bhat

$$\hat{B} = \begin{bmatrix} \delta & 1 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$= A\delta + B$$

So make

$$\delta = -\frac{B}{A}$$

The transformed cubic is now:

$$\hat{A}\hat{x}^3 + 3\hat{C}\hat{x}\hat{w}^2 + \hat{D}\hat{w}^3 = 0$$

Then apply the identity

$$\begin{aligned} T &= P - Q \\ T^3 &= P^3 - 3P^2Q + 3PQ^2 - Q^3 \\ &= P^3 - Q^3 - 3PQ(P-Q) \end{aligned}$$

$$T^3 + 3PQT - \left(P^3 - Q^3\right) = 0$$

Then match up with transformed cubic

$$T^3 = \hat{A}\hat{x}^3$$

$$PQT = \hat{C}\hat{x}\hat{w}^2$$

$$Q^3 - P^3 = \hat{D}\hat{w}^3$$

Solve first two for P

$$T = \hat{x}\sqrt[3]{\hat{A}}$$

$$P = \frac{\hat{C}\hat{x}\hat{w}^2}{QT} = \frac{\hat{C}\hat{x}\hat{w}^2}{Q\hat{x}\sqrt[3]{\hat{A}}} = \frac{\hat{C}}{Q\sqrt[3]{\hat{A}}}\hat{w}^2$$

Plug into third

$$Q^3 - \left(\frac{\hat{C}}{Q\sqrt[3]{\hat{A}}}\hat{w}^2\right)^3 = \hat{D}\hat{w}^3$$

$$Q^3 - \frac{\hat{C}^3}{Q^3\hat{A}}\hat{w}^6 = \hat{D}\hat{w}^3$$

$$Q^6\hat{A} - Q^3\hat{A}\hat{D}\hat{w}^3 - \hat{C}^3\hat{w}^6 = 0$$

Solve this as a quadratic in Q

$$\begin{aligned} Q^3 &= \frac{\hat{A}\hat{D}\hat{w}^3 \pm \sqrt{\left(\hat{A}\hat{D}\hat{w}^3\right)^2 + 4\hat{A}\hat{C}^3\hat{w}^6}}{2\hat{A}} \\ &= \hat{w}^3\,\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}} \end{aligned}$$

Take (three) cube roots of this,

$$Q = \hat{w}\sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}$$

Plug it in to find value of P,

$$P = \frac{\hat{C}}{Q\sqrt[3]{\hat{A}}}\hat{w}^2$$
$$= \frac{\hat{C}}{\sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}\sqrt[3]{\hat{A}}}\hat{w}$$

Subtract from Q to get values of T.

$$T = P - Q$$

$$\hat{x}\sqrt[3]{\hat{A}} = \frac{\hat{C}}{\sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}\sqrt[3]{\hat{A}}}\hat{w} - \hat{w}\sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}$$

$$\hat{x}\sqrt[3]{\hat{A}} = \hat{w}\left(\frac{\hat{C}}{\sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}\sqrt[3]{\hat{A}}} - \sqrt[3]{\frac{\hat{A}\hat{D} \pm \sqrt{\hat{A}^2\hat{D}^2 + 4\hat{A}\hat{C}^3}}{2\hat{A}}}\right)$$

Then transform back to original coordinate system to get roots of original cubic.

General Choice

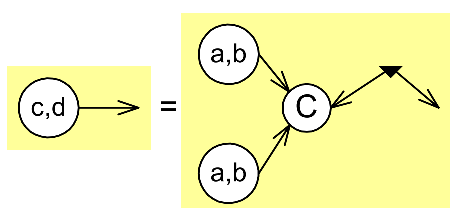
Let's now try what we did for the general solution to the quadratic polynomial. Given an arbitrary first row a,b we find the proper c,d that makes $\hat{B} = 0$. In algebraic form we want

$$\hat{B} = c\left(a^2A + 2abB + b^2C\right) + d\left(a^2B + 2abC + b^2D\right) = 0$$

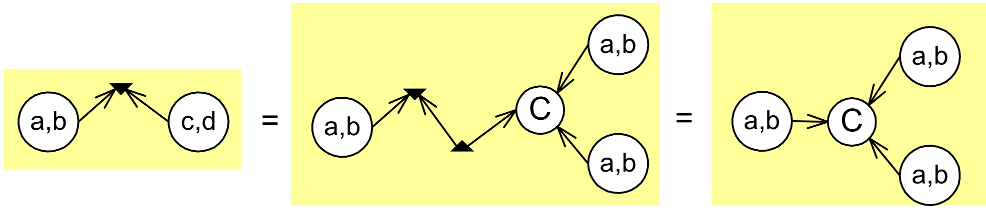
A choice for c,d that satisfies this is

$$c = -\left(a^2B + 2abC + b^2D\right)$$
$$d = +\left(a^2A + 2abB + b^2C\right)$$

In diagram form this will be:



The transformation this generates is fine as long as it isn't singular. Singularity happens if the following is zero.

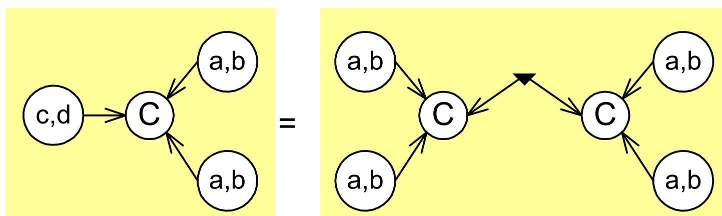


In other words, this will give us problems if the *a,b* we chose was already a root of **C**. This is similar to the constraint we had for quadratic polynomials.

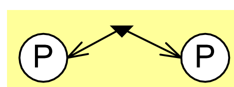
Now let's see what the other coefficients are as functions of a,b

Evaluate $\hat{B}$

We constructed c,d to make $\hat{B} = 0$. Another way to see this is to plug the diagram for *c,d* into the diagram for $\hat{B}$.



This is of form



Which is identically zero.

Evaluate $\hat{C}$

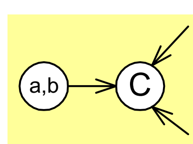
To show the power of diagram algebra I'll first do this using conventional algebra. (This is the first way I did this one, and it required monumental amounts of fiddling around).

Give names to some intermediate values

$$\lambda_1 = (aA + bB)$$
$$\lambda_2 = (aB + bC)$$
$$\lambda_3 = (aC + bD)$$

In diagram form we have

$$\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_3 \end{bmatrix} =$$



So some familiar quantities become:

$$\begin{aligned}\hat{A} &= a^3A + 3a^2bB + 3ab^2C + b^3D \\ &= a^2\left(aA+bB\right) + 2ab\left(aB+2bC\right) + b^2\left(aC+bD\right) \\ &= a^2\lambda_1 + 2ab\lambda_2 + b^2\lambda_3 \\ &= a\left(a\lambda_1 + b\lambda_2\right) + b\left(a\lambda_2 + b\lambda_1\right)\end{aligned}$$

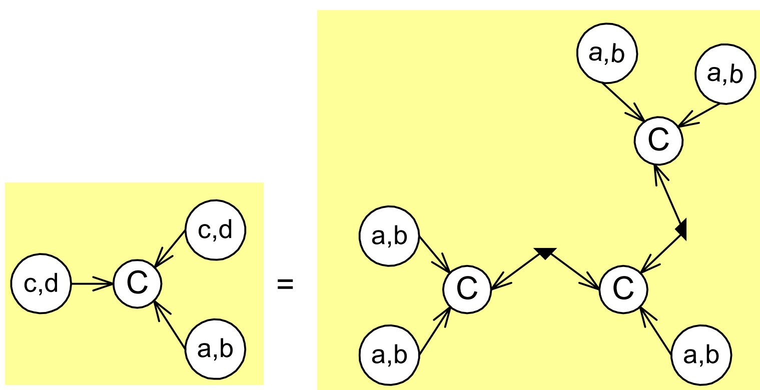
$$c = -\left(a\left(aB+bC\right) + b\left(aC+bD\right)\right) = -\left(a\lambda_2 + b\lambda_3\right)$$

$$d = \left(a\left(aA+bB\right) + b\left(aB+bC\right)\right) = \left(a\lambda_1 + b\lambda_2\right)$$

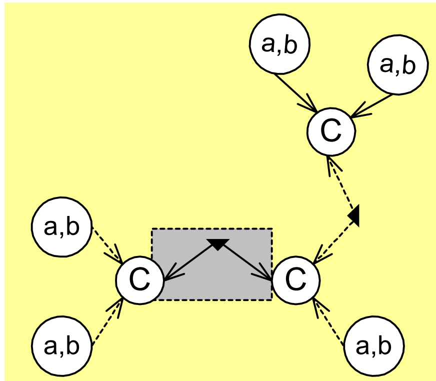
Then we can factor $\hat{C}$ by

$$\begin{aligned}\hat{C} &= c^2\lambda_1 + 2cd\,\lambda_2 + d^2\lambda_3 \\ &= c\left(c\lambda_1 + d\,\lambda_2\right) + d\left(c\lambda_2 + d\,\lambda_3\right) \\ &= c\left(-\left(a\lambda_2 + b\lambda_3\right)\lambda_1 + \left(a\lambda_1 + b\lambda_2\right)\lambda_2\right) + d\left(-\left(a\lambda_2 + b\lambda_3\right)\lambda_2 + \left(a\lambda_1 + b\lambda_2\right)\lambda_3\right) \\ &= c\left(-b\lambda_1\lambda_3 + b\lambda_2\lambda_2\right) + d\left(-a\lambda_2\lambda_2 + a\lambda_1\lambda_3\right) \\ &= cb\left(-\lambda_1\lambda_3 + \lambda_2\lambda_2\right) + da\left(-\lambda_2\lambda_2 + \lambda_1\lambda_3\right) \\ &= \left(ad - bc\right)\left(\lambda_1\lambda_3 - \lambda_2\lambda_2\right) \\ &= \hat{A}\left(\lambda_1\lambda_3 - \lambda_2\lambda_2\right) \\ &= \hat{A}\det\begin{bmatrix}\lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_3\end{bmatrix}\end{aligned}$$

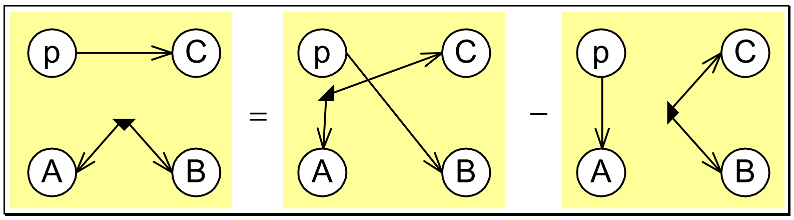
Now let's do it with diagrams. Plugging the definition of *c,d* into the definition of $\hat{C}$ we get



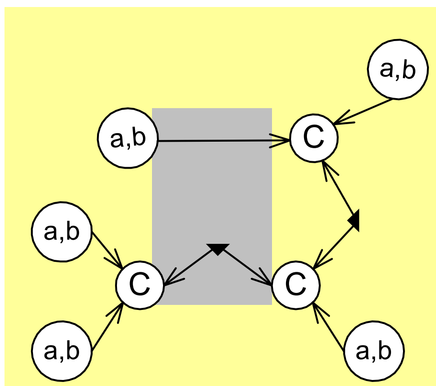
We can simplify this by applying an epsilon identity. How to pick? If we want to operate on the indicated epsilon below, the dotted ones (being only one node away) are unlikely to help:



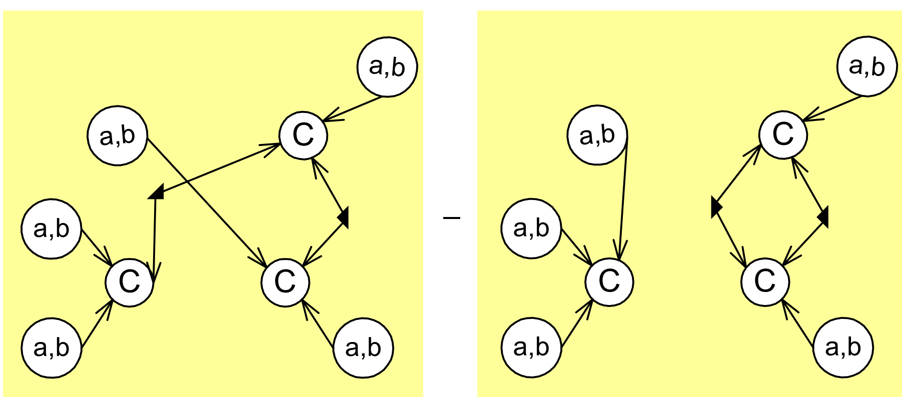
Let us, then, try applying an identity to an arc that is farther away. We use the following variant of the epsilon-delta identity from our catalog:



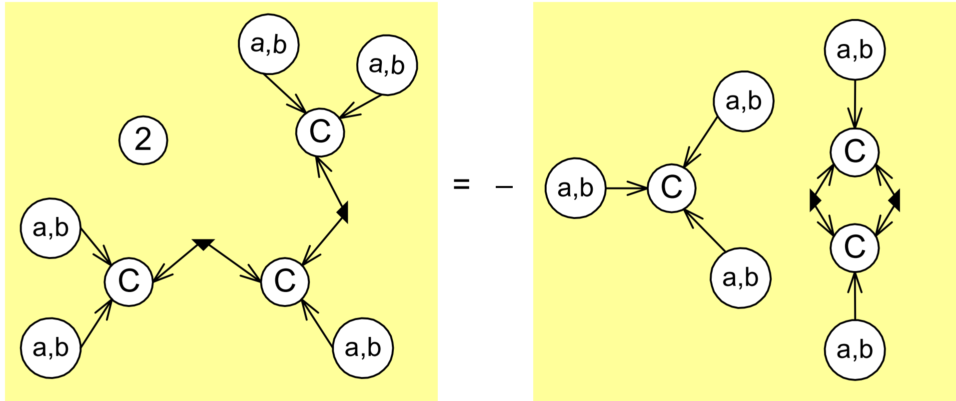
and apply it to the following arcs



We get



The first of these is just the original diagram with a sign flip (an epsilon mirrored). We can move this to the left side of the equals and get:



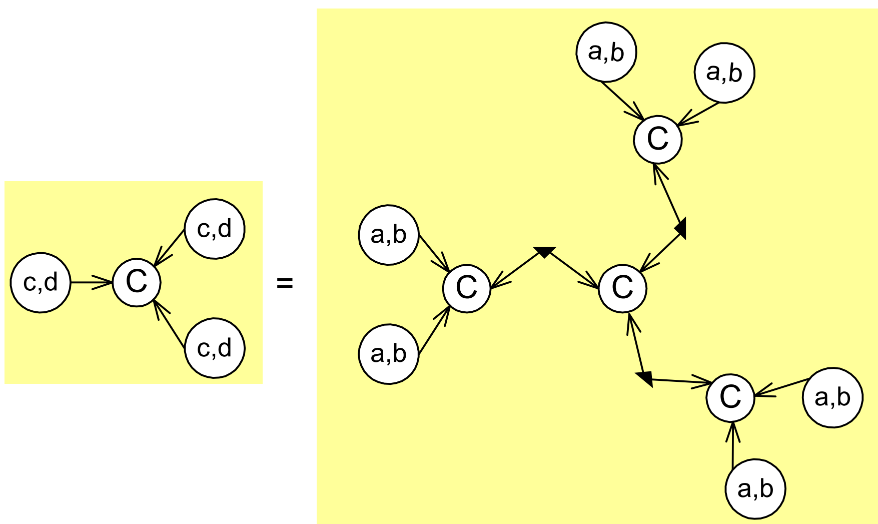
This checks with

$$\hat{C} = \hat{A}\det\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_3 \end{bmatrix}$$

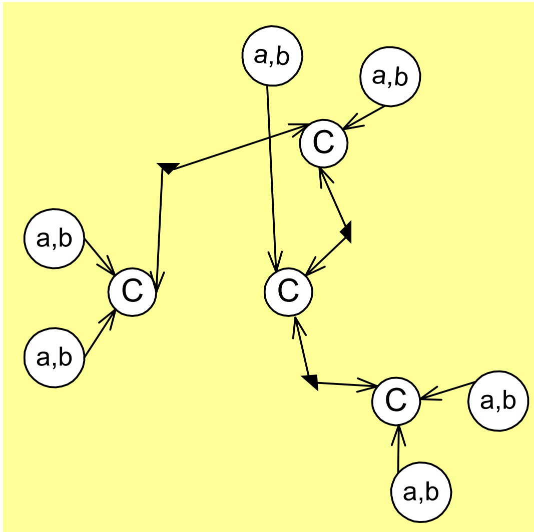
since the rightmost diagram fragment is –2 times the determinant of the matrix of lambdas:

Evaluate $\hat{D}$

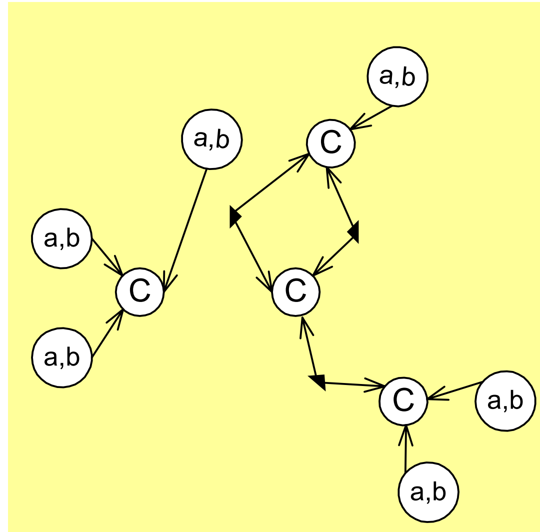
Now we're cookin'. We can do exactly the same thing with the definition of $\hat{D}$. The diagram is.



Apply the identity in the same way we did for $\hat{C}$.



$-$



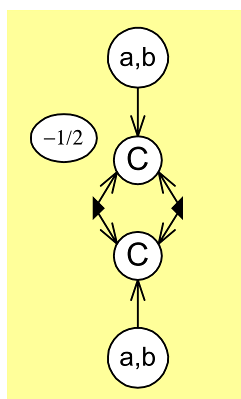
Now look at the left hand term. What do you see? Remarkably, this is two identical things on each side of an epsilon (the middle one of the three). So its… ZERO. And $\hat{D}$ is just the right hand term. And notice the factor of $\hat{A}$.

Punchline 1

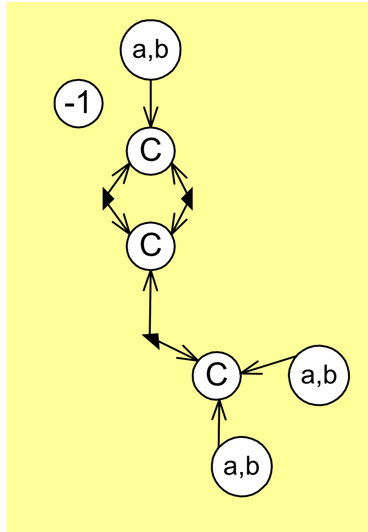
We have transformed our polynomial into the form:

$$\hat{A}x^3 + 3\hat{C}xw^2 + \hat{D}w^3 =$$
$$\hat{A}\left(x^3 + 3\tilde{C}xw^2 + \tilde{D}w^3\right)$$

where

$\tilde{C} =$ 

$\tilde{D} =$



and the previously derived solution simplifies to

$$\hat{x} = \hat{w}\left(\frac{\tilde{C}}{\sqrt[3]{\frac{\tilde{D} \pm \sqrt{\tilde{D}^2 + 4\tilde{C}^3}}{2}}} - \sqrt[3]{\frac{\tilde{D} \pm \sqrt{\tilde{D}^2 + 4\tilde{C}^3}}{2}} \right)$$

Note that the thing under the square root sign is the discriminant of the transformed cubic polynomial.

Another Interesting Transformation

The above discussion dealt with the result of picking an arbitrary (*a,b*) to generate a nice parameter transformation. The conventional choice is to pick $(a,b) = (1,0)$. Let's now be a bit more creative. Let's pick (*a,*b) to be a root to the equation

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 2\Delta_1 & \Delta_2 \\ \Delta_2 & 2\Delta_3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

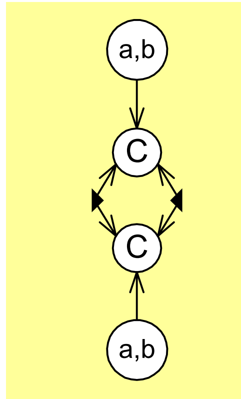
The matrix above is the one whose determinant is the discriminant of the original cubic polynomial. The desired roots are:

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} -\Delta_2 - \sqrt{-\Delta} & 2\Delta_1 \end{bmatrix}, \begin{bmatrix} 2\Delta_3 & -\Delta_2 - \sqrt{-\Delta} \end{bmatrix}$$

or (alternative notation)

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} 2\Delta_3 & -\Delta_2 + \sqrt{-\Delta} \end{bmatrix}, \begin{bmatrix} -\Delta_2 + \sqrt{-\Delta} & 2\Delta_1 \end{bmatrix}$$

By looking at the diagram form of the matrix above we see that this choice of (*a,b*) means:



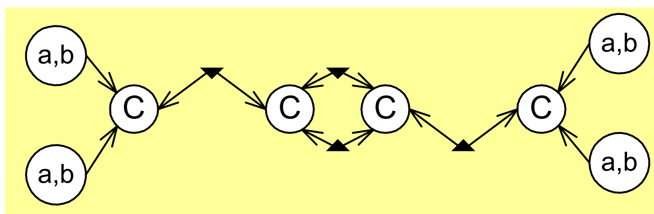
$= 0$

But that means that

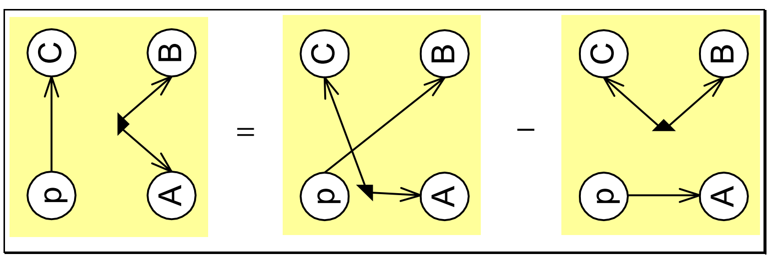
$$\tilde{C} = 0$$

Not bad for starters.

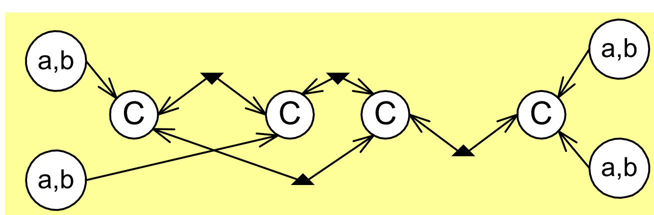
Now what about *c,d*? Let's see what happens if we plug *c,d* into the matrix. Using the tensor diagram definition of *c,d* and rotating the diagram sideways we get:



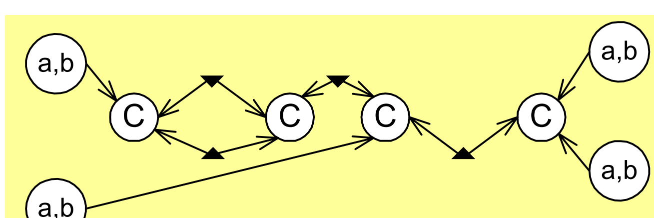
apply the identity



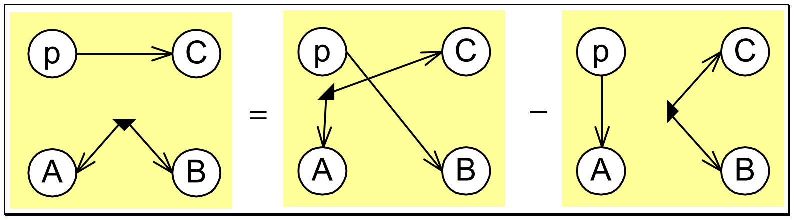
We get:



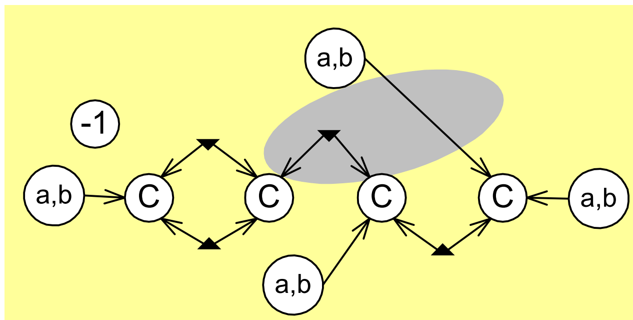
−



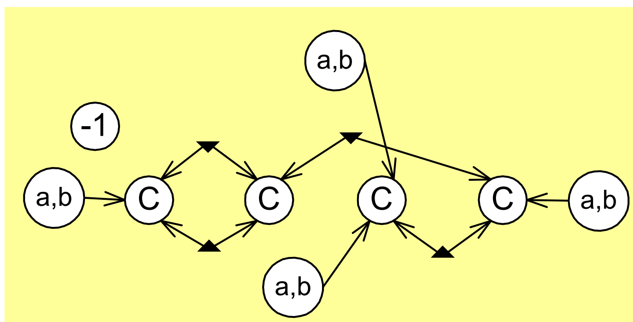
The first term of this has a ring of three **C**'s and so is identically zero. Now play with the second term. We will apply our favorite identity:



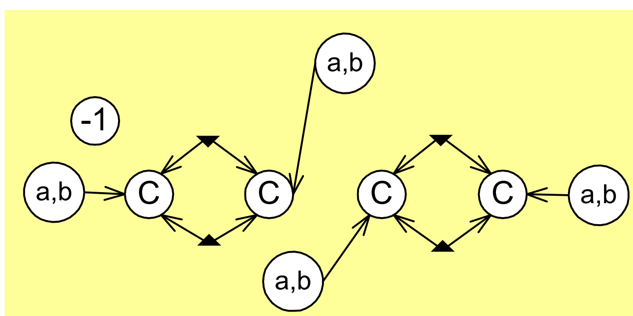
to the shaded arcs



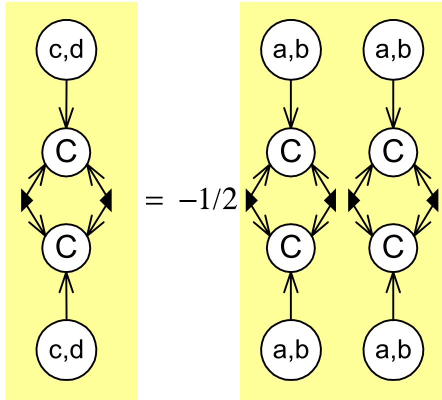
We get:



-



The first term is (surprise) minus the diagram we started with. So we have the net result:



In other words, if (*a,b*) is a root of the Delta matrix, then so is (*c,d*). In fact (*c,d*) is THE OTHER ROOT. But we MUST be careful here about homogeneous scale factors. If we recall that

$$c = -\left(a^2B + 2abC + b^2D\right)$$
$$d = +\left(a^2A + 2abB + b^2C\right)$$

The numerically nice solutions (for B>0) are

$$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -B - \sqrt{B^2 - AC} \end{bmatrix}$$
$$\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B - \sqrt{B^2 - AC} & A \end{bmatrix}$$

If we take (a,b) as the first one we get:

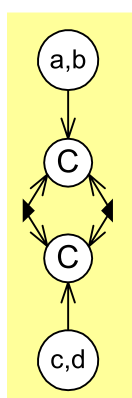
$$c = -\left(C^2B + 2C\left(-B - \sqrt{B^2 - AC}\right)C + \left(-B - \sqrt{B^2 - AC}\right)^2 D\right)$$
$$d = +\left(C^2A + 2C\left(-B - \sqrt{B^2 - AC}\right)B + \left(-B - \sqrt{B^2 - AC}\right)^2 C\right)$$

We therefore have transformed our cubic into:

$$\hat{x}^3 + \tilde{D}\hat{w}^3 = 0$$

where

$\tilde{D} =$

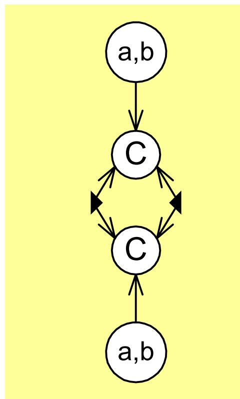


This is, again, much like the way we did it for quadratic polynomials. We have

$$\begin{bmatrix} \hat{x} & \hat{w} \end{bmatrix} = \begin{bmatrix} \sqrt[3]{-\tilde{D}} & 1 \end{bmatrix}$$

In general, we get three (complex) roots when we take the cube root. This gives us our three roots to the transformed cubic polynomial. Then, given that we solved a quadratic equation to find (*a,b*) and (*c,d*), we use the matrix composed of these two rows to transform back to the original $\begin{bmatrix} x & w \end{bmatrix}$.

Let's look further at this solution. We find (*a,b*) as the solution of:



$= 0$

Let's temporarily reuse the names *A,B,C* as the elements of the 2x2 matrix above and review the algorithm for solving this quadratic:

$B > 0$		$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -B-\sqrt{B^2-AC} & A \end{bmatrix}$
$B < 0$		$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} -B+\sqrt{B^2-AC} & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix}$
$B = 0$	$\lvert A \rvert \ge \lvert C \rvert$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} \sqrt{-AC} & A \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} -\sqrt{-AC} & A \end{bmatrix}$
	$\lvert A \rvert \le \lvert C \rvert$	$\begin{bmatrix} x_1 & w_1 \end{bmatrix} = \begin{bmatrix} C & -\sqrt{-AC} \end{bmatrix}$ $\begin{bmatrix} x_2 & w_2 \end{bmatrix} = \begin{bmatrix} C & \sqrt{-AC} \end{bmatrix}$

Our expression for D~ then becomes

If $B > 0$

$$\tilde{D} = \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} -B-\sqrt{B^2-AC} \\ A \end{bmatrix}$$

$$= \begin{bmatrix} AC-B^2-B\sqrt{B^2-AC} & BC-BC-C\sqrt{B^2-AC} \end{bmatrix} \begin{bmatrix} -B-\sqrt{B^2-AC} \\ A \end{bmatrix}$$

$$= \left(AC-B^2-B\sqrt{B^2-AC}\right)\left(-B-\sqrt{B^2-AC}\right) + A\left(-C\sqrt{B^2-AC}\right)$$

$$= -B\left(AC-B^2\right) + B^2\sqrt{B^2-AC} - \left(AC-B^2\right)\sqrt{B^2-AC} + B\left(B^2-AC\right) - AC\sqrt{B^2-AC}$$

$$= -B\left(AC-B^2\right) + B\left(B^2-AC\right) + \left(B^2-\left(AC-B^2\right)-AC\right)\sqrt{B^2-AC}$$

$$= +2B\left(B^2-AC\right) + 2\left(B^2-AC\right)\sqrt{B^2-AC}$$

$$= +2\left(B^2-AC\right)\left(B+\sqrt{B^2-AC}\right)$$

Other way to calculate: But there's a homogeneous constant factor in here.. be careful

$$\tilde{D} = \left\{ \begin{bmatrix} C & -B-\sqrt{B^2-AC} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} -B-\sqrt{B^2-AC} \\ A \end{bmatrix} \right\}^2$$

$$= \left\{ \begin{bmatrix} B+\sqrt{B^2-AC} & C \end{bmatrix} \begin{bmatrix} -B-\sqrt{B^2-AC} \\ A \end{bmatrix} \right\}^2$$

$$= \left\{ -\left(B+\sqrt{B^2-AC}\right)^2 + AC \right\}^2$$

$$= \left\{ -B^2 - 2B\sqrt{B^2-AC} - B^2 + AC + AC \right\}^2$$

$$= \left\{ -2B^2 - 2B\sqrt{B^2-AC} + 2AC \right\}^2$$

$B < 0$

$$\tilde{D}=\begin{bmatrix} C & -B+\sqrt{B^2-AC} \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} -B+\sqrt{B^2-AC} \\ A \end{bmatrix}$$
$$=\begin{bmatrix} AC-B^2+B\sqrt{B^2-AC} & BC-BC+C\sqrt{B^2-AC} \end{bmatrix}\begin{bmatrix} -B+\sqrt{B^2-AC} \\ A \end{bmatrix}$$
$$=\left(AC-B^2+B\sqrt{B^2-AC}\right)\left(-B+\sqrt{B^2-AC}\right)+A\left(+C\sqrt{B^2-AC}\right)$$
$$=-B\left(AC-B^2\right)-B^2\sqrt{B^2-AC}+\left(AC-B^2\right)\sqrt{B^2-AC}+B\left(B^2-AC\right)+AC\sqrt{B^2-AC}$$
$$=-B\left(AC-B^2\right)+B\left(B^2-AC\right)+\left(-B^2+\left(AC-B^2\right)+AC\right)\sqrt{B^2-AC}$$
$$=+2B\left(B^2-AC\right)+2\left(-B^2+AC\right)\sqrt{B^2-AC}$$
$$=+2\left(B^2-AC\right)\left(B-\sqrt{B^2-AC}\right)$$

$B=0,\left|A\right|\geq\left|C\right|$

$$\begin{bmatrix} \sqrt{-AC} & A \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}\begin{bmatrix} -\sqrt{-AC} \\ A \end{bmatrix}$$
$$=\begin{bmatrix} A\sqrt{-AC} & AC \end{bmatrix}\begin{bmatrix} -\sqrt{-AC} \\ A \end{bmatrix}$$
$$=-A\left(-AC\right)+A^2C$$
$$=2A^2C$$

How impressed should we be with this? It has two potential difficulties:

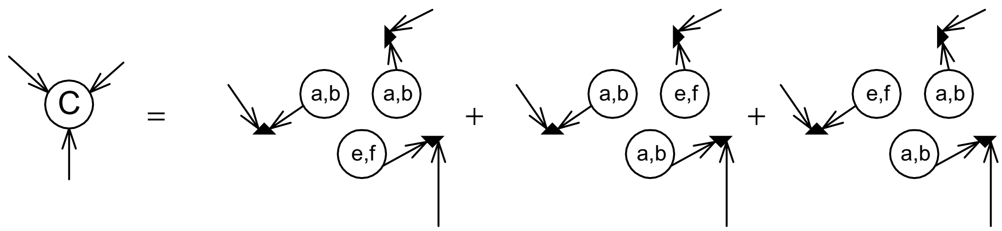
1. The parametric transformation matrix is singular if the quadratic polynomial generated by the Delta matrix has a double root. This occurs if the cubic polynomial has a double root. Even if the cubic has two very closely spaced discreet roots, the *a,b* and *c,d* might give a very ill conditioned matrix. The double root of the cubic is the same as the double root of the quadratic: *a,b*. The remaining single root can be found by… (what)
2. If the quadratic has imaginary roots we must deal with complex numbers in the transformation. Probably not disasterous. This happens when the cubic has 1 (or is it 3) real roots.

It also needs to be analyzed for numerical stability.

But, all in all, it might be useful.

Double Roots

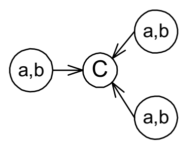
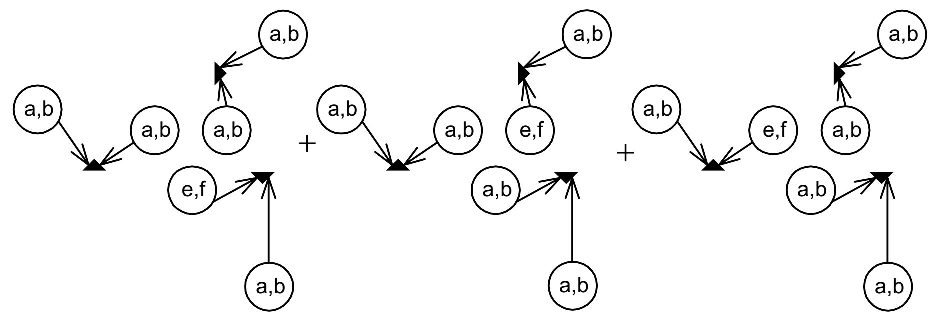
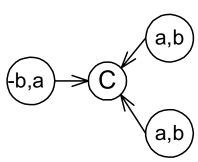
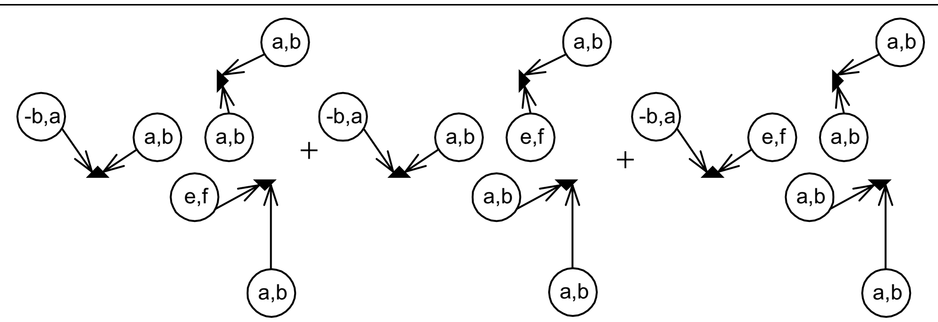
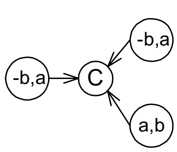
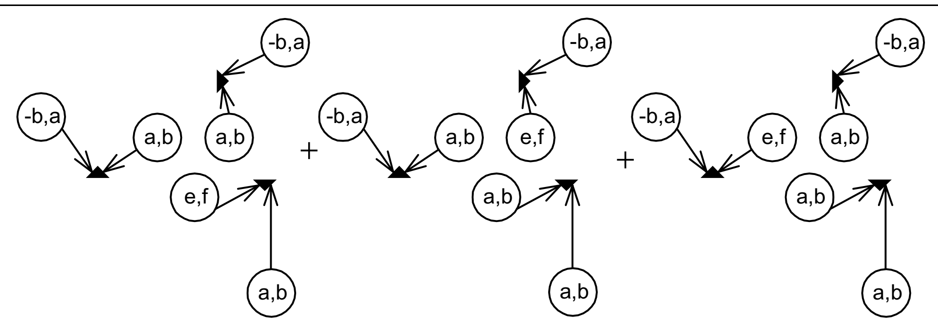
If the quadratic has a double root the above won't work because the transformation matrix (ab,cd) generated is singular. This is because the cubic itself has a double root. That is (ab) (prove this). Diagrammatically, the internal structure of the cubic tensor looks like:

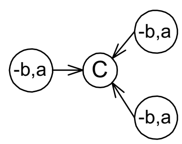
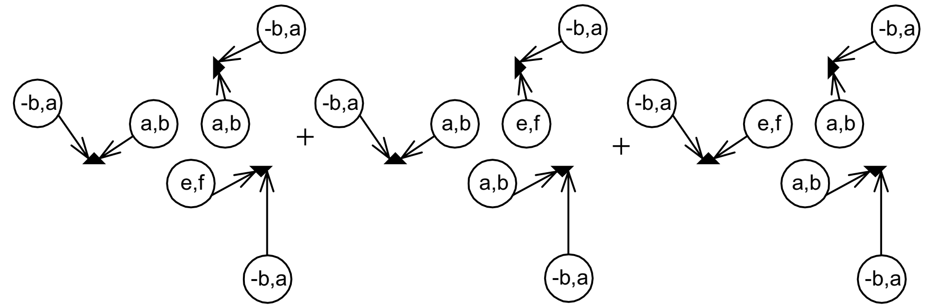


In this situation we have gotten the double root, we still need to extract the remaining single root: We form the transformation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

The components of the transformed cubic are then:

$\hat{A}$		
$\hat{B}$		
$\hat{C}$		

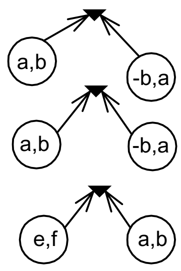
$\hat{D}$		

This gives:

A^=0

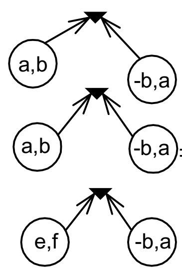
B^=0

$$\hat{C} =$$



$$= \left(a^2+b^2\right)^2\left(-af+be\right)$$

$$\hat{D} = 3$$



$$= \left(a^2+b^2\right)^2\left(3bf+3ae\right)$$

Throwing out the nonzero factor a^2+b^2 we get

$$\begin{bmatrix}\hat{C} & \hat{D}\end{bmatrix} = \begin{bmatrix}-af+be & 3bf+3ae\end{bmatrix}$$

$$= \begin{bmatrix}e & f\end{bmatrix}\begin{bmatrix}b & 3a \\ -a & 3b\end{bmatrix}$$

so we can use:

$$\begin{bmatrix}\hat{C} & \hat{D}\end{bmatrix}\begin{bmatrix}3b & -3a \\ a & b\end{bmatrix} = \begin{bmatrix}e & f\end{bmatrix}$$

$$\begin{bmatrix}3\hat{C} & \hat{D}\end{bmatrix}\begin{bmatrix}b & -a \\ a & b\end{bmatrix} = \begin{bmatrix}e & f\end{bmatrix}$$

This is tantamount to transforming the cubic to:

$$3\hat{C}xw^2 + \hat{D}w^3 = 0$$

$$\left(3\hat{C}x + \hat{D}w\right)w^2 = 0$$

Chapter 1-07

2D(1DH)
Resultant of a Quadratic and a Cubic

We now know the resultant of two quadratics. Let's now jump up an order and find the diagram form for the resultant of a quadratic and a cubic. We'll review a simpler problem first.

