Linear-Fractional / Bilinear-Rational / Möbius Functions

The function $M(z) := (\mu z + ß)/(b - mz)$ is determined by the values of four constant parameters ß, μ, b and m chosen almost arbitrarily subject to an inequality constraint $ç := ßm + b\mu \neq 0$ that prevents $M(z)$ from degenerating to a constant function. Actually just three parameters suffice to determine M since dividing all four of ß, μ, b and m by the same nonzero constant (and dividing ç by its square) leaves M unchanged. Therefore nothing important is lost by assuming that one parameter has the value 1 ; but we shall do this only when it is convenient.

The equation $w = M(z)$ can always be solved for $z = W(w) := (bw - ß)/(\mu + mw)$; the two *Bilinear–Rational* functions M and W are inverse to each other in the sense that $M(W(w)) = w$ and $W(M(z)) = z$. Their parameters determine the same nonzero value ç , which appears in the *Bilinear* equation $(b - mz)(\mu + mw) = ç$. This is more symmetrical than equations $w = M(z)$ and $z = W(w)$ but equivalent to them only if $m \neq 0$. Differentiate it to find ç in the derivatives

$$M'(z) = ç/(b - mz)^2 \quad \text{and} \quad W'(w) = ç/(\mu + mw)^2 ,$$

which vanish only at ∞ if anywhere. Evidently no bilinear-rational function has a *Critical Point* (a finite place where its derivative vanishes); otherwise it could have no inverse. (Recall that an analytic function cannot be a one-to-one *bijection* in the neighborhood of its critical point.)

The foregoing expressions for their derivatives show that *Real* bilinear-rational functions (those whose parameters and argument are all real) are all monotonic except across their poles and thus fall into two classes: those that are increasing ($ç > 0$), and those that are decreasing ($ç < 0$) functions of their real argument. Nothing quite like that segregates complex bilinear-rational functions; each can be treated as a point on a connected 3-dimensional quadric hypersurface embedded in a complex 4-dimensional (ß, μ, b, m)-space by restricting $ç = ßm + b\mu = 1$, say. Restricting $ç = \pm 1$ treats real bilinear-rational functions as points on two 3-dimensional hyperboloids embedded in a real 4-dimensional (ß, μ, b, m)-space; the two hyperboloids have no finite intersection though they approach each other arbitrarily closely out towards infinity where lie functions nearly constant nearly everywhere.

Exercise 1: Show that all bilinear rational functions M satisfy the same differential equation $2M' \cdot M''' = 3(M'')^2$, all of whose other solutions are constants. Hint: Consider $(1/\sqrt{M'})''$.

Stereographic Projection and the Riemann Sphere

The natural domain for bilinear rational functions like $M(z) := (\mu z + ß)/(b - mz)$ and its inverse $W(w) := (bw - ß)/(\mu + mw)$ includes a single point at ∞ ; $M(\infty) = -\mu/m$, $W(-\mu/m) = \infty$, $W(\infty) = b/m$, and $M(b/m) = \infty$. Adjoining this point at ∞ to the complex plane $\mathbb{C}$ turns it topologically into a spherical surface $\mathbb{S}$ called the *Riemann Sphere*. Let us imagine $\mathbb{S}$ to be a sphere of radius 1 centered at the origin in a 3-dimensional Euclidean space of points $[w, \Omega]$ in which Ω is the altitude above or below a copy of the complex plane in which w runs. Then $|w|^2 + \Omega^2 = 1$ when $[w, \Omega]$ lies on $\mathbb{S}$. *Stereographic Projection* identifies $[w, \Omega]$ on $\mathbb{S}$ with a complex z in $\mathbb{C}$ by joining them with a straight line through $[0, 1]$ at the North pole of $\mathbb{S}$. This maps z in $\mathbb{C}$ to $S(z) := [2z/(|z|^2+1), (|z|^2-1)/(|z|^2+1)]$ on $\mathbb{S}$, and maps $[w, \Omega]$ on $\mathbb{S}$ to $C(w, \Omega) := w/(1-\Omega)$ in $\mathbb{C}$, whose unit circle projects to itself unchanged as $\mathbb{S}$'s equator.

