Submitted exclusively to the London Mathematical Society
doi:10.1112/0000/000000

Norms of Möbius maps

Alan F. Beardon and Ian Short

Abstract

We derive inequalities between the matrix, chordal, hyperbolic, three-point, and unitary norms of a Möbius map. These extend inequalities proved earlier by Gehring and Martin.

1. *Introduction*

A Möbius transformation $z \mapsto (az+b)/(cz+d)$, where $ad-bc \neq 0$, is a homeomorphism of the extended complex plane $\mathbb{C}_\infty$ onto itself with the chordal metric q given by

$$q(z,w) = \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}}, \quad q(z,\infty) = \frac{2}{\sqrt{1+|z|^2}},$$

and also a conformal isometry of the upper half-space $\mathbb{H}^3$ of $\mathbb{R}^3$ endowed with the hyperbolic metric $ds = |dx|/x_3$. In [**3**, **4**] Gehring and Martin derived inequalities between the matrix norm, the chordal norm and the hyperbolic norm of a Möbius map (all of which are defined below). Here we introduce two more norms and study the relationships between these five norms. It is known that if a sequence of Möbius transformations converges at three distinct points to three distinct values, then it converges uniformly on $\mathbb{C}_\infty$ to a Möbius transformation. The work in this paper originated in an attempt to find a proof of this result which exhibits an explicit rate of convergence, and our inequalities provide such an estimate.

The group $\mathcal{M}$ of Möbius maps is equipped with the supremum metric d, where

$$d(f,g) = \sup\left\{ q\big(f(z), g(z)\big) : z \in \mathbb{C}_\infty \right\}, \quad f, g \in \mathcal{M},$$

so that $d(f_n, f) \to 0$ if and only if $f_n \to f$ uniformly on $\mathbb{C}_\infty$. Following Gehring and Martin ([**3**, **4**]), we define the *chordal norm* of a Möbius map f to be $d(f, I)$, where I denotes the identity map: thus

$$d(f,I) = \sup\{ q\big(f(z), z\big) : z \in \mathbb{C}_\infty \}.$$

Given a Möbius map f, we can write

$$f(z) = \frac{az+b}{cz+d}, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad-bc = 1, \tag{1.1}$$

where A is determined to within a factor ± 1. The *matrix norm* of f is $\|f\|$, or $\|A\|$, where

$$\|f\| = \|A\| = \sqrt{|a|^2 + |b|^2 + |c|^2 + |d|^2},$$

and this is independent of the factor ± 1. Gehring and Martin introduce the norm

$$m(f) = \|A - A^{-1}\| = \sqrt{2|a-d|^2 + 4|b|^2 + 4|c|^2},$$

2000 *Mathematics Subject Classification* 51B10 (primary), 30F45 (secondary).

The second author's work was supported by Science Foundation Ireland grant 05/RFP/MAT0003.

but more often than not use the expression $m(f)/\|A\|$. We shall combine these ideas and define

$$M(f) = \frac{\|A - \det(A)A^{-1}\|}{\|A\|},$$

where A is now any matrix which represents f as in (1.1), except that we no longer insist that $ad - bc = 1$. Note that $M(f)$ is independent of the choice of A from $\mathrm{GL}(2, \mathbb{C})$ which represents f, and if $\det(A) = 1$, then $M(f) = m(f)/\|A\|$. In this notation, Gehring and Martin prove (see [**4**, (3.35)])

$$M(f) \leqslant d(f, I) \leqslant \sqrt{2}\, M(f). \tag{1.2}$$

Both constants in these inequalities are best possible.

Guided by the fact that convergence at three points implies uniform convergence we now introduce the *three-point norm*

$$\varepsilon(f) = \max\left\{q\big(f(1), 1\big), q\big(f(\omega), \omega\big), q\big(f(\omega^2), \omega^2\big)\right\},$$

where $\omega = e^{2\pi i/3}$. In principle, we could take any three points here, and originally the authors proved an estimate similar to (1.3) below using the points 0, 1 and ∞ instead of 1, ω and ω^2, and with a constant on the right between 2 and 3 (and not best possible). The authors thank the referee for providing the better inequality $M(f) \leqslant \sqrt{2}\,\varepsilon(f)$ in (1.3) and its proof.