Linear and Cubic

Find resultant of

$$\mathbf{C}(x,w) = C_3x^3 + 3C_2x^2w + 3C_1xw^2 + C_0w^3$$
$$\mathbf{L}(x,w) = L_1x + L_0w$$

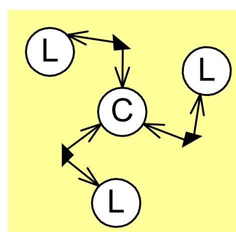
Generate the solution to the linear equation

$$\begin{bmatrix} x & w \end{bmatrix} = \begin{bmatrix} -L_0 & L_1 \end{bmatrix}$$

Plug into cubic and get

$$\boxed{\rho(L,C) = -C_3L_0^{\ 3} + 3C_2L_0^{\ 2}L_1 - 3C_1L_0L_1^{\ 2} + C_0L_1^{\ 3}}$$

The diagram is simply



Quadratic and Cubic

Now find resultant of the two homogeneous polynomials:

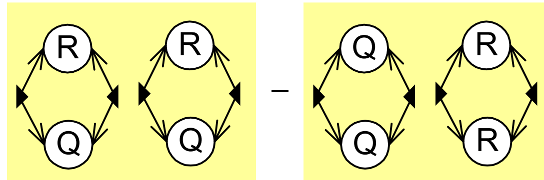
$$\mathbf{C}(x,w) = Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$$
$$\mathbf{Q}(x,w) = Ex^2 + 2Fxw + Gw^2$$

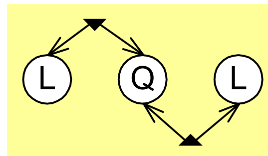
The non-homogeneous way of doing this would actually treat **Q** as a cubic with leading coefficient of zero. Then find resultant of the two cubics. We are not going to do this here since in homogeneous polynomial land a quadratic is not just a cubic with zero for the leading term. We want to get a diagram that has two-pronged **Q**'s and three pronged **C**'s in it. What do we expect that this will look like?

We first think of our cubic as the product of a quadratic R times a linear polynomial L. In a manner similar to what we did in the last chapter we can see that the resultant we want will then be:

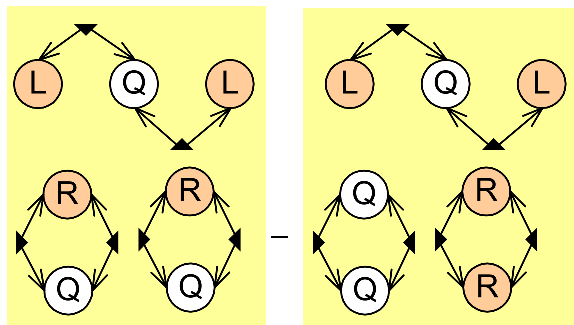
$$\rho(Q,C) = \rho(Q,R)\,\rho(Q,L)$$

We already have the diagrams for these latter two resultants.

$\rho(Q,R) =$ 

$\rho(Q,L) =$ 

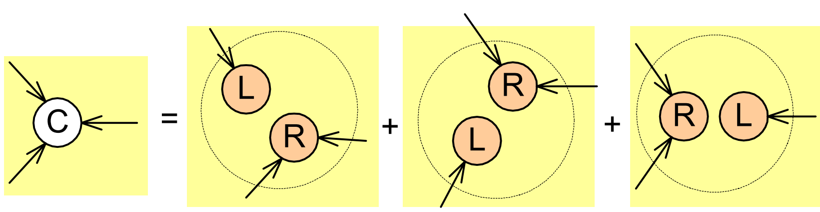
So the net resultant is

$\rho(Q,C) =$ 

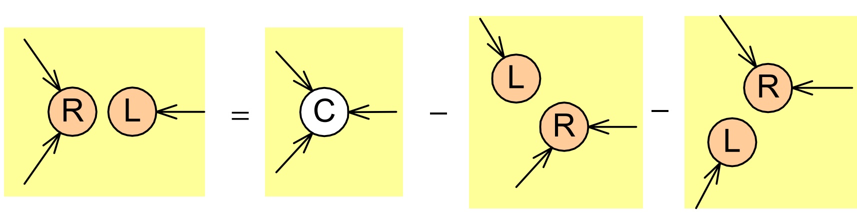
Now of course we ultimately want this in terms of the three-pronged node C instead of the nodes for R and L. I've indicated this by coloring in the undesirable R and L nodes.

General Strategy

First look at the "internal structure" of **C.**:



As we did in the Quadratic-Quadratic case we are going to merge RL pairs into C nodes by the following strategy: Whenever you see and R and L, plug in the following rearrangement of this internal structure:

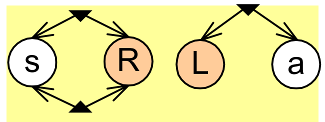


Each such substitution will introduce a C node but also two terms containing our bad R and L nodes. We hope that other simplifications will allow us to get rid of them too. After playing around with this a while I have found the following identities to be useful:

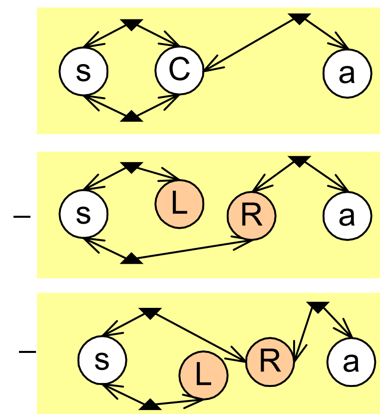
Some Identities

Identity 1

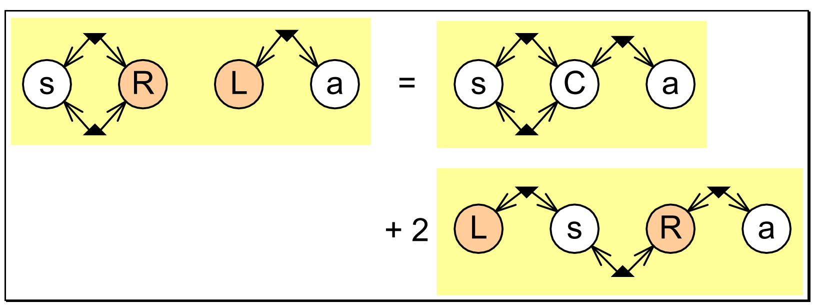
Start with any symmetric sub diagram s and arbitrary diagram fragment a:



Do the C substitution

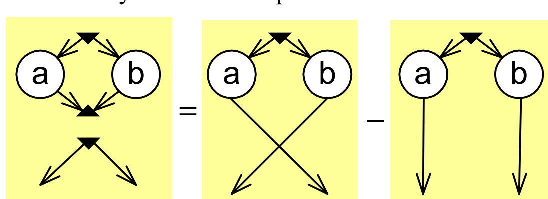


Rearrange and neaten and you find that the last two terms of this are equal, so we have

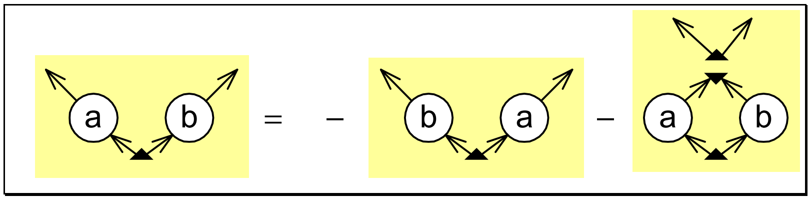


Identity 2

Plug the epsilon product of an arbitrary a and b on top of the standard $\varepsilon\delta$ identity

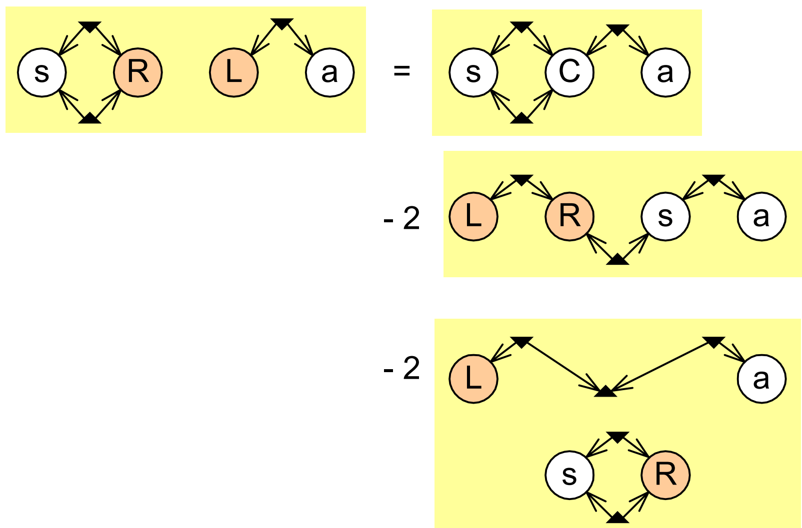


Rearrange, turn upside down, ane neaten

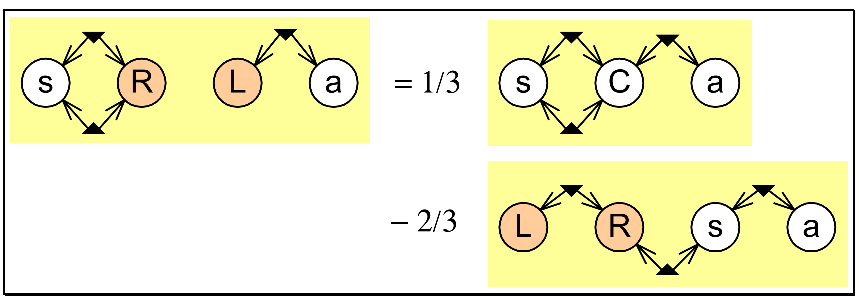


Identity 3

Combine Identity 2 with the last term of Identity 1 using a=s and b=R

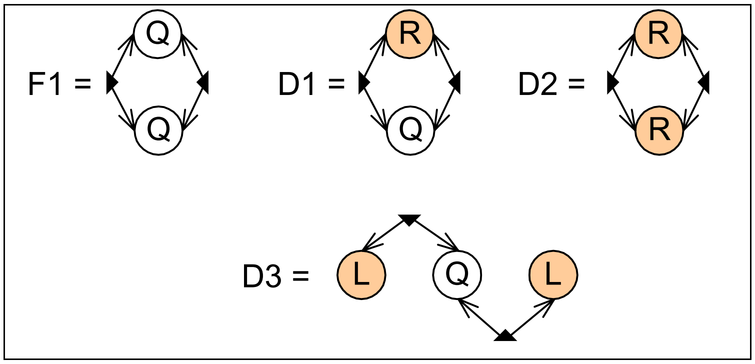


This last term is just the same as the left hand side. So



Lets go

Lets give the names to the diagram fragments we have seen so far. Any diagram fragment that contains only Q and C will have the name **Fn**, implying that it can be part of the final result. Any intermediate that still contains R and L will be called **Dn**.



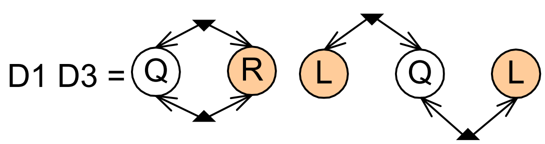
Our desired resultant is then

$$\rho(Q,C) = D_1^{\,2} D_3 - F_1 D_2 D_3$$

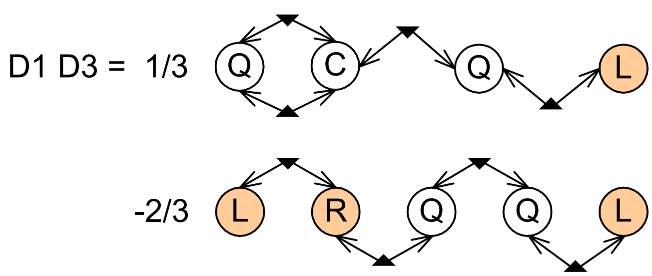
We will operate on the two terms separately

Step 1 —C substitution into D1 D3

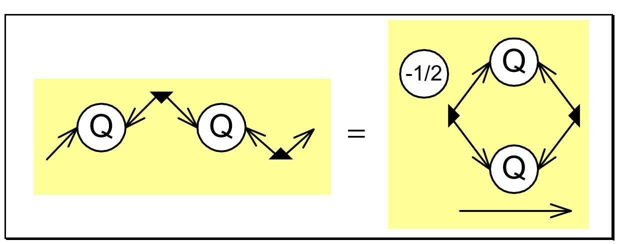
Look at D1 times D3



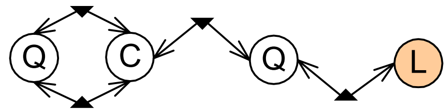
This is of the form of identities 1 and 3 with s=Q and a=QvL. Identity 3 brings s and a together so we chose it:

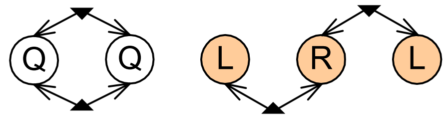


Now we can apply the standard -Q^Q- chain identity to the last term.

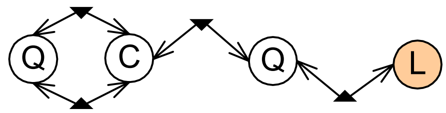


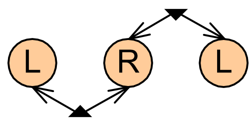
This gives

D1 D3 = +1/3 

+1/3 

Give names to these new diagram fragments

D4 = 

D5 = 

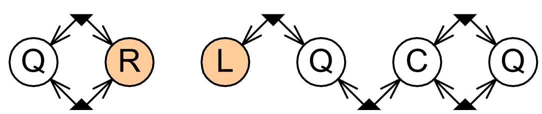
$$D_1 D_3 = \frac{1}{3} D_4 + \frac{1}{3} F_1 D_5$$

Step 2 — Multiply in D1

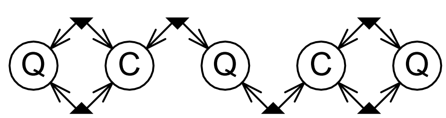
Now multiply in the other D1 factor

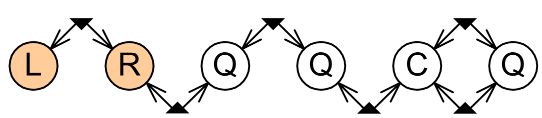
$$D_1{}^2 D_3 = \frac{1}{3}\left(D_1 D_4\right) + \frac{1}{3} F_1 \left(D_1 D_5\right)$$

Evaluate D1 D4

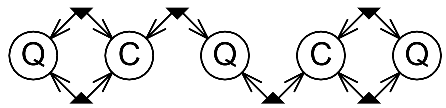
D4 D1 = 

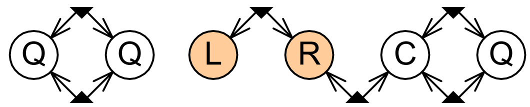
Apply identity 3

D4 D1 = 1/3 

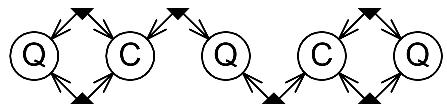
–2/3 

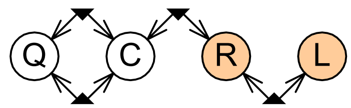
Apply Q^Q identity

D1 D4 = 1/3 

1/3 

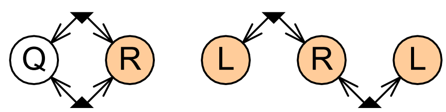
Give names

F2 = 

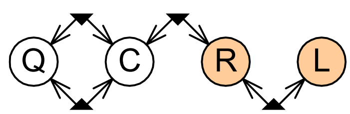
D6 = 

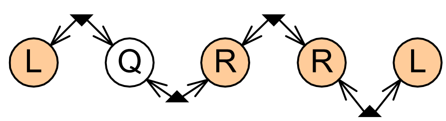
$$D_4 D_1 = \frac{1}{3} F_2 + \frac{1}{3} D_6 F_1$$

Evaluate D1 D5

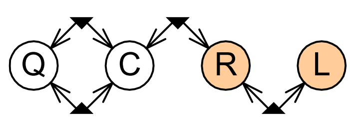
D1 D5 = 

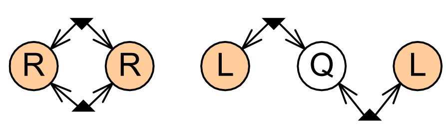
Use identity 1. The last two terms of this are the same.

D1 D5 = 

+ 2 

Apply R^R identity

D5 D1 = 

– 

We have already given names to all three of these diagram fragments.

$$D_5 D_1 = D_6 - D_2 D_3$$

Step 3 — Put together what we have so far

We have the following relationships

$$\rho(Q,C) = D_1^{\;2} D_3 - F_1 D_2 D_3$$

$$D_1 D_3 = \frac{1}{3} D_4 + \frac{1}{3} F_1 D_5$$

$$D_4 D_1 = \frac{1}{3} F_2 + \frac{1}{3} D_6 F_1$$

$$D_5 D_1 = D_6 - D_2 D_3$$

The first term of ρ is

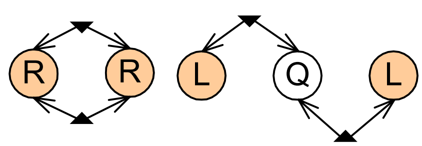
$$\begin{aligned} D_1^{\;2} D_3 &= \frac{1}{3} D_1 D_4 + \frac{1}{3} F_1 D_1 D_5 \\ &= \frac{1}{3}\left(\frac{1}{3} F_2 + \frac{1}{3} D_6 F_1\right) + \frac{1}{3} F_1 \left(D_6 - D_2 D_3\right) \\ &= \frac{1}{9} F_2 + \frac{4}{9} D_6 F_1 - \frac{1}{3} F_1 D_2 D_3 \end{aligned}$$

Note that the final term of this is (one third of) the second term in ρ

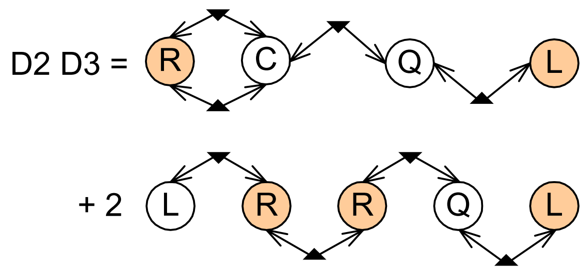
$$\begin{aligned} \rho(Q,C) &= D_1^{\;2} D_3 - F_1 D_2 D_3 \\ &= \left(\frac{1}{9} F_2 + \frac{4}{9} D_6 F_1 - \frac{1}{3} F_1 D_2 D_3\right) - F_1 D_2 D_3 \\ &= \frac{1}{9} F_2 + \frac{4}{9} D_6 F_1 - \frac{4}{3} F_1 D_2 D_3 \\ &= \frac{1}{9} F_2 + \frac{4}{3} F_1 \left(\frac{1}{3} D_6 - D_2 D_3\right) \end{aligned}$$

Now we just need to play with D6 and D2*D3.

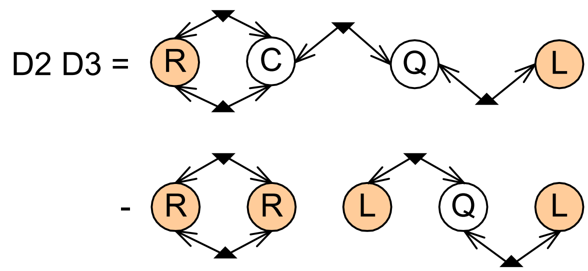
Step 4 — Evaluate D2*D3

D2 D3 =


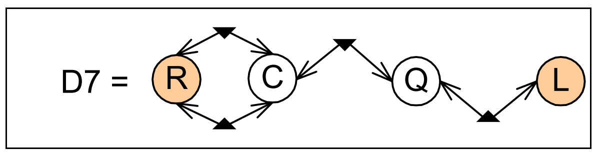
Apply Identity 1



Apply the –Q^Q-identity



The second half of this is just D2*D3 again. Name the first term



The we have

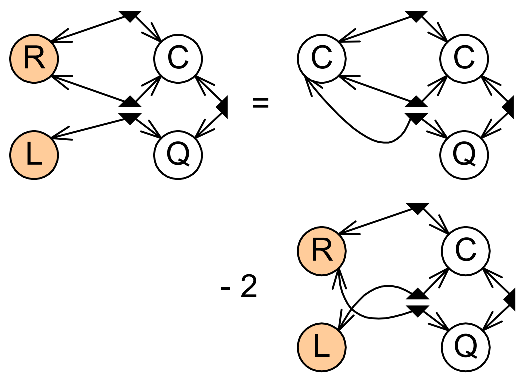
$$D_2 D_3 = \frac{1}{2} D_7$$

Putting this together we have the net resultant is

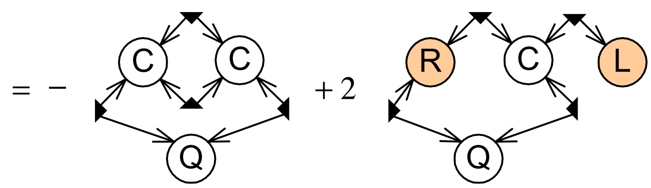
$$\rho(Q,C) = \frac{1}{9}F_2 + \frac{4}{3}F_1\left(\frac{1}{3}D_6 - \frac{1}{2}D_7\right)$$
$$= \frac{1}{9}F_2 + \frac{2}{9}F_1\left(2D_6 - 3D_7\right)$$

Step 5 — Work on D7

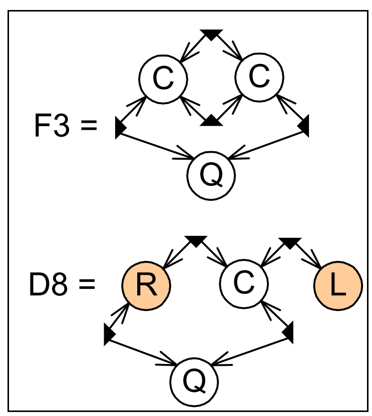
Do C substitution:



Clean up

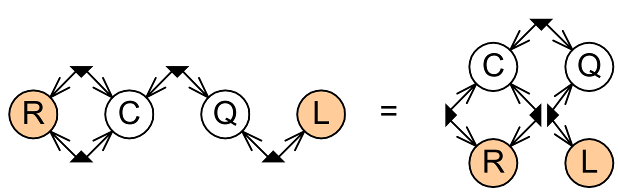


Name these new fragments

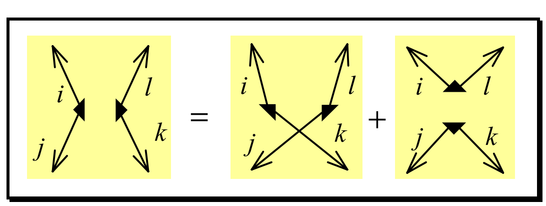


$$D_7 = -F_3 + 2D_8$$

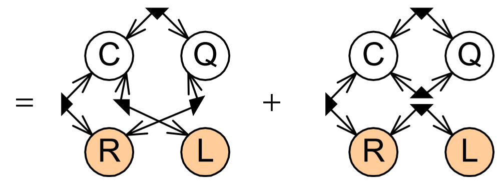
Step 6 — Apply $\varepsilon\varepsilon$ identity to D7



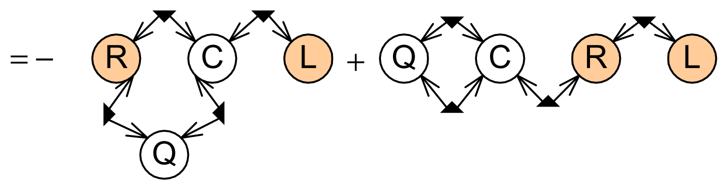
Apply the $\varepsilon\varepsilon$ identity



We get



Clean up



Name these

$$\boxed{D_7 = -D_8 + D_6}$$

Step 8 — Final wrapup

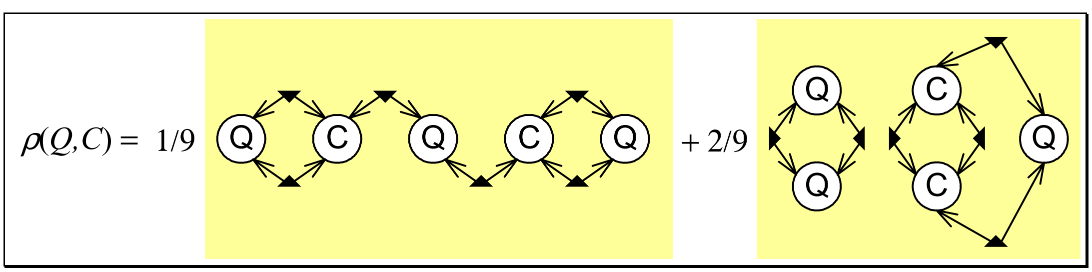
Putting together the parts we have so far

$$\begin{aligned}2D_6 &= 2D_7 + 2D_8\\ 2D_6 - 3D_7 &= -D_7 + 2D_8\\ &= F_3 - 2D_8 + 2D_8\\ &= F_3\end{aligned}$$

Then putting this into the equation for the resultant

$$\begin{aligned}\rho(Q,C) &= \frac{1}{9}F_2 + \frac{2}{9}F_1\left(2D_6 - 3D_7\right)\\ &= \frac{1}{9}F_2 + \frac{2}{9}F_1F_3\end{aligned}$$

In diagram terms, the net answer is

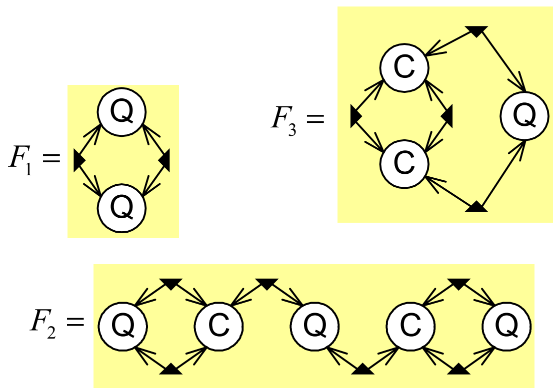


Post Mortem

This was admittedly pretty exhausting. The reader might well as the question of how, at each step, I was able to tell what to do next. The basic technique is to plug in a C substitution and play with the left over terms. Keep naming new diagram fragments and trying to relate them via standard identities like $\varepsilon\varepsilon$. After a while you will have discovered the relationships between all the possible fragments and you can eliminate the ones containing R and L. I hope to be able to simplify this process and make its application more obvious.

Other Answers

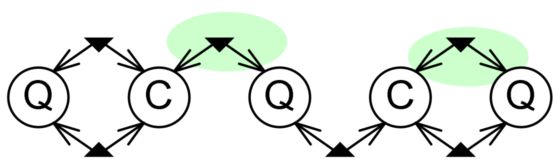
We can see that the resultant of Q and C will be composed of diagram fragments that contain two C nodes and three Q nodes and the required number of epsilons to glue them together. So far we have seen:



There happens to be one more possible diagram. We can see this by applying the $\varepsilon\varepsilon$ identity to various likely pairs of epsilons in our existing diagrams. As usual, doing this to epsilons that are linked to the same tensor node doesn't give us anything. The epsilons we use need to be at least two nodes apart. Thus F1 and F3 won't give us any new relations. But F2 will.

Attempt 1

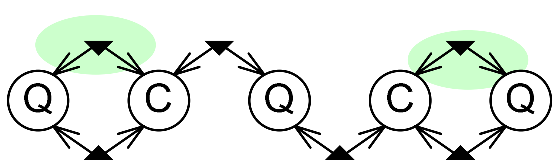
Apply the $\varepsilon\varepsilon$ identity to the two epsilons of F2 indicated



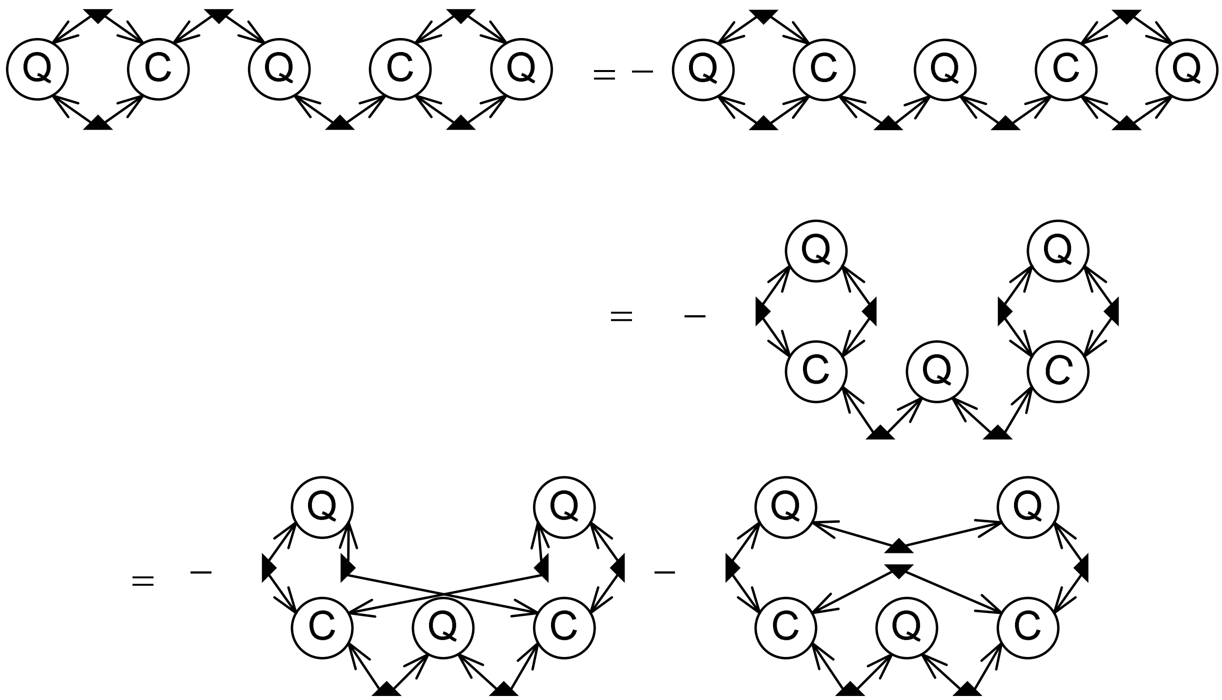
I'll leave as an exercise showing that this doesn't give us anything new.

Attempt 2

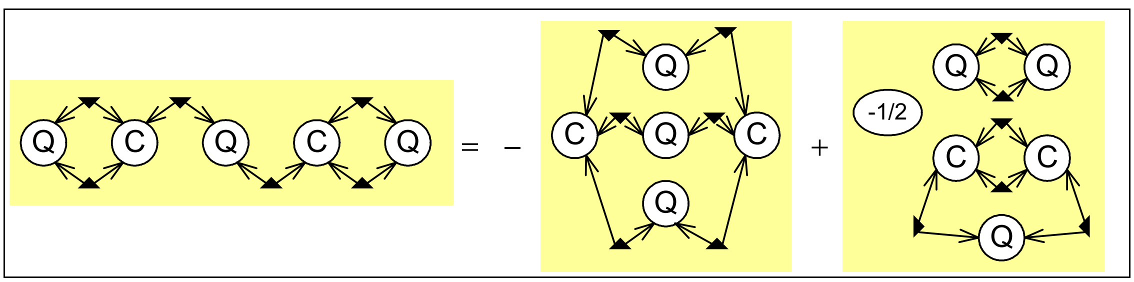
Try another pair:



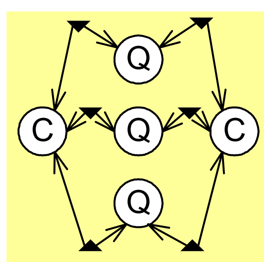
This gives us:



Applying the –Q^Q- identity and cleaning up gives us a new diagram and its relation to the existing ones:



$F_4 =$



$$F_2 = -F_4 - \frac{1}{2} F_1 F_3$$

and

$$-2\left(F_2 + F_4\right) = F_1 F_3$$

Resultant formulations

We originally got

$$\rho(Q,C)=\frac{1}{9}(F_2+2F_1F_3)$$

And in terms of F4 and F1F3

$$\rho(Q,C)=\frac{1}{9}\left(\left(-F_4-\frac{1}{2}F_1F_3\right)+2F_1F_3\right)$$
$$=\frac{1}{9}\left(-F_4-\frac{3}{2}F_1F_3\right)$$

An in terms of F4 and F2

$$\rho(Q,C)=\frac{1}{9}\big(F_2+2\big(-2(F_2+F_4)\big)\big)$$
$$=\frac{1}{9}(-3F_2-4F_4)$$

Appendix

For those who need to evaluate these diagram fragments I have done so with symbolic algebra. We have

$$\mathbf{F}_1=2\left(F^2-EG\right)$$

$$\begin{aligned}\mathbf{F}_2 = {} & A^2\left(G^3\right)-6AB\left(FG^2\right)+AC\left(2EG^2+4F^2G\right)-2AD(EFG)\\ & +B^2\left(EG^2+8F^2G\right)+BC\left(-10EFG-8F^3\right)+BD\left(2E^2G+4EF^2\right)\\ & +C^2\left(E^2G+8EF^2\right)-6CD\left(E^2F\right)\\ & +D^2\left(E^3\right)\end{aligned}$$

$$\mathbf{F}_3=2\left(AC(G)-AD(F)-B^2(G)+BC(F)+BD(E)-C^2(E)\right)$$

$$\begin{aligned}\mathbf{F}_4 = {} & -A^2\left(G^3\right)+6AB\left(FG^2\right)-6AC\left(F^2G\right)+2AD\left(F^3\right)\\ & +B^2\left(-3\left(EG^2\right)-6\left(F^2G\right)\right)+BC\left(12EFG+6\left(F^3\right)\right)-6BD\left(EF^2\right)\\ & +C^2\left(-3\left(E^2G\right)-6\left(EF^2\right)\right)+6CD\left(E^2F\right)\\ & -D^2\left(E^3\right)\end{aligned}$$

Chapter 1-08

1DH(2D)
The Resultant of Two Cubic Polynomials

Our next order jump is to find the resultant of two cubic polynomials. There are several general techniques for finding resultants so the problem solution is not new. I am trying to find the answer, of course, in terms of tensor diagrams. I have not been successful yet. This chapter shows some partial answers. I hope to get a general one soon. Note that our construction of the diagram of the resultant of a quadratic and a cubic was a bit complicated. It gets worse here. I hope that it can be simplified.

The Problem

Given two homogeneous cubic polynomials

$$\mathbf{C}(x,w) = Ax^3 + 3Bx^2w + 3Cxw^2 + Dw^3$$
$$\mathbf{D}(x,w) = ax^3 + 3bx^2w + 3cxw^2 + dw^3$$

find the polynomial function of (A,B,C,D,a,b,c,d) that is zero when **C** and **D** have a common root.