Exercise 2: Confirm that the three points $[0, 1]$, $S(z)$ and $[z, 0]$ are collinear in $[w, \Omega]$-space, and that the functions C and S are inverse to each other, and that they associate the North pole $[0, 1]$ of $\mathbb{S}$ with ∞ in $\mathbb{C}$, the South pole $[0, -1]$ of $\mathbb{S}$ with 0 in $\mathbb{C}$, and $\mathbb{S}$'s equator $[w, 0]$ (on which $|w| = 1$) with the unit circle $z = w$ in $\mathbb{C}$. Stereographic projection takes two finite points z_1 and z_2 in $\mathbb{C}$ to *antipodal* points on $\mathbb{S}$ if and only if $\bar{z}_1 \cdot z_2 = -1$; prove this.

Exercise 3: Show that $\mathrm{Re}(\overline{w}\cdot dw) = -\Omega\cdot d\Omega$ when $[w, \Omega]$ moves on $\mathbb{S}$, so moving $z = C(w, \Omega)$ that $dz = dw/(1-\Omega) + w\cdot d\Omega/(1-\Omega)^2$ satisfies $|dz|^2 = (|dw|^2 + d\Omega^2)/(1-\Omega)^2$. Show the *Chordal* distance between projections on $\mathbb{S}$ of points z_1 and z_2 in $\mathbb{C}$ to be $2|z_2-z_1|/\sqrt{((|z_1|^2+1)(|z_2|^2+1))}$.

In many a text the Riemann Sphere $\mathbb{S}$ is a sphere of radius 1/2 resting on top of $\mathbb{C}$ at its origin; this would halve the chordal distance and alter some of the algebraic details but not change the gist of our reasoning about $\mathbb{S}$ ..

A crucially important property of stereographic projection is that it maps circles on $\mathbb{S}$ to and from circles and straight lines in $\mathbb{C}$. To establish this property we characterize circles on $\mathbb{S}$ as intersections with planes in $[w, z]$-space. The equation of a plane $\mathbb{P}$ in $[w, \Omega]$-coordinates is $\mathrm{Re}(cw) + \alpha\Omega + \beta = 0$ for an arbitrary choice of complex constant c and real constants α and β. The distance from the origin $[0, 0]$ to $\mathbb{P}$ is $||\mathbb{P}|| := |\beta|/\sqrt{(\alpha^2 + |c|^2)}$, and must not exceed 1 if $\mathbb{P}$ is to intersect $\mathbb{S}$, where $|w|^2 + \Omega^2 = 1$. The points $[w, \Omega]$, in the circle where $\mathbb{S}$ and $\mathbb{P}$ intersect, project to $z = C(w, \Omega)$ in $\mathbb{C}$, so $[w, \Omega] = S(z) = [2z/(|z|^2+1), (|z|^2-1)/(|z|^2+1)]$ which, when substituted into $\mathbb{P}$'s equation, yields the equation $2\cdot\mathrm{Re}(cz) + \alpha(|z|^2-1) + \beta(|z|^2+1) = 0$ that z must satisfy. If $\alpha+\beta \neq 0$ this is the equation of a circle in $\mathbb{C}$ whose radius squared turns out to be $(\alpha^2 + |c|^2)(1 - ||\mathbb{P}||^2)/(\alpha+\beta)^2 \geq 0$; but if $\alpha+\beta = 0$ it is the equation of a straight line in $\mathbb{C}$ projected through $\mathbb{P}$ to its intersection with $\mathbb{S}$, a circle through $[0, 1]$ in $[w, \Omega]$-space.

Exercise 4: Confirm the previous four sentences' assertions. Where are the circles' centers?

