

Notes on Steiner's ellipse and related matters

Patrick D. F. Ion

20 October 2005

1 Basic Facts

Any three non-coincident points in the plane have a unique ellipse passing through them such that the center of gravity of the ellipse is the same as the center of gravity of the initial three points. This ellipse is called Steiner's ellipse, and was studied in classical plane geometry.

With a classical mindset as to geometry one can see that there is such an ellipse: For three points at the vertices of an equilateral triangle there's clearly a circle passing through them that has the property that the center is the centroid of the circle and of the triangle. Any three distinct points are related by an affine transformation to the standard equilateral triangle configuration. The image under the affine transformation of the circle is the ellipse sought.

There is an approach to this construction and to the geometry of triangles and ellipses that uses the notion of the discrete Fourier transform and complex coordinates in an interesting way. It also points to generalizations involving more points and can be useful in discussing the orbits of celestial mechanics.

Suppose one starts with three points $z_0, z_1, z_2 \in \mathbb{C}$. Consider them as a triple $z = (z_0, z_1, z_2)$, and think of that as a complex-valued function on the space of three points $\{0, 1, 2\}$, so z is an element of \mathbb{C}^3 . Go further and decide that the three-point space is actually the group $\mathbb{Z}/_3 := \mathbb{Z}/3\mathbb{Z}$ of integers modulo 3. Then we are looking at a function in the complex group algebra, $z \in \mathbb{C}\mathbb{Z}/_3$. But we could also consider our z as a function in the space of square summable functions with respect to a mass distribution on $\mathbb{Z}/_3$, say the uniform distribution $m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$; so we have $z \in L^2(\mathbb{Z}/_3, m)$.

The jump motion from point to point, z_0 to z_1 to z_2 to z_0 , is what we want to interpolate with a motion $z(t)$ smoothly adjusting all three points to get to the next one. We shall assume, without essential loss of generality that the triangle z_0, z_1, z_2 is positively oriented so that the motion we construct will be too; otherwise we can just reverse the time parameter.

The way to our goal is harmonic analysis on $\mathbb{Z}/_3$. Let ω and ω^2 be the two imaginary roots of unity in \mathbb{C} , with the first taken as that in the upper half plane; in fact, $\bar{\omega} = \omega^2$.

That means we have the identities $\omega^3 = 1$ and $1 + \omega + \omega^2 = 0$.

There are three basic one-dimensional representations of the cyclic group $\mathbb{Z}/3$ given by the characters

$$\chi_0 = (1, 1, 1); \quad \chi_1 = (1, \omega, \omega^2); \quad \chi_2 = (1, \omega^2, \omega).$$

These character functions are in fact orthogonal, and unit vectors, in the natural Hilbert space $L^2(\mathbb{Z}/3, m)$ whose sesquilinear inner product is given by

$$\langle u, v \rangle := \sum_{g \in \mathbb{Z}/3} \bar{u}(g)v(g)m(g) = \frac{1}{3}(\bar{u}_0v_0 + \bar{u}_1v_1 + \bar{u}_2v_2).$$

For instance, $\langle \chi_1, \chi_2 \rangle = 3^{-1}(1 \times 1 + \omega^2 \times \omega^2 + \omega \times \omega) = 3^{-1}(1 + \omega + \omega^2) = 0$; and so on. So we have an orthonormal basis for our Hilbert space; thus any z can be expanded in terms of it:

$$z = \langle \chi_0, z \rangle \chi_0 + \langle \chi_1, z \rangle \chi_1 + \langle \chi_2, z \rangle \chi_2 =: f_0 \chi_0 + f_1 \chi_1 + f_2 \chi_2.$$

We have hence a Fourier transform map, $z \mapsto f$, which in coordinate terms amounts to the equations

$$3f_0 = z_0 + z_1 + z_2; \quad 3f_1 = z_0 + \omega^2 z_1 + \omega z_2; \quad 3f_2 = z_0 + \omega z_1 + \omega^2 z_2,$$

and the inverse operation, given by

$$z_0 = f_0 + f_1 + f_2; \quad z_1 = f_0 + \omega f_1 + \omega^2 f_2; \quad z_2 = f_0 + \omega^2 f_1 + \omega f_2.$$

We also see that the coefficient f_0 of the trivial character χ_0 represents the center of gravity of the cluster of the three starting points.