Solution 1

I learned the following "conventional" solution I learned from Tom Sederberg. (It also works to find the resultant of two quadratics, or in fact any two polynomials of equal order. I'll leave that as an exercise.) It operates on nonhomogeneous polynomials so back off to:

$$\mathbf{C}(x) = Ax^3 + 3Bx^2 + 3Cx + D$$
$$\mathbf{D}(x) = ax^3 + 3bx^2 + 3cx + d$$

WE can also write this in matrix form

$$\begin{bmatrix} \mathbf{C}(x) & \mathbf{D}(x) \end{bmatrix} = \begin{bmatrix} x^3 & x^2 & x & 1 \end{bmatrix} \begin{bmatrix} A & a \\ 3B & 3b \\ 3C & 3c \\ D & d \end{bmatrix}$$

We then calculate the function

$$f(x,y) = \frac{\mathbf{C}(x)\mathbf{D}(y) - \mathbf{C}(t)\mathbf{D}(y)}{x-y}$$

We can do this as matrix operations by

$$\mathbf{C}(x)\mathbf{D}(y)-\mathbf{C}(y)\mathbf{D}(x)=$$
$$\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}A\\3B\\3C\\D\end{bmatrix}\begin{bmatrix}a & 3b & 3c & d\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}-\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}a\\3b\\3c\\d\end{bmatrix}\begin{bmatrix}A & 3B & 3C & D\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}$$

Expanding the outer products of (ABCD,abcd) we get

$$\mathbf{C}(x)\mathbf{D}(y)-\mathbf{C}(y)\mathbf{D}(x)=\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}Aa & A3b & A3c & Ad\\3Ba & 3B3b & 3B3c & 3Bd\\3Ca & 3C3b & 3C3c & 3Cd\\Da & D3b & D3c & Dd\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}$$
$$-\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}Aa & 3Ba & 3Ca & Da\\A3b & 3B3b & 3C3b & D3b\\A3c & 3B3c & 3C3c & D3c\\Ad & 3Bd & 3Cd & Dd\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}$$

Doing the subtraction

$$\mathbf{C}(x)\mathbf{D}(y)-\mathbf{C}(y)\mathbf{D}(x)=$$
$$\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}0 & 3(Ab-Ba) & 3(Ac-Ca) & Ad-Da\\3(Ba-Ab) & 0 & 9(Bc-Cb) & 3(Bd-Db)\\3(Ca-Ac) & 9(Cb-Bc) & 0 & 3(Cd-Dc)\\Da-Ad & 3(Db-Bd) & 3(Dc-Cd) & 0\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}$$

Give single letter names to the six unique elements of this matrix

$$p=3(Ab-Ba),\ q=3(Ca-Ac),\ r=Ad-Da$$
$$s=9(Bc-Cb),\ t=3(Db-Bd),\ u=3(Cd-Dc)$$

We can now see that the calculation takes the form of an anti-symmetric matrix

$$\mathbf{C}(x)\mathbf{D}(y)-\mathbf{C}(y)\mathbf{D}(x)=\begin{bmatrix}x^3 & x^2 & x & 1\end{bmatrix}\begin{bmatrix}0 & p & -q & r\\-p & 0 & s & -t\\q & -s & 0 & u\\-r & t & -u & 0\end{bmatrix}\begin{bmatrix}y^3\\y^2\\y\\1\end{bmatrix}$$

digression

There is an interesting connection here. This looks a lot like the 3D(4DH) calculation of a 3D line through two 3D points described in an earlier chapter. (What is the geometric connection?) In that chapter we showed any (*pqrstu*) derived in such a manner must satisfy the identity:

$$(pu-qt+rs)=0$$

The anti-symmetric matrix is rank 2 since eigenvalues are solutions to the characteristic equation

$$\lambda^4 + \lambda^2\left(t^2 + u^2 + s^2 + p^2 + q^2 + r\right) + \left(pu - qt + rs\right)^2 = 0$$

In other words we have two zero eigenvalues.

end digression

The matrix can be seen as the sum of a symmetric 3x3 matrix shifted in two directions, corresponding to multiplication by x and $-y$.:

$$\begin{bmatrix} 0 & p & -q & r \\ -p & 0 & s & -t \\ q & -s & 0 & u \\ -r & t & -u & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -p & q & -r & 0 \\ q & -s-r & t & 0 \\ -r & t & -u & 0 \end{bmatrix} + \begin{bmatrix} 0 & p & -q & r \\ 0 & -q & s+r & -t \\ 0 & r & -t & u \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So

$$\mathbf{C}(x)\mathbf{D}(y) - \mathbf{C}(y)\mathbf{D}(x) =$$

$$\begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} p & -q & r \\ -q & s+r & -t \\ r & -t & u \end{bmatrix} \begin{bmatrix} y^3 \\ y^2 \\ y \end{bmatrix} + \begin{bmatrix} x^3 & x^2 & x \end{bmatrix} \begin{bmatrix} p & -q & r \\ -q & s+r & -t \\ r & -t & u \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

when we divide by $x - y$ we get the desired result:

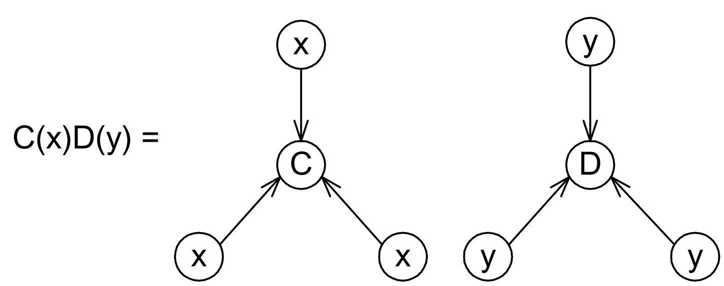
$$f(x, y) = \frac{\mathbf{C}(x)\mathbf{D}(y) - \mathbf{C}(y)\mathbf{D}(x)}{x - y} = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} p & -q & r \\ -q & s+r & -t \\ r & -t & u \end{bmatrix} \begin{bmatrix} y^2 \\ y \\ 1 \end{bmatrix}$$

Now for the punchline. If x is a root of both **C** and **D** then $f(x,y)$=0 no matter what y is (as long as its different from x). This can only happen if the matrix is singular. So setting the determinant of the matrix to zero gives us the resultant:

$$\boxed{\begin{aligned} \rho(\mathbf{C}, \mathbf{D}) &= pu(s+r) + 2qtr - r^2(s+r) - t^2 p - q^2 u \\ &= psu + pru - pt^2 + 2qrt - q^2 u - r^2 s - r^3 \end{aligned}}$$

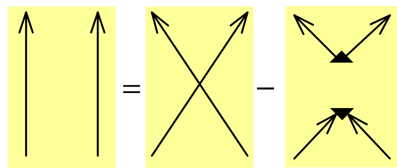
Diagram Version

The cubics **C** and **D** are three-pronged nodes. We first calculate

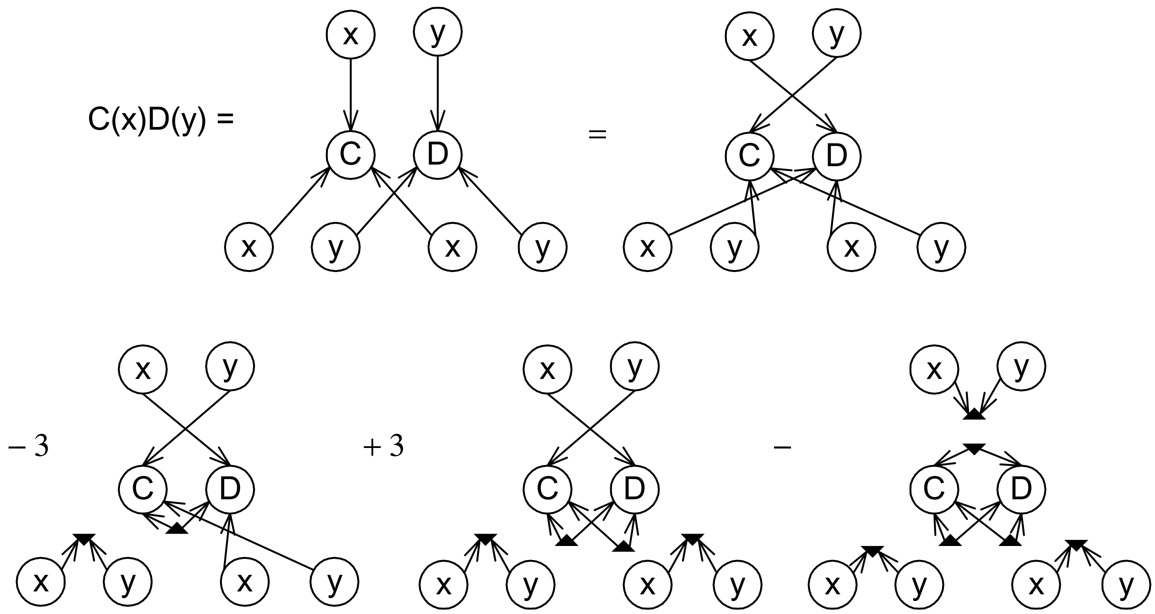


We automatically go to homogeneous land when we do this, so x and y now represent homogeneous 2-vectors.

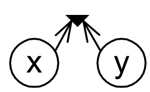
Use the $\varepsilon\delta$ identity



We us this identity three times, on each of the three pairs of parallel arcs. Each application will double the number of diagram fragments (to a total of 8), but many of these are equal to each other. Some playing around with this should convince you that:

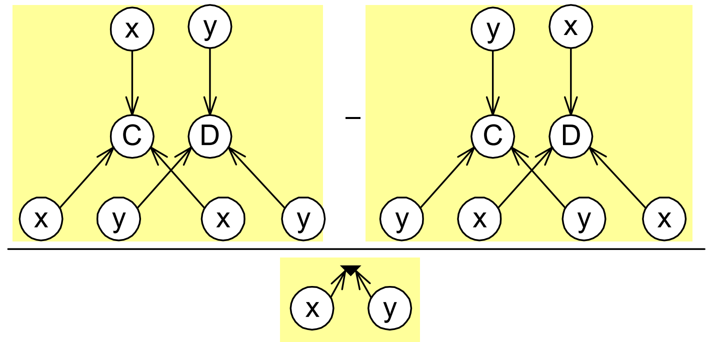


Notice two things about the right hand side of this equation. The first term is just **C**(y)**D**(x) and the remaining terms have a common factor of

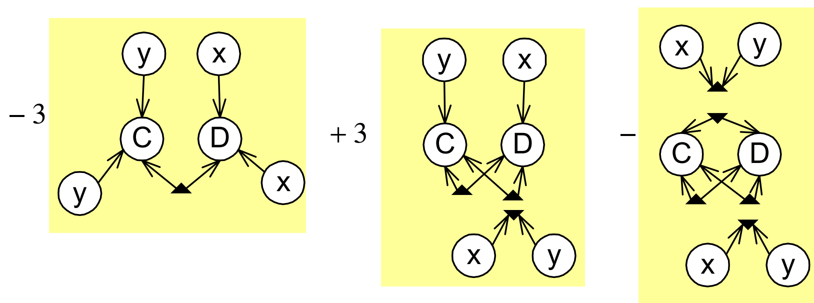


This is just the homogeneous formulation of *x-y*. We can put this all together to get:

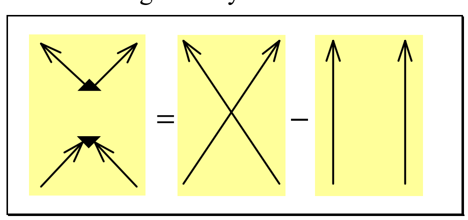
$$\frac{\mathbf{C}(x)\mathbf{D}(y)-\mathbf{C}(y)\mathbf{D}(x)}{x-y} =$$



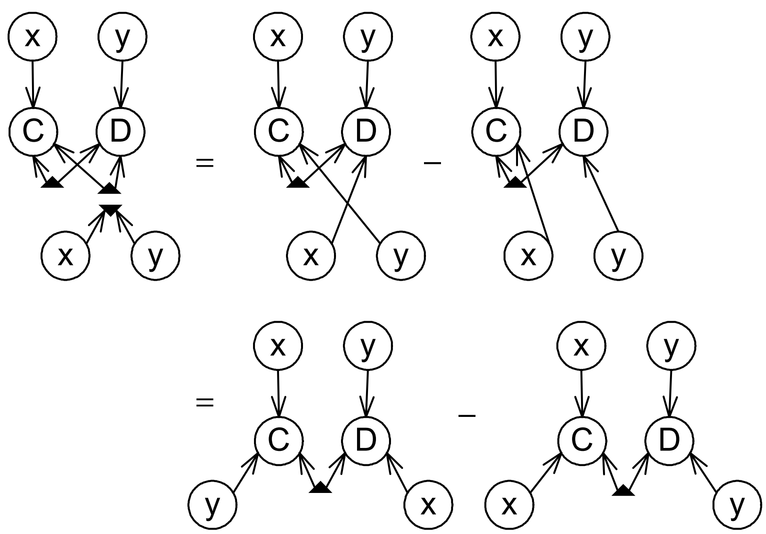
=



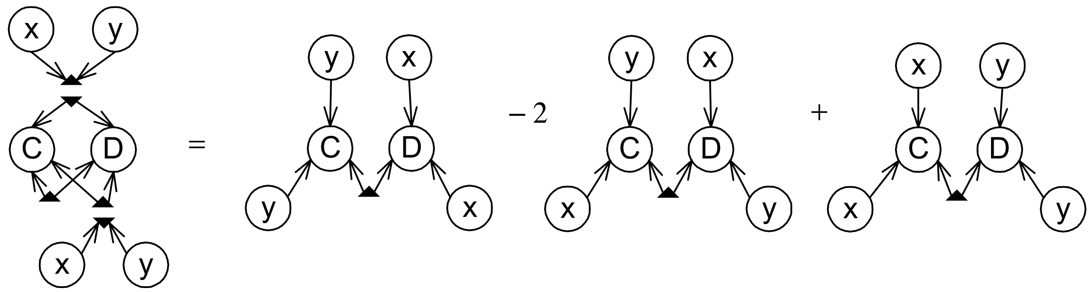
Now re-ravel the second and third terms using identity:



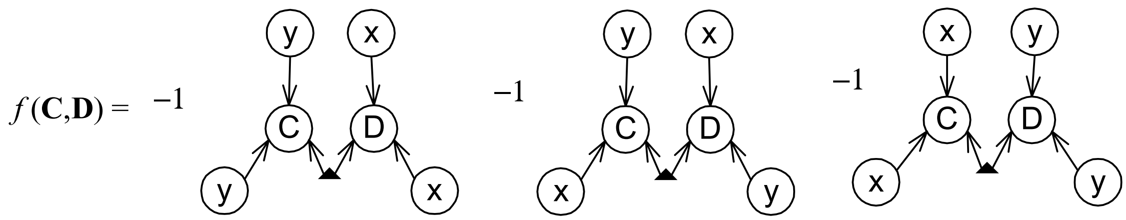
Second term gives:



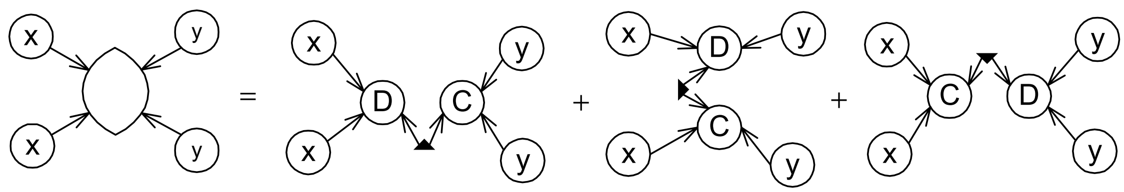
The third term gives, after some work:



Putting these all together with the proper factors of 3 etc, gives:



We can rearrange the x and y nodes to lump this together into a four-pronged node. Note that this is not a symmetrical with respect to exchange of a single x and y node. We indicate this by making the node oddly shaped. The idea is that you can exchange any arcs along a smooth side but not across an angular protrusion.



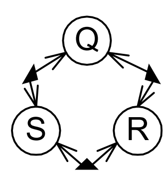
Now we need to figure out how to interpret this as a 3x3 matrix to take the determinant. This is tricky since our 1DH diagram technology only knows about 2x2 matrices.

3x3 determinant

Consider the following. We look at three 2x2 symmetric matrices. (I am recycling letters for the purposes of this section:

$$\mathbf{Q}=\begin{bmatrix} A & B \\ B & C \end{bmatrix},\ \mathbf{R}=\begin{bmatrix} D & E \\ E & F \end{bmatrix},\ \mathbf{S}=\begin{bmatrix} G & H \\ H & J \end{bmatrix}$$

Now evaluate the diagram:



The explicit evaluation of this gives

$$trace\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} D & E \\ E & F \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} G & H \\ H & J \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$$= trace\begin{bmatrix} -B & A \\ -C & B \end{bmatrix}\begin{bmatrix} -E & D \\ -F & E \end{bmatrix}\begin{bmatrix} -H & G \\ -J & H \end{bmatrix}$$

$$= trace\begin{bmatrix} BE-AF & AE-BD \\ CE-BF & BE-DC \end{bmatrix}\begin{bmatrix} -H & G \\ -J & H \end{bmatrix}$$

$$= trace\begin{bmatrix} -BEH+AFH-AEJ+BDJ & * \\ * & CEG-BFG+BEH-CDH \end{bmatrix}$$

$$= AFH-AEJ+BDJ+CEG-BFG-CDH$$

Now compare that with the determinant of the matrix:

$$\det\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & J \end{bmatrix} = AEJ+BFG+CDH-GEC-HFA-JDB$$

In other words, the ring shaped diagram of 2x2 matrices equals minus the determinant of the 3x3 matrix.

When we interpret the components of the 2x2 matrices as quadratic polynomials, converting to a 3-vector implies that we are adding the cross diagonal elements twice. That is:

$$\begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} A & 2B & C \end{bmatrix}\begin{bmatrix} x^2 \\ xw \\ w^2 \end{bmatrix}$$

This only generates a global constant however, since:

$$\det\begin{bmatrix} A & 2B & C \\ D & 2E & F \\ G & 2H & J \end{bmatrix} = 2\det\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & J \end{bmatrix}$$

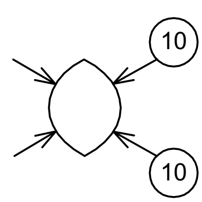
In addition, the 3x3 matrix we are dealing with will be symmetric. Again this only introduces another constant

$$\det\begin{bmatrix} A & 2B & C \\ 2B & 4E & 2H \\ C & 2H & J \end{bmatrix} = 4\det\begin{bmatrix} A & B & C \\ B & E & H \\ G & H & J \end{bmatrix}$$

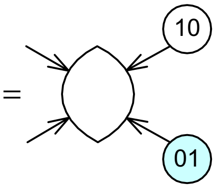
Back to 2x2

So what are the three 2x2 matrices. I believe the following is true

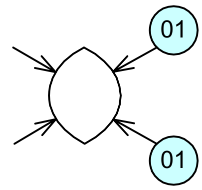
$$\begin{bmatrix} p & -q \\ -q & r \end{bmatrix} =$$



$$\begin{bmatrix} -q & s+r \\ s+r & -t \end{bmatrix} =$$

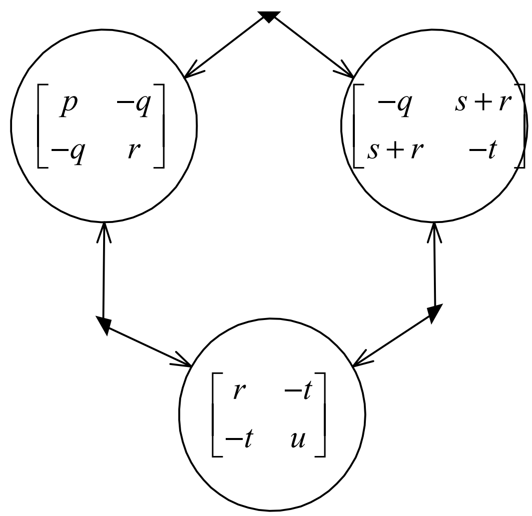


$$\begin{bmatrix} r & -t \\ -t & u \end{bmatrix} =$$

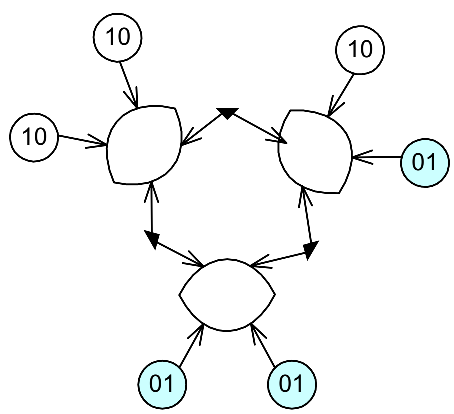


I have colored the (01) basis vector blue to make it stand out from the (1,0) vector.
So we have the resultant being:

$$\det\begin{bmatrix} p & -q & r \\ -q & s+r & -t \\ r & -t & u \end{bmatrix} =$$



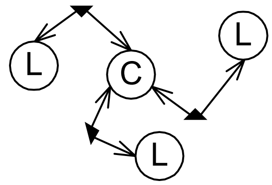
Substitiuting the above we get



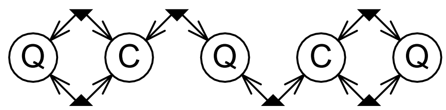
This still requires substituting the contents of the lemon shaped node into this ring. Each substitution generates three subdiagrams for a total of 3*3*3=27 diagrams we need to mash around. This complexity can probably be tamed by looking at various symmetries but I have't gotten it to work yet.

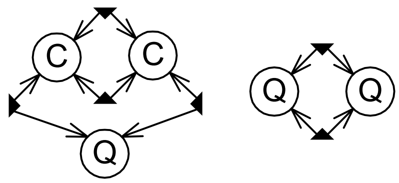
Solution 2

The approach I will describe in this section is similar to what we did to get the resultant of a Quadratic and Cubic given the diagrams for the resultant of Quadratic-Quadratic and the resultant of Linear-Quadratic. We start with

$\rho(C,L) =$ 

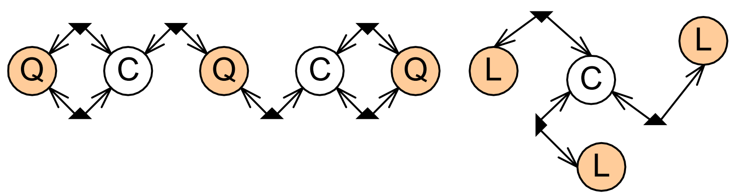
and

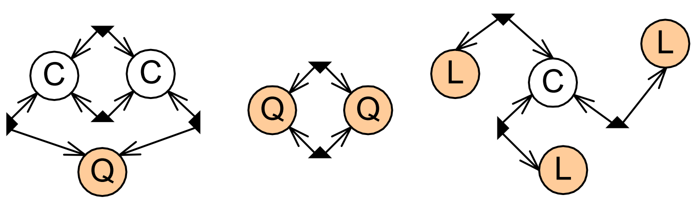
$\rho(C,Q) =$ 1/9 

+2/9 

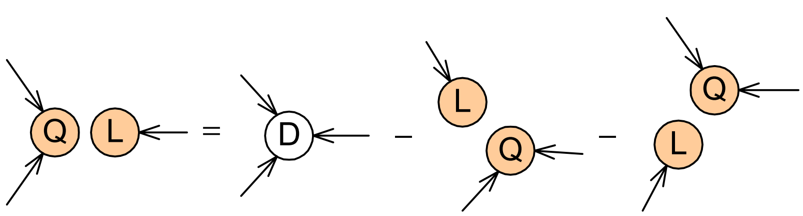
(I will toss out the factor of 1/9 for the rest of this discussion.) If we think of our new cubic **D** as being the product of linear **L** and quadratic **Q** we can say that the desired resultant is

$$\rho(C,D) = \rho(C,Q)\,\rho(C,L)$$

$\rho(C,D) =$ 

+ 2 

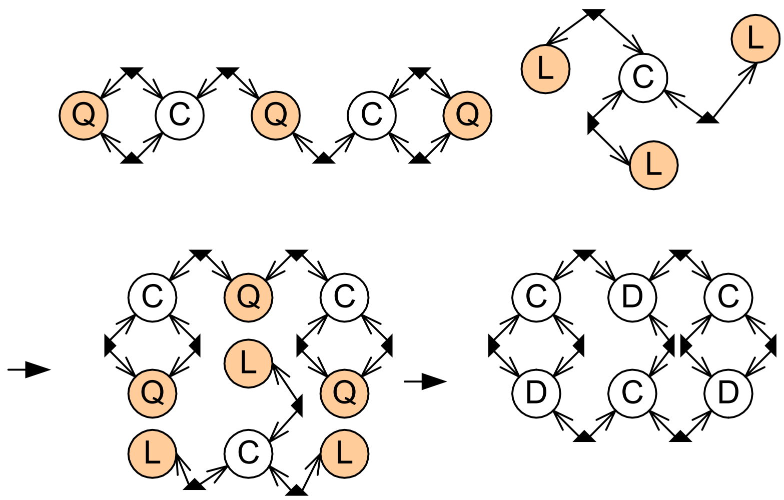
Once again, we have shaded the nodes that are factors of **D**. Now we operate on these diagram fragments to merge the shaded nodes into **D** nodes by using the relation



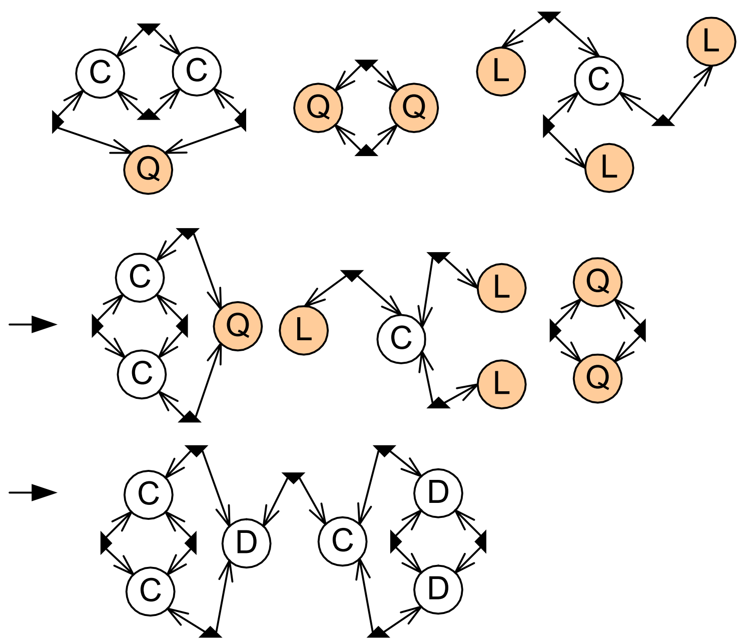
This unfortunately gets complicated rather fast.

What we expect

Rather than launching into this I want to preview the sorts of diagram fragments that are likely to show up. Looking at the above equation for the resultant we can see that one of the terms of the answer will be formed by just the first of the terms in the equation for D. Looking at just this term and neglecting signs and constants for now we see



And the second term will become (among other things)



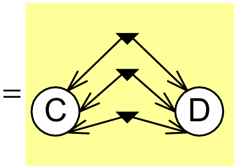
In other words, our resultant will be some sum of diagram fragments each containing three **C** nodes and three **D** nodes, with the appropriate number of epsilons (9) to glue them together.

Family Portraits

Lets spend the remainder of this chapter getting friendly with diagram fragments containing equal numbers of C and D nodes.

Order 2 Diagrams

There is one such diagram

$\mathbf{I}_2 =$ 

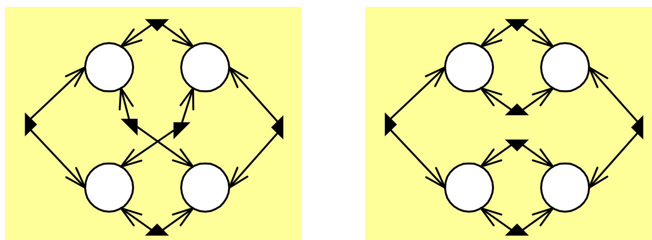
In terms of the components of C and D this happens to be

$$\mathbf{I}_2 = Ad - 3Bc + 3Cb - Da$$

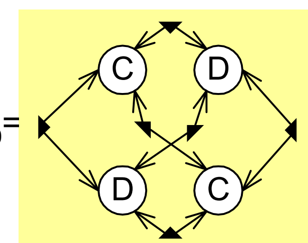
Note that, since it has an odd number of epsilons, swapping the **C** and **D** coefficients flips the sign.

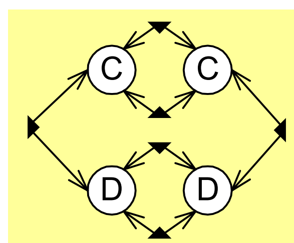
Order 4 Diagrams

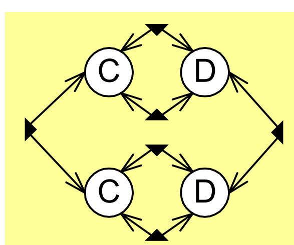
There are two ways to glue four nodes together

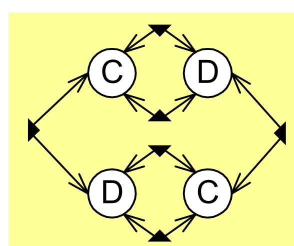


Distributing two each of **C** and **D** letters in these gives

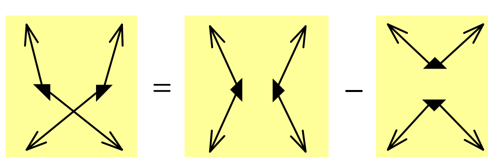
$D_0=$ 

$D_1=$ 

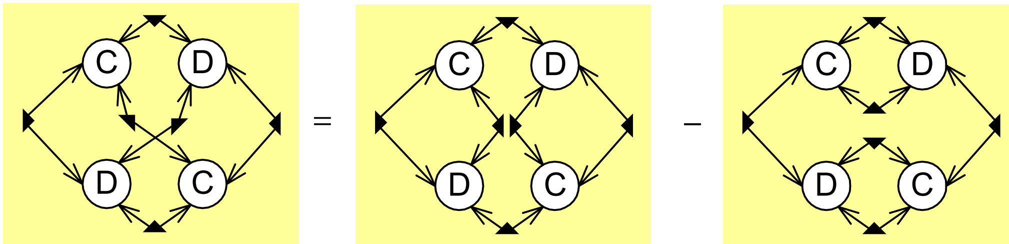
$D_2=$ 

$D_3=$ 

There are some interesting relationships here. We will apply the $\varepsilon\varepsilon$ identity



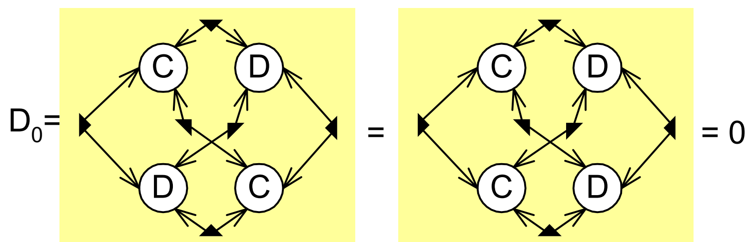
Apply this to D0



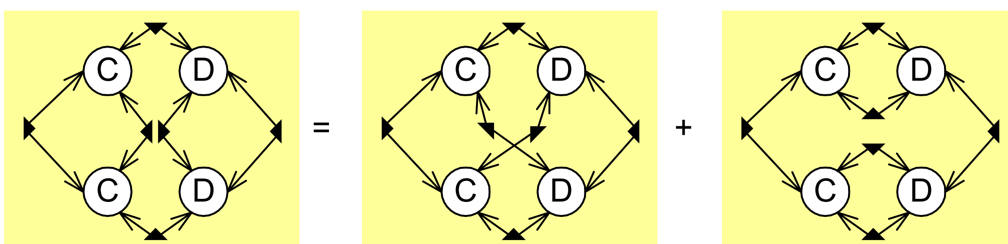
The two diagrams on the right are both equal to D3 so we have just shown that

$$\boxed{D_0 = D_3 - D_3 = 0}$$

Note that any permutation of the positions of C and D on the D0 diagram still equals D0 (the epsilon arcs basically connect the nodes into a tetrahedron. Any labeling of two vertices of a tetrahedron with a C will be the same.)



Next, lets apply $\varepsilon\varepsilon$ to diagram D1

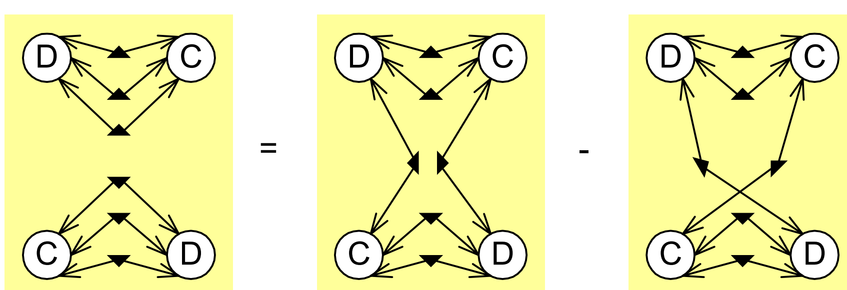


We have shown that

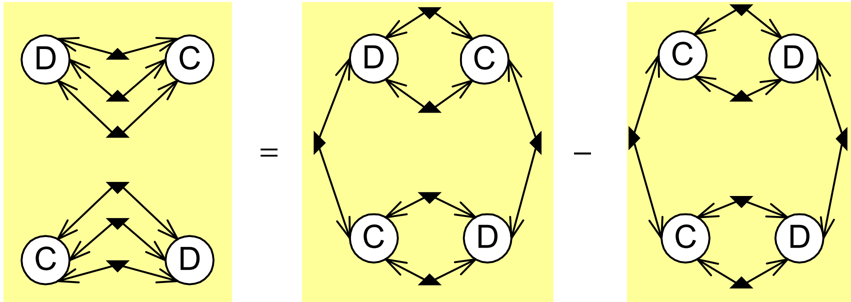
$$D_1 = D_0 + D_2$$
$$D_1 = D_2$$

Relate Order 2 and Order 4

Apply $\varepsilon\delta$ to the square of $\mathbf{I}_2$



This cleans up to



$$\mathbf{I}_2{}^2 = D_3 - D_2$$

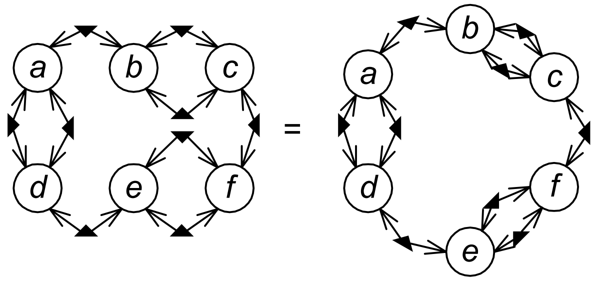
So we only have two unique invariants so far $\mathbf{I}_2$ and D_3. The others come from

$$D_0 = 0$$
$$D_1 = D_2$$
$$D_2 = D_3 - \mathbf{I}_2{}^2$$

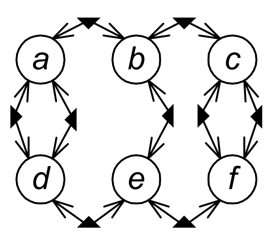
Order 6

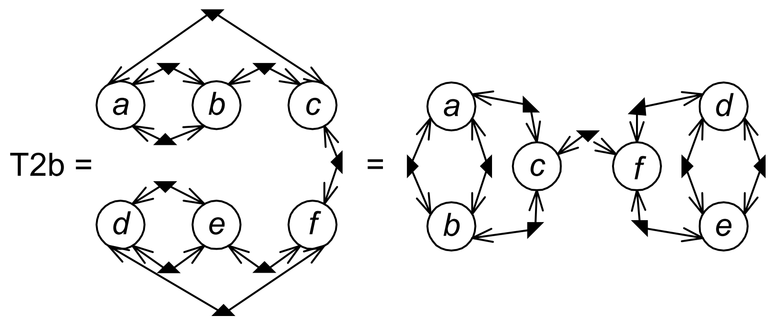
There are lots of Possible topological connections between 6 tensor nodes. The ones I have found so far are:

Three Double Bonds

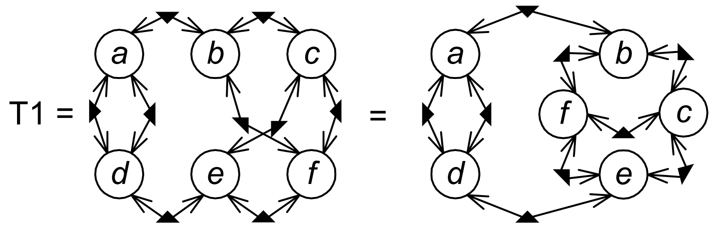
T3 =


Two Double Bonds

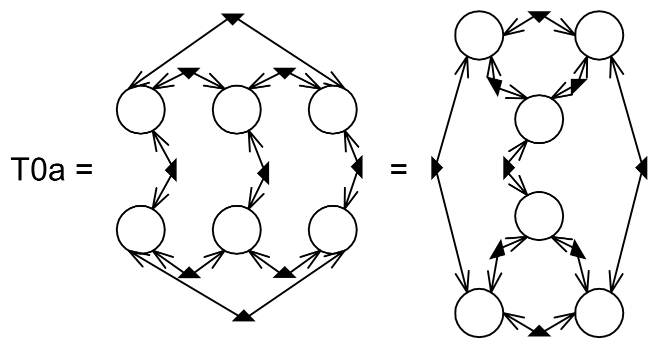
T2a =


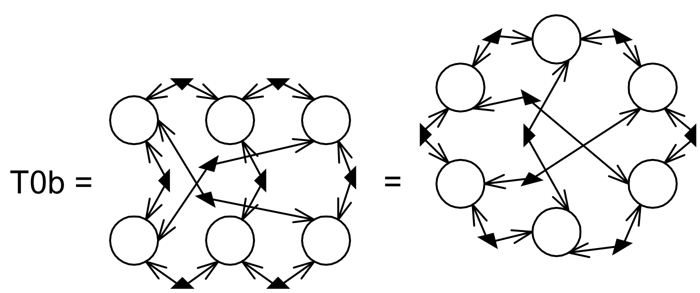


One Double Bond



Zero Double Bonds





Then counting the number of ways to distribute three C's and three D's among these gets even more complicated. I suspect however that there are many relations between them and several are identically zero.

Why

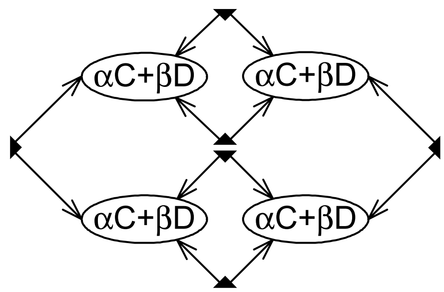
The reason to look at all these types of diagram fragments is simply this: The resultant (when we figure it out) does not contain all the information we need to truly express the transformationally invariant relationships between two cubic polynomials. Just as we saw in the Quadratic-Quadratic case, we need to look at the diagram fragments individually to get this extra information. It appears that we are going to have to get friendly with more of these diagram fragments to do this.

Solution 3

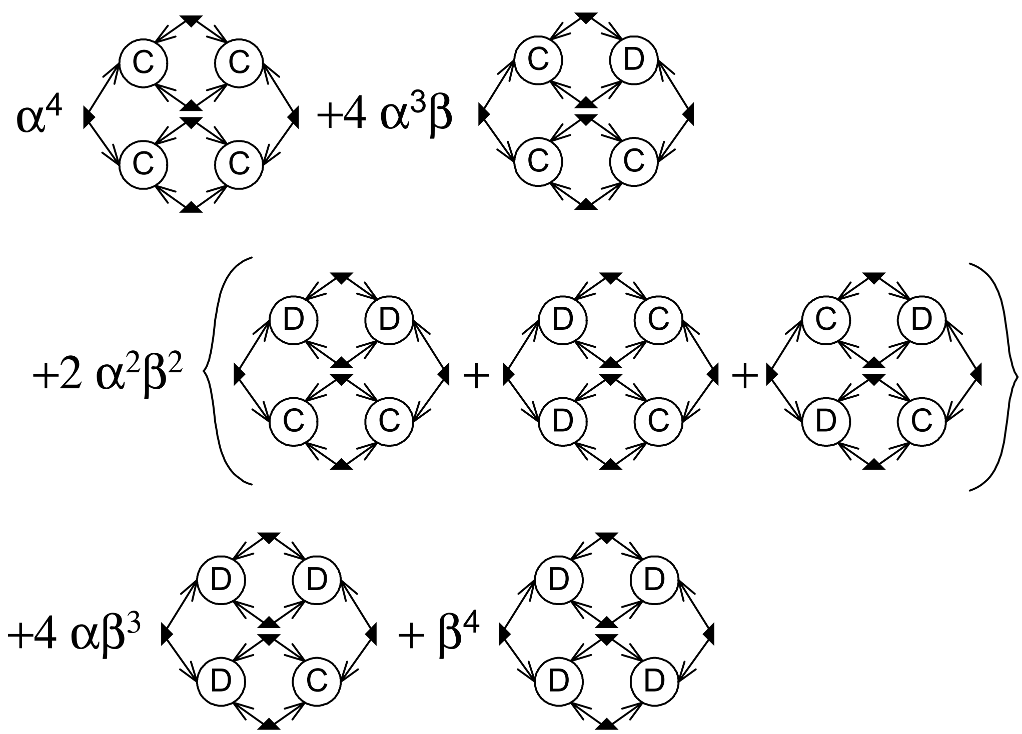
Here's another approach that shows great promise. It is based on the final technique shown in the chapter on the resultant of two quadratics. We look at the root structure of the linear combination of the two cubics:

$$\alpha \mathbf{C} + \beta \mathbf{D}$$

If C and D have a common root, then this will be a root of all the cubics along the blend axis. We will want to look at places where there are double roots as a function of (alpha,beta). This means evaluating the discriminant of the blend of the cubics. The diagram for this is:



Which expands out to:



This is a quartic polynomial in (alpha, beta). A later chapter tells us how to find discriminants of this. Setting this discriminant to zero should give us a resultant for C and D. The margins are too small to contain the complete proof. (Actually the Siggraph deadline is approaching so I wont be able to finish this here.)