Stereographic projection is *Conformal*: it preserves curves' angles of intersection. To see why this is so, consider two curves in the complex plane $\mathbb{C}$ that intersect at some angle Ø. This is the angle between our curves' tangents, two straight lines that project to two circles on $\mathbb{S}$ tangent to the projections onto $\mathbb{S}$ of our two curves. The two circles are cut from $\mathbb{S}$ by two planes, each a plane through one of the tangent lines in $\mathbb{C}$ and through the North pole of $\mathbb{S}$. Here the circles intersect at the same angle as do their tangent lines, which are the two planes' intersections with another plane parallel to $\mathbb{C}$ through the North pole; that angle is Ø again. The circles' other intersection, also at angle Ø, is at the projection onto $\mathbb{S}$ of our curves' intersection point, so the projections onto $\mathbb{S}$ of our curves intersect at angle Ø too. (Proof due to C. Carathéodory.)

Every bilinear rational function $M(z) := (\mu z + ß)/(b - mz)$ is construed simultaneously as a map of the *Completed* (at ∞) complex plane $\mathbb{C}$ to itself, and as a map of the compact Riemann Sphere $\mathbb{S}$ to itself thus: $[w, \Omega]$ maps to $S(M(C(w, \Omega)))$. Both ways, every such map is called a *Möbius Transformation*. As we shall see, the set of all Möbius transformations forms a *non-Abelian* (non-commutative) *Group* under composition, and every member of this group maps

- circles to circles on the Riemann Sphere $\mathbb{S}$, and
- "circles" to "circles" in the completed complex plane $\mathbb{C}$, subject to the understanding that if a "circle" passes through ∞ it is actually a straight line in the plane.

Conversely, any map of one simply connected open region on $\mathbb{S}$ to another that maps *all* circles inside the first region to circles inside the second must be a Möbius transformation or its complex conjugate. C. Carathéodory proved this in 1937 without presuming the map's continuity! His proof will not be reproduced here, alas.

Exercise 5: Extend analytic functions' domain $\mathbb{C}$ to $\mathbb{S}$ as follows: Show that every finite-ordered pole of an analytic function f is a removable singularity and a zero of $1/f$; and that if $S(f(z))$ approaches a path-independent limit-point on $\mathbb{S}$ as $z \to \infty$, then $f(1/z)$ has a pole of finite order or a removable singularity at $z = 0$. Hint: Recall the Casorati-Weierstrass theorem.

The Completion of $\mathbb{C}$ by adjoining one point at $\infty = 1/0$ is a mixed blessing. On one hand, it identifies $\mathbb{C}$ with a compact set $\mathbb{S}$ on which all continuous functions are uniformly continuous, and all rational functions not merely continuous but differentiable even at their poles, which seems advantageous. On the other hand it can render formerly innocuous algebraic operations *risqué* . For instance a cancellation law " $z/z = 1$ ", formerly violated only by $z = 0$, is now violated also by $z = \infty$, as are other cancellation laws " $z - z = 0$ " and " $z \cdot 0 = 0$ ". Do not confuse these operations invalid upon *numbers* (when ∞ is treated as a number) with operations that look the same in print but are actually operations upon *functions*: " $z/z = 1$ " and " $z - z = 0$ " and " $z \cdot 0 = z/\infty = 0$ " are valid operations upon analytic functions after removable singularities at $z = 0$ and at $z = \infty$ have been removed. This removal is a limiting process, not an arithmetic process that computers can carry out with mere numbers. For this reason computers (but not human mathematicians) require some other symbol besides " ∞ " to be adjoined to completions of the real and complex number systems; such a symbol is " NaN ", standing for " Not a Number." Without it, any convention that assigned numerical values like 1 to $0/0$ and ∞/∞ , or 0 to $\infty \cdot 0$ and $\infty - \infty$, would merely propagate the cancellation laws' failures to other laws like the associative and distributive laws: Examples like " $1 = 0/0 = (3 \cdot 0)/0 = 3 \cdot (0/0) = 3 \cdot 1 = 3$ " and "$\infty = (3-2) \cdot \infty = 3 \cdot \infty - 2 \cdot \infty = \infty - \infty = 0$ " and " $0 = \infty \cdot 0 = \infty/\infty = 1$ " would undermine our faith in computation (as the first example did in *APL* and the others do in *MathCad*). With NaN, and with conventions that assign it (instead of something more confusing) to otherwise invalid arithmetic operations and that propagate it through all algebraic operations and most (but not all) other operations, the order in which cancellation, associative, distributive and other algebraic laws are applied during the evaluation of an *unconditional* (no comparisons) arithmetic expression can affect its result in only this way:

If different orders can produce different results, just two results are possible and one of them must be NaN .