Let us return to constructing the claimed ellipse interpolating the points of the triple z . The cyclic permutation $(0, 1, 2) \mapsto (1, 2, 0)$, the generator of the group $\mathbb{Z}/3$, extended to act on functions, maps $(z_0, z_1, z_2) \mapsto (z_1, z_2, z_0)$, thus moving the points one around the putative curve. If we write our coordinates out as column vectors (so as to end up with matrix actions on the left, which is probably more common) then we can write

$$\begin{pmatrix} z_1 \\ z_2 \\ z_0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ z_2 \end{pmatrix},$$

or (using Matlab notation in which ‘,’ is a horizontal spacer and ‘;’ is a vertical spacer) the result is Jz with $z := (z_0; z_1; z_2)$ and $J := (0, 1, 0; 0, 0, 1; 1, 0, 0)$. What we need to get smoothly from z_0 to z_1 could then be just the curve $[0, 1] \ni t \mapsto J^t z$. But this we can define neatly.

First note $J^3 = 1$, so that J 's eigenvalue list is $(1, \omega, \omega^2)$. Then note that we know from above the corresponding eigenvectors to be $(\chi_0^\top, \chi_1^\top, \chi_2^\top)$ ($^\top$ is the notation for the

transpose); for instance, $J\chi_1^\top = (\omega; \omega^2; 1) = \omega(1; \omega; \omega^2)$. Thus we can see an explicit diagonalization of J using $(\chi_0; \chi_1; \chi_2)$

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega^2 & \omega \\ 1 & \omega & \omega^2 \end{pmatrix},$$

or with $U^* := \frac{1}{\sqrt{3}}(\chi_0; \chi_1; \chi_2)$ we have $UJU^* = (1: \omega: \omega^2)$, where we have used ‘:’ as a diagonal separator to supplement our horizontal and vertical ones, and * denotes Hermitian conjugation. We note U is unitary: $U^*U = 1$. But this means we can readily evolve J , writing

$$J^t = U^*(1: \omega^t: \omega^{2t})U = U^*(1: \omega^t: \bar{\omega}^t)U.$$

And we can recognize U^* as the (normalized) matrix of the discrete Fourier transform above, since it clear that

$$\frac{1}{\sqrt{3}}U^*(z_0; z_1; z_2) = (f_0; f_1; f_2).$$

We could have written the Fourier transform equations in matrix form before as

$$3 \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega^2 & \omega \\ 1 & \omega & \omega^2 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ z_2 \end{pmatrix},$$

and

$$3 \begin{pmatrix} z_0 \\ z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}.$$

If we single out the presence of the 3×3 Vandermonde matrix $V = (\chi_0; \chi_1; \chi_2)$ in these formulas, we can write them as

$$3z = V^*f \quad \text{and} \quad 3f = Vz,$$

and we have $U = \frac{1}{\sqrt{3}}V$.

Hence, if we write our interpolating curve as

$$[0, 1] \ni t \mapsto z(t) =: (z_0(t), z_1(t), z_2(t)) \in \mathbb{C}^3.$$

we have

$$z(t) = U^*(1: \omega^t: \bar{\omega}^t)Uz = U^*(1: \omega^t: \bar{\omega}^t)Uz = V^*(1: \omega^t: \omega^{-t})f.$$

In other words, to find the interpolating curve:

- Do the discrete Fourier transform.
- Leave the center of gravity alone.

- Evolve the two non-trivial coefficients rotationally in opposite directions with a rotation of $2\pi t/3$.
- Fourier transform back to get the new configuration.

If we consider $z(t)$ as evolving over the whole of $t \in [0, 3]$ then we can note that

$$z_0(3) = z_1(2) = z_2(1) = z_0(0),$$

and other such periodicities. We also see that the center of gravity of the complete curve has to be the original f_0 , since the curve's points come in sets of 3 balanced about that point.

To verify that the curve is an ellipse it makes sense to do some preliminary simplification. Note that the transformations $t \mapsto z(t)$ are linear in z for fixed t , so that we can transform the curve back to one whose center of gravity is the origin without changing its algebraic type. Therefore let us assume that $f_0 = 0$. Then we must look at

$$z_0(t) = \frac{1}{3}(0 + \omega^t f_1 + \omega^{-t} f_2).$$

Suppose, in polar coordinates (ρ, θ) , that $f_1 = \rho_1 \mathbf{e}^{\theta_1}$ and $f_2 = \rho_2 \mathbf{e}^{\theta_2}$, and write $\omega = \mathbf{e}^{i\varpi}$, i.e., $\varpi = 2\pi/3$. Then