Chapter 1-09

1DH(2D)
Homogeneous Quartic Polynomials

This chapter discusses quartic polynomials. These have the form

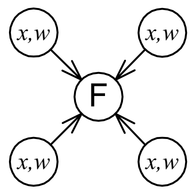
$$F(x,w) = Ax^4 + 4Bx^3w + 6Cx^2w^2 + 4Dxw^3 + Ew^4$$

Writing this as a matrix gives

$$F(x,w) = \begin{bmatrix} x & w \end{bmatrix} \left\{ \begin{bmatrix} x & w \end{bmatrix} \begin{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} & \begin{bmatrix} B & C \\ C & D \end{bmatrix} \\ \begin{bmatrix} B & C \\ C & D \end{bmatrix} & \begin{bmatrix} C & D \\ D & E \end{bmatrix} \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} \right\} \begin{bmatrix} x \\ w \end{bmatrix}$$

and as a tensor diagram

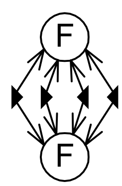
$\mathbf{F}(x,w) =$



Hilberts Invariants

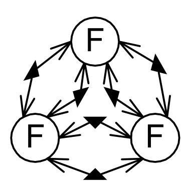
About a hundred years ago, David Hilbert gave a series of lectures on invariants of binary forms. These were recently reprinted in {D. Hilbert, *Theory of Algebraic Invariants,* Cambridge University Press, 1993}. Binary forms are just what we are dealing with here in the 1DH(2D) domain. Hilbert found some rather complex rules for generating invariants from these forms that, in our notation, translate readily into tensor diagrams. I have already mentioned how he found two invariant quantities that, translated into tensor diagrams are:

$F_2 \equiv$



and

$F_3 \equiv$

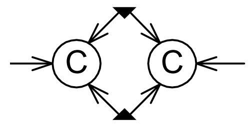


The discriminant of the quartic is then:

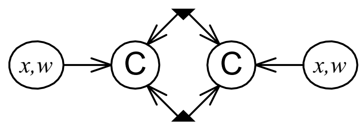
$$\Delta_4 = \tfrac{1}{8}\left(6\left(F_3\right)^2 - F_2^{\ 3}\right)$$

Hilbert's Covariants

Hilbert also showed that there are several other quantities which he called *covariants*. In diagram terms, these are diagrams that don't have only covariant tensor nodes, but still have some contravariant parameter nodes. (There might be a co/contra notational mismatch here). We have already seen one of these in the cubic polynomial analysis. The cubic has a double root if the discriminant is zero. The discriminant is the determinant of the matrix:



If, however, the entire matrix is zero, the polynomial has a triple root. We can state this another way by saying that the polynomial has a triple root if the quantity

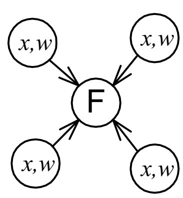


is zero for all values of (x,w). This latter diagram is what Hilbert calls a covariant. The idea is that a when you transform C by the matrix **T**, you also transform (x,w) by the matrix $\mathbf{T}^{-1}$. Plugging these into the covariant diagram shows that the cancel out. For our purposes it means that the polynomial in (x,w) represented by the covariant will be identically zero both before and after the coordinate transformation. In practical terms, a Hilbert covariant being zero gives several polynomials in the coefficients of C that all must be zero simultaneously (one for each coefficient of the covariant). This situation can represent other properties of the root structure of C than the simple existence of a double root. We will now see how it represents more properties for the quartic.

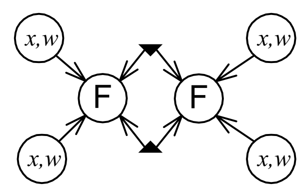
Covariants of a Quartic

Hilbert finds the following three covariants for a quartic

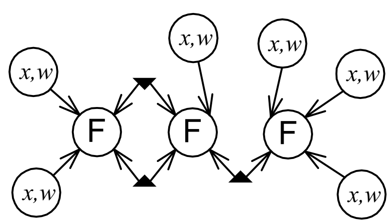
$C_1 \equiv$



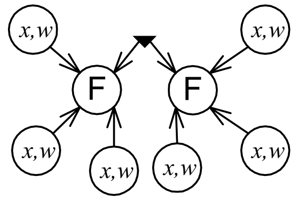
$C_2 \equiv$



$C_3 \equiv$



Hilbert's claim is that these are the only covariants. Any other covariant expression is either identically zero always, or is a function of these and the F2 and F3 invariants. I haven't played with this a lot yet to see how to verify it with tensor diagrams, but there certainly are other possible diagrams you could imagine. For example



This is obviously identically zero since it is identical stuff on either side of an epsilon. Etc..

Meaning of the Covariants

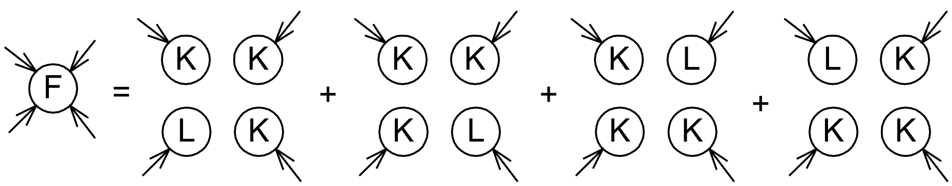
Lets see what setting these covariants to zero says about the root structure of F.

C1

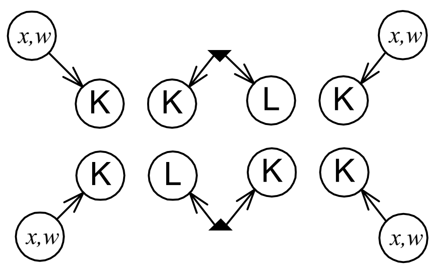
this one may seem a bit trivial, but it does satisfy the requirements for a covariant. Its geometric meaning is simply that, if it is identically zero, then the polynomial **F** is identically zero.

C2

C2 is another quartic polynomial. If it is identically zero it means that F has a quadruple root. You can see this by plugging in the following hypothetical factorization for F;



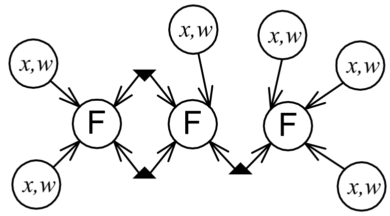
This factorization shows that F has at least a triple root. Plugging this in to the C2 definition gives a possible 16 terms for the two copies of F. Most are zero with a K^K term or an L^L term. The only ones that are not immediately zero are:



This tells us two things. If C2 identically equals zero then K must equal L and we have a quadruple root. If C2 is not identically zero and F has a triple root K then the quartic polynomial C2 has K as a quadruple root.

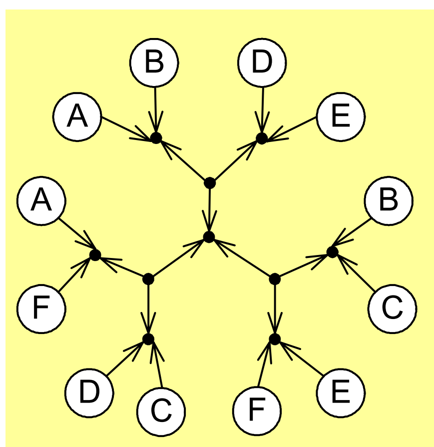
C3

If this covariant is identically zero for all x,y then the quartic F has two double roots. We can see this by first noting an interesting fact. The covariant C3 being equal to zero means that the covariant quartic C2 is just a homogeneous scale of the original quartic F. Why is this? Consider C3 as two subdiagrams on either side of an epsilon:

$C_3 \equiv$ 

I’ll leave the rest as an exercise if I don’t get to write it up in time for the course notes deadline.

PART 2
2DH (3D)



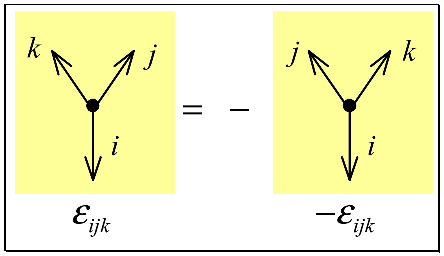
The statement of the Theorem of Pappus

Chapter 2-00

2DH(3D)
Diagram Identity Catalog

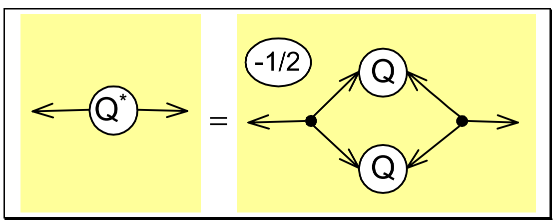
Epsilon

Antisymmetry of epsilon: mirroring (or reversing order of traversal) negates the quantity.

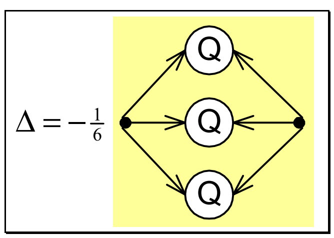


Determinants etc

adjoint

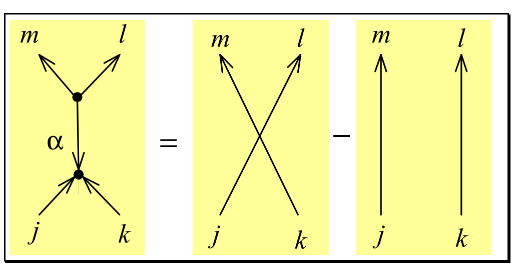


Determinant



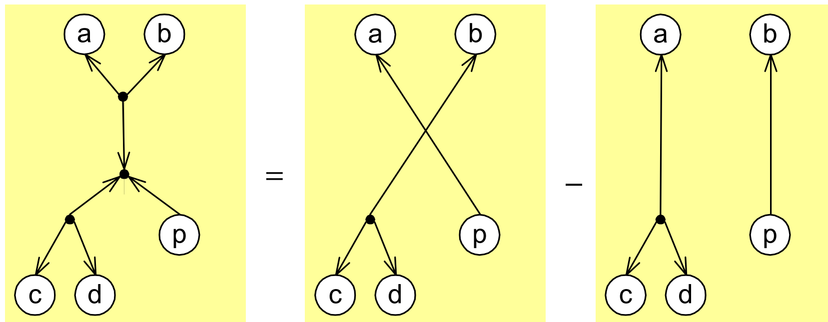
Epsilon Delta Identity

As shown earlier

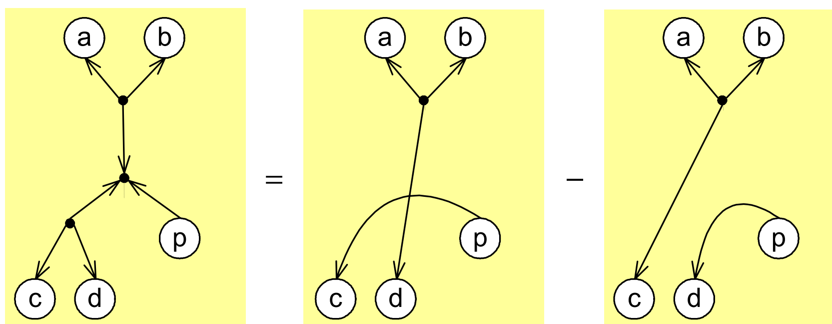


A variant

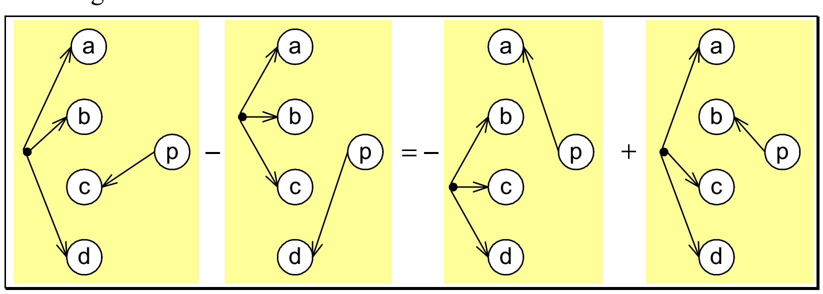
Stick an epsilon onto the left bottom arc of the $\varepsilon\delta$ identity and on nodes a,b,c,d,p to keep track of free indices



Apply $\varepsilon\delta$ to the left diagram fragment



put together and rearrange



Interpret it more matrix terms: p is a row vector, a,b,c,d are column vectors. Write a,b,c,d as a matrix:

$$\begin{bmatrix}\cdots\mathbf{p}\cdots\end{bmatrix}\begin{bmatrix}\vdots & \vdots & \vdots & \vdots\\ \mathbf{a} & \mathbf{b} & \mathbf{c} & \mathbf{d}\\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}=\begin{bmatrix}\mathbf{p}\cdot\mathbf{a} & \mathbf{p}\cdot\mathbf{b} & \mathbf{p}\cdot\mathbf{c} & \mathbf{p}\cdot\mathbf{d}\end{bmatrix}$$

The identity says that:

$$(\mathbf{p}\cdot\mathbf{c})\det\begin{bmatrix}\vdots & \vdots & \vdots\\ \mathbf{a} & \mathbf{d} & \mathbf{b}\\ \vdots & \vdots & \vdots\end{bmatrix}-(\mathbf{p}\cdot\mathbf{d})\det\begin{bmatrix}\vdots & \vdots & \vdots\\ \mathbf{a} & \mathbf{c} & \mathbf{b}\\ \vdots & \vdots & \vdots\end{bmatrix}+(\mathbf{p}\cdot\mathbf{a})\det\begin{bmatrix}\vdots & \vdots & \vdots\\ \mathbf{b} & \mathbf{d} & \mathbf{c}\\ \vdots & \vdots & \vdots\end{bmatrix}-(\mathbf{p}\cdot\mathbf{b})\det\begin{bmatrix}\vdots & \vdots & \vdots\\ \mathbf{a} & \mathbf{d} & \mathbf{c}\\ \vdots & \vdots & \vdots\end{bmatrix}=0$$

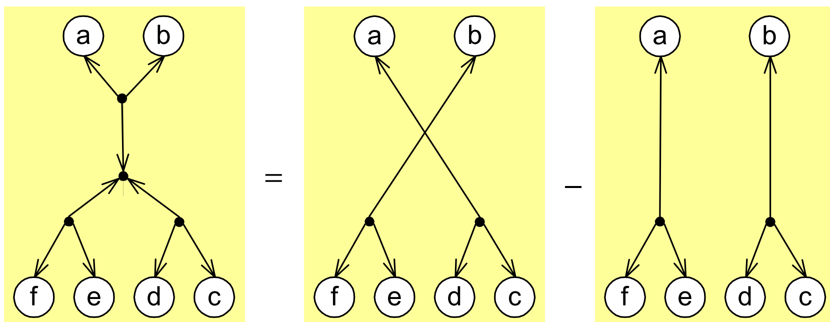
Or more succinctly, eliminating **p** as a free variable:

$$+\det\begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{a} & \mathbf{b} & \mathbf{c} \\ \vdots & \vdots & \vdots \end{bmatrix}\mathbf{d} - \det\begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{b} & \mathbf{c} & \mathbf{d} \\ \vdots & \vdots & \vdots \end{bmatrix}\mathbf{a} - \det\begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{d} & \mathbf{c} & \mathbf{a} \\ \vdots & \vdots & \vdots \end{bmatrix}\mathbf{b} - \det\begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{d} & \mathbf{a} & \mathbf{b} \\ \vdots & \vdots & \vdots \end{bmatrix}\mathbf{c} = 0$$

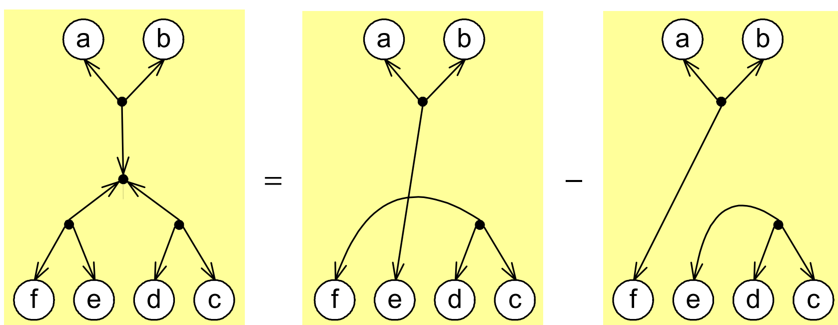
(Check Signs)

Epsilon Epsilon Identity

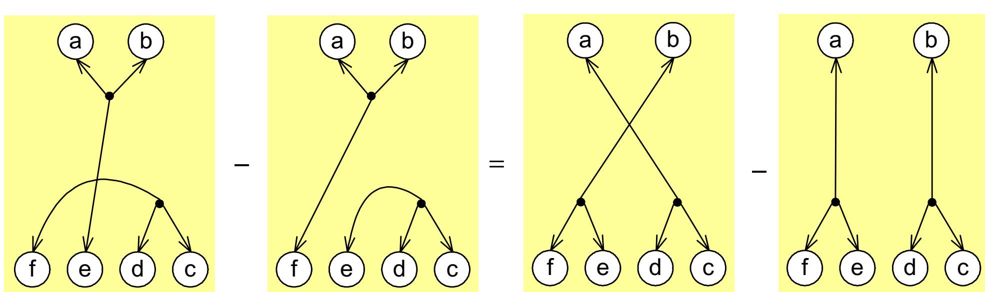
Stick an epsilon onto each of the bottom arcs of the $\varepsilon\delta$ identity. We will also stick on arbitrary nodes a,b,c,d,e,f to keep track of free indices:



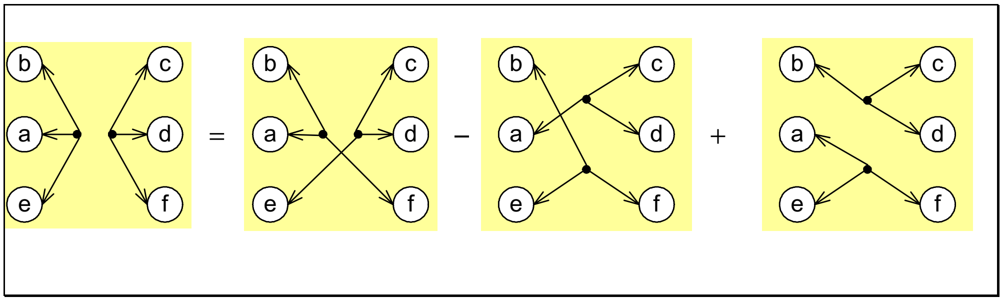
Now apply $\varepsilon\delta$ to the left hand diagram:



So we have:



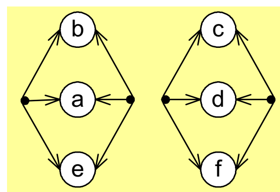
Rearrange



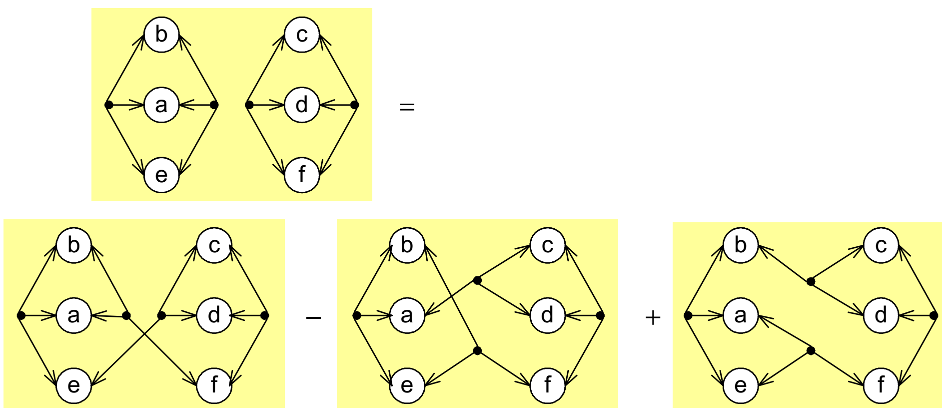
(Check Signs)

A use

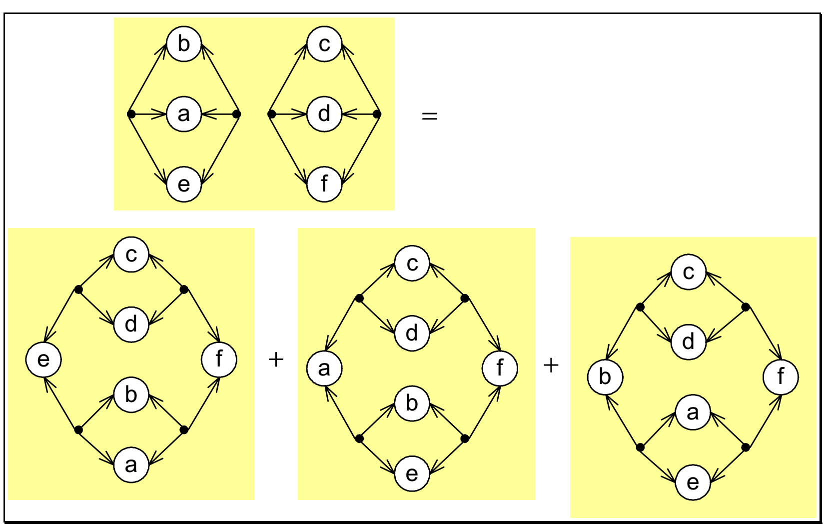
Suppose we have the product of two cross determinants:



Then this is equal to:



Rearrange to



Chapter 2-01

2DH(3D)
Quadratic Curves

Definition

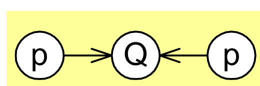
A second order (quadratic) plane curve has the equation:

$$\begin{aligned} &Ax^2 + 2Bxy + Cy^2 \\ &\quad +2Dxw + 2Eyw \\ &\qquad\qquad +Fw^2 = 0 \end{aligned}$$

All homogeneous points $\begin{bmatrix} x & y & w \end{bmatrix}$ that satisfy the equation lie on the curve. This can be written as a matrix:

$$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} A & B & D \\ B & C & E \\ D & E & F \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \mathbf{pQp}^T = 0$$

The tensor diagram looks like:



= 0

I am going to examine such curves both for analysis (given **Q** find its geometric properties) and for synthesis (given desired geometric properties, construct **Q**)

The Catalog

How many different curves can this generate? This, of course, depends on what you mean by different. In Euclidean geometry we are used to the idea that all equilateral triangles are the same shape, no matter how they are placed or oriented. Likewise, in projective geometry, an ellipse and a hyperbola are the same shape. In fact, geometry can be described as the study of those properties of a shape that remain unchanged even if it's subjected to some transformation. Different "don't care" types of transformations generate different geometries. I am going to deal with two-dimensional homogeneous coordinates with the transformation being the standard homogeneous projective transformation representable by a 3×3 matrix. Any two shapes that can transform into each other via such a matrix are counted as the same shape.

It turns out that there are exactly 5 unique quadratic curves. Why? As you might expect, it can generate conic sections. But all conic sections are really the same shape, so that's 1. Anything else? Remember, two shapes are considered the same if there is some homogeneous transformation (possibly containing perspective) that can change one into another.

To answer the question I will invoke several standard results from matrix theory. We can geometrically transform the curve by multiplying its **Q** matrix by an arbitrary 3x3 transform:

$$\mathbf{TQT}^t = \mathbf{Q}'$$

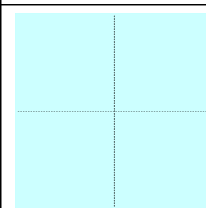
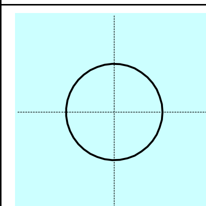
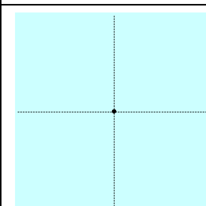
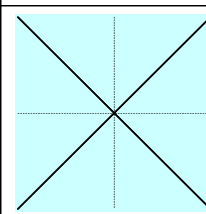
where **T** is the inverse of the matrix that transforms points.

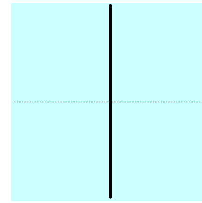
It is always possible to find a transformation that turns **Q** into a diagonal matrix. This just involves finding the eigenvectors of **Q.** Since **Q** is symmetric, all the eigenvalues will turn out to be real numbers, and all the eigenvectors will be perpendicular to each other. So we can make up the transformation **T** by stacking up the eigenvectors. What we get is

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} \mathbf{Q} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

Then just multiply on the right by the inverse of the matrix formed by the column of V's and you get **Q** transformed into a diagonal matrix with the eigenvalues on the diagonal. This is why eigenvectors are so interesting. They allow you to see the essence of a matrix without getting confused by the coordinate system it's in.

We can further simplify this by applying a scaling transformation that will scale the diagonal elements to be either +1, −1, or 0. (Notice that you can't change the sign of an eigenvalue even with a mirror transformation, since the transformation is multiplied in twice.) Thus the only types of second order curves are those that represent unique combinations of these values on the diagonal. Only one of them generates a true non-degenerate curve, the others are singular and represent such things as the products of two lines. The five possible combinations along with the eigenvalue signs are:

Signs of eigenvalues	name	Homogeneous example	Projects to	picture
+++ ---	(N)Null	$x^2+y^2+w^2=0$	$X^2+Y^2=-1$	
++- --+	(C)Conic	$x^2+y^2-w^2=0$	$X^2+Y^2=1$	
++0 --0	(P)Single Point	$x^2+y^2=0$	$X=0, Y=0$	
+-0	(L2)Two distinct lines	$x^2-y^2=0$	$(X+Y)(X-Y)=0$	

+00	(L1)Double line	$x^2 = 0$	$X^2 = 0$	

Rank and Invariants of a Symmetric 3x3 Matrix

How do we find this, given an arbitrary matrix, which of these we have? We want to find the signs of the eigenvalues. The eigenvalues themselves come as roots to the characteristic equation formed from the determinant

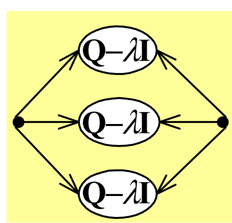
$$\det\{\mathbf{Q} - \lambda \mathbf{I}\}$$

Explicitly, this gives us

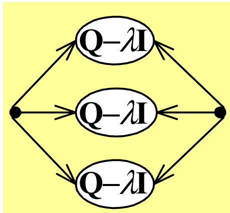
$$\det \begin{bmatrix} A-\lambda & B & C \\ B & D-\lambda & E \\ C & E & F-\lambda \end{bmatrix} = 0$$

$$\begin{aligned} &-\lambda^3 \\ &+\lambda^2 (A+D+F) \\ &+\lambda \left(E^2 - DF + B^2 - AD + C^2 - AF\right) \\ &\left(ADF + 2BEC - B^2F - C^2D - E^2A\right) \\ &= 0 \end{aligned}$$

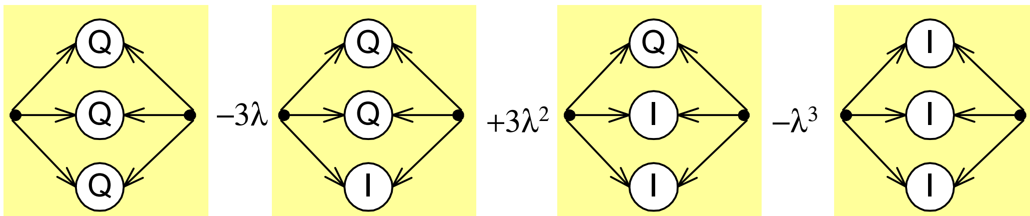
Each coefficient can be represented in diagram notation. The Rank is 3 minus the number of these (successively) that are zero. Expressing this in diagram notation shows us the following. The determinant is:



We learned how to take determinants of sums of matrices in an earlier chapter. We get eight terms all together, but a little though shows that these boil down to the following, which looks kind of like the cube of a binomial:



=



Each of the four diagrams is the diagram representation of the appropriate coefficient in the characteristic equation. We can recognize them in more conventional language as (a constant multiple times):

1. The determinant of Q
2. The trace of the adjoint of Q
3. The trace of Q
4. A constant

These are invariant properties of **Q** under Euclidean 3x3 transformation matrices (pure 3D rotations). For the more general case of general 3x3 matrices, only their sign (or zeroness) is invariant.

Degenerate Curves

In this section I will relate the properties of the three degenerate curves to their representation in diagram notation. This will include extraction of the geometric properties as well as construction of **Q** given some desired properties. I will also talk about stationary transforms. These are nonsingular transformation matrices that leave the shape of the object unchanged, while moving around all the other points that are not on the object. (These will be useful when we consider intersections of quadratics with other shapes.)

We start with the simplest.

Double Line

This is doubly degenerate. **Q** has two zero eigenvalues. The adjoint of **Q** is all zeroes.

Construction

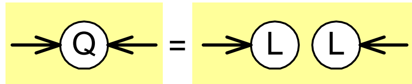
Given the vector for the desired line,

$$\mathbf{L} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Form:

$$\mathbf{Q} = \mathbf{L}\mathbf{L}^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

The diagram looks like



Analysis

Given **Q** how do we find **L**? Note that each row or column contains the three components of **L** times a uniform homogeneous factor. Any of *a,b,c* might be zero however. To ensure that we don't get nailed by this we must take the row or column with largest absolute length.

Stationary transform

Any transformation **T** that has **L** as one of its eigenvectors will leave **L** unchanged.

Two Distinct Lines

Singly degenerate. The adjoint of **Q** is doubly degenerate.

Construction

Given the vectors for the lines,

$$\mathbf{L}_1 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \mathbf{L}_2 = \begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

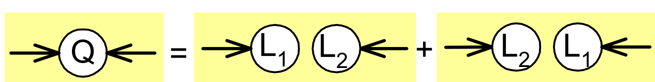
form:

$$\mathbf{L}_1\mathbf{L}_2{}^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix}\begin{bmatrix} p & q & r \end{bmatrix} = \begin{bmatrix} ap & aq & ar \\ bp & bq & br \\ cp & cq & cr \end{bmatrix}$$

Then must symmetrize by adding it to its transpose

$$\mathbf{Q} = \mathbf{L}_1\mathbf{L}_2{}^T + \mathbf{L}_2\mathbf{L}_1{}^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix}\begin{bmatrix} p & q & r \end{bmatrix} = \begin{bmatrix} 2ap & bp+aq & cp+ar \\ bp+aq & 2bq & cq+br \\ cp+ar & cB+br & 2cr \end{bmatrix}$$

Tensor diagram is

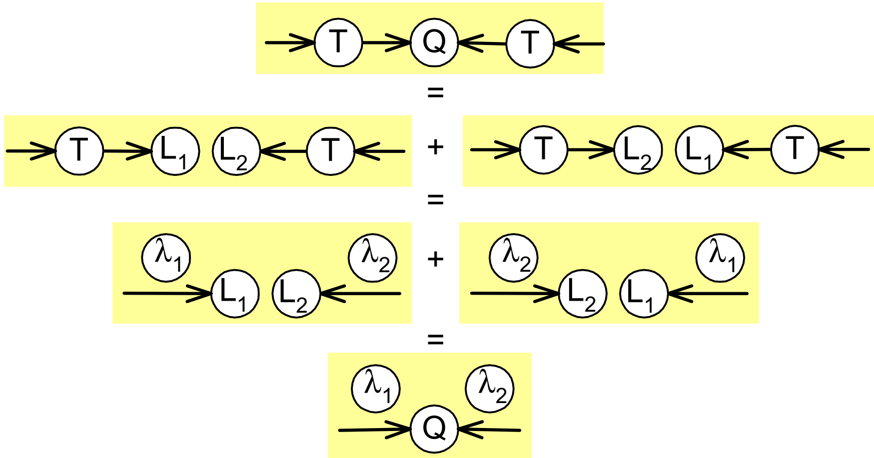


Stationary Transform

Any transformation matrix that has L_1 and L_2 as eigenvectors will, when applied to **Q**, leave **Q** unchanged (except for a homogeneous global scale). Here's why. The transform **T** satisfies:

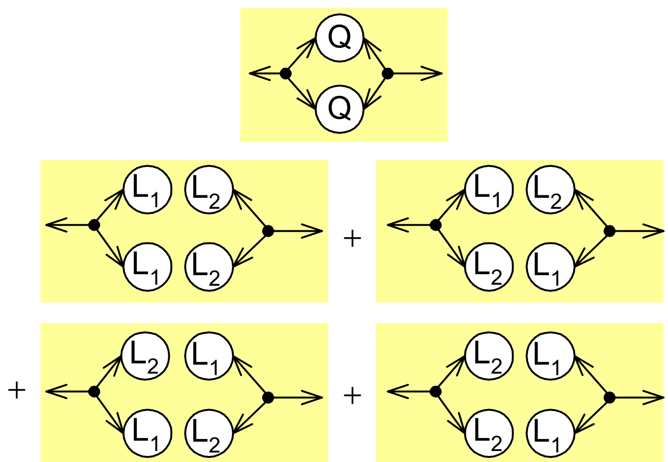
$$\mathbf{T}\mathbf{L}_1 = \lambda_1\mathbf{L}_1$$
$$\mathbf{T}\mathbf{L}_2 = \lambda_2\mathbf{L}_2$$

So apply **T** to the diagram for **Q**



Analysis

Given the above "internal structure" we can construct the adjoint of **Q**. Recall this is



Two of these are zero, and the other two are equal. In fact they are the outer product of the point of intersection of the two lines. So our first analysis is to extract this [*x,y,w*] in a manner similar to the line extraction process for the doubly degenerate line. We will name this:

$$\mathbf{P}_c = \begin{bmatrix} x & y & w \end{bmatrix}$$

Given this, we still need to find the **L** vectors. We build a general transformation that puts $\mathbf{P}_c$ at the origin in a "homogeneously safe" manner. This means that there are no singularities if, say, $\mathbf{P}_c$ is at infinity.

To do this we select two out of the three following points:

$$\mathbf{P}_x = \begin{bmatrix} 0 & -w & y \end{bmatrix}$$
$$\mathbf{P}_y = \begin{bmatrix} w & 0 & -x \end{bmatrix}$$
$$\mathbf{P}_w = \begin{bmatrix} -y & x & 0 \end{bmatrix}$$

Again, we select the two points that are largest in absolute magnitude. (Better way is to discard the point containing the two *x,y,w* coordinates that are smallest in magnitude.) This gives us three points, $\mathbf{P}_c$ and the two we selected, that are guaranteed to be nonzero, no matter what.

Stacking these up into a transformation and applying to **Q** gives

$$\begin{bmatrix} \mathbf{P}_X \\ \mathbf{P}_Y \\ \mathbf{P}_c \end{bmatrix} \mathbf{Q} \begin{bmatrix} \mathbf{P}_X & \mathbf{P}_Y & \mathbf{P}_c \end{bmatrix} = \hat{\mathbf{Q}}$$

$$\begin{bmatrix} \mathbf{P}_X\mathbf{QP}_X & \mathbf{P}_X\mathbf{QP}_Y & \mathbf{P}_X\mathbf{QP}_c \\ \mathbf{P}_Y\mathbf{QP}_X & \mathbf{P}_Y\mathbf{QP}_Y & \mathbf{P}_Y\mathbf{QP}_c \\ \mathbf{P}_c\mathbf{QP}_X & \mathbf{P}_c\mathbf{QP}_Y & \mathbf{P}_c\mathbf{QP}_c \end{bmatrix} = \hat{\mathbf{Q}}$$

$$\begin{bmatrix} A & B & 0 \\ B & D & 0 \\ 0 & 0 & 0 \end{bmatrix} = \hat{\mathbf{Q}}$$

This is a degenerate quadratic consisting of two lines that intersect at the origin
Solve

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} A & B \\ B & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} =$$

$$Ax^2 + 2Bxy + Dy^2 = 0$$

Solution pairs are

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} B + \sqrt{B^2 - AD} & -A \end{bmatrix}, \begin{bmatrix} -D & B + \sqrt{B^2 - AD} \end{bmatrix}$$

We know that the solutions are real because the eigenvectors of the original Q were 0+-. This means that the determinant of the 2x2 matrix is negative (AD-B^2) or that the quantity under the root is positive. So, the two lines in Qhat space are formed from the two cross products

$$\hat{\mathbf{L}}_1 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} B + \sqrt{B^2 - AD} & -A & w \end{bmatrix}$$

$$= \begin{bmatrix} A \\ B + \sqrt{B^2 - AD} \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{L}}_2 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} -D & B + \sqrt{B^2 - AD} & w \end{bmatrix}$$

$$= \begin{bmatrix} B + \sqrt{B^2 - AD} \\ D \\ 0 \end{bmatrix}$$

Now to get these back to **Q** space we must multiply by the adjoint of what got us here. This is formed from the cross products of the rows of the original matrix.

$$\begin{bmatrix} \mathbf{P}_X \\ \mathbf{P}_Y \\ \mathbf{P}_c \end{bmatrix}^* = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{P}_Y \times \mathbf{P}_c & \mathbf{P}_c \times \mathbf{P}_X & \mathbf{P}_X \times \mathbf{P}_Y \\ \vdots & \vdots & \vdots \end{bmatrix}$$

So we have

$$\mathbf{L}_1 = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{P}_Y \\ \mathbf{P}_c \end{bmatrix}^* \begin{bmatrix} A \\ B + \sqrt{B^2 - AD} \\ 0 \end{bmatrix}$$

$$\mathbf{L}_2 = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{P}_Y \\ \mathbf{P}_c \end{bmatrix}^* \begin{bmatrix} B + \sqrt{B^2 - AD} \\ D \\ 0 \end{bmatrix}$$

There is some more simplification that can be done here that I haven't finished.

The Single Point

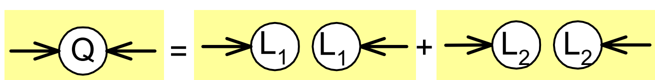
The eigenvalues are ++0 or --0

Construction

Take any two distinct lines $\mathbf{L}_1$ and $\mathbf{L}_2$ through the given point. Form the matrix:

$$\mathbf{Q} = \mathbf{L}_1\mathbf{L}_1^T + \mathbf{L}_2\mathbf{L}_2^T$$

Tensor diagram



Note that multiplying a point $\mathbf{P}$ by this matrix gives us:

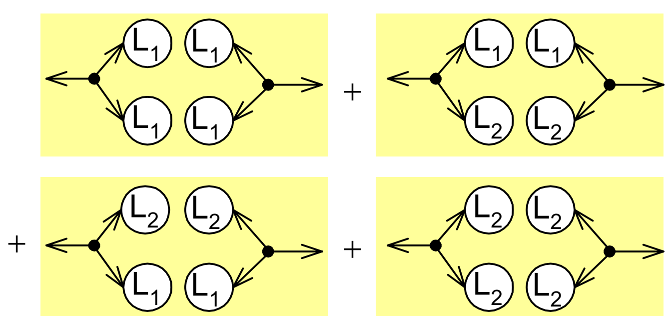
$$\mathbf{PQP}^T = (\mathbf{P} \cdot \mathbf{L}_1)^2 + (\mathbf{P} \cdot \mathbf{L}_2)^2$$

So this will give zero iff $\mathbf{P}$ is on *both* lines $\mathbf{L}_1$ and $\mathbf{L}_2$.

Note that the matrix $\mathbf{Q}$ is not unique. For a given $\mathbf{P}$ there are many different (and not just by a homogeneous scale) matrices $\mathbf{Q}$ that have only $\mathbf{P}$ as a solution.

Analysis

At first it might look like $\mathbf{Q}$ is doubly degenerate (it's the sum of two doubly degenerate matrices). But if you construct the adjoint you find:



The first and last terms are zero, the second and third are just the outer product of $\mathbf{P}$ with itself. In other words we can recover $\mathbf{P}$ by extraction of the longest row or column of the adjoint of $\mathbf{Q}$.

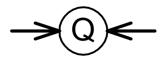
The Conic Section

For conic construction hints see the chapter on the Theorem of Pappus.

For analysis hints I first want to discuss polar lines and points to a quadratic.

