

Numerical solution of the omitted area problem of univalent function theory

Lehel Banjai and Lloyd N. Trefethen

Abstract

The omitted area problem was posed by Goodman in 1949: what is the maximum area \mathcal{A}^* of the unit disk D that can be omitted by the image of the unit disk under a univalent function normalized by $f(0) = 0$ and $f'(0) = 1$? The previous best bounds were $0.240005\pi < \mathcal{A}^* \leq .31\pi$. Here the problem is addressed numerically and it is found that these estimates are slightly in error. To ten digits, the correct value appears to be $\mathcal{A}^* = 0.2385813248\pi$.

Keywords: omitted area problem, numerical conformal mapping, univalent function theory.

Mathematics Subject Classification: 30C30, 30C75, 65E05.

1 Introduction

One of the major branches of complex analysis is univalent function theory: the study of one-to-one analytic functions f of the unit disk $D = \{z : |z| < 1\}$ conventionally normalized to have Taylor series

$$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots. \quad (1)$$

Many papers and books have been written about the properties of the class S of such functions. A celebrated result in this area is *Bieberbach's Conjecture* (1916), which became *de Branges Theorem* (1985): for any $f \in S$, the Taylor coefficients satisfy $|a_k| \leq k$ [5, 9, 10].

Our subject here is another well-known problem of univalent function theory. For each $f \in S$, let $\mathcal{A}(f)$ denote the area of $D \setminus f(D)$. The example

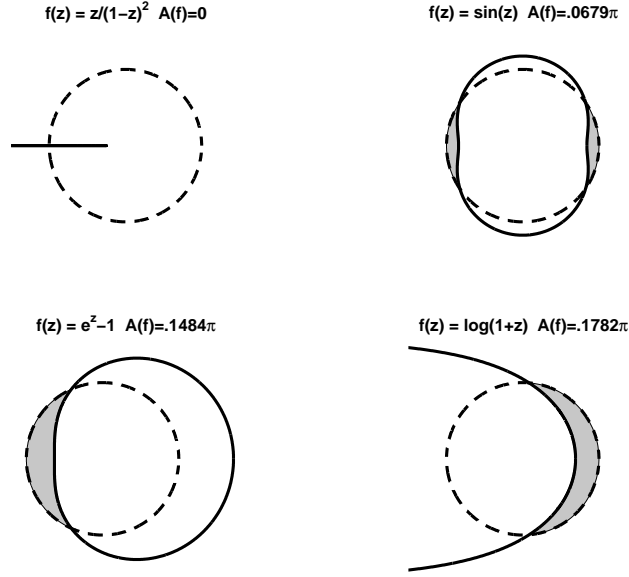


Figure 1: Four examples of functions $f \in S$. The dashed curve is the unit circle.

$f(z) = z$ shows that $\mathcal{A}(f)$ can be as small as 0. Since $f(D)$ always contains the disk about 0 of radius $\frac{1}{4}$ [9], it can be no larger than $15\pi/16$. How large can it be? That is, what is the value of the constant

$$\mathcal{A}^* = \sup_{f \in S} \mathcal{A}(f) ? \quad (2)$$

Moreover, what can be said about the function or functions f^* , if any, that achieve this supremum? This is the *omitted area problem*.

Figure 1 shows four functions $f \in S$ and the corresponding values of $\mathcal{A}(f)$. From these examples we see that \mathcal{A}^* is at least as large as 0.1782π . Here are all the results we know of that have been published bounding \mathcal{A}^* from above and below:

Goodman 1949	[7]	$.2272\pi < A^* \leq .5\pi$
Jenkins 1953	[11]	$.2272\pi < A^* \leq .4613\pi$
Goodman & Reich 1955	[8]	$.2272\pi < A^* \leq .38\pi$
Barnard & Pearce 1985	[4]	$.240005\pi < A^* \leq .38\pi$
Barnard & Lewis 1987	[3]	$.240005\pi < A^* \leq .31\pi$

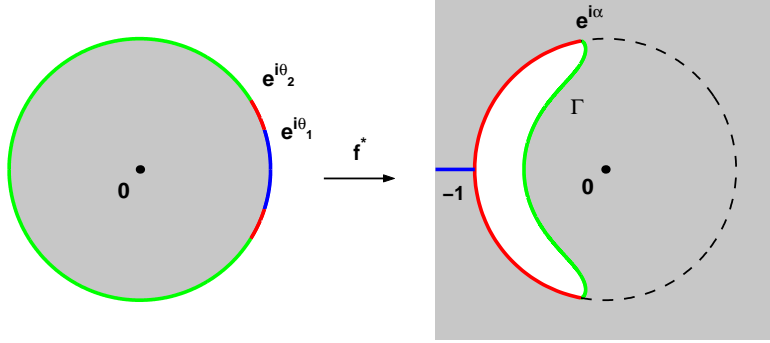


Figure 2: The form of the extremal function f^* , according to Theorem 1. f^* is symmetric about the real axis, and coloured parts of the unit circle are mapped to the parts of the boundary of the image domain of matching colour.