$$\begin{aligned} z_0(t) &= \rho_1 \mathbf{e}^{i(\varpi t + \theta_1)} + \rho_2 \mathbf{e}^{i(-\varpi t + \theta_2)} \\ &= \mathbf{e}^{i(\theta_1 + \theta_2)/2} [\rho_1 \mathbf{e}^{i(\varpi t + (\theta_1 - \theta_2)/2)} + \rho_2 \mathbf{e}^{-i(\varpi t + (\theta_1 - \theta_2)/2)}] \\ &= \mathbf{e}^{i(\theta_1 + \theta_2)/2} [(\rho_1 + \rho_2) \cos(\varpi t + (\theta_1 - \theta_2)/2) + i(\rho_1 - \rho_2) \sin(\varpi t + (\theta_1 - \theta_2)/2)]. \end{aligned}$$

The first term in the last line is a pure rotation about the origin, which is the center of mass by assumption, and the second is the standard expression for a point on an ellipse with semi-major axis $(\rho_1 + \rho_2)$ and semi-minor axis $(\rho_1 - \rho_2)$. We have thus proved by explicit construction the existence of the Steiner ellipse for any three distinct points in the Euclidean plane, doing it by simple use of complex variables.

Conversely, given a non-degenerate ellipse in the plane we can find a triple of points whose Steiner ellipse it will be, but this the triple is not unique, of course. We can use these Fourier transform coordinates to give coordinates to the space of plane ellipses.

The easiest way to see this seems to be to consider a standard sort of ellipse. Any ellipse in the plane can be brought to have center at the origin by a simple translation that will only be reflected in the value of f_0 . Any ellipse centered at the origin can be turned so that the semi-major axis lies along the x -axis; this rotation is only reflected in corresponding rotations of f_1 and f_2 .

So let us assume a standard ellipse whose major axis extends from $-a$ to a , with a real, and whose minor axis runs from $-b\mathbf{i}$ to $b\mathbf{i}$, with b real; clearly, a and b may not be zero. The standard equation of our ellipse will thus be

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Consider the following three points: the right-hand end of the major axis, $a \in \mathbb{C}$, and the upper and lower ends of the chord of the ellipse perpendicular to the major axis at the point half way to the left from the center, $-\frac{a}{2}$, which are $\frac{1}{2}(-a + \sqrt{3}b\mathbf{i})$ and $\frac{1}{2}(-a - \sqrt{3}b\mathbf{i})$, respectively, as one may verify from the equation of the ellipse. From the triple

$$z = (a, \frac{1}{2}(-a + \sqrt{3}b\mathbf{i}), \frac{1}{2}(-a - \sqrt{3}b\mathbf{i}))$$

we form the discrete Fourier transform

$$f = \frac{1}{3}(0, a + \omega^2(\frac{1}{2}(-a + \sqrt{3}b\mathbf{i})) + \omega\frac{1}{2}(-a - \sqrt{3}b\mathbf{i}), a + \omega(\frac{1}{2}(-a + \sqrt{3}b\mathbf{i})) + \omega^2\frac{1}{2}(-a - \sqrt{3}b\mathbf{i})).$$

Consider the term

$$\begin{aligned} f_1 &= \frac{1}{3}[a + \omega^2(\frac{1}{2}(-a + \sqrt{3}b\mathbf{i})) + \omega(\frac{1}{2}(-a - \sqrt{3}b\mathbf{i}))] \\ &= \frac{1}{3}[\frac{3}{2}a + \mathbf{i}\frac{\sqrt{3}}{2}b(\omega^2 - \omega)] \\ &= \frac{1}{3}[\frac{3}{2}a - \mathbf{i}\frac{\sqrt{3}}{2}b(2\omega + 1)] \\ &= \frac{1}{2}[a + b] \end{aligned}$$

since $(2\omega + 1)^2 = 4\omega^2 + 4\omega + 1 = -3$. Similarly, since only a change of sign of the b term results from a switch of the powers of ω ,

$$f_2 = \frac{1}{2}(a - b).$$

So the semi-major axis is $a = f_1 + f_2$, and the semi-minor axis is $b = f_1 - f_2$.

A rotation of the axis of the ellipse by $\mathbf{e}^{i\theta}$ would result in a similar rotation of f_1 with respect to f_2 . In fact f_1 would be displaced positively to $\mathbf{e}^{i\theta/2}f_1$ and f_2 negatively to $\mathbf{e}^{-i\theta/2}f_2$.