Theorem 1.1. *For any Möbius transformation f,*

$$\frac{\varepsilon(f)}{\sqrt{2}} \leqslant M(f) \leqslant \sqrt{2}\,\varepsilon(f), \tag{1.3}$$

and

$$\varepsilon(f) \leqslant d(f, I) \leqslant 2\,\varepsilon(f). \tag{1.4}$$

Further, each constant in each of the four inequalities is best possible.

In the case of parabolic Möbius maps, we can say more than Theorem 1.1. Let φ be the stereographic projection of $\mathbb{C}_\infty$ onto the unit sphere $\mathbb{S}$ in $\mathbb{R}^3$. Then p and $\hat{p}$ in $\mathbb{C}_\infty$ are *antipodal points* if and only if $\varphi(p)$ and $\varphi(\hat{p})$ are the end-points of a diameter of $\mathbb{S}$ (that is, $q(p, \hat{p}) = 2$). In particular, the points 0 and ∞ are antipodal, and we define

$$\varepsilon_0(f) = \max\left\{q\left(f(0), 0\right), q\left(f(\infty), \infty\right)\right\}.$$

The choice of the pair of antipodal points used to define ε_0 is insignificant (see Section 2).

Theorem 1.2. *For any parabolic Möbius transformation f,*

$$\frac{\varepsilon_0(f)}{\sqrt{2}} \leqslant M(f) \leqslant \sqrt{2}\,\varepsilon_0(f), \tag{1.5}$$

and

$$\varepsilon_0(f) \leqslant d(f, I) \leqslant 2\,\varepsilon_0(f). \tag{1.6}$$

Further, each constant in each of the four inequalities is best possible.

Next, the *hyperbolic norm* of f is $\rho\big(j, f(j)\big)$, where ρ is the hyperbolic metric on $\mathbb{H}^3$, and $j = (0, 0, 1)$, and Gehring and Martin obtained an inequality that is equivalent to

$$2 \tanh \tfrac{1}{2}\rho(j, f(j)) \leqslant d(f, I). \tag{1.7}$$

(see [**4**, (1.19) and Theorem 3.19]).

A Möbius map u is a *unitary map* if its conjugate $\varphi u \varphi^{-1}$ is a rotation of the sphere $\mathbb{S}$ or, equivalently, if $u(j) = j$. Now suppose that u is unitary, and apply (1.7) to fu^{-1} instead of f; this gives

$$2 \tanh \tfrac{1}{2}\rho(j, f(j)) \leqslant d(f, u).$$

We let $\mathcal{U}$ be the subgroup of unitary maps, and we call $d(f, \mathcal{U})$ the *unitary norm* of f, where

$$d(f, \mathcal{U}) = \inf \{d(f, u) : u \in \mathcal{U}\}\,;$$

this measures how far f is from the subgroup $\mathcal{U}$. The Gehring Martin inequality (1.7) implies that

$$2 \tanh \tfrac{1}{2}\rho(j, f(j)) \leqslant d(f, \mathcal{U}), \tag{1.8}$$

and our last result is that this inequality is, in fact, an equality.

THEOREM 1.3. *In the notation above,*

$$d(f, \mathcal{U}) = 2 \tanh \tfrac{1}{2}\rho(j, f(j)) = 2\sqrt{\frac{\|f\|^2 - 2}{\|f\|^2 + 2}}.$$

We discuss unitary maps in Section 2. The proofs of Theorems 1.1, 1.2 and 1.3 are given in Sections 3, 4 and 5, respectively. Finally, in Section 6, we return to the original motivation for this paper and obtain a quantitative version of the theorem that says convergence at three points implies uniform convergence.

2. *Unitary maps*

We devote this section to a brief discussion of unitary maps, and the reader is referred to [**1**] for details. A Möbius map u is a *unitary* map if and only if $\varphi u \varphi^{-1}$ is a rotation of the sphere $\mathbb{S}$, and a unitary Möbius map is clearly a chordal isometry. Also, a Möbius map is unitary if and only if, in its action on $\mathbb{H}^3$, it fixes $j = (0, 0, 1)$. Since

$$\|f\|^2 = 2\, \cosh \rho\big(j, f(j)\big) \tag{2.1}$$

for any Möbius map f, we also see that f is unitary if and only if $\|f\|^2 = 2$.

The metric d is *right invariant*: for all Möbius f, g and h,

$$d(fh, gh) = d(f, g).$$

The Möbius map h induces the left invariance property $d(hf, hg) = d(f, g)$ for all f and g if and only if h is unitary.