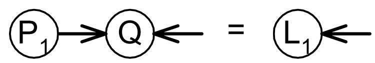
Polar Lines and Points

Start with a quadratic curve **Q**

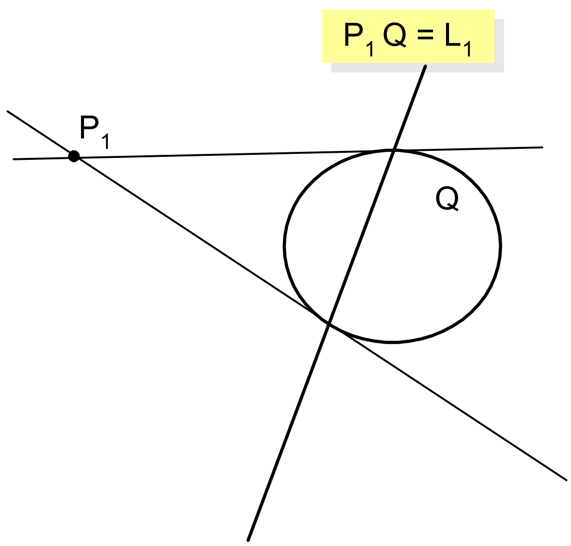


Point P_1

Pick any point P_1 not on Q. Form the line P_1 Q. This is called the polar line to P_1

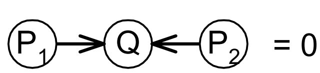


Geometrically, this is the line connecting the two points of tangency from P_1 to Q

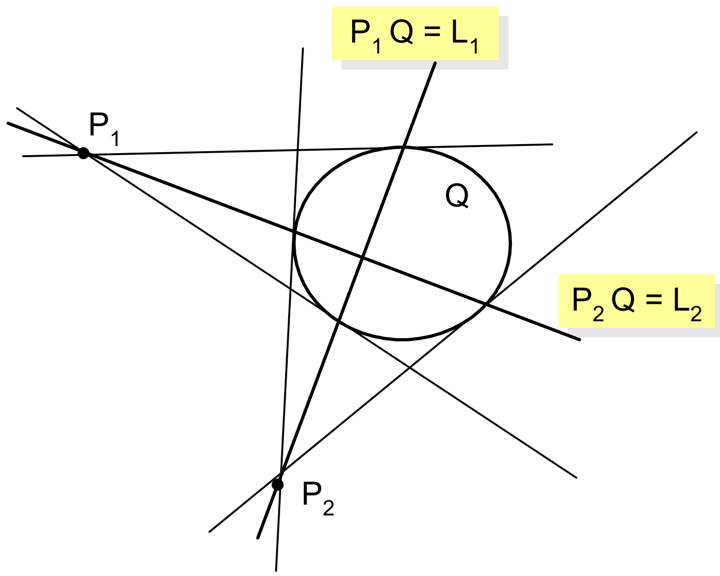


Point P_2

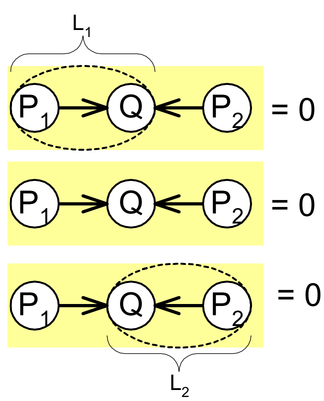
Pick any other point P_2 on this polar line. Then



Now construct the polar line to P_2 in a similar manner; it's the line connecting the points of tangency from P_2 to Q. If P_2 is on the polar line of P_1 then P_1 is on the polar line of P_2. Geometrically it looks like

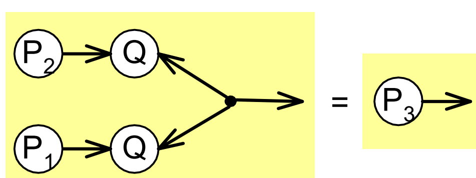


The proof of this is just associativity on the tensor diagram:

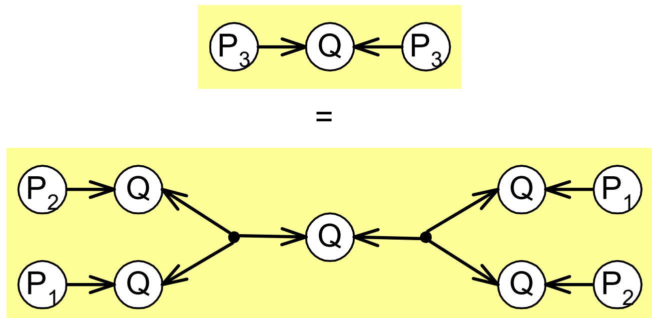


Point P_3

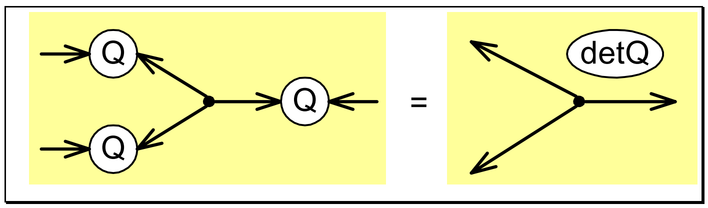
Now construct P_3 as the intersection of the two polar lines: P_1 Q and P_2 Q. Diagrammatically:



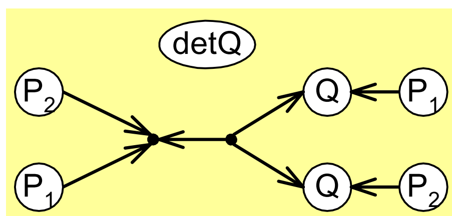
If we have P_1 and P_2 not on Q, what about P_3? Calculate:



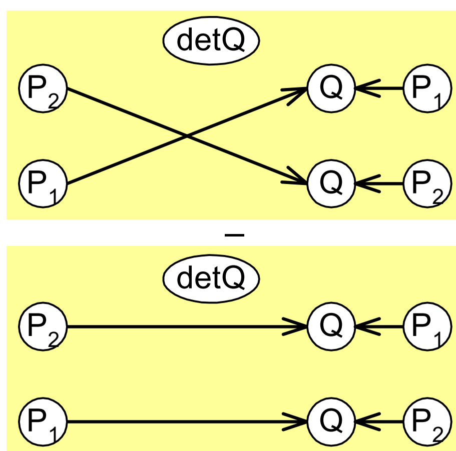
Apply the identity



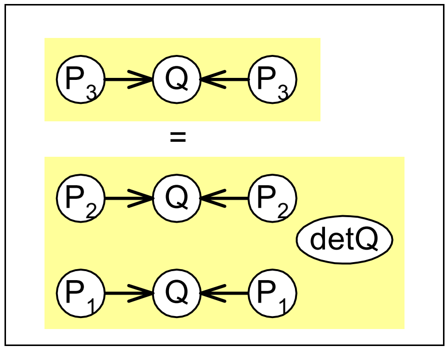
and get:



Apply epsilon/delta and get



The second term is zero from above construction so we have, finally:

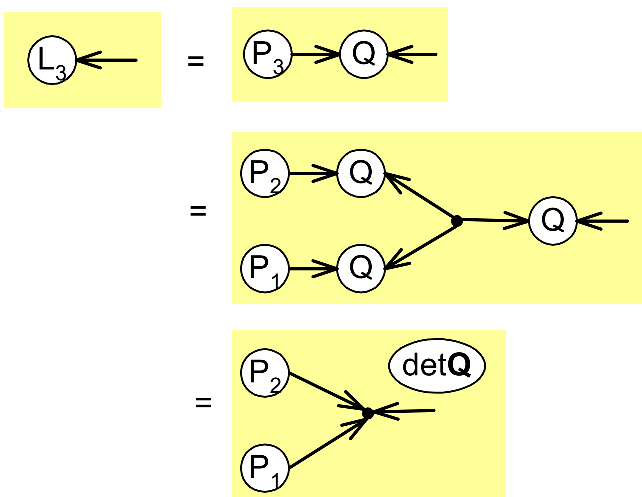


Now what does this mean? If det**Q** is negative it means that points on the inside of the conic give a negative PQP and points on the outside give positive PQP. This means that

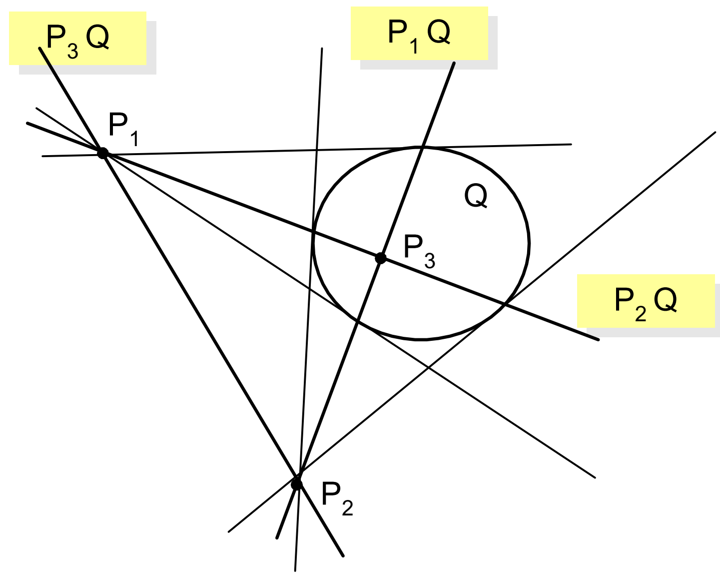
If P_1 and P_2 are outside the conic, then P_3 is inside

A similar argument holds if det**Q** is positive.

Now for the kicker. Form the polar line of P_3 as the product $\mathbf{L}_3 = \mathbf{P}_3\mathbf{Q}$. Then we have the diagram



In other words, the polar line of P_3 is the line through points P_1 and P_2. Geometrically this is:



This also gives us a purely geometric construction for the polar line of a point inside a conic. Take any line through the point. This hits the conic at two points. Take the tangents at those points and plot their intersection. As the line through the inside point changes, the intersection of the tangents traces out its polar line.

Standard Position

The above discussion gives us a way to construct a transformation matrix that diagonalizes **Q**.

1. Pick any point P_1 not on **Q**
2. Pick any point P_2 on the polar line P_1**Q** but not on **Q**
3. Construct P_3 as above.
4. Stack these points into a matrix

This transforms **Q** as follows:

$$\begin{bmatrix} \cdots \mathbf{P}_1 \cdots \\ \cdots \mathbf{P}_2 \cdots \\ \cdots \mathbf{P}_3 \cdots \end{bmatrix} \mathbf{Q} \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{P}_1 & \mathbf{P}_2 & \mathbf{P}_3 \\ \vdots & \vdots & \vdots \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{P}_1\mathbf{Q}\mathbf{P}_1^T & \mathbf{P}_1\mathbf{Q}\mathbf{P}_2^T & \mathbf{P}_1\mathbf{Q}\mathbf{P}_3^T \\ \mathbf{P}_2\mathbf{Q}\mathbf{P}_1^T & \mathbf{P}_2\mathbf{Q}\mathbf{P}_2^T & \mathbf{P}_2\mathbf{Q}\mathbf{P}_3^T \\ \mathbf{P}_3\mathbf{Q}\mathbf{P}_1^T & \mathbf{P}_3\mathbf{Q}\mathbf{P}_2^T & \mathbf{P}_3\mathbf{Q}\mathbf{P}_3^T \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{P}_1\mathbf{Q}\mathbf{P}_1^T & 0 & 0 \\ 0 & \mathbf{P}_2\mathbf{Q}\mathbf{P}_2^T & 0 \\ 0 & 0 & \mathbf{P}_3\mathbf{Q}\mathbf{P}_3^T \end{bmatrix} =$$

Dual form of Polars

There is a dual form for the above discussion, basically swapping the words “point” and “line”.

Given a line **L** and a quadratic **Q**, we form the dual point by… mumble

Chapter 2-02

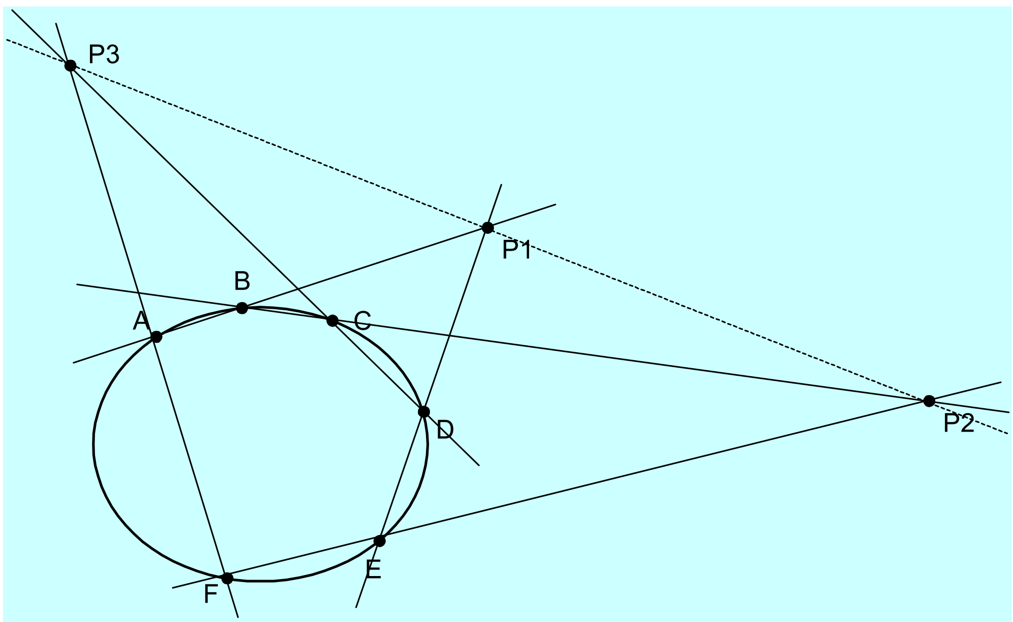
2DH(3D)
The Theorem of Pappus

The Theorem of Pappus is an old chestnut in the field of projective geometry. Any algebraic framework needs to be able to address it.

Statement of Theorem

Conventional form

Start with six points that lie on a conic section: A…F



Construct Line AB, and Line DE, and their intersection P1

Construct Line BC, and Line EF, and their intersection P2

Construct Line CD, and Line FA, and their intersection P3

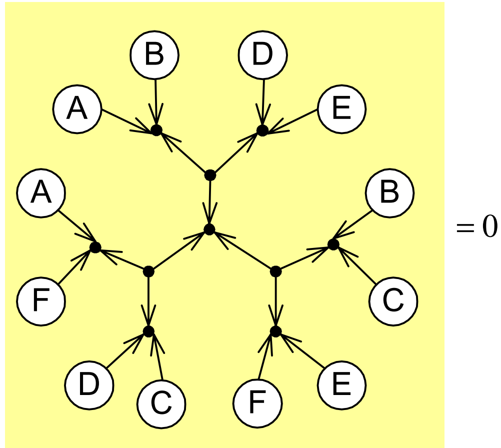
Then Points P1, P2, P3 are collinear

There is a dual form of this formed by swapping the words "line" and "point" in the above statement.

Statement in Diagram form

The above is a bit tricky to keep track of. The diagram form, thought, is quite pretty:

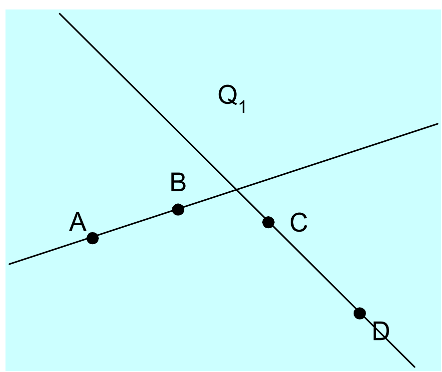
Given six points A…F on a conic section, then



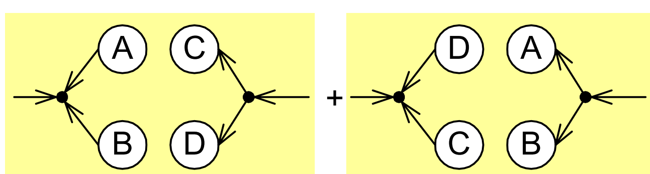
Defining a Conic Section

In order to prove the theorem we must first work out what it means (in diagram form) for 6 points to lie on a conic section. In general five points A,B,C,D,E will uniquely determine a conic section. The conventional technique to generate the six elements of **Q** from this involves finding six determinants of 5x5 matrices. Let's try something different.

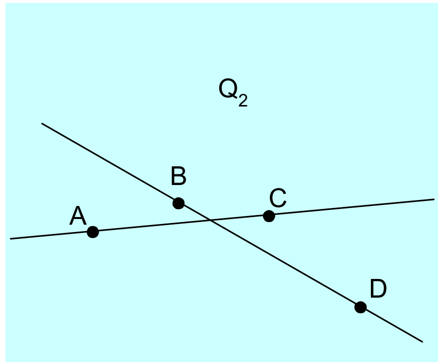
Let's just consider the first 4 points. We can construct a degenerate conic through these points that is the union of the line AB and the line CD. Call this $\mathbf{Q}_1$.



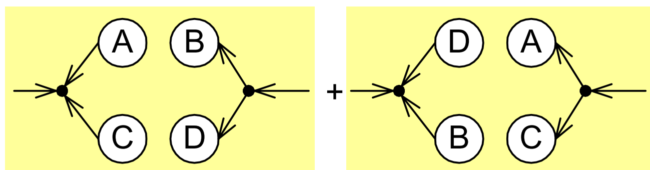
The diagram for the matrix of $\mathbf{Q}_1$ is



Now construct the degenerate conic through the points that is the union of the line AC and the line BD. Call this $\mathbf{Q}_2$.



The diagram for the matrix of $\mathbf{Q}_2$ is



Now consider all linear combinations of quadratics $\mathbf{Q_1}$ and $\mathbf{Q_2}$.

$$\mathbf{Q} = \alpha \mathbf{Q}_1 + \beta \mathbf{Q}_2$$

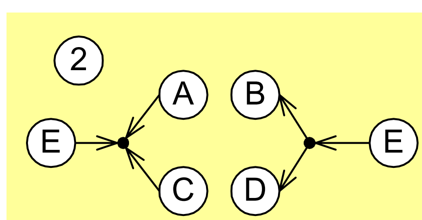
Points ABCD are all on both $\mathbf{Q_1}$ and $\mathbf{Q_2}$, so all four points are on the more general quadratic **Q**. Now let's pick the parameters (α, β) that generate the particular **Q** that also passes through point E. That is, we want

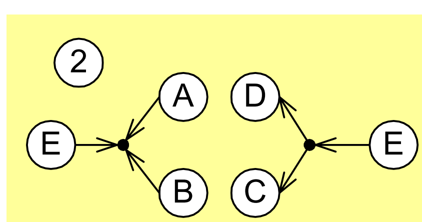
$$\begin{aligned}\mathbf{EQE}^T &= \mathbf{E}\left(\alpha \mathbf{Q}_1 + \beta \mathbf{Q}_2\right)\mathbf{E}^T \\ &= \alpha\left(\mathbf{EQ}_1\mathbf{E}^T\right) + \beta\left(\mathbf{EQ}_2\mathbf{E}^T\right) \\ &= 0\end{aligned}$$

The solution is:

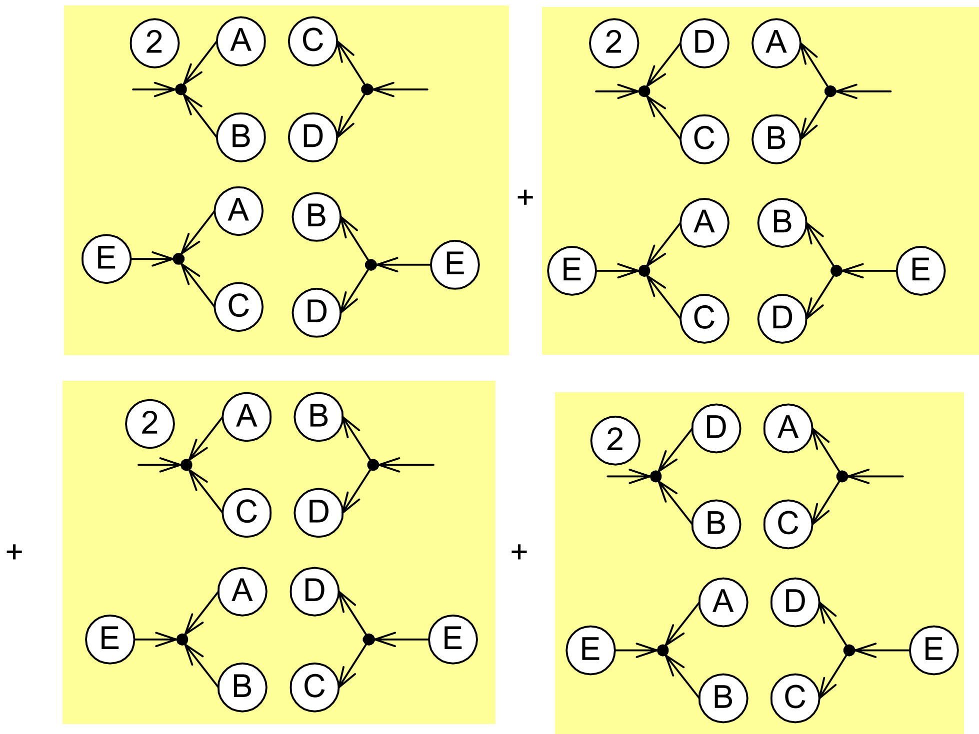
$$\begin{aligned}\alpha &= \mathbf{EQ}_2\mathbf{E}^T \\ \beta &= -\mathbf{EQ}_1\mathbf{E}^T\end{aligned}$$

In diagram form we have, after some straightforward merging and mirroring/sign flipping:

$\alpha =$ 

$\beta =$ 

Now let's generate the diagram for the whole matrix **Q.**:



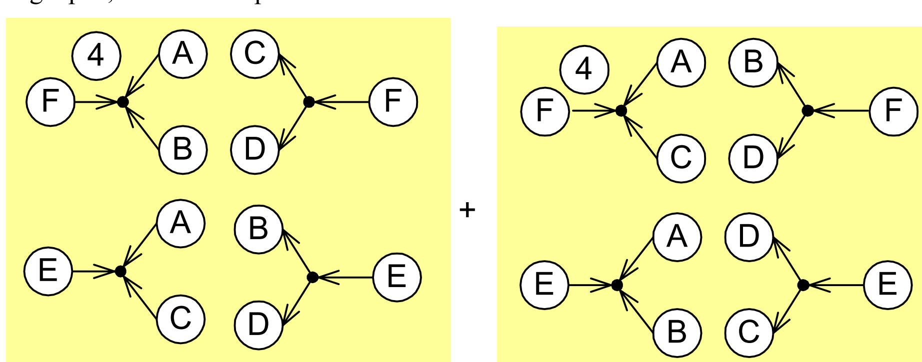
Arithmetically this is a lot better than taking six 5x5 determinants. Close scrutiny shows that it requires

1) four triple products, (ACE), (BDE), (ABE), (CDE)
2) four cross products; (AB), (CD), (AC), (BD)

plus some arithmetic glue. It also affords us some control of numerical error. This is because we can permute the usage of the five points in this expression and still get the same quadratic curve **Q**. So we might figure out which four selections of three points to take the triple product of that generates the least round off error.

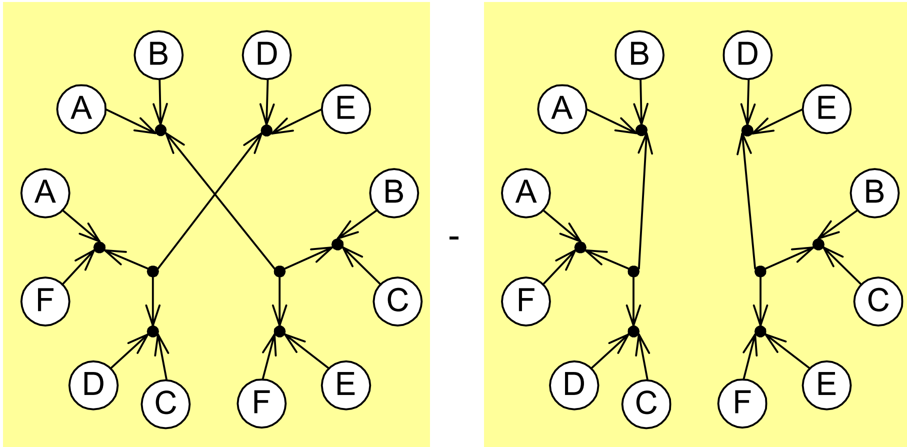
Proof of Pappus

Given the form for the conic through five points ABCDE, the condition of a sixth point F being on the conic arrived at by simply plugging F into the above diagram. The first two and second two subdiagrams wind up being equal, so the net expression is

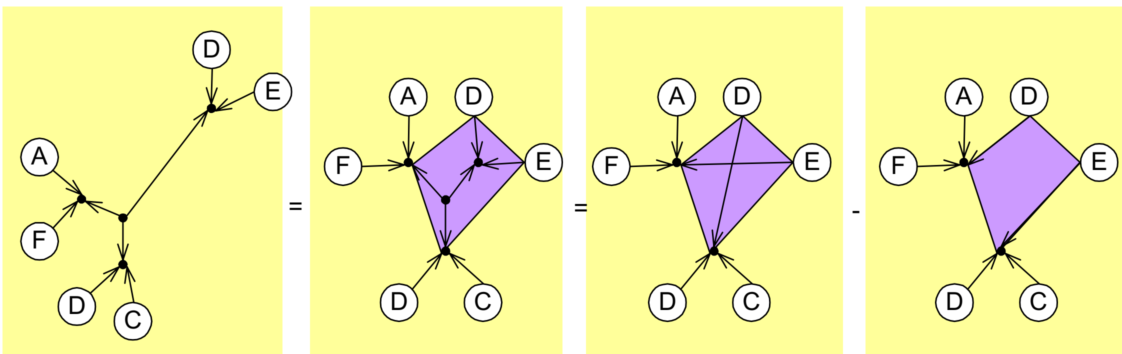


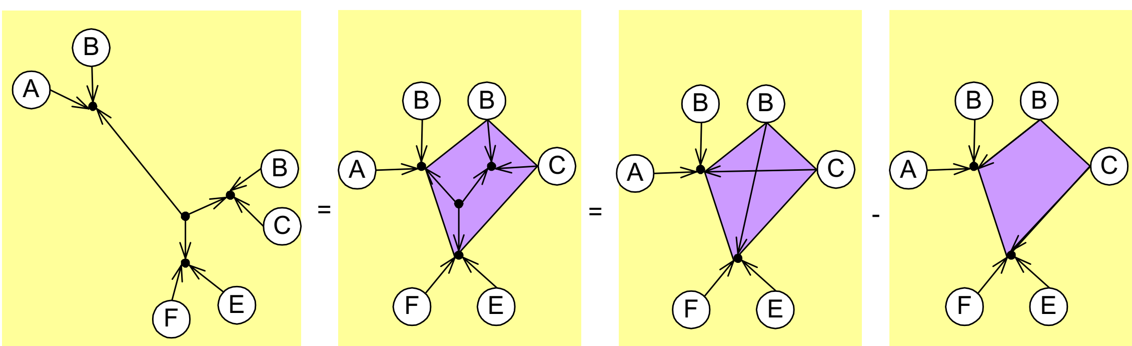
Now compare this with the original statement of Pappus. We see that a massive orgy of epsilon/delta is in order. Here we go.

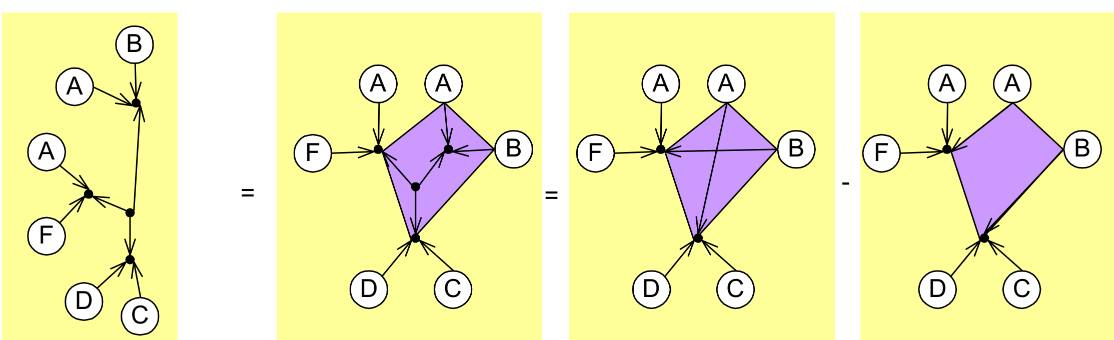
Apply Epsilon/Delta to original diagram:

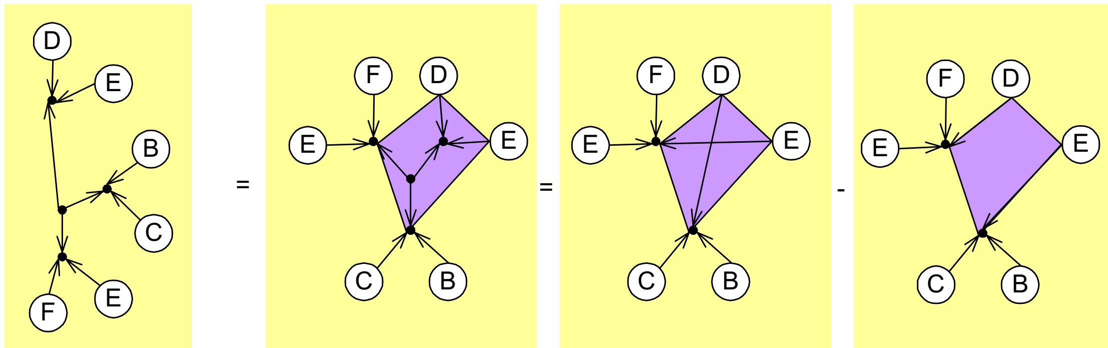


The four terms can be rubber stamped:

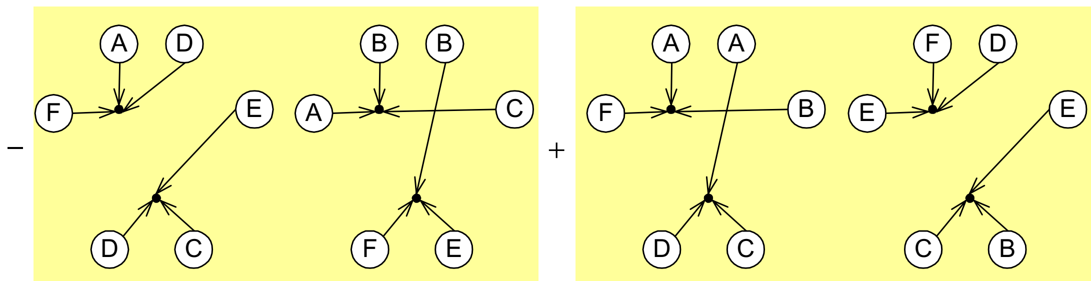








Collecting these together we get the net diagram of Pappus equal to



This is presumably the same as the expression we got by plugging F into our formula for **Q**. Since I did these two halves of the proof at two different times, there is of course, a permutation of the letters ABCDEF necessary to make them actually equal. But you get the idea.

Chapter 2-03

2DH(3D)
Intersecting A Line with a Quadratic

The Problem

Given line **L** and quadratic curve **Q**, find the common points in a homogeneously safe way.

Way 1

Find two points on L. Any other point on L is a linear combination of these. (We are basically parametrizing the implicit curve for L)

$$\begin{aligned}\mathbf{P} &= \alpha\mathbf{P}_1 + \beta\mathbf{P}_2 \\ &= \begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}\mathbf{P}_1 \\ \mathbf{P}_2\end{bmatrix}\end{aligned}$$

Then plug this into find points on **Q**

$$\begin{aligned}\mathbf{PQP}^T &= \begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}\mathbf{P}_1\mathbf{QP}_1^T & \mathbf{P}_1\mathbf{QP}_2^T \\ \mathbf{P}_2\mathbf{QP}_1^T & \mathbf{P}_2\mathbf{QP}_2^T\end{bmatrix}\begin{bmatrix}\alpha \\ \beta\end{bmatrix} = 0 \\ &= \begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}A & B \\ B & C\end{bmatrix}\begin{bmatrix}\alpha \\ \beta\end{bmatrix} = 0\end{aligned}$$

Then solve this homogeneous equation for alpha, beta.

B>0	$\begin{bmatrix}-C & B+\sqrt{B^2-AC}\end{bmatrix}, \begin{bmatrix}B+\sqrt{B^2-AC} & -A\end{bmatrix}$
B<0	$\begin{bmatrix}B-\sqrt{B^2-AC} & -A\end{bmatrix}, \begin{bmatrix}-C & B-\sqrt{B^2-AC}\end{bmatrix}$

Discriminant is

(DIAGRAM) figure out signs/factors

call this Delta

Solutions are

$$\begin{bmatrix}\alpha & \beta\end{bmatrix} = \begin{bmatrix}B & -A \\ -C & B\end{bmatrix} + \Delta\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}$$

Plug into original equation for **P** to get coordinates of points.

Finding points on L

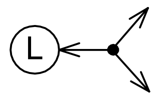
Given **L** how can we find two distinct points on it?

First form the anti-symmetric matrix from the elements of **L**

$$\begin{bmatrix}0 & c & -b \\ -c & 0 & a \\ b & -a & 0\end{bmatrix}$$

Each row/column of this matrix is a point on **L**.

Diagrammatically, this is



How to pick which two to use? To avoid degeneracies we want to pick two that cannot all be zeroes. Find the absolute maximum of (a,b,c) and pick the two points that contain that coordinate. To visualize this, project abc onto an origin centered cube. Then each face of the cube represents a situation where one of the coordinates is greater than the other two. Each face-pair represents a choice for which two points to use..

Possible problems

What if one or both of the points so chosen is on **Q**. Are we in trouble? The worst-case scenario: the line has a=b=0 (the line at infinity). The matrix is

$$\begin{bmatrix} 0 & c & 0 \\ -c & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{P}_2 = \begin{bmatrix} -1 & 0 & 0 \end{bmatrix}$$

Our algorithm picks the first two rows. Each of these is a point at infinity. So far so good. Now what if **Q** passes thru these two points. This would happen if the **Q** matrix were:

$$\mathbf{Q} = \begin{bmatrix} 0 & B & C \\ B & 0 & E \\ C & E & F \end{bmatrix} \quad \det \mathbf{Q} = B\left(2EC - BF\right)$$

The three unique components of the

$$\mathbf{P}_1\mathbf{Q}\mathbf{P}_1^T = 0$$

$$\mathbf{P}_1\mathbf{Q}\mathbf{P}_2^T = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & B & C \\ B & 0 & E \\ C & E & F \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} = -B$$

$$\mathbf{P}_2\mathbf{Q}\mathbf{P}_2^T = 0$$

The quadratic equation is:

$$-2B\alpha\beta = 0$$

and the two homogeneous solutions are:

$$\begin{bmatrix} \alpha & \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \end{bmatrix}$$

No problem

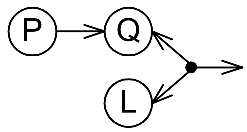
Way 2 - Better?

Pick one point $\mathbf{P}_1$ on **L** and not on **Q**. (Expand on how later.)

Find polar line of $\mathbf{P}_1$ with respect to **Q**. Intersect this with L to generate second point on L.

$$\mathbf{P}_2 = \mathbf{P}_1\mathbf{Q}\times\mathbf{L}$$

Diagram

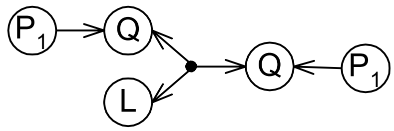


Then when we form the coefficient matrix for the quadratic equation

$$\begin{bmatrix} \mathbf{P}_1\mathbf{Q}\mathbf{P}_1^T & \mathbf{P}_1\mathbf{Q}\mathbf{P}_2^T \\ \mathbf{P}_2\mathbf{Q}\mathbf{P}_1^T & \mathbf{P}_2\mathbf{Q}\mathbf{P}_2^T \end{bmatrix}$$

we get the diagrams

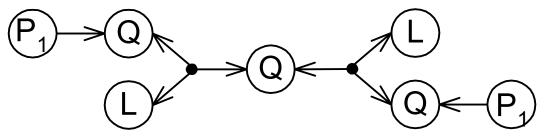
$\mathbf{P}_2\mathbf{Q}\mathbf{P}_1^T$ =



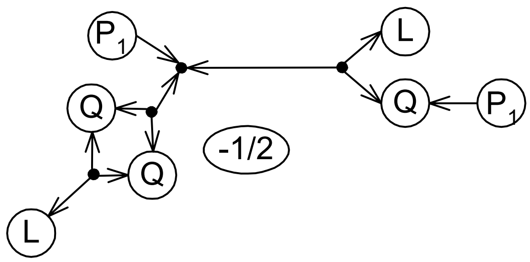
= 0

And

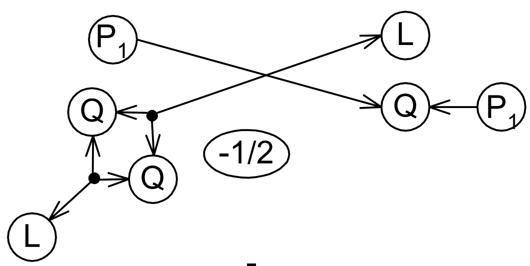
$\mathbf{P}_2\mathbf{Q}\mathbf{P}_2^T$ =



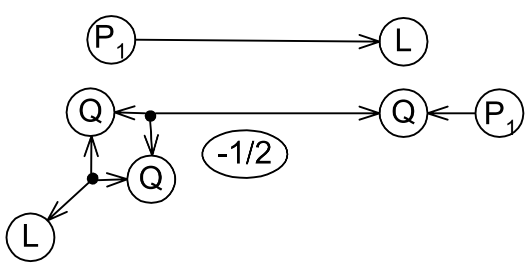
Apply identity



Now apply epsilon delta:

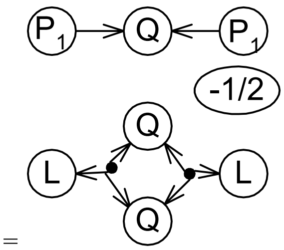


-



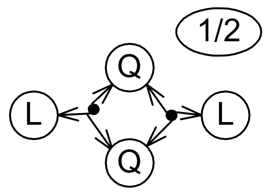
The second of these is zero since P_1 is on L. Neatening up the first we get:

$\mathbf{P}_2\mathbf{Q}\mathbf{P}_2{}^T =$



Define

D =



So, the quadratic equation becomes.

$$\left(\mathbf{P}_1\mathbf{Q}\mathbf{P}_1^T\right)\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -D\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=0$$

Can throw out the common factor of P_1QP_1 and the solution is:

$$\alpha^2 = D\beta^2$$

$$\alpha = \pm\sqrt{D}, \beta = 1$$

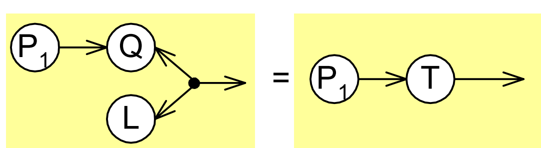
The final equation for the intersection points is

$$\mathbf{P} = \alpha\mathbf{P}_1 + \beta\mathbf{P}_2$$

$$= \pm\sqrt{D}\mathbf{P}_1 + \mathbf{P}_2$$

Note that the value D is the quantity we have previously identified as zero if **L** is tangent to **Q** and positive if **L** intersects **Q** (and negative if **L** doesn't intersect **Q**).