If we consider the standard ellipse as sitting within a standard circle of radius a and the same center, then we see that the three points we started with are the intersections of the lines through the three cube roots of unity with the ellipse.

If we consider rotating the three cubic root lines by $\mathbf{e}^{i\psi}$, say, and taking the intersections with the ellipse as the starting points to reckon the Steiner evolution, then we are in fact back at the situation analyzed above.

Remark that the discrete Fourier transform, which does map triples of points to other triples, is far from volume-preserving on the triangles the points subtend. It happily maps completely non-degenerate triangles to linear ones.

Finally remark that, with hindsight, we might have guessed the Fourier transform form of the ellipse fairly directly. The parametric expression for a point on the standard ellipse is $z(t) = a \cos t + \mathbf{i}b \sin t$. Expanding this we have $z(t) = \frac{1}{2}[a(\mathbf{e}^{it} + \mathbf{e}^{-it}) + \mathbf{i}b(-\mathbf{i})(\mathbf{e}^{it} - \mathbf{e}^{-it})] = (a + b)\mathbf{e}^{it} + (a - b)\mathbf{e}^{-it}$, which is supposed to be the evolution path of the point $z(0)$. So it looks as though we might have been able to read off that $f_1 = a + b$ and $f_2 = a - b$, since each is, respectively, what gets multiplied by \mathbf{e}^{it} and \mathbf{e}^{-it} .

2 More points than 3; higher cyclic groups

One can carry through most of the constructions given above in the case of three points in the plane for any number N of points, but the special properties of ellipses do not have clear generalizations at the moment. What does remain true is that there is a natural global interpolating curve that passes through a list of N given points in the plane, that it is smooth, and that it has the same center of gravity as the original cluster of points. The adjective global is suggestive of the fact that there is a clear sense in which the position of every initial point has generally an effect on the form of the curve as it passes between any two of the given points. A subtlety does arise, however, depending on the divisibility properties of N .

So consider a list of N points in \mathbb{C} , $z = (z_0, \dots, z_{N-1}) = (z_n, n \in 0..N-1) \in \mathbb{C}^N$.

Here we shall use the list notation $(x_n, n \in B)$ to mean a list of objects x_n indexed by values of n lying in an ordered set B ; of course, this is just a notation for a particular function x on the domain B . We shall think of the lists as row vectors when the separator is a comma. Similarly then $(x_n; n \in B)$ will mean a vertical or column list, and $(x_n : n \in B)$ a list running down the usual diagonal descending to the right. The discrete interval notation $i..j$ denotes the range of integers running from i to j , inclusive of the end values.

Again consider $z \in L^2(\mathbb{Z}/N, m)$ with constant weights $m(n) = N^{-1}$. The inner product in this Hilbert space is again

$$\langle u, v \rangle := \sum_{g \in \mathbb{Z}/N} \bar{u}(g)v(g)m(g) = \frac{1}{N}(\bar{u}_0v_0 + \bar{u}_1v_1 + \dots + \bar{u}_{N-1}v_{N-1}).$$

Let $\omega = \omega_N = e^{i2\pi/N}$ be N th root of unity in the upper half-plane with the smallest argument. We have $\omega^N = 1$ and

$$1 + \omega + \omega^2 + \dots + \omega^{N-1} = \sum(\omega^k, k \in 0..N-1) = 0,$$

where summing over a finite list means adding all the elements together, with the understanding that an empty sum is 0. Similarly a product \prod over a finite list will mean complete multiplication, but an empty product gives 1.

The additive character list of the abelian cyclic group \mathbb{Z}/N is $(\chi_n, n \in 0..N-1)$ where

$$\chi_n = (\omega_N^{nk}, k \in 0..N-1).$$

Proceeding as before, we consider the inner products for $j, k \in 0..N-1$

$$\begin{aligned} \langle \chi_j, \chi_k \rangle &= N^{-1} \sum(\bar{\omega}^j \omega^{kl}, l \in 0..N-1) \\ &= N^{-1} \sum(\omega^{(N-j+k)l}, l \in 0..N-1) \\ &= \langle j = k \rangle \end{aligned}$$

where the logical characteristic function ($j = k$) is 1 if $j = k$ and 0 if $j \neq k$. This is because each character of a finite cyclic group corresponds to a one-dimensional representation and so is absolutely irreducible [M. Hall, *Theory of Groups*, Macmillan, New York, 1959, p. 271].