Unitary maps have other useful invariance properties; for example, if X is any 2×2 complex matrix (singular or non-singular), and if U is a unitary matrix (corresponding to a unitary Möbius map), then $\|UX\| = \|X\| = \|XU\|$. Now consider any Möbius map f, and any unitary Möbius map u. Then, $\|ufu^{-1}\| = \|f\|$. Also, since u is a chordal isometry,

$$d(ufu^{-1}, I) = d(ufu^{-1}, uu^{-1}) = d(f, I).$$

These facts imply that any relationship between $d(f, I)$ and $\|f\|$ is invariant under conjugation by a unitary map, and this leads to a considerable simplification of our arguments. Since a unitary map u fixes j and is a hyperbolic isometry, we see that, for any Möbius map f,

$$\rho\big(ufu^{-1}(j), j\big) = \rho\big(f(j), j\big),$$

so similar comments apply to this norm too.

In conclusion, these remarks show that we could replace the three points 1, ω and ω^2 in the definition of ε by any three points equally spaced around a great circle and Theorem 1.1 would remain true. Similarly, Theorem 1.2 would remain true if 0 and ∞ are replaced by any pair of antipodal points.

3. *The proof of Theorem 1.1*

To establish the four inequalities in (1.3) and (1.4), it suffices to prove the *right hand* inequality of (1.3) and the *left hand* inequality of (1.4), because the remaining inequalities follow from these two inequalities together with $d(f,I) \leqslant \sqrt{2}M(f)$ from (1.2). The left hand inequality of (1.4), $\varepsilon(f) \leqslant d(f,I)$, follows immediately from the definition of ε, so we have only to prove the right hand inequality $M(f) \leqslant \sqrt{2}\varepsilon(f)$ of (1.3). In fact, we shall prove the following slightly stronger result.

PROPOSITION 3.1. *There are positive numbers μ_0, μ_1 and μ_2, with $\mu_0+\mu_1+\mu_2=1$, such that*

$$M(f)^2 \leqslant 2\left[\mu_0\, q\left(f(1),1\right)^2 + \mu_1\, q\left(f(\omega),\omega\right)^2 + \mu_2\, q\left(f(\omega^2),\omega^2\right)^2\right] \leqslant 2\,\varepsilon(f)^2. \tag{3.1}$$

Proof. If $|z|=1$ then

$$\left(|az+b|^2+|cz+d|^2\right) = \|A\|^2 + 2\mathrm{Re}[(a\bar{b}+c\bar{d})z],$$

and since $1+\omega+\omega^2=0$, this gives

$$\sum_{z^3=1}\left(|az+b|^2+|cz+d|^2\right) = 3\|A\|^2. \tag{3.2}$$

In a similar way we get

$$\sum_{z^3=1}|(az+b)-z(cz+d)|^2 = 3\left(|a-d|^2+|b|^2+|c|^2\right).$$

Now define

$$\mu_j = \frac{|a\omega^j+b|^2+|c\omega^j+d|^2}{3\|A\|^2}, \quad j=0,1,2. \tag{3.3}$$

We see from (3.2) that $\mu_0+\mu_1+\mu_2=1$. Observe that

$$2|(a\omega^j+b)-\omega^j(c\omega^j+d)|^2 = 3\mu_j\|A\|^2\, q\left(f(\omega^j),\omega^j\right)^2,$$

which means that

$$\sum_{j=0}^{2}\mu_j\, q\left(f(\omega^j),\omega^j\right)^2 = \frac{2}{\|A\|^2}\left(|a-d|^2+|b|^2+|c|^2\right),$$

and this gives (3.1) since

$$M(f)^2 = \frac{2}{\|A\|^2}\left(|a-d|^2+2|b|^2+2|c|^2\right).$$

□

To show that the constants in the four inequalities from (1.3) and (1.4) are best possible it suffices to prove that the constants in the *left hand* inequality of (1.3) and the *right hand* inequality of (1.4) are best possible, because then, using the inequality $d(f,I) \leqslant \sqrt{2}M(f)$ from (1.2), we see that all four constants are best possible. For example, if we show that the constant

2 in $d(f,I) \leqslant 2\varepsilon(f)$ is best possible, then the constants in the two inequalities $d(f,I) \leqslant \sqrt{2}M(f)$ and $M(f) \leqslant \sqrt{2}\varepsilon(f)$ must also be best possible.