These estimates result from theoretical arguments supported to varying degrees by numerical computations, and one could discuss to what extent each should or should not be regarded as a theorem. Our own results to be presented here are purely numerical, and give an estimate of \mathcal{A}^* that we do not claim as a theorem but believe is correct to many digits. Our number falls a fraction of a percent below the final lower bound in the table above, which we conclude is in error.

An almost full characterization is known of the extremal functions for the omitted area problem. We trust that the following notation involving functions of intervals $[\cdot, \cdot]$ and (\cdot, \cdot) is self-explanatory.

Theorem 1. *There exist functions $f^* \in S$ that attain the supremum in (2). If a function $f \in S$ has this property, then it takes the form illustrated in Figure 2: f is symmetric about the real axis and maps three arcs $\exp(i[0, \theta_1])$, $\exp(i[\theta_1, \theta_2])$, and $\exp(i[\theta_2, \pi])$ of the unit circle to $[-\infty, -1]$, to the circular arc $\exp(i[\alpha, \pi])$ for some $\alpha \in (0, \pi)$, and to a curve Γ interior to the unit disk characterized by a condition of constant modulus of the derivative:*

$$|f'(z)| = \text{const.}, \quad z \in \exp(i(\theta_2, \pi)). \quad (3)$$

Moreover, f' is continuous along the arc $\exp(i[\theta_2, \pi])$, in particular at the point $\exp(i\theta_2)$ that maps to $e^{i\alpha}$.

This theorem is due to Barnard and Lewis [2, 3, 12]¹. All that is missing is a statement about uniqueness. It seems likely that f^* is unique, but it appears that this has not been proved. Nevertheless in what follows, for simplicity, we shall speak of “the function f^* ”.

2 Simplification by conformal maps

Theorem 1 establishes so much about f^* that one might imagine that it would have led quickly to a numerical solution of the omitted area problem. However, it is not a straightforward matter to construct functions to high accuracy that satisfy the required conditions. Indeed, we approached this problem with the confidence of hands-on numerical analysts and were surprised at how challenging it proved. Along the way we tried a method that we shall not describe, which involved the numerical construction of Schwarz-Christoffel maps onto polygons with many vertices by means of Driscoll’s Schwarz-Christoffel Toolbox for MATLAB [6]. We were unable to obtain more than four or five digits of accuracy by that method, but those digits agreed with the results we will now present from our more efficient method. This lends us confidence that our final value is probably accurate.

Our computation is based on the construction of f as a composition of conformal maps, each symmetric with respect to the real axis:

$$f(z) = f_3(f_2(f_1(z))). \quad (4)$$

The functions f_1 and f_3 contain boundary singularities, but are known analytically. The function f_2 is unknown, but smooth, and can be represented by a rapidly converging Taylor series. We now describe f_1 , f_2 and f_3 in turn.

As sketched in Figure 3, f_1 maps D onto the left half of the unit disk, which we denote by C (“crescent”). In view of the real symmetry condition, there is just one free parameter in this map, which we take to be the angle θ_2 , $0 < \theta_2 < \pi$, such that $f_1(e^{i\theta_2}) = i$. The formula is

$$f_1(z) = i \frac{\sqrt{1 - ze^{i\theta_2}} - \sqrt{z - e^{i\theta_2}}}{\sqrt{1 - ze^{i\theta_2}} + \sqrt{z - e^{i\theta_2}}}$$

¹In fact, not all of Theorem 1 as we have stated it appears in the published papers [2, 3, 12], for these papers leave open the possibility that Γ may contain slits along the real axis. We are informed privately by Barnard, however, that this possibility has been excluded in subsequent work.