Can think of diagram that defines P_2 as

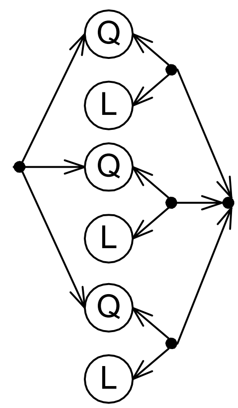


$$\mathbf{P}_2 = \mathbf{P}_1\mathbf{T}$$

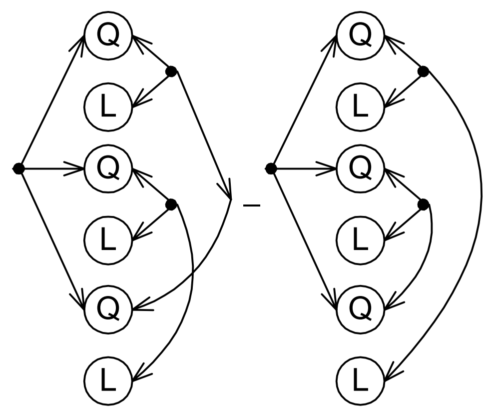
Then you can think of the solutions as multiplication of P_1 by a transformation matrix:

$$\mathbf{P} = \mathbf{P}_1\left(\mathbf{T} \pm \sqrt{D}\mathbf{I}\right)$$

What does this really mean? First let's look more closely at matrix **T**. It is singular; we see this by taking the determinant



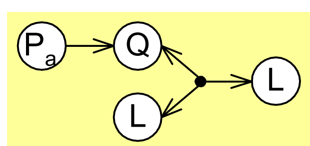
and applying epsilon/delta



Both these terms are zero since they each have two **L**'s going into an epsilon

Next, we see that **T** maps all points onto the line **L**. We can see this immediately by finding where **T** maps the arbitrary point $\mathbf{P_a}$ and dotting it with **L**:

$$(\mathbf{P}_a\mathbf{T})\mathbf{L} =$$

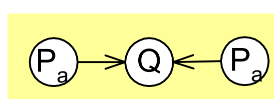


$= 0$

Furthermore it maps points that are already on **L** onto other different points on **L**. We see this by comparing the dot product of the arbitrary point with **Q** both before and after being transformed by **T**.

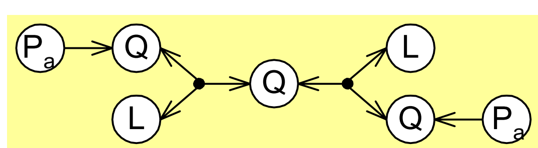
Before:

$$(\mathbf{P}_a)\mathbf{Q}(\mathbf{P}_a)^T =$$



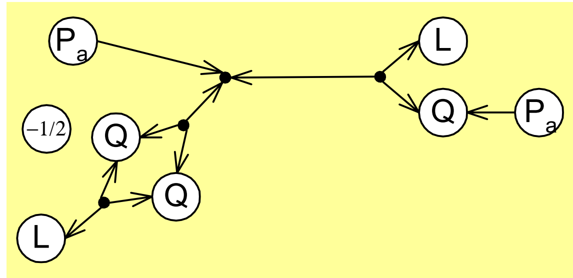
After:

$$(\mathbf{P}_a\mathbf{T})\mathbf{Q}(\mathbf{P}_a\mathbf{T})^T =$$

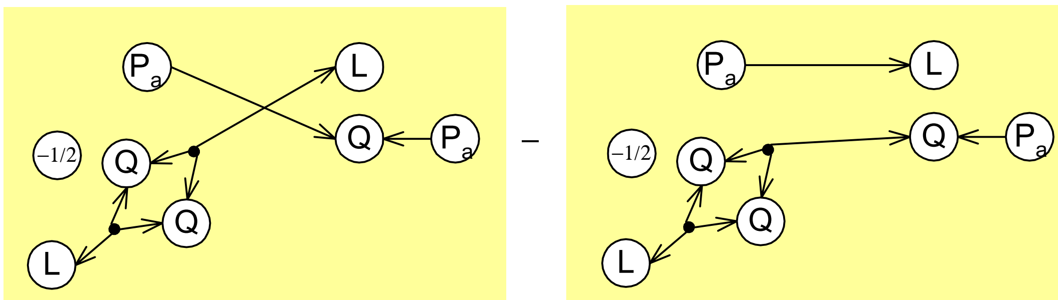


apply identity in appendix:

$$(\mathbf{P}_a\mathbf{T})\mathbf{Q}(\mathbf{P}_a\mathbf{T})^T =$$

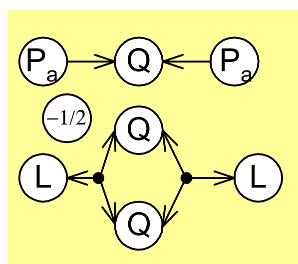


Apply eps delta.



If Pa is on L the second term above is zero so we have:

$$(\mathbf{P}_a\mathbf{T})\mathbf{Q}(\mathbf{P}_a\mathbf{T})^T =$$



So if Pa is a point on L and the value D is positive then Pa having a positive product with Q before the transform, has a negative product after the transform. If Pa is on Q the product is zero both times.

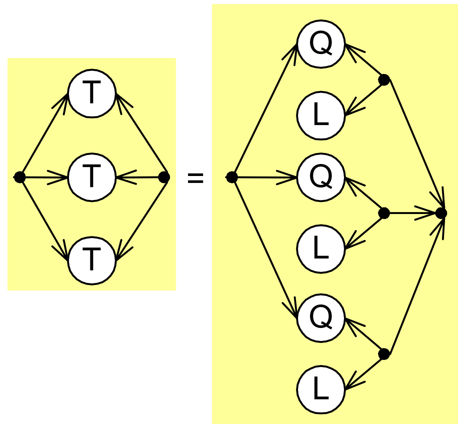
So according the equation

$$\mathbf{P} = \mathbf{P}_1\left(\mathbf{T} \pm \sqrt{D}\mathbf{I}\right)$$

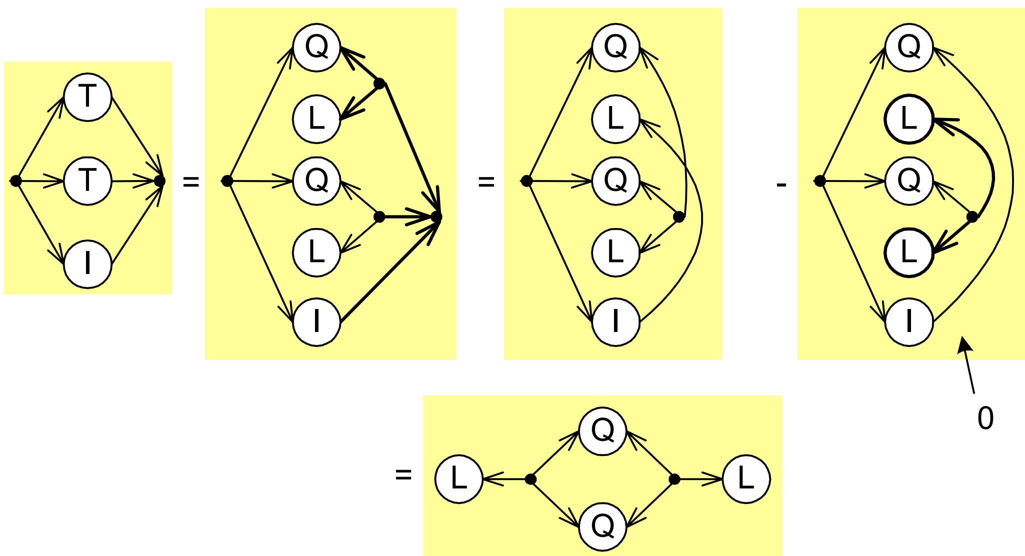
The solution point P we just solved for is the sum of a point on ,L P1 T, with the original point P1 (also on L).

Eigenvalues of T

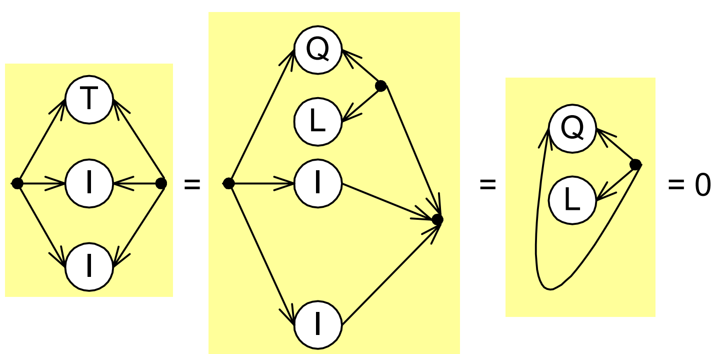
The three coefficients of the characteristic equation are:



=0



=2D



So the characteristic equation and eigenvalues are

$$(-6)\lambda^3 + 3(0)\lambda^2 + 3(2D)\lambda + 1(0) = 0$$

$$\lambda\left(-6\lambda^2 + 6D\right) = 0$$

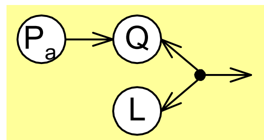
$$\lambda = 0, \pm\sqrt{D}$$

T is singular and maps all points onto L. Has eigenvalues as above:

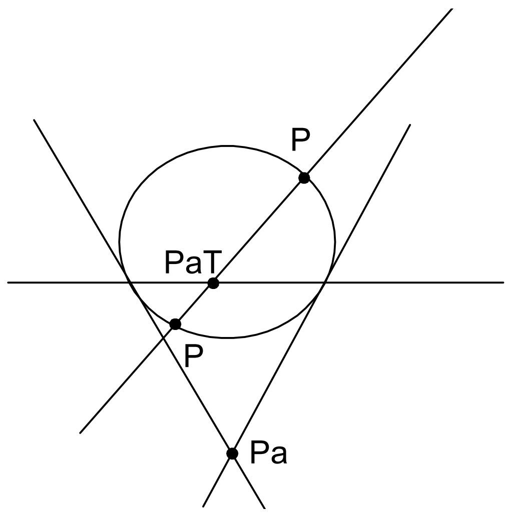
The matrix $\left(T \pm \sqrt{D}I\right)$ has eigenvalues either $\left(0, \sqrt{D}, 2\sqrt{D}\right)$ or $\left(0, -\sqrt{D}, -2\sqrt{D}\right)$. The determinant of this matrix is indeed zero, from plugging into the diagrams etc from above.

Effect T on arbitrary points

If a point P1 is on L it will map to a solution. If a point Pa is not on L where does it go?

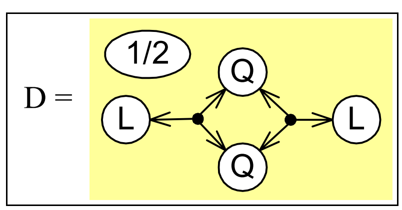
$(\mathbf{P}_a\mathbf{T}) =$ 

Pa maps to the intersection of the polar of Pa wrt Q with L.

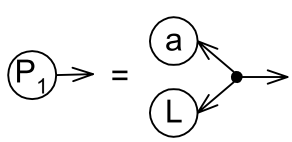


Punch line

Given **Q** and **L** calculate

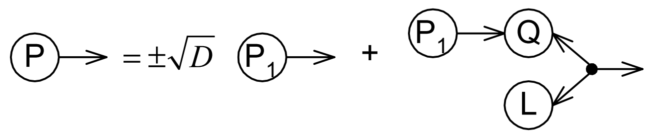


Find arbitrary point on **L** by:

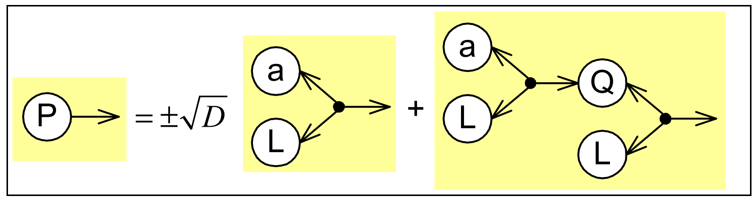


where line **a** is any line other than **L**.
Intersections **P** are then:

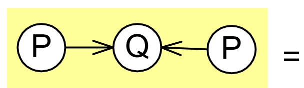
$$\mathbf{P} = \pm\sqrt{D}\mathbf{P}_1 + \mathbf{P}_2$$

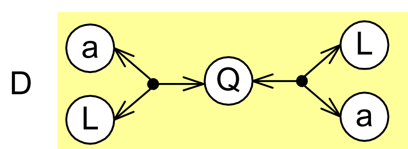


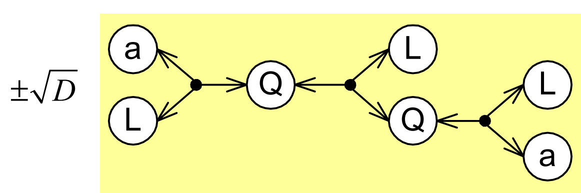
so plugging in diagram for $\mathbf{P_1}$ where **a** is any line not equal to **L**

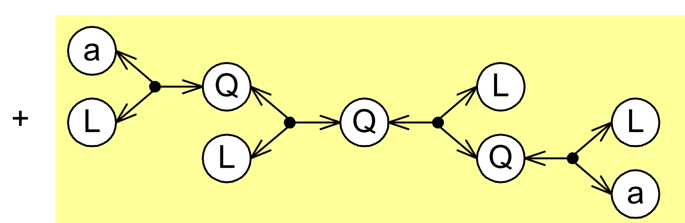


This point **P** is obviously on **L**. Lets prove that its on **Q** too. Plug in:

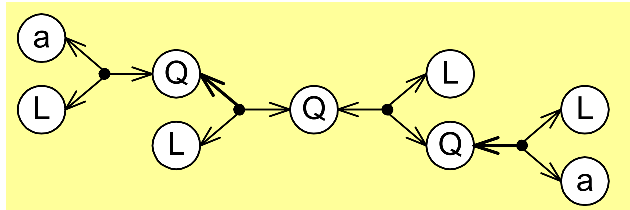




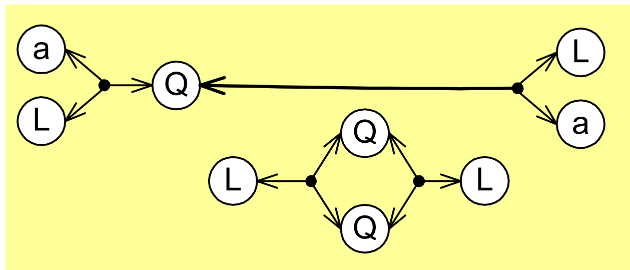




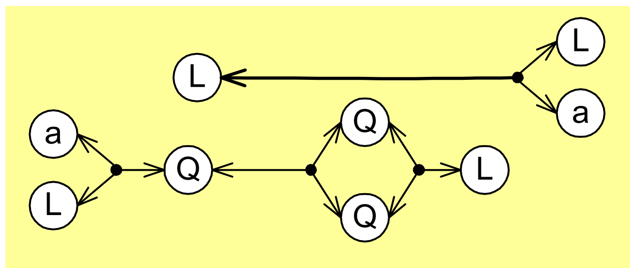
Middle term is zero. Apply second identity to final term:



$= -1/2$

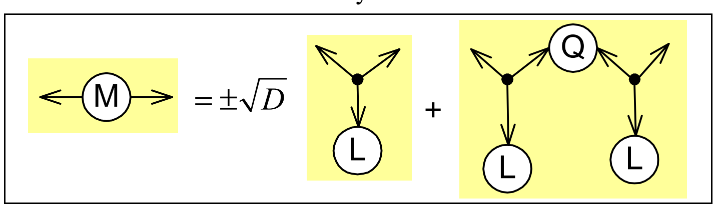


$+1/2$



the second term of this is zero. The first term cancels the first term of original equation. Point is on **Q**
QED.

Thinking of this as a matrix transformation of arbitrary line **a**



Note that this matrix is not quite symmetrical (the first term is antisymmetric). (??This is a doubly degenerate matrix (rank 1) that is the dual analog of a double line matrix. It is the outer product of a solution point P with itself…??)

Explicitly it will be the matrix;

$$\pm\sqrt{D}\begin{bmatrix}0 & c & -b\\ -c & 0 & a\\ b & -a & 0\end{bmatrix}+\begin{bmatrix}0 & c & -b\\ -c & 0 & a\\ b & -a & 0\end{bmatrix}\begin{bmatrix}A & B & C\\ B & D & E\\ D & E & F\end{bmatrix}\begin{bmatrix}0 & -c & b\\ c & 0 & -a\\ -b & a & 0\end{bmatrix}$$

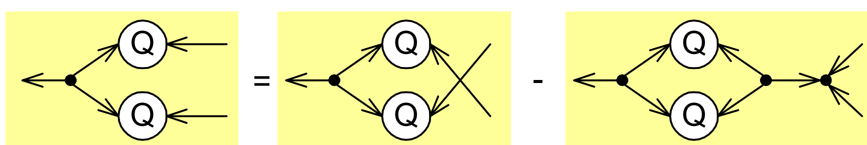
$$=\pm\sqrt{D}\begin{bmatrix}0 & c & -b\\ -c & 0 & a\\ b & -a & 0\end{bmatrix}+\begin{bmatrix}cB-bD & cD-bE & cE-bF\\ -cA+aD & -cB+aE & -cC+aF\\ bA-aB & bB-aD & bC-aE\end{bmatrix}\begin{bmatrix}0 & -c & b\\ c & 0 & -a\\ -b & a & 0\end{bmatrix}$$

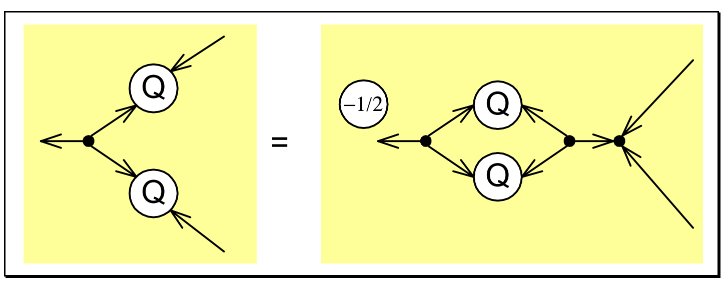
$$=\pm\sqrt{D}\begin{bmatrix}0 & c & -b\\ -c & 0 & a\\ b & -a & 0\end{bmatrix}+\begin{bmatrix}ccB-cbD-bcE+bbF & -ccB+acE-abF+bcD & cbB-bbD-acD+abE\\ -ccB+acE-abF+bcC & & \\ cbB-caD-bbC+baE & & \end{bmatrix}$$

Appendix - Identities

first

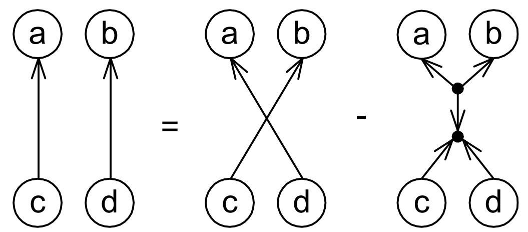
apply eps delta to:

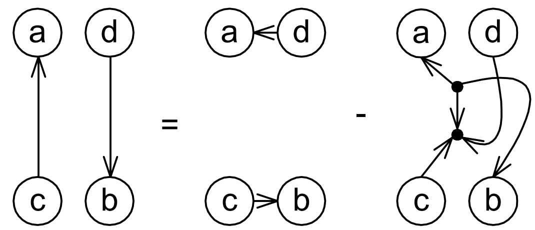




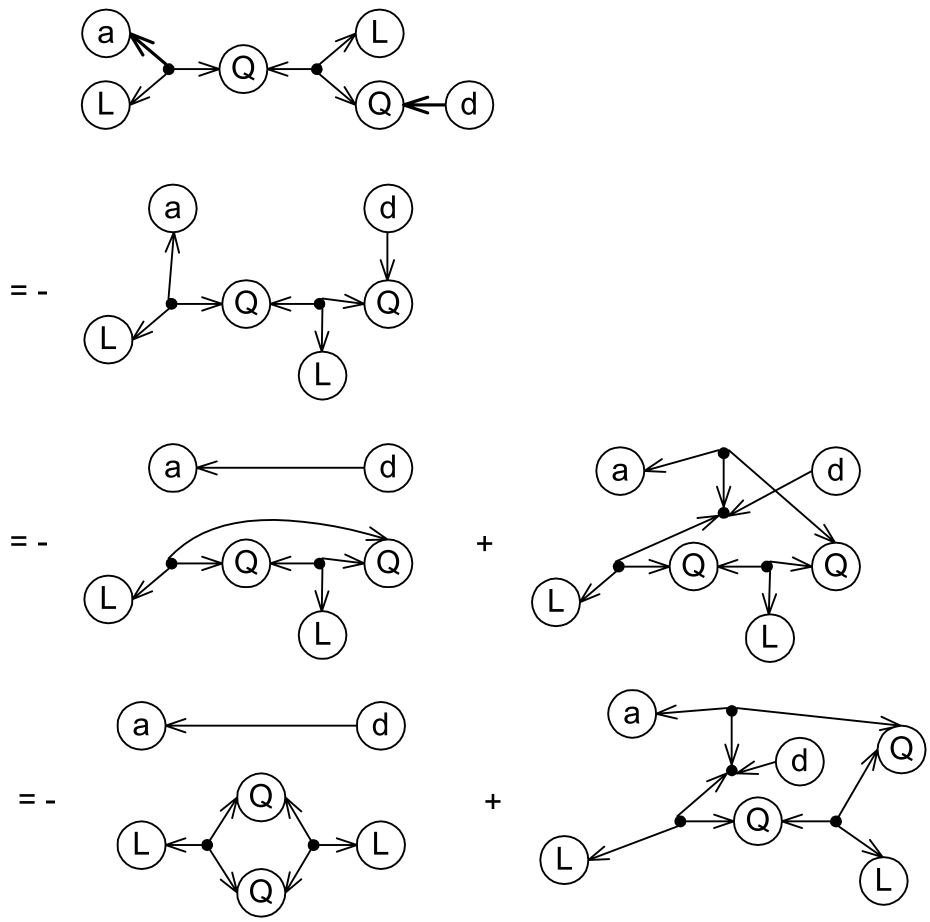
second

First make variant of eps delta

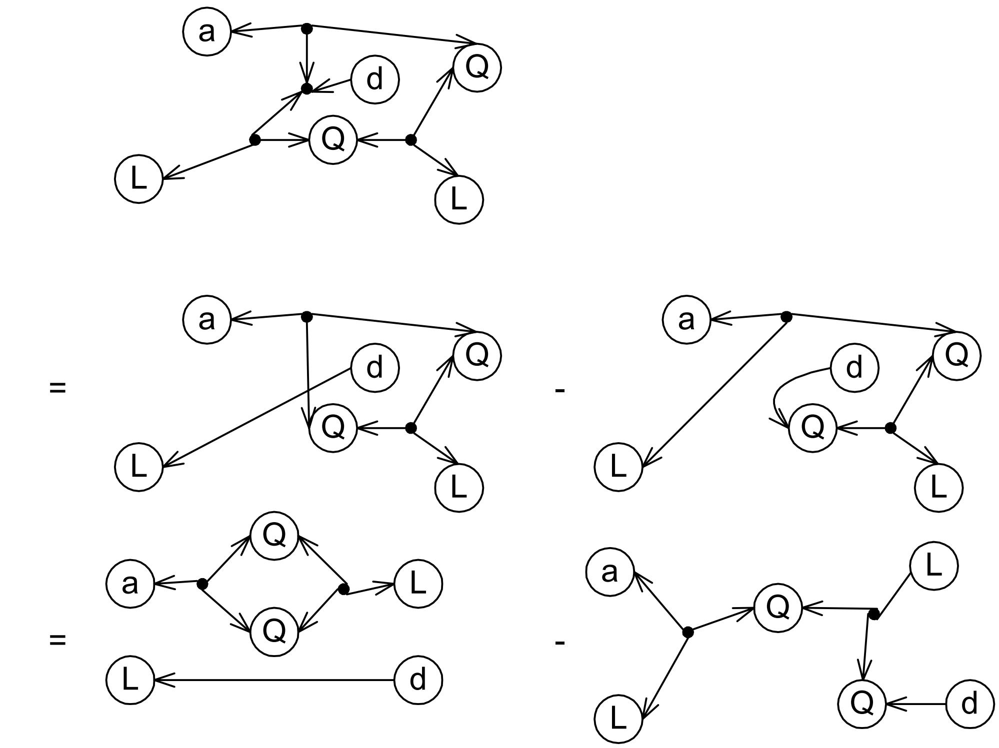




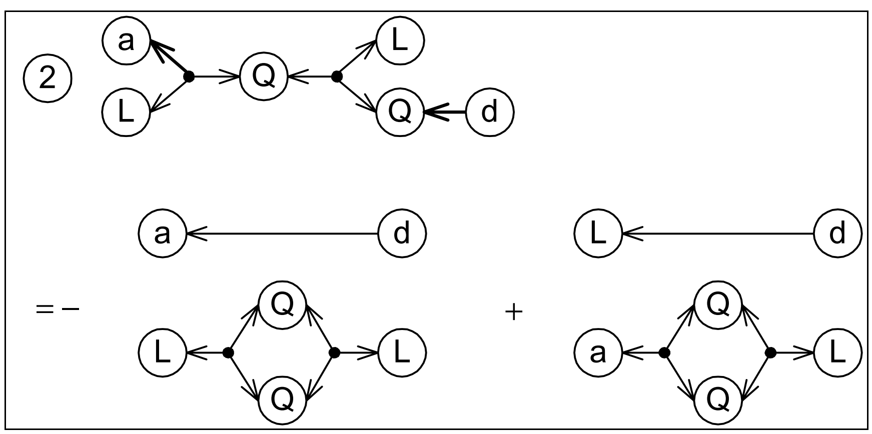
Now apply to chain of two QL matrices



Operate on this last term:



Note that final term is minus the one we started with. Plug in, move over equals and Net identity is:



Symbolic

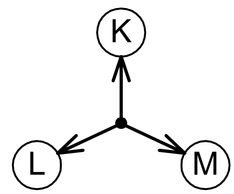
Chapter 2-04

2DH
Resultants of Linear and Quadratic Tensors

In each case setting the diagram equal to zero gives the condition that there is a point in common to the all tensors involved (generally there will be three). If there are only two, we are talking about a tangency relation of some sort.

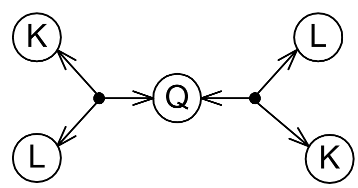
3 Linears

This is just the condition that three lines have a common point



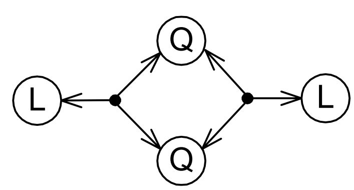
2 Linears, 1 Quadratic

Condition that intersection of lines lies on quadratic



1 Line, 1 Quadratic

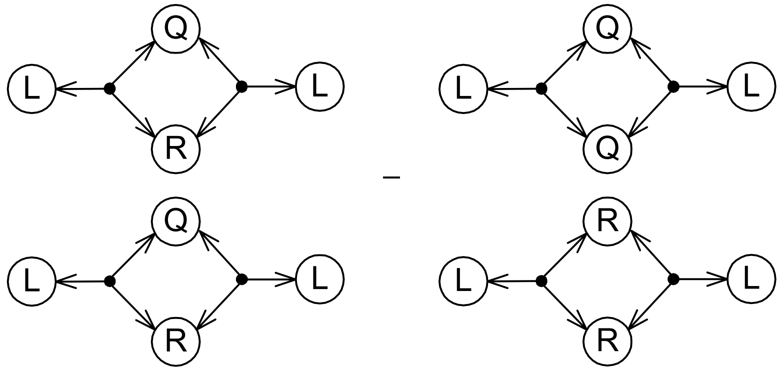
This is the tangency condition. The intersection must be a double root. Note only two tensors involve



1 Linear, 2 Quadratics

Line passes through intersection of quadratics:

Speculation that this will be analog of resultant of 2 1DH quadratic polynomials



(x)

Usage

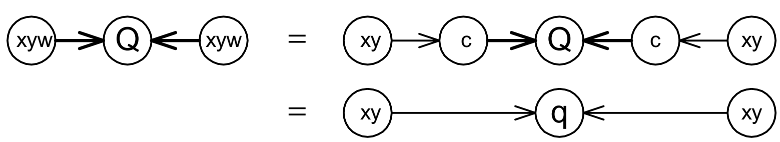
This is the traditional way to intersect two quadratic curves. The idea is to compress a 2DH problem into a 1DH problem by changing from (x,y,w) to (x,p) where p is a polynomial in (y,w). This essentially means that we will multiply each 3x3 matrix by

$$\begin{bmatrix} x & y & w \end{bmatrix}\begin{bmatrix} A & B & C \\ B & D & E \\ C & E & F \end{bmatrix}\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} x & \begin{bmatrix} y & w \end{bmatrix} \end{bmatrix}\begin{bmatrix} A & \begin{bmatrix} B & C \end{bmatrix} \\ \begin{bmatrix} B \\ C \end{bmatrix} & \begin{bmatrix} D & E \\ E & F \end{bmatrix} \end{bmatrix}\begin{bmatrix} x \\ \begin{bmatrix} y \\ w \end{bmatrix} \end{bmatrix}$$

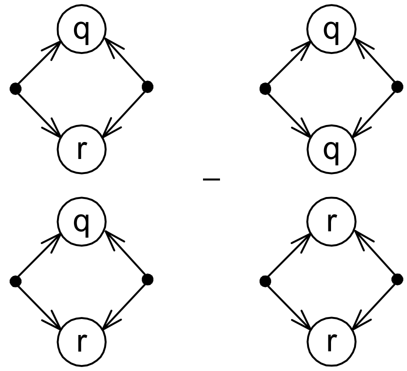
$$= \begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} A & B\frac{y}{w}+C \\ B\frac{y}{w}+C & D\frac{y^2}{w^2}+2E\frac{y}{w}+F \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}$$

$$= \begin{bmatrix} x & w \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{y}{w} & 1 \end{bmatrix}\begin{bmatrix} A & B & C \\ B & D & E \\ C & E & F \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{w} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ w \end{bmatrix}$$

This is tantamount to intersecting the quadratic with the line thru the 2 points [1,0,0] and [0,y/w,1]. This is the horizontal line at Y=y/w. The [x,w] now become the 1DH coordinates of points on this line. In diagram notation this is (like chapter x)

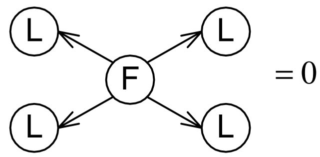


For the two 2DH quadratic curves Q and R, this projects them into two 1DH polynomials q and r. The resultant of these two polynomials is



Replacing each polynomial from the above equation gives: yadda yadda

Anyway, this ultimately has us solving a 4th order thing to find L since equation x looks like:



where F is the symmetrized innards of (x). In the most general case there will be 4 intersections between Q and R, and for each such intersection there will be a collection of lines L that pass through it. So F is a collection of bunches of lines that pass through 4 points. This is the dual of the situation where F consists of bunches of points that are on 4 different lines. That is, we expect F to be factorable into the product of 4 lines…

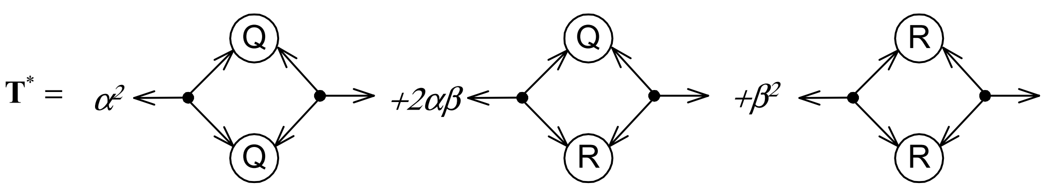
Generalization to other orders

This shows how to intersect any two curves given the resultant of the lower dimensional version of each curve tensor.

another thing

This can also come from line L tangent to linear combo of Q and R

$$\mathbf{T} = \alpha\mathbf{Q} + \beta\mathbf{R}$$

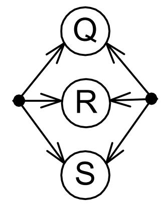


find condition that this is singular…, that is there is a double root (two possible linear combos??) gives above

3 Quadratics

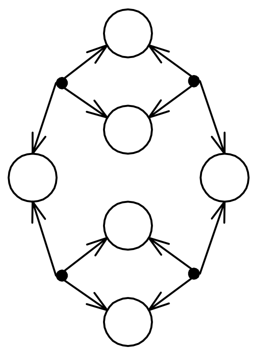
Possible diagrams that are symmetric in QRS implies: equal numbers of each

1 each

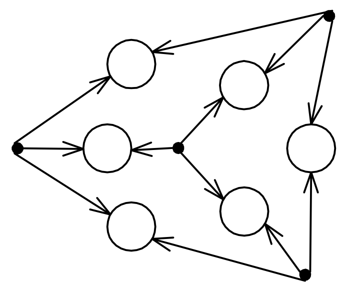


2 each

Two possible topologies:



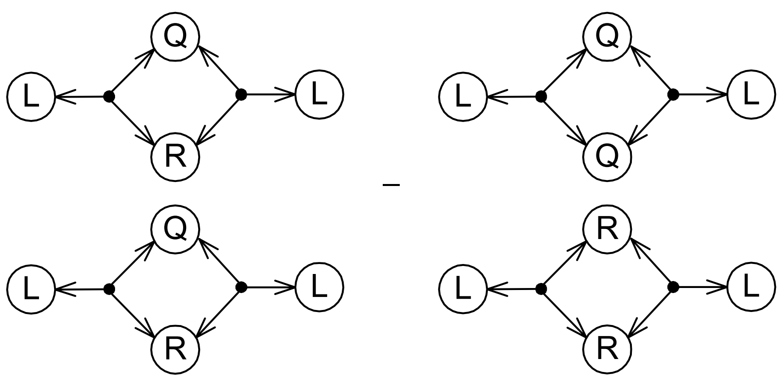
and



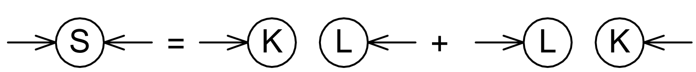
Then there are several distributions of QQRRSS into these slots

Construction Idea

Start with generic resultant of linear, L, and 2 quadratics, Q and R



If third quadratic is product of two linears

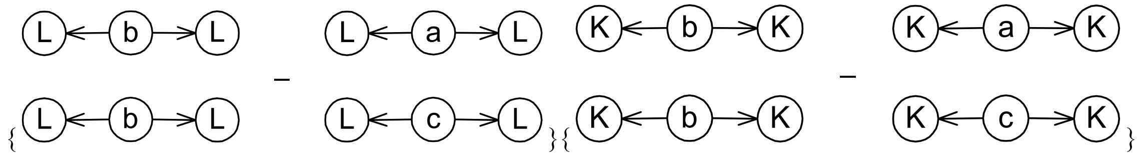


Then desired Resultant of three quadratics is

$$\mathbf{R}(Q,R,S)=\mathbf{R}(Q,R,L)\,\mathbf{R}(Q,R,K)$$

First, abbreviate QQ=a, QR=b RR=c

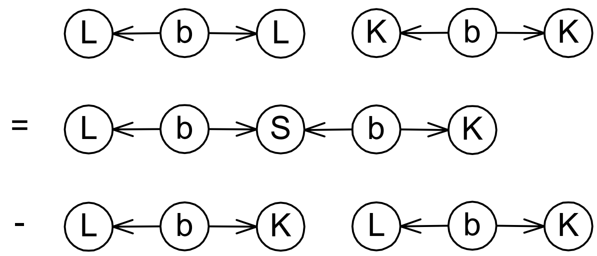
So have



four terms are

$$
\begin{aligned}
&= T_1 - T_2 \\
&-T_3 + T_4
\end{aligned}
$$

Term 1a



LbK is bS in similar manner that we did resultant of two quadratic polys (1DH).

Chapter 2-05

2DH(3D)
Cubic Curves

This is cannibalized from
How Many Different Cubic Curves Are There?
and
Cubic Curve Update
Which are chapters 4 and 6 of
Jim Blinn's Corner: Dirty Pixels

Definition

$$Ax^3 + 3Bx^2y + 3Cxy^2 + Dy^3$$
$$+3Ex^2w + 6Fxyw + 3Gy^2w$$
$$+3Hxw^2 + 3Jyw^2$$
$$+Kw^3 = 0$$

It's possible to write the cubic equation as a sort of cubical matrix of coefficients. Sliced up and laid end to end this would look something like

$$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} A & B & E \\ B & C & F \\ E & F & H \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} x +$$
$$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} B & C & F \\ C & D & G \\ F & G & J \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} y +$$
$$\begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} E & F & H \\ F & G & J \\ H & J & K \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} w = 0$$

The Catalog

Given various values for the A's and B's and so forth, what is the zoology of shapes this can generate?

Also remember that any two curves that can be made to match by any homogeneous transformation (possibly containing perspective) count as the same shape.

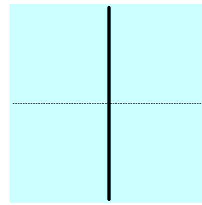
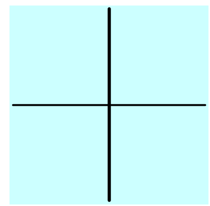
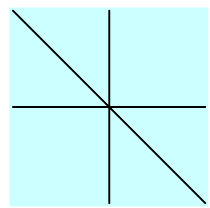
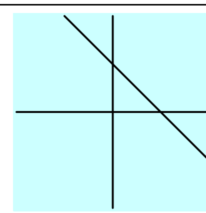
Degenerate Cubics

First of all, let's go through the various degenerate shapes. These are shapes formed when the cubic equation can be factored. This can happen two ways.

First let's talk about the degenerate curves. Visually, these look like two or more low order curves drawn on top of each other. Algebraically it means that the cubic expression can be factored into the product of two or three lower order expressions.

Doubly Degenerate

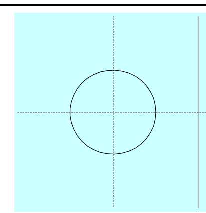
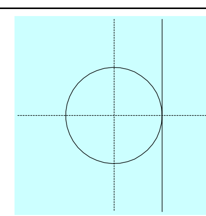
First, there are shapes formed if the cubic is factorable into 3 linear equations. I'll list the possible combinations along with an example equation.

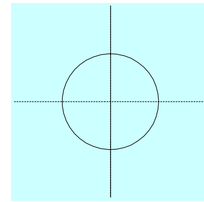
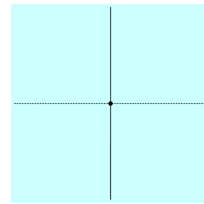
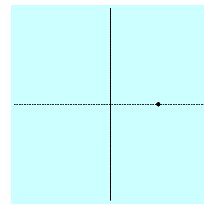
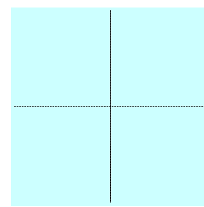
1.	3 coincident lines	$x^3 = 0$	
2.	2 coincident lines and a third distinct one	$x^2 y = 0$	
3.	3 distinct lines that intersect at the same point	$xy(x+y) = 0$	
4.	3 distinct lines that intersect at 3 distinct points	$xy\left(x+y-w\right) = 0$	

Each of these counts as one shape since it's possible to take any transformed version of the above and transform it into the "standard form" shown. This is not completely trivial but you can figure it out.