For example, one application apiece of the cancellation and distributive laws turn $1 - 1/z$ into $(z-1)/z$, so they are the same analytic function after the latter expression's singularity at $z = \infty$ has been removed; but this removal cannot be accomplished by mere evaluation, which yields NaN instead of 1 for the latter expression at $z = \infty$.

NaN and ∞ offer no panacea. Aside from Riemann's removal of isolated singularities by a limiting process, no methods exist in general to deduce uniquely correct extensions of familiar functions of real and complex arguments to cope with ∞ and NaN. This is partly because the completion of the complex plane $\mathbb{C}$ by a single point at ∞ is not the only possible completion; a line or circle at infinity might serve better in some circumstances. Similarly, the real line can be completed as well by two infinities $\pm\infty$ as by one unsigned ∞ . Different completions generate NaN differently, though they must always conform to the requirement displayed bold-faced above.

If computations had no way to get rid of NaN it would be useless; computation would be better stopped before so persistent a NaN were created. This means that some operations upon NaN must produce something else without exacerbating confusion. For example, $\text{NaN}^0 := 1$ because $z^0 := 1$ by convention for all real and complex values of z including $z = 0$ and $z = \infty$. More generally, suppose a function $f(z, w)$ is so defined for all finite and infinite arguments that $f(z, 0)$ is a constant independent of z ; a plausible principle requires $f(\text{NaN}, 0)$ to take the same constant value. But this plausible principle can be ineffective. For example, predicate " $\text{NaN} \neq z$ " is *True* and "$z \neq z$ " *False* for all finite and infinite z ; what should " $\text{NaN} \neq \text{NaN}$ " be? It is *True* by definition, and so " $\text{NaN} = \text{NaN}$ " is *False*. These astonish implementors of programming languages. The exigencies of computer programming are driven sometimes less by logic than by accidents of history; one of these forced " $\text{NaN} \neq \text{NaN}$ " to be declared *True*. Another example: $\text{Max}\{x, +\infty\} = +\infty$ for $-\infty \le x \le +\infty$, implying that $\text{Max}\{\text{NaN}, +\infty\} := +\infty$ too and suggesting that $\text{Max}\{\text{NaN}, x\}$ is better defined to be x than to be NaN even though " $\text{NaN} \le x$ " is *False* for all x , as is " $\text{NaN} > x$ ". Thus ∞ that begets NaN may postpone or worsen complications, not eliminate them.

Computer systems are expected to raise a *Flag* to signal the creation of ∞ from finite arguments, and to raise another flag when NaN is created instead of merely propagated. Such a signal is obligatory because NaN and ∞ can destroy the validity of programs designed without them in mind. Programmers determined to prevent ∞ or NaN from being created would have to scrutinize the inputs to every operation or subprogram that had not been proved incapable of creating these unwanted entities, and then figure out what to do with possibly dangerous inputs. Such scrutiny could be onerous, time-consuming for both programmers and computers, even if nothing dangerous ever materialized. In most cases the speedier way on average is to check a flag afterwards and, on the rare occasion (if any) when a NaN or ∞ created unwittingly has corrupted computed results, supplant these by something else.

Sometimes Topology annoys Algebra; the open plane $\mathbb{C}$ is a *Field*, the closed sphere $\mathbb{S}$ is not. Sometimes Computing annoys Mathematics; signal flags and NaN are mathematically ugly computational details. Much as we yearn to avoid annoyances, some of them are provably unavoidable; and they put bread on many of our tables.