To see that the constant $\sqrt{2}$ in the left hand inequality of (1.3), $\varepsilon(f) \leqslant \sqrt{2}M(f)$, is best possible, consider the following sequence of Möbius transformations:

$$f_n(z) = \frac{nz-(n-1)}{-(n+1)z+n}, \quad n = 1, 2, \dots .$$

Then $f_n(1) = -1$, so that $\varepsilon(f_n) = 2$, and one can check that $M(f_n) \to \sqrt{2}$ as $n \to \infty$.

To see that the constant 2 in the right hand inequality of (1.4), $d(f,I) \leqslant 2\varepsilon(f)$, is best possible, consider the following one parameter group of Möbius transformations with fixed points -1 and 1:

$$f_t(z) = \frac{(1+t)z+(1-t)}{(1-t)z+(1+t)}, \quad t \in \mathbb{R}.$$

Notice that $f_t(z) \to -1$ as $t \to \infty$ for all points z other than 1. Therefore

$$\varepsilon(f_t) \to 1, \quad d(f_t, I) \to 2,$$

as $t \to \infty$, and this shows that the constant 2 is best possible.

4. *The proof of Theorem 1.2*

To establish the four inequalities in (1.5) and (1.6), it suffices to prove the *right hand* inequality of (1.5) and the *left hand* inequality of (1.6), because the remaining inequalities follow from these two inequalities together with $d(f,I) \leqslant \sqrt{2}M(f)$ from (1.2). The left hand inequality of (1.6), $\varepsilon_0(f) \leqslant d(f,I)$, follows immediately from the definition of ε_0, so we have only to prove the right hand inequality of (1.5), $M(f) \leqslant \sqrt{2}\varepsilon_0(f)$.

Since f is parabolic we may assume, with the notation of (1.1), that $ad - bc = 1$ and $a + d = 2$. This means that

$$4bc = 4(ad-1) = -4(a-1)^2 = -(a-d)^2.$$

Hence

$$\begin{aligned} M(f)^2 &= \frac{2|a-d|^2 + 4|b|^2 + 4|c|^2}{|a|^2+|b|^2+|c|^2+|d|^2} \\ &\leqslant 2\left(\frac{4|b|^2+4|c|^2}{|a|^2+|b|^2+|c|^2+|d|^2}\right) \\ &\leqslant 2\max\left\{\frac{4|b|^2}{|b|^2+|d|^2}, \frac{4|c|^2}{|a|^2+|c|^2}\right\} \\ &= 2\epsilon_0(f)^2, \end{aligned}$$

as required.

To show that the constants in the four inequalities in (1.5) and (1.6) are best possible it suffices to prove that the constants in the *left hand* inequality of (1.5) and the *right hand* inequality of (1.6) are best possible, because then, using the inequality $d(f,I) \leqslant \sqrt{2}M(f)$ from (1.2), we see that all four constants are best possible.

To see that the constant $\sqrt{2}$ in the left hand inequality of (1.3), $\varepsilon_0(f) \leqslant \sqrt{2}M(f)$, is best possible, consider the maps $f_t(z) = z + t$ for $t > 0$. One can check that $\varepsilon_0(f_t)/M(f_t) \to \sqrt{2}$ as $t \to 0$. To see that the constant 2 in the right hand inequality of (1.4), $d(f,I) \leqslant 2\varepsilon_0(f)$, is best possible, consider the following one parameter group of parabolic Möbius transformations with fixed point i:

$$f_t(z) = \frac{(1+it)z+t}{tz+(1-it)}, \quad t \in \mathbb{R}.$$

As $t \to 0$ we find that $q(f_t(0), 0) \sim 2t$, $q(f_t(\infty), \infty) \sim 2t$ and $q(f_t(-i), -i) \sim 4t$. This means that $\limsup_{t\to 0} d(f_t, I)/\epsilon_0(f_t) \geqslant 2$. Therefore the constant 2 in $d(f, I) \leqslant 2\varepsilon_0(f)$ is best possible.

5. *The proof of Theorem 1.3*

We begin with a decomposition result for a general Möbius f; this is a straightforward consequence of the standard results on isometric spheres.