Singly Degenerate

Second, there are the shapes formed from a linear term times a second order term. There are:

5.	A conic and a line that is disjoint	$\left(x^2+y^2-w^2\right)\left(x-2w\right)=0$	
6.	A conic and a line that is tangent to it;	$(x^2+y^2-w^2)(x-w)=0$	

7.	A conic and a line that intersects it	$\left(x^2+y^2-w^2\right)x=0$	
8.	A single point and a line that is on the point;	$(x^2+y^2)x=0$	
9.	A single point and a line that is not on the point;	$(x^2+y^2)(x-w)=0$	
10.	A null curve and a line;	$(x^2+y^2+w^2)x=0$	

All these examples represent single shapes under perspective transformations. For example, any conic-and-tangent (Type 6) can be transformed to the above standard position since we can transform the conic into a unit circle (taking the tangent line with it) and then rotate the tangent line to be, say $x=1$. Type 8 is similarly a single shape since we can transform the point to the origin and rotate the line to be vertical. Type 9 is a single shape since we can transform the point to the origin, rotate the line vertical, and scale the whole thing in x to put it at $x=1$. Type 10 is a single shape since the line can be transformed anywhere. Types 5 and 7 are also single shapes, but it might take some perspective intuition to see it. To see why, let's try to transform a circle and line into the same circle with the line at a different position. First, do some sort of perspective transform on the circle, turning it into an ellipse. The line goes along for the ride and winds up somewhere new. Now perform a non-uniform scaling and translation of the ellipse to turn it back into the original circle. The line moves again, but now it's in a different position. To be explicit, the transform might be

$$\begin{bmatrix} x & y & w \end{bmatrix}\begin{bmatrix} 5 & 0 & 3 \\ 0 & 4 & 0 \\ 3 & 0 & 5 \end{bmatrix}=\begin{bmatrix} x' & y' & w' \end{bmatrix}$$

You can verify for yourself that if $x^2+y^2=w^2$ then $x'^2+y'^2=w'^2$, so this transforms the circle to itself. But the vertical line $x/w=k$ transforms into

$$x/w=\frac{5k+3}{3k+5}$$

Notice that the lines $x/w=-1$ and $x/w=1$ stay put; they are the tangents. Only the other vertical lines move—pretty magical.

NonDegenerate

Preview

Of course the interesting cubics are the non-degenerate ones. Before we get into these it will be useful to preview some things that cubics can do that we haven't seen before with lower order curves.

The most important of these is the inflection point, indicated in most of the figures below with an *x*. This occurs when the tangent to the curve has the curve lying on both sides of it.

The other interesting thing is the double point, a point of self-intersection, see figure 5b. We have already sort of seen this in degenerate conics consisting of two intersecting lines.

Some General Formulae

Let's review some algebraic formulas that work for any implicitly defined curve.

The line tangent to any curve given by $f(x, y, w) = 0$ is the column vector formed by

$$\begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}$$

where the elements are the partial derivatives of the function *f* evaluated at the point on the curve where the tangent is desired. If it happens that all three derivatives are zero at a point on the curve, then that point is a double point.

Inflection points can be found by means of an auxiliary curve called the Hessian curve. The equation for this curve is constructed by the determinant

$$\det\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} = 0$$

where the elements of the matrix are the various second derivatives of the function *f*. If *f* is a cubic function, the Hessian will generate another cubic curve. The usefulness of this curve comes from the fact that it intersects the original curve at all double points and at all inflection points.

Thus we can find double points of *f* by seeing where all three of the first derivatives are zero, and we can find all inflection points by intersecting the curve with its Hessian and throwing out all the already identified double points. Of course in a real application this might turn into a lot of algebra; but we will find these equations useful.

By the way, if *f* is a second order function, the Hessian works out to be a constant, the determinant of the original matrix. If the determinant (Hessian) is zero it means that the second order curve is degenerate (lines or a single point).

Standard Position 1

We can now attempt to find a standard position for a non-degenerate cubic curve that will let us see its geometric properties best. I'm going to be pretty glib about saying what's possible without explicitly telling you how to calculate the transformation. But that's the nature of existence proofs.

First, let's transform the curve so that it passes through the point $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, the point at infinity on the *y* axis. Plugging this into the general cubic equation it means that the coefficient *D* has to be zero.

Now let's rotate it about that point so that the tangent coincides with the line at infinity, that is the line $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$. Evaluating the first derivatives at this point we get

$$\begin{bmatrix} f_x \\ f_y \\ f_w \end{bmatrix} = 3\begin{bmatrix} C \\ D \\ G \end{bmatrix}$$

so we must have $C = 0$ and $G \neq 0$.

We have to specify that the point we started with was not a double point, or G would be zero. This isn't hard; there are lots of non-double points. But while we're at it, let's arrange it so the point we started with *was* an inflection point. (All non-degenerate cubics have at least one inflection point.) Evaluating the second derivatives at $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ and constructing the Hessian gives:

$$\det\begin{bmatrix} B & 0 & F \\ 0 & 0 & G \\ F & G & J \end{bmatrix} = -G^2 B$$

so $B = 0$.

What we have said is that *any* non-degenerate cubic can be transformed so that the equation looks like

$$\begin{aligned} &Ax^3 \\ &+3Ex^2w + 6Fxyw + 3Gy^2w \\ &+3Hxw^2 + 3Jyw^2 \\ &+Kw^3 = 0 \end{aligned}$$

But we're not done yet. Let's now scale and skew the y coordinate via the transformation

$$y \mapsto -\frac{F}{G}x + y - \frac{J}{2G}w$$

Plug this in, turn the crank, and you will discover that the F and J terms disappear. The result is that all non-degenerate cubics can be transformed into the form

$$y^2w = ax^3 + bx^2w + cxw^2 + dw^3$$

or in non-homogeneous form (with $X = x/w$ and $Y = y/w$)

$$Y^2 = aX^3 + bX^2 + cX + d$$

That is, the set of possible shapes is constructed by taking the square root of all possible cubic polynomials in X. If the parameter a is zero it means we have a degenerate curve. We can see this by going back to the homogeneous form and putting the y back on the right hand side.

$$\begin{aligned} 0 &= bx^2w + cxw^2 + dw^3 - y^2w \\ &= w\left(bx^2 + cxw + dw^2 - y^2\right) \end{aligned}$$

In other words we have the line-at-infinity times a quadratic curve (which may itself be factorable.)

Back to the case where $a \neq 0$. The equation $y^2w = ax^3 + bx^2w + cxw^2 + dw^3$ isn't really a four-parameter set of shapes. We can boil this down still further by translating in x to "center" the cubic polynomial (so that its second derivative is zero at $x = 0$). This is similar to what we do in the conventional algorithm for solving cubic polynomials; here it will make the b coefficient zero. We now have:

$$y^2w = ax^3 + cxw^2 + dw^3$$

(The new *a,c,d* coefficients will be some function of the old *a,b,c,d* . I've just recycled the letters *a,c,d* here.)

Next, we can scale in *x* to get $a = 1$ and we have (again recycling the letters *c* and *d*):

$$y^2 w = x^3 + cxw^2 + dw^3$$

This looks, at first, like it might be a two-parameter family of curves, but it's not. A further scale in x and y can transform the curve as follows

$$0 = (s_x x)^3 + c(s_x x)w^2 + dw^3 - (s_y y)^2 w$$
$$0 = s_x^{\,3} x^3 + cs_x xw^2 + dw^3 - s_y^{\,2} y^2 w$$

If we pick scales that satisfy $s_x^{\,3} = s_y^{\,2}$ (note that s_x must be positive) we can write:

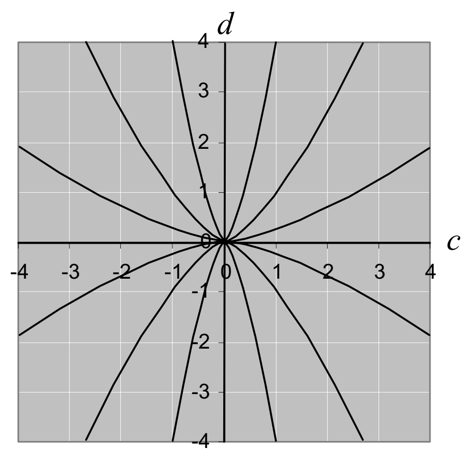
$$0 = s_x^{\,3} x^3 + cs_x xw^2 + dw^3 - s_x^{\,3} y^2 w$$
$$= s_x^{\,3}\left(x^3 + \left(\frac{c}{s_x^{\,2}}\right) xw^2 + \left(\frac{d}{s_x^{\,3}}\right) w^3 - y^2 w \right)$$

This is back to our canonical form but with different values of *c* and *d*.

$$\hat{c} = \frac{c}{s_x^{\,2}}$$

$$\hat{d} = \frac{c}{s_x^{\,3}}$$

In other words, the two curves with parameters (c, d) and $(\hat{c}, \hat{d})$ are the same curve (modulo some scaling transformation). All the cubic curves generated by (c, d) values along any of the curved lines in the following *c,d* parameter space graph are the "same shape":

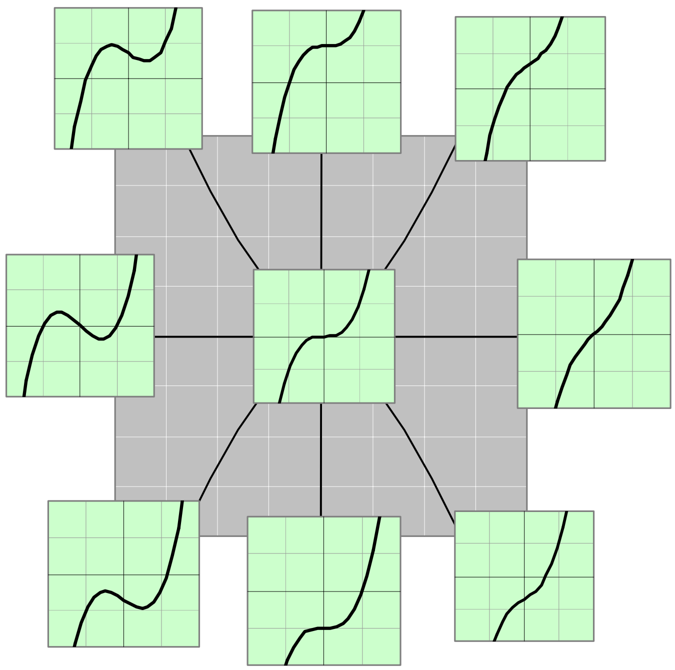


All the *different* cubic curves then form a single parameter family represented by this parameter-space curve sweeping around the origin, with all cubic curves within each family satisfying:

$$\frac{c^3}{d^2} = \text{constant}$$

Note, however, that the two curves with parameter (c,d) and $(c,-d)$ are different curves, so just using the constant c^3/d^2 does not give an unambiguous parameterization of the space of cubic curves.

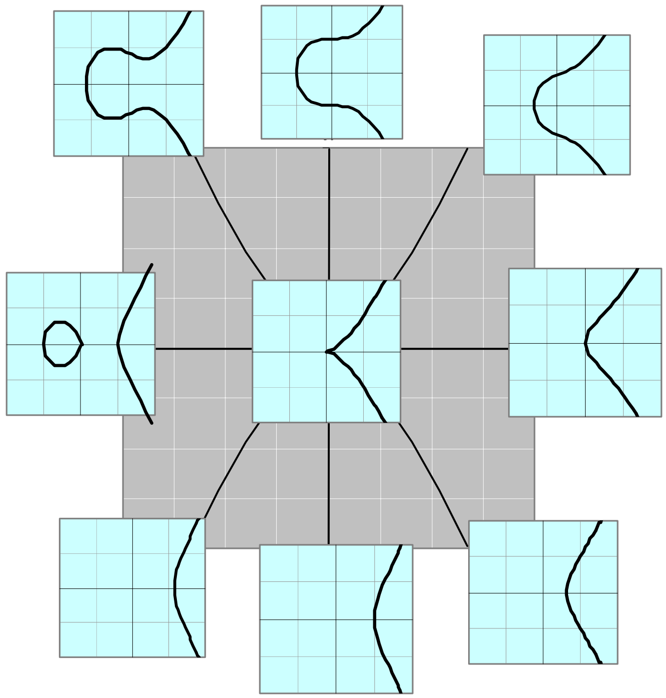
To get a sense of the possible shapes of cubic polynomials we can have I'll generate a few of them and place the plots on the (c,d) diagram. In addition to the ring of polynomial shapes, don't forget the single cubic curve shape represented by the origin $(c,d)=(0,0)$.



The actual cubic curve itself is

$$Y=\sqrt{X^3+cX+d}$$

and is defined where the cubic in X is positive. These shapes appear in the plot below;



That is, we have a continuum of shapes plus one distinct shape at the origin of (*c,d*) space. Each of the shapes shown above is distinct from the others. Despite the fact that several of them look very similar, they cannot be transformed into each other by a homogeneous perspective transformation.

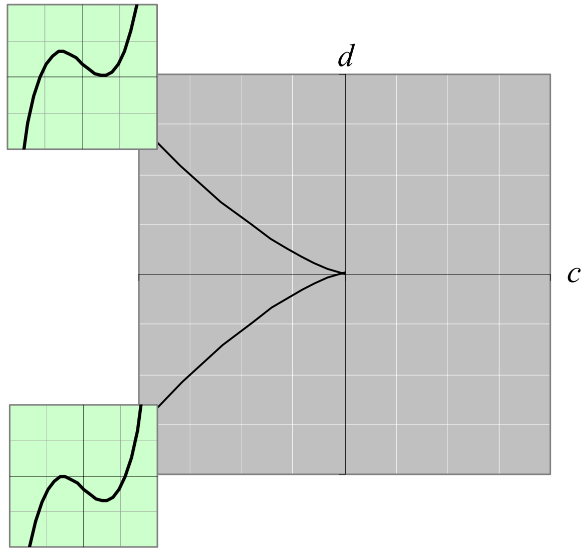
The shapes in the continuum can have either one or two disjoint pieces. The transition point happens when the cubic polynomial $x^3 + cxw^2 + dw^3$ has a double root. This happens when the polynomial is

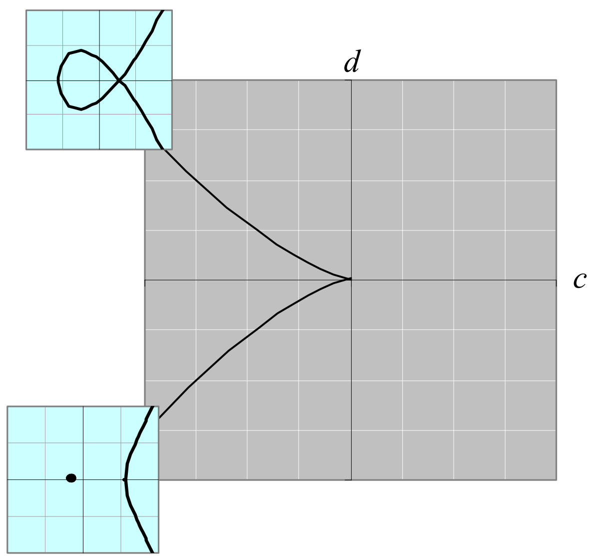
$$x^3 - 3xw^2 + 2w^3 = (x + 2w)(x - w)^2$$

and when the polynomial is

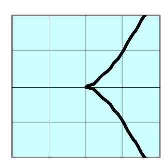
$$x^3 - 3xw^2 - 2w^3 = (x - 2w)(x + w)^2$$

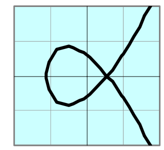
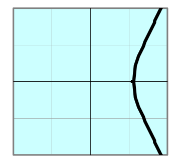
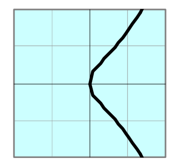
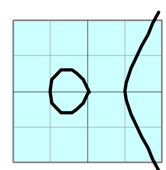
These two situations generate the additional curves





We will then tabulate this as three distinct curve types and two continuous families of topologically similar types:

10.	A cusp	$x^3 - y^2 w = 0$	

11.	A loop	$x^3 - 3xw^2 + 2w^3 - y^2w = 0$	
12.	A wiggle and a dot	$x^3 - 3xw^2 - 2w^3 - y^2w = 0$	
13.	A continuum of wiggles	$x^3 + cxw^2 + dw^3 - y^2w = 0$	
14	A continuum of egg-and-wiggles	$x^3 + cxw^2 + dw^3 - y^2w = 0$	

Encompassing all cubics

The form we boiled down to is able to represent all cubics that are irreducible, that is they are not factorable into products of lower order curves. Sometimes we will need a general canonical form that can represent *all* cubics. In that case we rewind back to an earlier form that just transform to make coefficients B,C,D,F,J=0 and get:

$$Ax^3 + 3Ex^2w + 3Gy^2w + 3Hxw^2 + Kw^3 = 0$$

The reducible curves happen if

G=0

Then we have

$$Ax^3 + 3Ex^2w + 3Hxw^2 + Kw^3 = 0$$

This is a homogeneous cubic polynomial in (x,w). Each real solution gives a vertical line, and all these lines pass through the infinite point [0 1 0]. This is a doubly degenerate curve of types 1, 2, or 3.

A=0, G!=0

Then we have

$$w\left(3Ex^2 + 3Gy^2 + 3Hxw + Kw^2\right) = 0$$

This is the line-at-infinity times a quadratic curve. (We might need E!=0 when A=0. We cannot assume that we have translated to make E=0.) The quadratic in question looks like:

$$\mathbf{Q} = \begin{bmatrix} 3E & 0 & \frac{3H}{2} \\ 0 & 3G & 0 \\ \frac{3H}{2} & 0 & K \end{bmatrix}$$

The determinant of this gives us more information

$$\begin{aligned} \det \mathbf{Q} &= (9EGK) - 3G\frac{9H^2}{4} \\ &= 9EGK - \frac{3G9H^2}{4} \\ &= \frac{9}{4}G\left(4EK - 3H^2\right) \end{aligned}$$

The quadratic degenerates into two lines if $4EK - 3H^2 = 0$. These will not pass through the point [0 1 0] since G is not zero. Therefore, this form gives us the type 4 doubly degenerate cubic (three lines that do not intersect at a common point).

If the determinant is nonzero we either have a null curve or an ellipse. The tangency of that ellipse to the line at infinity come from the adjoint is of **Q**

$$\mathbf{Q}^* = \begin{bmatrix} 3GK & 0 & -\frac{9GH}{2} \\ 0 & 3EK - \frac{9H^2}{4} & 0 \\ -\frac{9GH}{2} & 0 & 9GE \end{bmatrix}$$

The line at infinity [0 0 1]^T is tangent to the quadratic if E=0, it intersects if GE is negative and is disjoint if GE is positive (or vice versa, check). This gives us types 5,6,7.

In sum, the three discriminating quantities that determine the shape of Q are:

$$\begin{aligned} &\det \mathbf{Q} = \frac{9}{4}G\left(4EK - 3H^2\right) \\ &trace\mathbf{Q}^* = 3GK + 3EK - \frac{9H^2}{4} + 9GE \\ &trace\mathbf{Q} = 3E + 3G + K \end{aligned}$$

Standard Position 2

I would like to thank William Waterhouse of Penn State for making me aware of some of the analysis here. Waterhouse points out that there is a prettier formulation of the cubic equation devised by Hesse, viz.

$$x^3 + y^3 + w^3 - \lambda xyw = 0$$

This form is nice and symmetric; each coordinate is treated the same as any other. It also has some nice geometric properties. Notice, for example, that all the curves in this formulation pass through the three points $\begin{bmatrix}-1 & 0 & 1\end{bmatrix}$, $\begin{bmatrix}0 & -1 & 1\end{bmatrix}$ and $\begin{bmatrix}1 & -1 & 0\end{bmatrix}$. Furthermore, these three points happen to be the three inflection points of the curve, and they are of course collinear.

This equation has two degenerate special cases. When $\lambda = 3$ it degenerates to a line and a dot (type 8 from above); the line is $X + Y = -1$ and dot is at $\begin{bmatrix}1 & 1\end{bmatrix}$. I originally saw this by plugging in numbers and plotting them; with some work you can also see it algebraically by factoring:

$$x^3 + y^3 + w^3 - 3xyw = (x + y + w)\left(x^2 + y^2 + w^2 - xy - yw - wx\right)$$

When $\lambda = \infty$ the curve degenerates into three distinct lines (type 4 from above). How do we to deal algebraically with an infinite value for λ? Just divide the equation by λ and *then* let it go to infinity. The equation you get is:

$$xyw = 0$$

This gives the three lines: the X-axis, the Y-axis, and the line at infinity.

I actually played with this symmetric form a lot when writing the first version of this as article as an IEEE column, trying to see if I could get *all* of the nondegenerate cubics to be expressible in this form. Unfortunately, I started with the cusp (it's the simplest algebraically), and found that it *cannot* be transformed into this symmetric form. I got discouraged and gave up. Well guess what… all of them *can* be transformed into this form *except* for the cusp (type 10), the loop (type 11) and the wiggle-and-dot (also called the serpentine) (type 12). So I gave up too soon. I guess there's a lesson here.

Correspondence

Anyway, what I want to do now is to show the correspondence between the $Y^2 = \ldots$ form and Hesse's symmetric form. To make the correlation, let's start with the symmetric form and try to transform it to the Y^2 form. First for some visual inspiration. Look at a sampling of the sorts of shapes it can generate in the right column of figure 1. They are all tilted by 45 degrees, so our first guess is to tilt the symmetric equation back by 45 degrees. Actually it's simpler to use the following transform:

$$\begin{aligned} x &\mapsto x + y \\ y &\mapsto x - y \end{aligned}$$

The symmetric equation turns into

$$y^2(\lambda w + 6x) + 2x^3 - \lambda x^2 w + w^3 = 0$$

Now remember, we're trying to put this into the form $Y^2 = (\text{a cubic polynomial in X})$. So let's try to turn the factor $(\lambda w + 6x)$ into a simple w. With blinding inspiration we try replacing

$$w \mapsto (w - 6x)/\lambda$$

Warning! Danger! What happens if $\lambda = 0$? If this happens, our intermediate form is

$$y^2(6x) + 2x^3 + w^3 = 0$$

and we can bash it into the desired form by simply exchanging x and w. After fiddling we get the equation for a particular type 14 curve with the c,d parameters equal to (0,-1):

$$Y^2 = X^3 - 1$$

our first match between the old world and the new.

Whew! Back to $\lambda \neq 0$. Make the $(w-6x)/\lambda$ substitution, algebrize a bit, go to non-homogeneous coordinates and you wind up with:

$$Y^2 = X^3\left(\frac{8\left(3^3-\lambda^3\right)}{\lambda^3}\right) + X^2\left(\frac{\lambda^3-4\cdot 3^3}{\lambda^3}\right) + X\frac{18}{\lambda^3} + \frac{-1}{\lambda^3}$$

so

$$a = 8\left(3^3-\lambda^3\right)/\lambda^3$$
$$b = \left(\lambda^3-4\cdot 3^3\right)/\lambda^3$$
$$c = 18/\lambda^3$$
$$d = -1/\lambda^3$$

Now for a reality check. If $\lambda = 3$ the cubic term evaporates and we have a degenerate curve. Check. If $3ac = b^2$ we should get a type 14 curve. So substitute in and solve for λ. You get two solutions: $\lambda = 0$ and $\lambda = -6$. We already saw that $\lambda = 0$ leads to (c,d)=(0,-1). Double check. $\lambda = -6$ turns out to give (c,d) = (0,+1). Triple check. Other values of λ give various cases of type 14 and 15.

Orphans

There are some curves left out in the cold though: the cusp (type 11), the loop (type 12), and the wiggle-dot (type 13). These cannot be generated by just finding a magic value for λ. Their algebraic form just won't be symmetric. So how close can we come?

After some tinkering I was able to transform the loop into the approximately symmetrical form

$$x^3 + y^3 + xyw = 0$$

Notice that there's no parameter here. The λ can be absorbed by scaling x and y. So all loops are the same shape (projectively).

Is there a sort-of-symmetric formulation for the cusp and wiggle-dot? I've flailed around for a while but haven't come up with one. (You can make the wiggle-dot a bit prettier by translating it so that it's $Y^2 = X^3 - X^2$)

Questions

Is there a way, given the original coefficients, to find which of these shapes it generates? For quadratic curves we looked at the eigenvalues of the coefficient matrix to catalog the possible shapes. Is there a generalization of eigenvalues the 3x3x3 tensor of cubic curve coefficients that would give us a similar indicator? This would presumably take the form of a tensor diagram consisting of several copies of the coefficient tensor.

Some of the curves plotted look suspiciously similar. What geometric property that is preserved under homogeneous perspective transformations makes them different? This would be a cross ratio generalization in a similar manner to the quadratic polynomial cross ratio generalization.

Given the original coefficients, what formulas calculate the following interesting geometric properties: (These are properties that remain invariant under perspective transformation.)

1. Number and location of inflection points
2. Number and location of double points
3. Intersection properties of the three tangent lines at the three inflection points

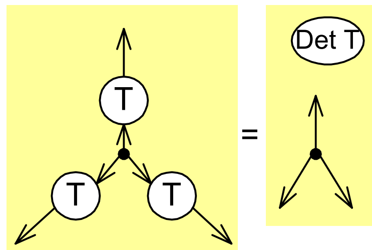
Some initial ideas appear in the next chapter.

Chapter 2-06

2DH(3D)
Cubic Curve Invariants

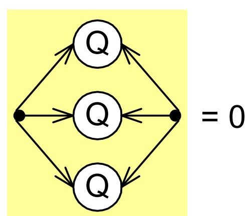
Here are some rough ideas for useful invariants for cubic curves.

Review: We formulate various invariants for 2DH by constructing a network connecting multiple copies of a tensor with epsilons. Any such diagram will be invariant due to the identity



Thus, plugging in a transformed **C** we can shove all the **T**'s over to the epsilons and they will turn into factors of det**T**. If we have an even number of epsilons, the sign won't change regardless of the sign of det**T**.

One such invariant gave the condition that a quadratic curve is factorable



This is just the determinant of **Q** (times some scale factor).

Cubic Curve invariants

What sorts of diagrams represent invariants for cubic curves? They will consist of some number of C nodes connected together with an equal number of epsilon nodes. I have experimented with the following two diagrams. Anything simpler (with fewer C's) appears to be identically zero.

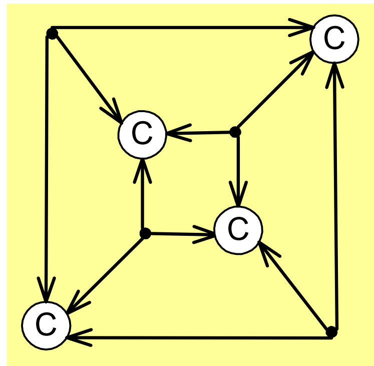
We will explicitly calculate the invariants for two different canonical forms of the cubic:

Y squared	$Ax^3 + 3Ex^2w + 3Hxw^2 + Kw^3 + 3Gy^2w = 0$
Lambda	$x^3 + y^3 + w^3 + 6Fxyw = 0$

Note that Salmon uses this latter form but uses the coefficient name $m = F$

Cube Invariant

The first arranges the **C** and epsilons topologically on the vertices of a cube.



Applying this to the canonical forms:

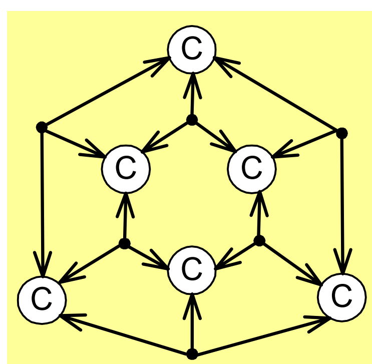
Y squared	$I_{cube} = 24G^2\left(E^2 - AH\right)$
Lambda	$I_{cube} = -24\left(F - F^4\right)$
Salmon	$S = m - m^4$

From this we conclude

$$I_{cube} = -24S$$

Hexagon invariant

The second is similar but is more like a hexagonal prism



Y squared	$I_{hexagon} = 24G^3\left(A\left(EH - AK\right) + 2E\left(AH - E^2\right)\right)$
Lambda	$I_{hexagon} = -6\left(1 - 20F^3 - 8F^6\right)$
Salmon	$T = 1 - 20m^3 - 8m^6$

$$I_{hexagon} = -6T$$

Octahedron invariant

The obvious extension of this diagram machinery to an octahedral extrusion doesn't buy us anything. I have convinced myself that it just gives a value that equals the square of the cube invariant. (But I don't have a solid proof).

Salmons Discriminant

This is given as

$$\Delta = T^2 + 64S^3$$

In terms of the diagram invariants:

$$\begin{aligned}\Delta &= \left(\frac{I_{hex}}{-6}\right)^2 + 64\left(\frac{I_{cube}}{-24}\right)^3 \\ &= \frac{1}{216}\left(6I_{hex}{}^2 - I_{cube}{}^3\right)\end{aligned}$$

In Salmons Canonical form

Using the lambda canonical form (but using Salmon's name for lambda: *m*) we have

$$\begin{aligned}\Delta &= \left(1-20m^3-8m^6\right)^2 + 64\left(m-m^4\right)^3 \\ &= 1-40m^3-16m^6+400m^6+320m^9+64m^{12} \\ &\quad +64m^3\left(1-3m^3+3m^6-m^9\right) \\ &= 1-40m^3-16m^6+400m^6+320m^9+64m^{12} \\ &\quad +64m^3-192m^6+192m^9-64m^{12} \\ &= 1-40m^3+64m^3+400m^6-16m^6-192m^6+192m^9+320m^9+64m^{12}-64m^{12} \\ &= 1+24m^3+192m^6+512m^9 \\ &= 1+3(2m)^3+3(2m)^6+(2m)^9 \\ &= \left(1+(2m)^3\right)^3\end{aligned}$$

This is zero only when:

$$\begin{aligned}(2m)^3 &= -1 \\ 2m &= -1 \\ m &= -\frac{1}{2}\end{aligned}$$

This is just the case where this form degenerates to a line and a disjoint point.

Also need to figure this out when m=infinity (the other degenerate case for this form, it being three disjoint lines.)

Y squared version

This form can encompass ALL cubics

$$216\Delta = 6I_{hex}{}^2 - I_{cube}{}^3$$

$$I_{cube} = 24G^2\left(E^2 - AH\right)$$

$$I_{hexagon} = 24G^3\left(A\left(EH - AK\right) + 2E\left(AH - E^2\right)\right)$$

We have the following subcases

G=0

This condition gives is three intersecting lines

$$I_{cube} = 0$$

$$I_{hexagon} = 0$$

$$\Delta = 0$$

Both invariants being zero imply three lines.

A=0 G!=0

$$I_{cube} = 24G^2E^2$$

$$I_{hexagon} = -48G^3E^3$$

$$\Delta = 0$$

Geometrically this is a quadratic and a line. If E=0 the quadratic is tangent to the line. In this case both invariants are zero. This still doesn't distinguish between lines disjoint from the ellipse and the more general cubic curve.

A!=0 G!=0

We can translate to make E=0 and scale to make A=1 and G=-1. Then

$$I_{cube} = -24H$$

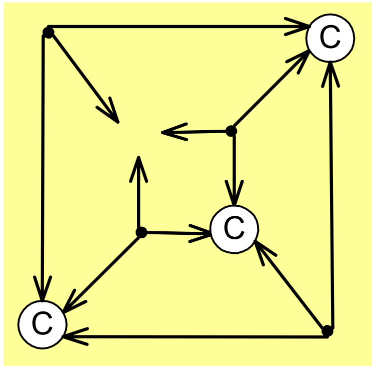
$$I_{hexagon} = 24K$$

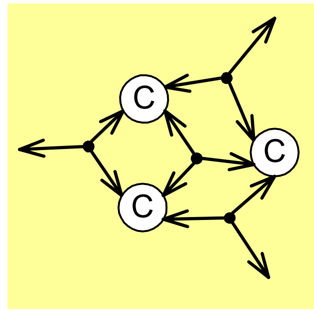
$$\Delta = 16K^2 + 64H^3 \quad \text{(check constants)}$$

The invariants give the coordinates in H,K space of the irreducible cubic (as shown in previous chapter). The discriminant will be zero only on the line that gives a loop or a wiggle-dot. In addition, if both H and K are zero we have a cusp.

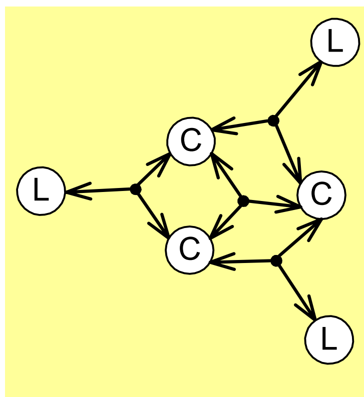
Cubic Adjoint (Cayleyan)

There is a quantity called by some authors the "Cayleyan" of the cubic. This turns out to be, in diagram form, equivalent to the cube invariant with one C removed.





This gives another cubic but it is a covariant cubic (arrows out instead of in). Therefore it is suitable for multiplying by L



I'm not yet sure what this means.

Chapter 2-07

2DH(3D)
The Hessian of a Cubic Curve

This chapter consists of various ideas on using the Hessian curve (defined below) to identify interesting properties of a cubic curve. In particular it finds inflection points. The ultimate goal is to be given a cubic tensor C and have an equation for the three inflection points and the line they lie on. This chapter is contains a few dead ends and sketchy ideas. Beware especially of missing constants and flipped signs.

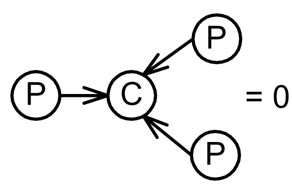
Cubic Curve

$$\begin{aligned} &Ax^3+3Bx^2y+3Cxy^2+Dy^3\\ &+3Ex^2w+6Fxyw+3Gy^2w\\ &+3Hxw^2+3Jyw^2\\ &+Kw^3=0 \end{aligned}$$

Matrix Form is 3x3x3 tensor

$$\begin{bmatrix} x & y & w \end{bmatrix} \left\{ \left\{ \begin{bmatrix} A & B & E \\ B & C & F \\ E & F & H \end{bmatrix} \begin{bmatrix} B & C & F \\ C & D & G \\ F & G & J \end{bmatrix} \begin{bmatrix} E & F & H \\ F & G & J \\ H & J & K \end{bmatrix} \right\} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \right\} \begin{bmatrix} x \\ y \\ w \end{bmatrix} = 0$$

Diagram form



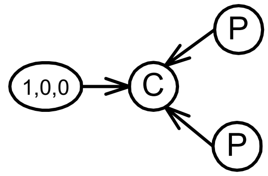
Derivatives

The function is

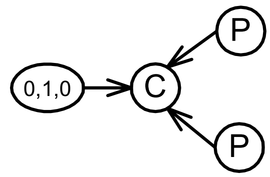
$$\begin{aligned} f(x,y,w) = &Ax^3+3Bx^2y+3Cxy^2+Dy^3\\ &+3Ex^2w+6Fxyw+3Gy^2w\\ &+3Hxw^2+3Jyw^2\\ &+Kw^3 \end{aligned}$$

First derivatives

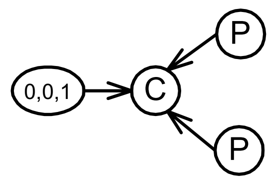
$$\frac{\partial f}{\partial x} = f_x = 3Ax^2 + 6Bxy + 3Cy^2 + 6Exw + 6Fyw + 3Hw^2$$ = first slice =



$$\frac{\partial f}{\partial y} = f_y = 3Bx^2 + 6Cxy + 3Dy^2 + 6Fxw + 6Gyw + 3Jw^2$$ = second slice =

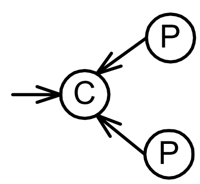


$$\frac{\partial f}{\partial w} = f_w = +3Ex^2 + 6Fxy + 3Gy^2 + 6Hxw + 6Jyw + 3Kw^2$$ = third slice =



More generally, can say the derivative of the function at point P is a contravariant vector that is:

$$\begin{bmatrix} f_x \\ f_y \\ f_w \end{bmatrix} = 3$$



Second derivatives

$$f_{xx} = 6Ax + 6By + 6Ew$$

$$f_{xy} = 6Bx + 6Cy + 6Fw$$

$$f_{xw} = 6Ex + 6Fy + 6Hw^2$$

$$f_{yy} = +6Cx + 6Dy + 6Gw$$

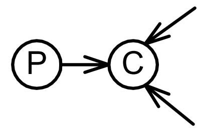
$$f_{yw} = +6Fx + 6Gy + 6Jw$$

$$f_{ww} = +6Hx + 6Jy + 6Kw$$

etc.

Similarly can construct matrix of second derivatives at point P as:

$$\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} = 6$$



The Hessian

The Hessian curve is a new cubic curve defined by

$$\mathbf{H}(x, y, w) = \det \begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} = 0$$

This is interesting because where the curve intersects its Hessian there is an inflection point.

Explicitly, the Hessian is

$$\det \begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} =$$

$$\det \begin{bmatrix} 6Ax + 6By + 6Ew & 6Bx + 6Cy + 6Fw & 6Ex + 6Fy + 6Hw^2 \\ 6Bx + 6Cy + 6Fw & 6Cx + 6Dy + 6Gw & 6Fx + 6Gy + 6Jw \\ 6Ex + 6Fy + 6Hw^2 & 6Fx + 6Gy + 6Jw & +6Hx + 6Jy + 6Kw \end{bmatrix} =$$

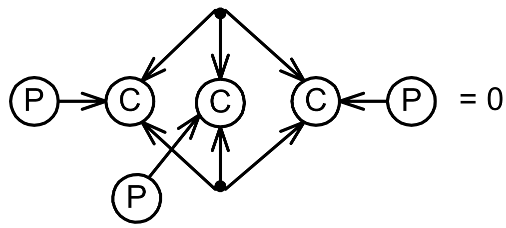
$$6^3 \det \begin{bmatrix} Ax + By + Ew & Bx + Cy + Fw & Ex + Fy + Hw \\ Bx + Cy + Fw & Cx + Dy + Gw & Fx + Gy + Jw \\ Ex + Fy + Hw & Fx + Gy + Jw & Hx + Jy + Kw \end{bmatrix} =$$

$$6^3 \det \left\{ x \begin{bmatrix} A & B & E \\ B & C & F \\ E & F & H \end{bmatrix} + y \begin{bmatrix} B & C & F \\ C & D & G \\ F & G & J \end{bmatrix} + w \begin{bmatrix} E & F & H \\ F & G & J \\ H & J & K \end{bmatrix} \right\} =$$

$$6^3 \det \left\{ x\mathbf{C}_x + y\mathbf{C}_y + w\mathbf{C}_w \right\} =$$

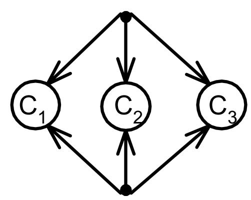
Diagram Form

The ways to evaluate this are best done in diagram form. The diagram is



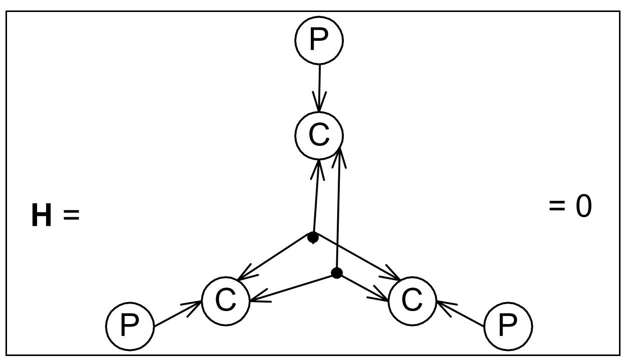
You can see that transforming C by the transformation T will transform the Hessian by the same transformation: (show diagram).