The Group of Möbius Transformations

Under composition, Möbius transformations form a non-commutative group identifiable with a quotient of subgroups of the multiplicative group of invertible 2-by-2 matrices. To see how this works let $M_j(z) := (\mu_j z + ß_j)/(b_j - m_j z)$ and $ç_j := b_j\mu_j + ß_j m_j \neq 0$ for $j = 1, 2$ and 3 ; then see that, for any such choices of M_1 and M_2 , their composition $M_2(M_1(z)) = M_3(z)$ is another Möbius transformation whose parameters can be computed by matrix multiplication thus:

$$\begin{bmatrix} b_3 & -m_3 \\ \beta_3 & \mu_3 \end{bmatrix} = \begin{bmatrix} b_2 & -m_2 \\ \beta_2 & \mu_2 \end{bmatrix} \cdot \begin{bmatrix} b_1 & -m_1 \\ \beta_1 & \mu_1 \end{bmatrix} \quad \text{and} \quad ç_3 = \det(\begin{bmatrix} b_3 & -m_3 \\ \beta_3 & \mu_3 \end{bmatrix}) = ç_2\, ç_1 \neq 0 .$$

Moreover the Möbius transformation M whose parameters are $\{ß, \mu, b, m; ç := b\mu + ßm\}$ has an inverse W that may be assigned parameters respectively $\{-ß/ç, b/ç, \mu/ç, -m/ç; 1/ç\}$, just as the inverse matrix $\begin{bmatrix} b & -m \\ \beta & \mu \end{bmatrix}^{-1} = \begin{bmatrix} \mu & m \\ -\beta & b \end{bmatrix}/ç$. But we cannot identify each Möbius transformation M with one uniquely determined matrix since multiplying M's parameters $\{ß, \mu, b, m; ç\}$ by a nonzero constant (and ç by its square) changes the matrix without changing M . Instead we can identify each Möbius transformation with a *pair* of matrices, but first we must restrict ç :

- If M is real, restrict ç to the set $\{1, -1\}$; if M is complex, fix $ç := 1$.

Doing so identifies each $M(z) := (\mu z + ß)/(b - mz)$ with the pair of matrices $\pm\mathbf{M} := \pm\begin{bmatrix} b & -m \\ \beta & \mu \end{bmatrix}$ whose determinants $\det(\pm\mathbf{M}) = b\mu + ßm = ç$. The set of all such matrices includes their inverses and thus constitutes a non-commutative multiplicative group consisting of the quotient

(the group of all 2-by-2 matrices with determinants 1 , or ±1)/(its subgroup {I, –I}) .

This multiplicative quotient group is *Isomorphic* to the group under composition of Möbius transformations. These groups factor naturally into the three subgroups tabulated here:

Table 1: Subgroups of Möbius Transformations M(z)

Subgroup Name	M(z)	ß	μ	b	m	ç	Matrix ±**M**
Dilation, or Scaling†	$\pm\mu^2\cdot z$	0	$\mu \neq 0$	$\pm 1/\mu$	0	±1	$\begin{bmatrix} \pm 1/\mu & 0 \\ 0 & \mu \end{bmatrix}$
Translation	z + ß	ß	1	1	0	1	$\begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix}$
Inversion*	–1/z , z	1 , 0	0 , 1	0 , 1	1 , 0	1 , 1	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(† The "±" sign is only for *Real* Möbius transformations; supplant it by "+" for Complex.)
(* The Inversion subgroup has two elements, one of them the Identity M(z) = z that changes nothing.)

Theorem: Every Möbius transformation is a composition of at most four transformations each selected from a subgroup tabulated above; and some Möbius transformations require all four.

Proof: $M(z) = (\mu z + ß)/(b - mz) = -1/((m^2/ç)z - mb/ç) - \mu/m$ is a translation of an inversion of a translation of a dilation unless $m = 0$, in which case M is a translation of a dilation. Either way, no more than four selections from the subgroups suffice to reproduce M . Three selections may be too few if $m \neq 0$ because then M requires an inversion to create the pole at $z = b/m$, and then two more selections are too few to determine three parameters. In other words, three selections suffice to generate only a finite number of two-dimensional surfaces inside the three-dimensional space of Möbius transformations. End of proof.