THEOREM 5.1. *Each Möbius map f can be represented in the form $f = ug$, where u is a unitary map, and g is a hyperbolic map with antipodal fixed points (or I).*

Proof. We may assume that f is not unitary (else we take $u = f$ and $g = I$). Then the action of the conjugate map $f^* = \varphi f \varphi^{-1}$ on the unit ball is given by $f^* = \alpha\beta$, where β is the inversion in the isometric sphere $\mathcal{S}$ of f, and α is some orthogonal map. Let ℓ be the Euclidean line that passes through 0 and the centre of $\mathcal{S}$, and let γ be the reflection in the plane through 0 that is orthogonal to ℓ. Then $f = (\alpha\gamma)(\gamma\beta)$, where $\alpha\gamma$ is unitary and $\gamma\beta$ is hyperbolic with antipodal fixed points. □

We can now complete the proof of Theorem 1.3, and we first prove this in the case of a hyperbolic map with antipodal fixed points.

LEMMA 5.2. *If g is hyperbolic with antipodal fixed points then*

$$d(g, \mathcal{U}) = 2\tanh \tfrac{1}{2}\rho(j, g(j)).$$

Proof. We know from (1.8) that

$$2\tanh \tfrac{1}{2}\rho(j, g(j)) \leqslant d(g, \mathcal{U}).$$

Also, Gehring and Martin prove in [**4**, Theorem 3.19] that equality holds in (1.7) when f is hyperbolic with antipodal fixed points. Thus $2\tanh \frac{1}{2}\rho\big(j, g(j)\big) = d(g, I) \geqslant d(g, \mathcal{U})$. □

Now let f be a general Möbius map, and write $f = ug$, where u is unitary and g is hyperbolic with antipodal fixed points (or I). Then

$$d(f, \mathcal{U}) = d(g, \mathcal{U}) = 2\tanh \tfrac{1}{2}\rho(j, g(j)) = 2\tanh \tfrac{1}{2}\rho(j, f(j))$$

as required. Finally, from (2.1) we obtain

$$d(f, \mathcal{U}) = 2\sqrt{\frac{\|f\|^2 - 2}{\|f\|^2 + 2}}.$$

6. *The convergence theorem*

We finish by returning to the original motivation for this paper, namely that if a sequence of Möbius maps converges at three distinct points to three distinct values, then it converges uniformly on $\mathbb{C}_\infty$ to a Möbius map. Theorem 1.1 implies that if a sequence of Möbius maps converges to I on $\{1, \omega, \omega^2\}$, then it converges to I uniformly on $\mathbb{C}_\infty$. The extension to the general case is easy because, for any Möbius map h,

$$\|h\|^{-2} q(z, w) \leqslant q\big(h(z), h(w)\big) \leqslant \|h\|^2\, q(z, w)$$

[**2**, pages 543–544]. Suppose that z_1, z_2 and z_3 are distinct, and that a sequence g_n of Möbius maps satisfies $g_n(z_j) \to w_j$, $j = 1, 2, 3$, where the w_j are distinct. We can choose Möbius maps r and s that map 1, ω and ω^2 to z_1, z_2 and z_3, and w_1, w_2 and w_3, respectively, and then

$$\begin{aligned} d\big(g_n, sr^{-1}\big) &\leqslant \|s\|^2\, d\big(s^{-1}g_n, s^{-1}sr^{-1}\big) \\ &= \|s\|^2\, d\big(s^{-1}g_n r, I\big) \\ &\leqslant 2\, \|s\|^2\, \varepsilon(s^{-1}g_n r) \\ &\leqslant 2\, \|s\|^4 \max\big\{q\big(g_n(z_j), w_j\big) : j = 1, 2, 3\big\}. \end{aligned}$$

We deduce that $g_n \to sr^{-1}$ with the given rate of convergence.

References

1. A. F. Beardon, *The geometry of discrete groups* (Graduate Texts in Mathematics, 91, Springer-Verlag, New York, 1983).
2. A. F. Beardon, 'Continued fractions, discrete groups and complex dynamics', *Comput. Methods and Funct. Theory* 1 (2001), no. 2, 535–594.
3. F. W. Gehring and G. J. Martin, 'The matrix and chordal norms of Möbius transformations', *Complex Analysis*, 51–59 (Birkhäuser, Basel, 1988).
4. F. W. Gehring and G. J. Martin, 'Inequalities for Möbius transformations and discrete groups', *J. Reine Angew. Math.* 418 (1991), 31–76.

Alan F. Beardon and Ian Short
Department of Pure Mathematics and Mathematical Statistics
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0JB
United Kingdom

a.f.beardon@dpmms.cam.ac.uk
ims25@cam.ac.uk