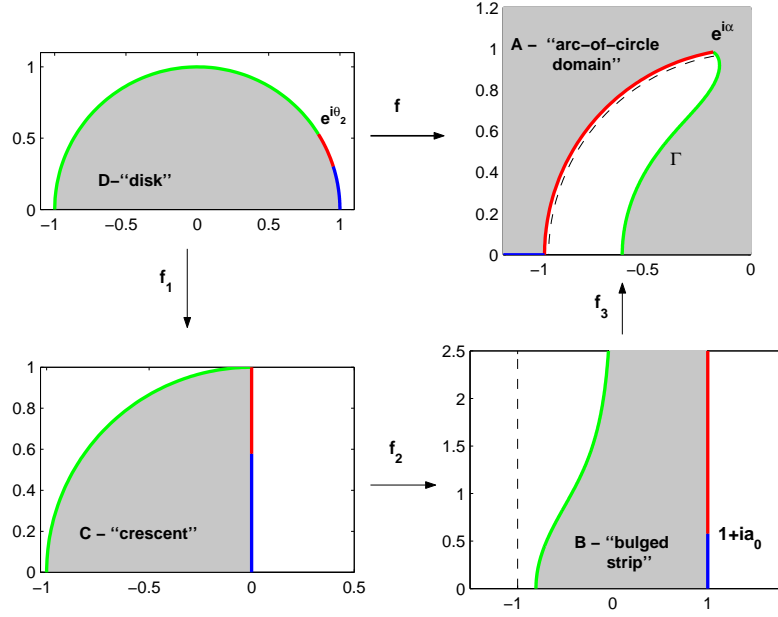


Figure 3: Construction of f as a composition of three maps.

with appropriate choices of branches.

The function f_2 maps C to a region B (“bulged strip”) bounded on the right by the line $\operatorname{Re} z = 1$ and on the left by a smooth curve whose real part approaches 0 as $z \rightarrow \pm i\infty$. We represent f_2 in the form

$$f_2(z) = p(z) + \frac{2i}{\pi} \log \left(\frac{z+i}{z-i} \right) - 1$$

for some function $p(z)$ analytic in D . The term containing the logarithm maps the half-disk to the infinite strip bounded by $\operatorname{Re} z = 0$ and $\operatorname{Re} z = 1$, and p provides the bulge on the left. Since p maps the imaginary axis into itself, its Taylor series takes the form

$$p(z) = d_1 z + d_3 z^3 + d_5 z^5 + \cdots \quad (5)$$

for some real coefficients d_j . Our construction guarantees that this representation is valid, with the Taylor series converging throughout D , and in section 4 we shall present numerical evidence that the convergence is geometric (so that the radius of convergence is in fact > 1).

Finally, the function f_3 is chosen to map the half-strip $-1 < \operatorname{Re} z < 1$, $\operatorname{Im} z > 0$ to the upper half-plane minus the circular arc $\exp(i[\alpha, \pi])$, with $\operatorname{Re} z = -1$ mapping to the inside of the arc and $\operatorname{Re} z = 1$ to the outside of the arc together with the interval $[-\infty, -1]$. Note that this map is again determined by just one parameter, for example α , but it turns out to be more convenient to have a number $a_0 > 0$ as the parameter, so that f_3 maps $1 + ia_0$ to -1 . The full strip contains B as a subset, and thus f_3 maps the upper half of B to a subset of the upper half-plane minus the arc. This subset is a domain A (“arc-of-circle domain”) of the form described by Theorem 1, except not in general satisfying the constant-modulus condition along Γ . The curved left boundary of the upper half of B maps to Γ , and as just mentioned, the straight right boundary maps to $[-\infty, -1]$ and $\exp(i[\alpha, \pi])$.

In addition these maps must satisfy the two normalizing conditions:

$$f(0) = 0, \quad f'(0) = 1. \quad (6)$$

With θ_2 and a_0 fixed, the functions f_1 and f_2 are determined, and hence the above two conditions reduce to two linear equations on the coefficients d_j :

$$f_2(f_1(0)) = f_3^{-1}(0), \quad (7)$$

$$f_2'(f_1(0)) = \frac{1}{f_3'(f_2(f_1(0))) f_1'(0)}. \quad (8)$$

To satisfy these conditions, we need two free parameters, and these we choose to be d_1 and d_3 .

An explicit specification of the function f_3 is elementary, but tricky. One could write down a single formula, but as a practical matter, for maps like this that are to be implemented on the computer without mistakes involving branches, we find it safest to work with compositions of elementary maps. This could be done in various ways. Our choice has been to take

$$f_3 = g_5 \circ g_4 \circ g_3 \circ g_2 \circ g_1$$

with

$$g_1(z) = \sin\left(\frac{\pi z}{2}\right),$$

$$g_2(z) = \left(\frac{2(z+1)}{a_1+1} - 1\right)^2,$$

$$g_3(z) = \frac{z(b-1)}{b(z-1)},$$

$$g_4(z) = z^{1/2},$$

$$g_5(z) = \frac{1+z}{1-z},$$

where $a_1 = g_1(1 + ia_0)$ and $b = g_2(1)$.