Exercise 6: State the foregoing theorem in matrix terms and prove it those terms.

Exercise 7: Verify that the Möbius transformations z , 1/z , 1–z and all their compositions constitute a subgroup with six members, the other three being 1/(1–z) , 1–1/z and 1/(1–1/z) .

Often an easy way to prove an assertion for all Möbius transformations is to do so separately for each tabulated subgroup and then infer it for their compositions. Often the assertion is obvious for all but the Inversion subgroup. Such is the case for the next three assertions offered as …

Exercise 8: Prove that …

- Möbius transformations map circles on the Riemann sphere $\mathbb{S}$ to circles on $\mathbb{S}$.
- Any two given circles on $\mathbb{S}$ can be mapped one to another by Möbius transformations.
- Möbius transformations map *Reflections* in a circle to reflections in its image circle.

Reflections in a circle are defined thus: Given the center c and radius r > 0 of a circle in the complex plane $\mathbb{C}$, two points p and q are said to be reflections of each other in that circle just when $\overline{(p-c)}(q-c) = r^2$ or, equivalently, (p–c)/(q–c) > 0 and $|p-c|\cdot|q-c| = r^2$. Thus c, p and q are collinear, and the latter two are one inside and the other outside the circle but both on the same side of c . For example, when $c = \imath r$ the reflection of p is $q = \bar{p}/(1 - \imath\bar{p}/r)$; now fix p while letting $r \to +\infty$ to find that $q \to \bar{p}$ while the circle approaches the real axis. In general, two points are reflections of each other in a straight line when it is the right bisector of the line segment joining the points, as one might expect.

Exercise 9: Show that all circles through two points, each the reflection of the other in another circle O , cut it orthogonally (at right angles). Hint: Möbius transform O to the real axis.

Fixed-Points

The only Möbius transformation with more than two fixed-points is the identity, which leaves all of $\mathbb{S}$ fixed. Every other M(z) := (µz + ß)/(b – mz) has one or two *Fixed-Points* $f = M(f)$; they are the zeros of the quadratic polynomial $mz^2 - (b-\mu)z + ß$ subject to the understanding that ∞ is a fixed-point just when m = 0 . When $(b-\mu)^2 - 4mß \ne 0$, two distinct points are fixed; call them f_1 and f_2 . If both are finite, M is defined by a more symmetrical equation $(M(z) - f_1)/(M(z) - f_2) = k\cdot(z - f_1)/(z - f_2)$; all Möbius transformations M with the same two finite distinct fixed-points constitute a subgroup isomorphic to the multiplicative group of finite nonzero constants k ; the identity transformation has k = 1 . If one fixed-point f is finite and the other is ∞ then $M(z) - f = k\cdot(z - f)$ for the same group of constants k .

M has only one fixed-point $f = M(f)$ just when $(b-\mu)^2 - 4mß = 0$. If this one fixed-point f is finite then the more symmetrical equation is $1/(M(z) - f) = K + 1/(z - f)$, and the subgroup of such Möbius transformations is isomorphic to the additive group of finite constants K with K = 0 for the identity. If the one $f = \infty$ then M(z) = z + K for the same group of constants K .

Exercise 10: Derive the "more symmetrical" equations from appropriate hypotheses about the quadratic, namely that the fixed-points satisfy $f_1 + f_2 = (b-\mu)/m$ and $f_1\cdot f_2 = ß/m$ when they are finite, or appropriate other equations when one or both fixed-points are infinite. Then use the more symmetrical equations to prove that a Möbius transformation M not the identity maps no circle onto itself on $\mathbb{S}$ just when M has two distinct fixed-points and $|k| \ne 1$ and $\mathrm{Im}(k) \ne 0$; in this case M is called "Loxodromic". Otherwise a one-real-parameter family of infinitely many circles each mapped onto itself by M covers $\mathbb{S}$; identify the three different kinds of family.