The 10 different coefficients of the Hessian cubic curve are various determinant/mutual determinants of the three "slices" of the C tensor, expressed by plugging in x,y,w in various permutations for 1,2,3 below:

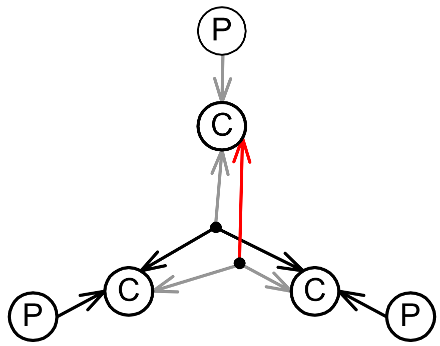


Another view of the Hessian diagram

We can unfold the above to:

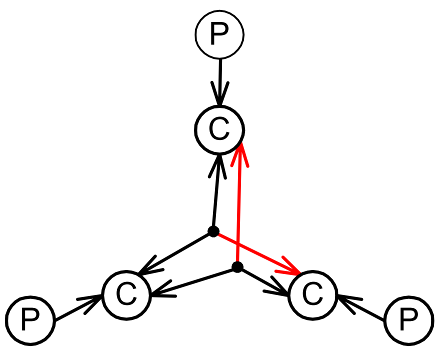


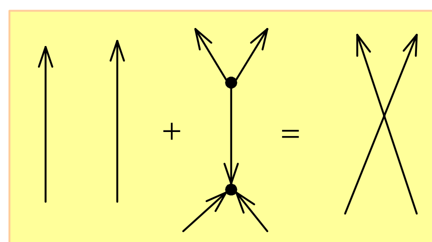
Can epsilon/delta simplify this? For the red arc below, we expect not to get anything useful by combining with the gray arcs.

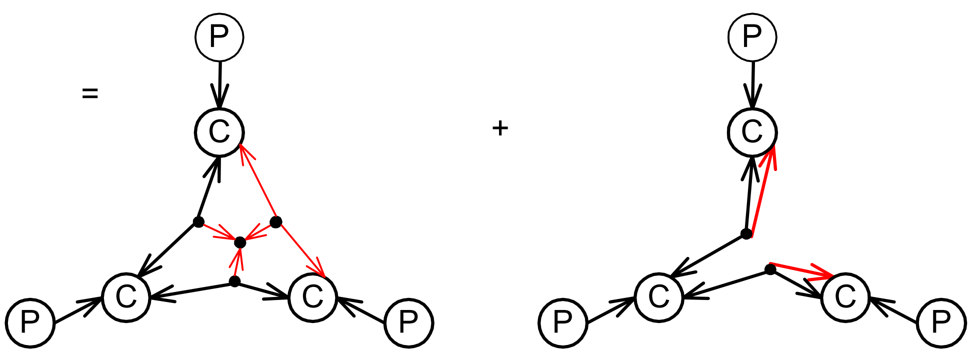


There are (modulo symmetry) two other choices, one attached to P and one attached to C

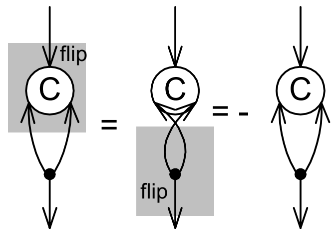
Choice 1



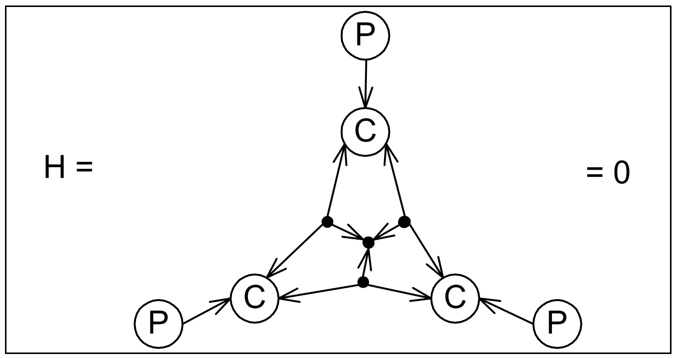




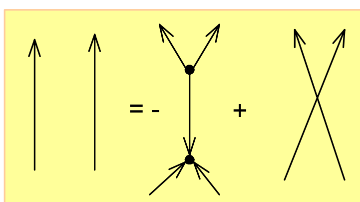
The second term is zero, since each arm is zero, since it's an anti-symmetric thing contracted with a symmetric thing.

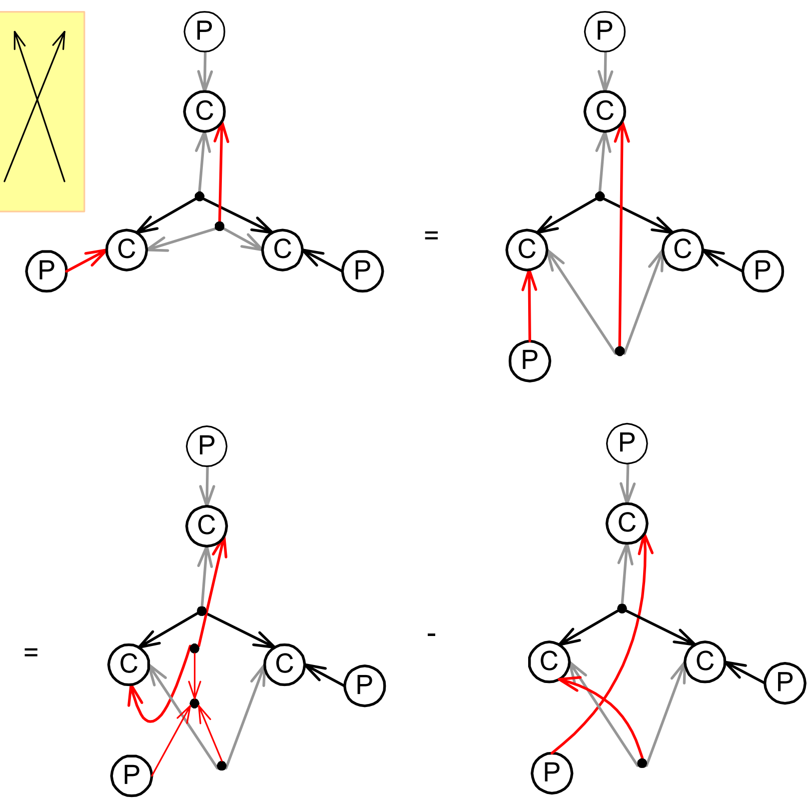


So we have the alternate form:

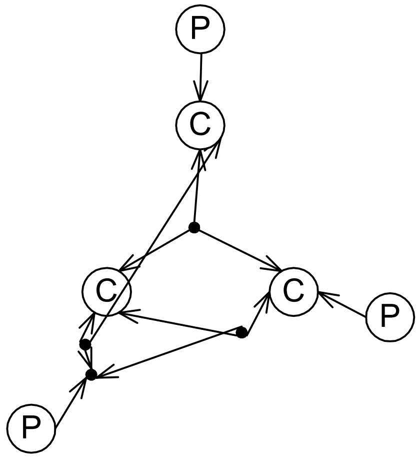


Choice 2

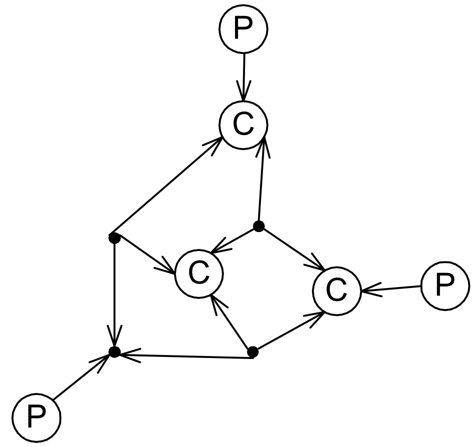




second term is identically zero. First term rearranges to;



which equals



Not sure if this is useful…

Hessian of Standard Reducible Curves

For the cubic in standard position

$$Ax^3 + 3Ex^2w + 3Hxw^2 + Kw^3 + 3Gy^2w = 0$$

The two reducible subtypes are:

G=0

$$f(x,y,w) = Ax^3 + 3Ex^2w + 3Hxw^2 + Kw^3$$
$$f_x = 3Ax^2 + 6Exw + 3Hw^2$$
$$f_y = 0$$
$$f_w = 3Ex^2 + 6Hxw + 3Kw^2$$

$$\mathbf{H}(x,y,w) = \det\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix}$$
$$= \det\begin{bmatrix} f_{xx} & 0 & f_{xw} \\ 0 & 0 & 0 \\ f_{xw} & 0 & f_{ww} \end{bmatrix} = 0$$

In other words the Hessian curve is identically zero for double degenerate curves.

A=0, G!=0,

$$f(x,y,w) = 3Ex^2w + 3Hxw^2 + Kw^3 + 3Gy^2w$$
$$f_x = 6Exw + 3Hw^2$$
$$f_y = 6Gyw$$
$$f_w = 3Ex^2 + 6Hxw + 3Kw^2 + 3Gy^2$$

$$\mathbf{H}(x,y,w)=\det\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix}$$

$$=\det\begin{bmatrix} 6Ew & 0 & 6Ex+6Hw \\ 0 & 6Gw & 6Gy \\ 6Ex+6Hw & 6Gy & 6Hx+6Kw \end{bmatrix}$$

$$\begin{aligned}
\frac{1}{6^3}\mathbf{H} &= EwGw\left(Hx+Kw\right)-Gw\left(Ex+Hw\right)^2-EwGyGy \\
&= EGHxw^2+EGKw^3-Gw\left(E^2x^2+2EHxw+H^2w^2\right)-EG^2y^2w \\
&= EGHxw^2+EGKw^3-GE^2x^2w-2GEHxw^2-GH^2w^3-EG^2y^2w \\
&= Gw\left(-E^2x^2-EHxw+\left(EK-H^2\right)w^2-EGy^2\right)
\end{aligned}$$

In other words, this is the line-at-infinity and another quadratic. The Hessian of a line times a quadratic is the same line times another quadratic. Investigate further…

Hessian of Standard irreducible Curve

The standard irreducible curve

$$3y^2w=x^3+3Hxw^2+Kw^3$$

$$f\left(x,y,w\right)=x^3+3Hxw^2+Kw^3-3y^2w$$

$$\begin{aligned}
f_x &= 3x^2+3Hw^2 \\
f_y &= -6yw \\
f_w &= 6Hxw+3Kw^2-3y^2
\end{aligned}$$

$$\mathbf{H}(x,y,w)=\det\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix}$$

$$=\det\begin{bmatrix} 6x & 0 & 6Hw \\ 0 & -6w & -6y \\ 6Hw & -6y & 6Hx+6Kw \end{bmatrix}$$

$$=-xw\left(Hx+Kw\right)+wH^2w^2-y^2x$$

$$=-Hx^2w-Kxw^2+H^2w^3-y^2x$$

This is another irreducible cubic. What relation does it have with the original? In particular, what shape does it have for various special cases (like the parametrizable curves cusp, loop, and serpentine)?

Cusp H=K=0

$$\mathbf{H}(C) = -y^2 x$$

So the Hessian of a cusp is a double line (y=0) and single line (x=0) intersecting at the cusp point.

Loop

Equation is

$$3y^2 w = x^3 - 3xw^2 + 2w^3$$
$$H = -1$$
$$K = 2$$

Hessian is

$$\mathbf{H}(C) = x^2 w - 2xw^2 + w^3 - y^2 x$$

Nonhomogeneous form of this curve is

$$Y^2 = \frac{(X-1)^2}{X}$$

Serpentine

$$3y^2 w = x^3 - 3xw^2 - 2w^3$$
$$H = -1$$
$$K = -2$$

Hessian is

$$\mathbf{H}(C) = x^2 w + 2xw^2 + w^3 - y^2 x$$

Nonhomogeneous form of this curve is

$$Y^2 = \frac{(X+1)^2}{X}$$

H=0 (mutually intersecting tangents at inflections)

This is a part of the continuum of wiggles

$$3y^2 w = x^3 + w^3$$
$$\mathbf{H}(C) = -Kxw^2 - y^2 x$$
$$= -x\left(Kw^2 + y^2\right)$$

This is the y axis (x=0) times the point at infinity on the x axis (1,0,0). In other words, this particular member of the continuum has a degenerate Hessian.

Inflection Points

Inflection points happen at intersections of the original curve with its Hessian curve. So we require both

$$0 = x^3 + 3Hxw^2 + Kw^3 - 3y^2 w$$
$$0 = -Hx^2 w - Kxw^2 + w^3 H^2 - y^2 x$$

(One point that satisfies this is [x y w]=[0 1 0].)
Solve both these equations for y

$$y^2 = \frac{x^3 + 3Hxw^2 + Kw^3}{3w}$$
$$y^2 = \frac{-Hx^2w - Kxw^2 + w^3H^2}{x}$$

equate

$$\frac{x^3 + 3Hxw^2 + Kw^3}{3w} = \frac{-Hx^2w - Kxw^2 + w^3H^2}{x}$$
$$x^4 + 3Hx^2w^2 + Kxw^3 = -3Hx^2w^2 - 3Kxw^3 + 3w^4H^2$$
$$x^4 + 3Hx^2w^2 + 3Hx^2w^2 + Kxw^3 + 3Kxw^3 - 3w^4H^2 = 0$$
$$x^4 + 6Hx^2w^2 + 4Kxw^3 - 3H^2w^4 = 0$$
$$X^4 + 6HX^2 + 4KX - 3H^2 = 0$$
$$\left(X^3 + 6HX + 4K\right)X - 3H^2 = 0$$

Solve this to find X of inflection points

Special cases

H=0

$$X^4 + 4KX =$$
$$X\left(X^3 + 4K\right) = 0$$
$$X = 0, \sqrt[3]{-4K}$$

(H,K) = (0,1/4) X=0,-1
(H,K) = (0,-1/4) X=0,+1

K=0

$$X^4 + 6HX^2 - 3H^2 = 0$$

$$X^2 = \frac{-6H \pm \sqrt{36H^2 - 4*(-3)H^2}}{2}$$
$$= H\frac{-6 \pm \sqrt{48}}{2}$$
$$= H\left(-3 \pm 2\sqrt{3}\right)$$

Linear combo of curve with its Hessian

One strategy for finding inflection points is to note that, since they are on both the curve and its Hessian, they are also on any linear combination of the two. Maybe the proper linear combination can generate something that is more easily solved. To investigate what happens here, let’s assume that we are in stand ard position and look at the linear combination:

$$\alpha\left(x^3+3Hxw^2+Kw^3-3y^2w\right)+\beta\left(-Hx^2w-Kxw^2+w^3H^2-y^2x\right)=$$
$$\alpha x^3+3\alpha Hxw^2+\alpha Kw^3-3\alpha y^2w-\beta Hx^2w-\beta Kxw^2+\beta w^3H^2-\beta y^2x=$$
$$(\alpha)x^3+(3\alpha H-\beta K)xw^2+\left(\alpha K+\beta H^2\right)w^3-(3\alpha)y^2w-(\beta H)x^2w-(\beta)y^2x=$$

Two interesting choices for alpha,beta

$$\alpha=K$$
$$\beta=3H$$

$$(\alpha)x^3+(3\alpha H-\beta K)xw^2+\left(\alpha K+\beta H^2\right)w^3-(3\alpha)y^2w-(\beta H)x^2w-(\beta)y^2x=$$
$$(K)x^3+(0)xw^2+\left(K^2+3H^3\right)w^3-(3K)y^2w-\left(3H^2\right)x^2w-(3H)y^2x=$$
$$(K)x^3+\left(K^2\right)w^3-(3K)y^2w+\left(3H^3\right)w^3-\left(3H^2\right)x^2w-(3H)y^2x=$$

If we are one of the paraetrizable curves (loop, cusp, serpentine) we have

$$K^2+3H^3=0$$
$$(K)x^3+(0)w^3-(3K)y^2w-\left(3H^2\right)x^2w-(3H)y^2x=$$
$$(K)x^3-(3K)y^2w-\left(3H^2\right)x^2w-(3H)y^2x=$$
$$\left((K)x^2-(3H)y^2\right)x-\left((3K)y^2+\left(3H^2\right)x^2\right)w=$$

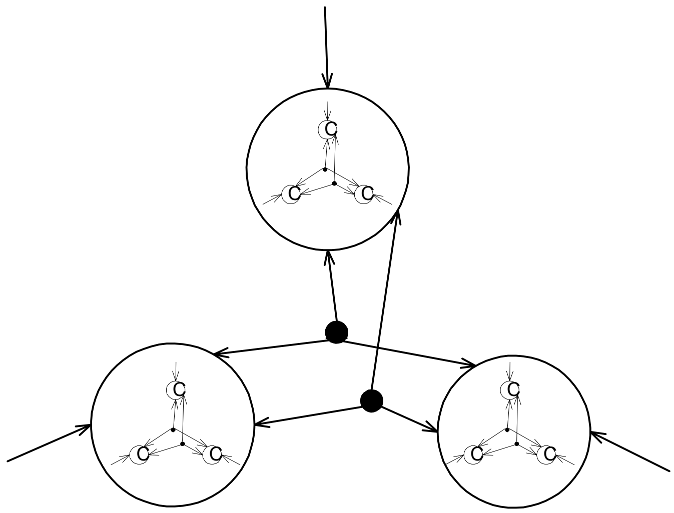
Other choice

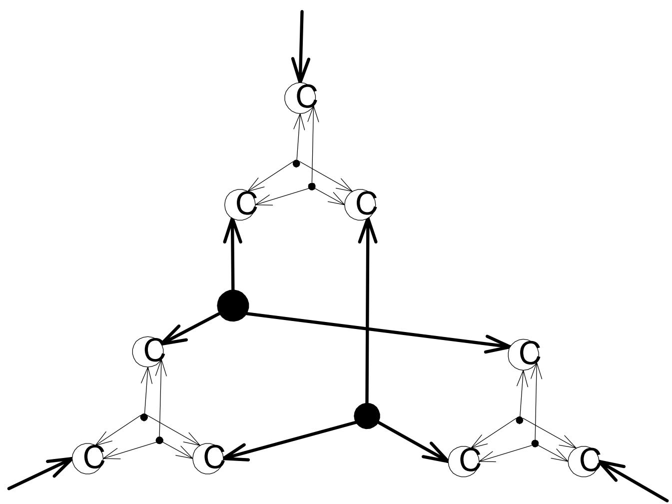
$$\alpha=H^2$$
$$\beta=-K$$
$$(\alpha)x^3+(3\alpha H-\beta K)xw^2+\left(\alpha K+\beta H^2\right)w^3-(3\alpha)y^2w-(\beta H)x^2w-(\beta)y^2x=$$
$$\left(H^2\right)x^3+\left(3H^2H+KK\right)xw^2+(0)w^3-\left(3H^2\right)y^2w-(-KH)x^2w-(-K)y^2x=$$
$$\left(H^2\right)x^3+\left(3H^3+K^2\right)xw^2-\left(3H^2\right)y^2w+(KH)x^2w+(K)y^2x=$$

needs work…

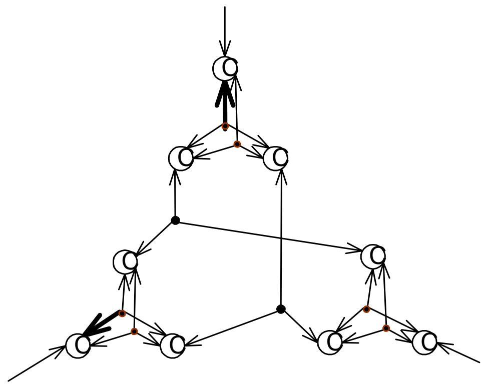
What is the Hessian of the Hessian?

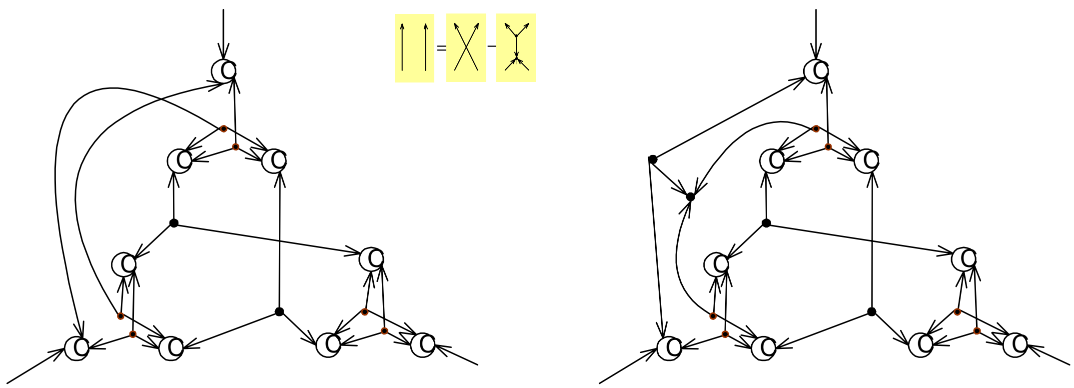
Diagram



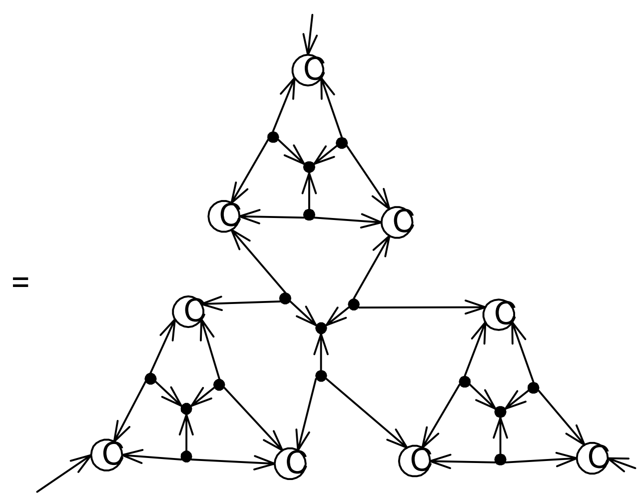


Try to simplify with $\varepsilon\delta$





Yikes.



Needs work…

Punchline

The easiest way to treat the Hessian that I have found so far is to consider the original curve in the lambda canonical form.

$$f(x, y, w) = x^3 + y^3 + w^3 - \lambda xyw = 0$$

This form can only express cubic curves that are irreducible (not the product of lower order curves) and that do not have singularities (cusp, loop or single isolated point). These latter three are the only cubics that can be generated parametrically.

The first derivatives of the function are:

$$\begin{bmatrix} f_x \\ f_y \\ f_w \end{bmatrix} = \begin{bmatrix} 3x^2 - \lambda yw \\ 3y^2 - \lambda xw \\ 3w^2 - \lambda xy \end{bmatrix}$$

and the second derivatives are:

$$\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} = \begin{bmatrix} 6x & -\lambda w & -\lambda y \\ -\lambda w & 6y & -\lambda x \\ -\lambda y & -\lambda x & 6w \end{bmatrix}$$

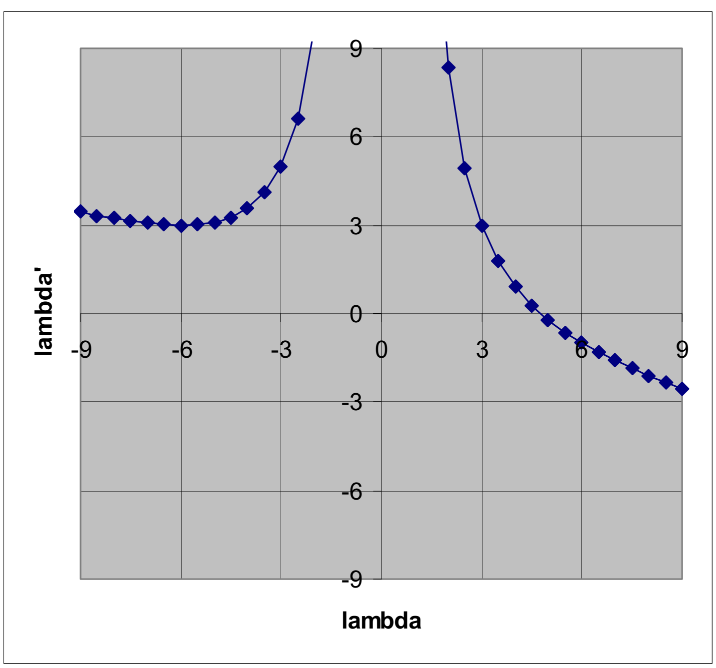
The determinant of this gives the Hessian curve

$$H(x,y,w) = 6^3 xyw - 2\lambda^3 xyw - 6\lambda^2 y^3 - 6\lambda^2 x^3 - 6\lambda^2 w^3$$
$$= -6\lambda^2 \left(x^2 + y^2 + w^2 + \left(\frac{6^3 - 2\lambda^3}{-6\lambda^2} \right) xyw \right)$$

In other words, the Hessian of such a curve with parameter λ is another curve that is still in our canonical form but which has a different lamba;

$$\lambda' = -\frac{6^3 - 2\lambda^3}{-6\lambda^2}$$

This function looks like



Iteration on this function generates a fractal sequence except for the special cases of lambda=3, 0, and infinity.

lamda=3

This is the degenerate case

$$\begin{aligned} f(x,y,w) &= x^3+y^3+w^3-3xyw \\ &= (x+y+w)\left(x^2+y^2+w^2-xy-xw-yw\right)=0 \end{aligned}$$

This is a line and the isolated point (1,1,1). The Hessian of this curve is the same curve.

lamda=0

This is the curve

$$f(x,y,w)=x^3+y^3+w^3=0$$

A perfectly good irreducible cubic curve. The Hessian is

$$H(x,y,w)=6^3xyw$$

This is a reducible product of three lines.

lamda=-6

This is the curve

$$f(x,y,w)=x^3+y^3+w^3+6xyw=0$$

This is another of the continuum of irreducible cubic curves. It has the special property that it is the single curve where the tangents to all three inflection points meet at one point. Calculating the lambda parameter of its Hessian gives:

$$\lambda' = -\frac{6^3-2\lambda^3}{-6\lambda^2}=3$$

lamda=infinity

This is effectively the curve

$$f(x,y,w)=xyw=0$$

we go back to first principles and calculate

$$\begin{bmatrix} f_x \\ f_y \\ f_w \end{bmatrix} = \begin{bmatrix} yw \\ xw \\ xy \end{bmatrix}$$

$$\begin{bmatrix} f_{xx} & f_{xy} & f_{xw} \\ f_{xy} & f_{yy} & f_{yw} \\ f_{xw} & f_{yw} & f_{ww} \end{bmatrix} = \begin{bmatrix} 0 & w & y \\ w & 0 & x \\ y & x & 0 \end{bmatrix}$$

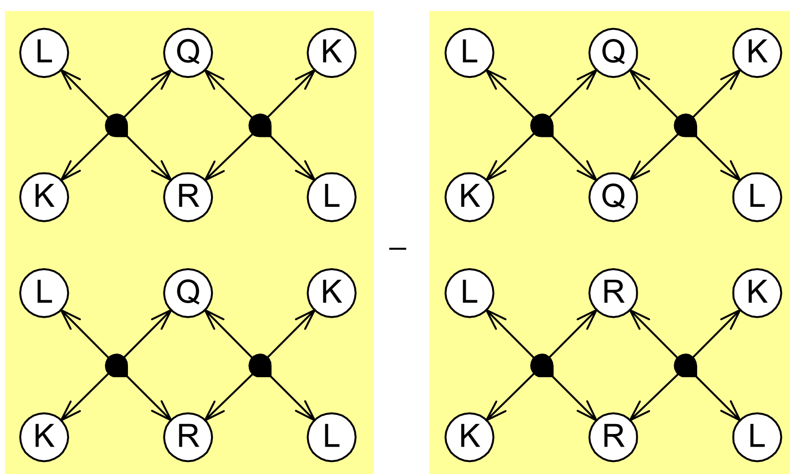
The determinant of this gives us back our original form:

$$\mathbf{H}(xyw)=2xyw$$

Question

Can we do the above with Tensor Diagrams?

PART 3
3DH (4D)



The resultant of two planes and two quadratic surfaces

Chapter 3-00

3DH(4D)
Diagram Identity Catalog

Epsilon

$\varepsilon_{ijkl} = 1$ if *ijkl* is an even permutation of 1234

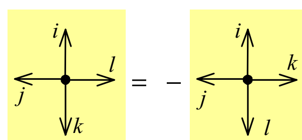
$\varepsilon_{ijkl} = -1$ if *ijkl* is an odd permutation of 1234

$\varepsilon_{ijkl} = 0$ otherwise

In contrast to 3D, in 4D a cyclic permutation *does* change the sign, that is

$$\varepsilon^{ijkl} = -\varepsilon^{jkli}$$

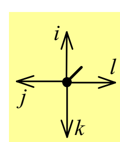
In 4D diagram notation the 4D epsilon is simply be a four-pronged node. An odd permutation of indices for the 4D epsilon is not so geometrically obvious if the diagram form is simply a node with four lines, as in



ε_{ijkl} $\qquad$ $\varepsilon_{ijlk} = -\varepsilon_{ijkl}$

To keep things straight we need to make the epsilon node shape asymmetrical either by adding a tick mark

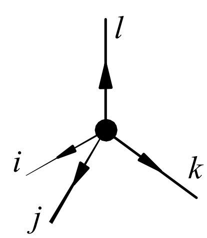
ε_{ijkl}



ε_{ijkl}

or by making it in 3D

ε_{ijkl}



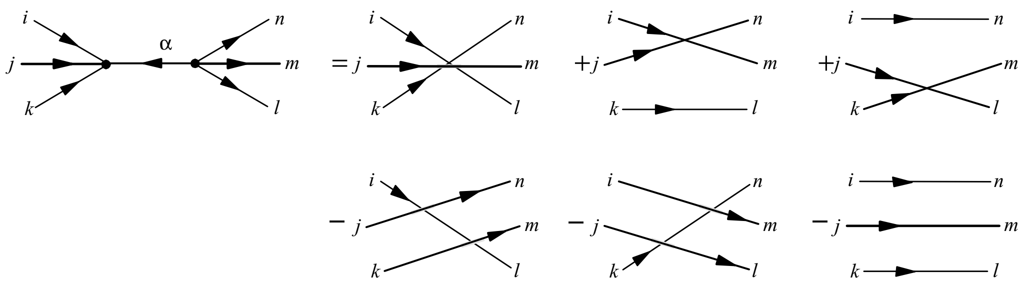
or something…

Epsilon Delta Identity

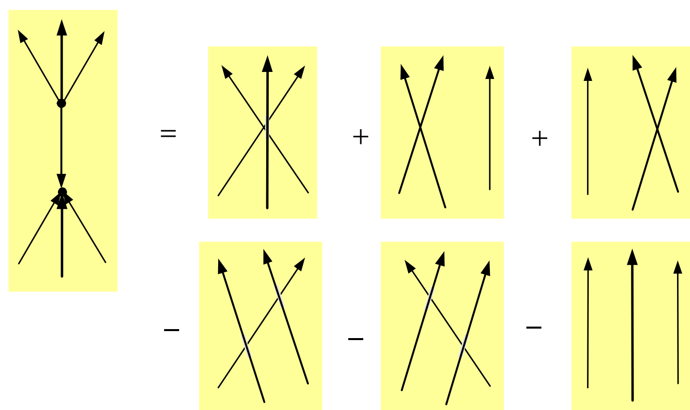
The 4D epsilon delta rule in 4D is

$$\varepsilon_{\alpha ijk}\varepsilon^{\alpha lmn} = \delta_i^l\delta_j^m\delta_k^n + \delta_i^m\delta_j^n\delta_k^l + \delta_i^n\delta_j^l\delta_k^m$$
$$-\delta_i^l\delta_j^n\delta_k^m - \delta_i^m\delta_j^l\delta_k^n - \delta_i^n\delta_j^m\delta_k^l$$

The diagram looks like.



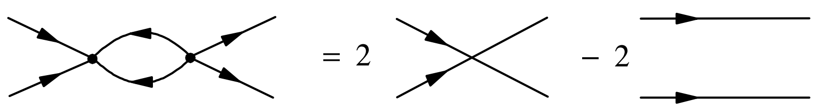
Or vertical orientation



Variant 1

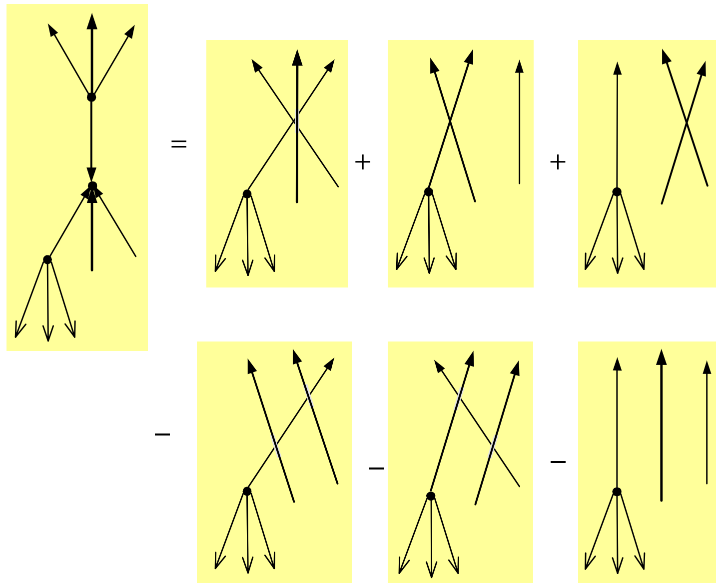
Summing over one index

$$\varepsilon_{\alpha\beta jk}\varepsilon^{\alpha\beta mn} = 2\left(\delta_j^m\delta_k^n - \delta_j^n\delta_k^m\right)$$



Variant 2

Plugging one, two or three epsilons onto one branch:



note arrows are all supposed to have same shaped heads (fix later)

Etc like we did in 2DH case…

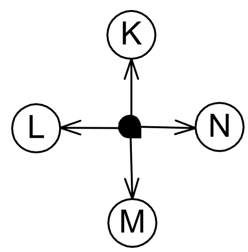
Chapter 3-01

3DH(4D)
Resultants of Linear and Quadratic Tensors

In each case setting the diagram equal to zero gives the condition that there is a point in common to the all tensors involved (generally there will be three). If there are only two, we are talking about a tangency relation of some sort.

4 Linears(planes)

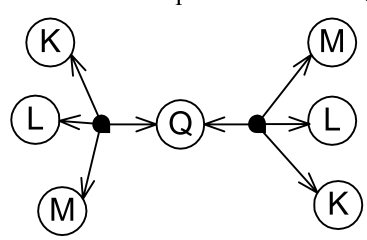
This is just the condition that four planes have a common point



3 Linears(planes) and…

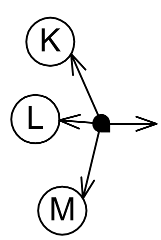
1 Quadratic

Condition that intersection of 3 planes KLM lies on quadratic surface Q



Any higher order

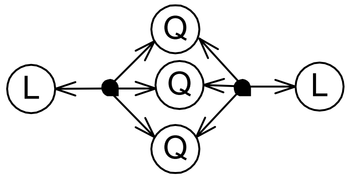
Just plug as many copies as necessary of the following



into each input of the higher order node.

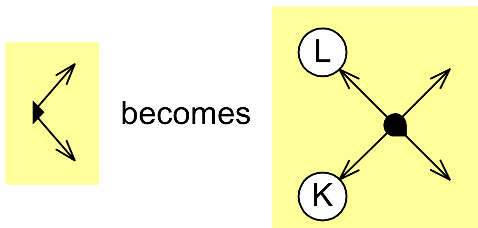
1 Plane, 1 Quadratic

This is the tangency condition. The intersection must be a double root. Note only two tensors involve



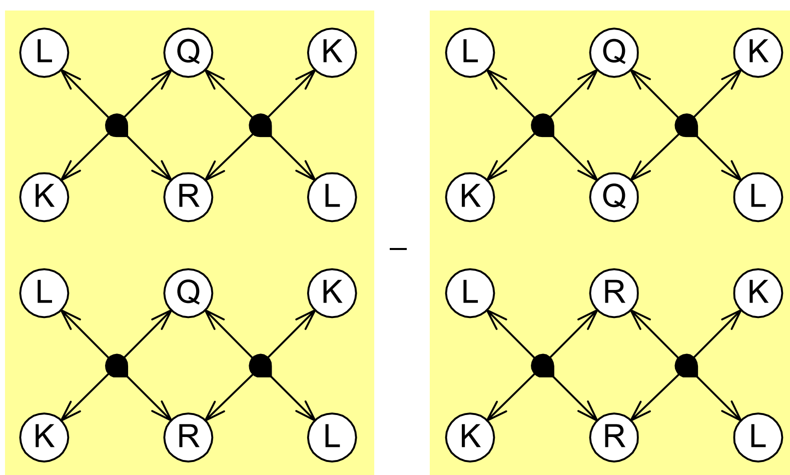
2 Linears(planes) and…

Line formed by intersection of two planes K,L passes through intersection of two higher order surfaces. That this will be analog of resultant of two 1DH quadratic polynomials but replacing



2 Quadratics

The line intersects the intersection of quadratics Q and R



and any combo of 2 higher orders

Again, by analogy, take the 1DH resultant diagram of the higher order curves and make the same epsilon replacement above:

1 Linear and…

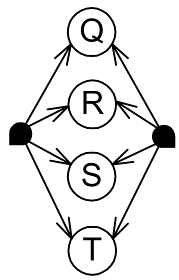
Any 3 higher orders

Take the diagram for the resultant of the higher order curves in 2DH and plug a copy of the linear into the extra socket on the epsilon created when you change from 2DH to 3DH epsilons

4 Quadratics

Possible diagrams that are symmetric in QRST implies: equal numbers of each

1 each



etc for diagrams with higher numbers of QRST nodes…

Summary

Benefits of Tensor Diagrams

Tensor diagrams can represent complex polynomial expressions in a way that we can see their internal structure. We can operate on small parts of this expression (as with applications of epsilon/delta) without getting overburdened with the whole rest of the diagram. Conventional notation does not give us this ability.

Limitations of Tensor Diagrams

I promised a discussion of the limitations of tensor diagrams. Here are some thoughts.

The main limitation seems to be the combinatorial explosion of terms in some applications. When tensor diagrams are applied to high degree polynomials the nodes have more and more arcs radiating from them. When they are applied to higher dimensionality objects the epsilon node has more and more arcs. Going much past 4 in either axis gets rather messy.

Some of the derivations shown here seem a bit brute force. This may be a problem for generalizations, but I think that a lot of this is just scaffolding. When we all understand diagram manipulation better many of these problems can be solved directly with diagrams without also having to check our work with conventional notation.

References

M. Paluszny and R Patterson, *A Family of Tangent Continuous Cubic Algebraic Splines*, ACM Transactions on Graphics, Vol. 12, No. 3, July 1993, page 212

G. Salmon, *A Treatise on the Higher Plane Curves*, Published by G. E. Stechert & Co., New York, 1934, A Photographic reprint of the Third Edition of 1879, pages 191, 192 and 199

G. Salmon, *Modern Higher Algebra*, Chelsea Publishing Company. (reprinted from the 1885 original)

Newman, W. and Sproull, R. *Principles of Interactive Computer Graphics*, McGraw-Hill, 1973, pp. 467.

Hodge, W.V.D. and Pedoe, D. *Methods of Algebraic Geometry*, Cambridge University Press, 1968, pp. 286.1990

Kuperberg, Greg, *Involutory Hopf algebras and 3-manifold invariants* International Journal of Mathematics, Vol 2, no. 1, (1991) pp 41-66, World Scientific Publishing Company

Stedman, Geoffrey E., *Diagram Techniques in Group Theory*, Cambridge University Press, United Kingdom, 1990.

A. Kurosh, *Higher Algebra*, MIR publishers, Moscow, 1975.

T. W. Sederberg, D. C. Anderson and R. N. Goldman, *Implicit representation of parametric curves and surfaces*, CVGIP 28, 72-84, 1984.

J. Kajiya, *Ray Tracing Parametric Patches*, *Proc.* Siggraph *82*, ACM Press, New York, 1982, p. 248.

Wee, C. and Goldman, R, *Elimination and Resultants Part 1*, IEEE CG&A January 1995