As before then we may decompose any list of points z , which subtend an N -gon, into harmonic components according to the characters and form a discrete Fourier transform

$$\mathbb{C}^N \ni (z_n, n \in 0..N-1) = z \mapsto f = (\langle \chi_n, z \rangle, n \in 0..N-1) \in \mathbb{C}^N,$$

implemented by a multiple of the Vandermonde matrix

$$V = (\chi_n; n \in 0..N-1) = (\omega_N^{jk}, j \in 0..N-1; k \in 0..N-1),$$

$$f = N^{-1}Vz.$$

We then can have an interpolating curve obtained as before

$$z(t) = V^*(\omega_N^{kt} : k \in 0..N-1)f.$$

The center of gravity remains fixed for the same reason as previously. We shall call the interpolating curve the Steiner interpolant of the N initial points.

We remark that the curve

$$\mathbb{R} \supset [0, N] \ni t \mapsto z_0(t) \in \mathbb{C}$$

is plainly a closed smooth curve passing through through all the points of $(z_n, n \in 0..N-1)$. Smoothness is clear since all the dependence on t is built up from powers like ω_N^{kt} , each of which is infinitely differentiable with respect to t .

Before discussing general properties of the Steiner interpolant, such as conditions of uniqueness, let us examine some other cases of small N .

2.1 $N = 2$

The two-element additive group $\mathbb{Z}/2 = \{0, 1\}$ has only one nontrivial character, namely the alternating character:

$$\chi_0 = (1, 1) \quad ; \quad \chi_1 = (1, -1),$$

so the Fourier transform of (a, b) is $\frac{1}{2}(a + b, a - b)$. The evolution on the interpolant is mediated by the power $(-1)^t = \mathbf{e}^{i\pi t}$. Note that we start the simplest example with the center of gravity at 0, so we should be considering the symmetric case $(a, b) = (-c, c)$ for some $c \in \mathbb{C}$. Then the Steiner interpolant will be the curve

$$[0, 2] \ni t \mapsto \mathbf{e}^{i\pi t} c \in \mathbb{C},$$

which is a circle that runs through c and $-c$. In this case there does not seem much more simple to say.

2.2 $N = 4$

The four-element additive group $\mathbb{Z}/4 = \{0, 1, 2, 3\}$ has four additive characters:

$$\chi_0 = (1, 1, 1, 1) \quad ; \quad \chi_1 = (1, \mathbf{i}, -1, -\mathbf{i}), \quad ; \quad \chi_2 = (1, -1, 1, -1), \quad ; \quad \chi_3 = (1, -\mathbf{i}, -1, \mathbf{i}).$$

The standard centered example is a quadrilateral (a, b, c, d) with $a + b + c + d = 0$. Then correspondingly

$$f_1 = \frac{1}{4}[(a - c) - \mathbf{i}(b - d)] \quad , \quad f_2 = \frac{1}{4}[(a - b) + (c - d)] \quad , \quad f_3 = \frac{1}{4}[(a - c) + \mathbf{i}(b - d)],$$

so

$$\begin{aligned} 4z_0(t) &= \mathbf{e}^{\mathbf{i}\pi t/2}[(a - c) - \mathbf{i}(b - d)] + \mathbf{e}^{\mathbf{i}\pi t}[(a - b) + (c - d)] + \mathbf{e}^{-\mathbf{i}\pi t/2}[(a - c) + \mathbf{i}(b - d)] \\ &= 2(a - c) \cos(\frac{1}{2}\pi t) + 2(b - d) \sin(\frac{1}{2}\pi t) + [(a - b) + (c - d)][\cos(\pi t) + \mathbf{i} \sin(\pi t)] \\ &= 2(a - c) \cos(\frac{1}{2}\pi t) + 2(b - d) \sin(\frac{1}{2}\pi t) + \\ &\quad [(a - b) + (c - d)][(\cos(\frac{1}{2}\pi t))^2 - (\sin(\frac{1}{2}\pi t))^2 + \mathbf{i}2 \sin(\frac{1}{2}\pi t) \cos(\frac{1}{2}\pi t)] \\ &= 2(a - c) \cos(\frac{1}{2}\pi t) + 2(b - d) \sin(\frac{1}{2}\pi t) + [(a - b) + (c - d)][2(\cos(\frac{1}{2}\pi t))^2 - 1] + \\ &\quad \mathbf{i}2[(a - b) + (c - d)] \sin(\frac{1}{2}\pi t) \cos(\frac{1}{2}\pi t) \end{aligned}$$