Cross-Ratios

Any quadruple {z, p, q, r} of four *distinct* points on $\mathbb{S}$ determine a *Cross-Ratio* $\frac{(z-p)}{(z-r)}\cdot\frac{(q-r)}{(q-p)}$.
If one member of the quadruple is ∞ merely replace this cross-ratio by an appropriate limit, which is tantamount to discarding the factors (…) containing ∞ . All cross-ratios are finite; and they are preserved by all Möbius transformations M(z) := (µz + ß)/(b – mz) , which means that

$$\frac{(z-p)}{(z-r)}\cdot\frac{(q-r)}{(q-p)} = \frac{(M(z)-M(p))}{(M(z)-M(r))}\cdot\frac{(M(q)-M(r))}{(M(q)-M(p))}$$

holds for *every* such M .

Exercise 16: Confirm this equation for each subgroup in Table 1, and then for the whole group.

Exercise 17: Since there are 24 permutations of the points in a quadruple {z, p, q, r} they could determine 24 cross-ratios, but these take at most six values; show that if X is one of those six values the others are 1/X , 1–X , 1/(1–X) , 1–1/X and 1/(1–1/X) . Hint: Exercise 7.

A Möbius transformation w = M(z) = (µz + ß)/(b – mz) is determined *completely* by what it does to *any* three distinct values p, q, r of z . It must map them to some three distinct values P := M(p) , Q := M(q) , R := M(r) respectively of w . Given the triples {p, q, r} and {P, Q, R} we may construct M by solving a bilinear cross-ratio equation like

$$(z-p)(q-r)(w-R)(Q-P) = (w-P)(Q-R)(z-r)(q-p)$$

for either w = M(z) or its inverse z = W(w) , thereby determining the constants ß, µ, b and m except for a common factor. One member of the triple {p, q, r} can be ∞ if the cross-ratio equation is replaced by an appropriate limit, which is tantamount to discarding the factors (…) containing ∞ ; similarly for {P, Q, R} . With that understanding, we can rewrite the cross-ratio equation in a more symmetrical form

$$¥ = \frac{(z-p)}{(z-r)}\cdot\frac{(q-r)}{(q-p)} = \frac{(w-P)}{(w-R)}\cdot\frac{(Q-R)}{(Q-P)}$$

that exhibits explicitly the cross-ratios in question and also exhibits explicitly the construction of M as a composition of two Möbius transformations, the first from z to ¥ that takes (p, q, r} to {0, 1, ∞} respectively, and the second from ¥ to w that takes {0, 1, ∞} respectively to {P, Q, R} .

Given triples {p, q, r} and {P, Q, R} determine M (but not its parameters) uniquely. To see why this is so suppose there were two Möbius transformations taking the first triple to the second; composing one with the other's inverse would create a transformation which, with three distinct fixed-points, would have to be the identity.

The triple of *distinct* points {p, q, r} determines uniquely a circle Ô (or a straight line in $\mathbb{C}$) through all three points; likewise {P, Q, R} determines another circle M(Ô) as well as the Möbius transformation M that maps one circular boundary to the other. What does M do to the circles' interiors? It maps the interior of one either to the other's interior or else to its exterior; how do we tell which? The question seems moot when either triple of points lies in a straight line, which occurs when either (q-r)/(q-p) or (Q-R)/(Q-P) is real. Otherwise …

Exercise 18: Show why M maps the interior of circle Ô to the interior of circle M(Ô) if and only if the imaginary parts of (q-r)/(q-p) and of (Q-R)/(Q-P) have the same nonzero signs.

Exercise 19: Why is it worth knowing that, provided {p, q, r} and {P, Q, R} are entirely finite,

$$(p-q)(q-r)(r-p)/\big((P-Q)(Q-R)(R-P)\big) = \big((b - mp)(b - mq)(b - mr)\big)^2/ç^3 \ ?$$