Here is a brief description of the behaviour of these maps, taking the argument of g_1 to be a point z in the upper half-plane; the lower half is obtained by reflection (though this is never needed in our numerical solution). The function g_1 maps the upper half-strip bounded by $\operatorname{Re} z = -1$ and $\operatorname{Re} z = 1$ to the upper half-plane; three particular boundary values are $g_1(-1, 1, \infty) = (-1, 1, \infty)$. The function g_2 maps the half-plane to the whole complex plane minus the slit $[0, \infty)$, with $g_2(-1, a_1, \infty) = (1^-, 1^+, \infty)$, where 1^- and 1^+ denote the numbers 1 on the lower and upper side of the slit, respectively. The function g_3 is a Möbius transformation that maps the slit plane to the plane with two slits extending to ∞ with a finite gap between; we have $g_3(0, b, 1) = (0, 1, \infty)$, where $b = g_2(1)$. The function g_4 folds the plane with two slits back to the upper half-plane with a single slit extending from a point along the positive imaginary axis to ∞ . Finally, g_5 is another Möbius transformation that maps the slit upper half-plane to the arc-of-circle domain A .

3 Computing the area of an arc-of-circle domain

From Theorem 1 it follows that the area of the region bounded by $f(\exp(i[\theta_1, 2\pi - \theta_1]))$ and the unit circle, not containing 0, is equal to the omitted area $\mathcal{A}(f)$. Hence we can use the following formula [13, p.5]:

$$\mathcal{A}(f) = -\frac{1}{2} \int_{\theta_1}^{2\pi} \operatorname{Im} \left\{ \frac{\partial f(z)}{\partial \theta} \overline{f(z)} \right\} d\theta \quad (9)$$

$$= -\frac{1}{2} \int_{\theta_1}^{2\pi} \operatorname{Re} \left\{ z f'(z) \overline{f(z)} \right\} d\theta, \quad z = e^{i\theta}. \quad (10)$$

The minus sign is present since when the unit circle is traced in the positive direction, the image under f traces a curve around the omitted region in the negative direction.

We need to be able to compute $\mathcal{A}(f)$ quickly and to high accuracy. Using the symmetry property of f we get that

$$\mathcal{A}(f) = - \int_{\theta_1}^{\theta_2} \operatorname{Re} \left\{ z f'(z) \overline{f(z)} \right\} d\theta - \int_{\theta_2}^{\pi} \operatorname{Re} \left\{ z f'(z) \overline{f(z)} \right\} d\theta. \quad (11)$$

It can be seen that the first integral is equal to twice the area of the slice of the unit disk $\{re^{i\theta} : \alpha \leq \theta \leq \pi, r \leq 1\}$, i.e., $\pi - \alpha$. For the second integral, we expect that the integrand is analytic along $z = \exp(i(\theta_2, \pi])$ but has a singularity at $z = \exp(i\theta_2)$. Hence if we choose some $\beta > \theta_2$ we can split the integral into two integrals

$$I_1 = - \int_{\theta_2}^{\beta} \operatorname{Re} \left\{ z f'(z) \overline{f(z)} \right\} d\theta, \quad (12)$$

$$I_2 = - \int_{\beta}^{\pi} \operatorname{Re} \left\{ z f'(z) \overline{f(z)} \right\} d\theta, \quad (13)$$

where I_1 can be evaluated using standard adaptive quadrature software and I_2 can be efficiently evaluated using Gauss-Legendre quadrature. Hence

$$\mathcal{A}(f) = \pi - \alpha + I_1 + I_2. \quad (14)$$

We choose $\beta = \theta_2 + 0.01$ and evaluate I_1 using adaptive Simpson quadrature (the MATLAB function `quad`) with tolerance set to 10^{-14} , which requires about 70 evaluations of the integrand. We evaluate I_2 using Gauss-Legendre quadrature with 80 nodes which appears to give accuracy on the order of 10^{-15} .

4 Numerical methods and results

For any real parameters $d_5, d_7, \dots, d_{2N-1}, \theta_2$ and $a_0, 0 < \theta_2 < \pi, a_0 > 0$, we have constructed an analytic function f in the unit disk with the properties (6). The construction is such that, provided the coefficients d_j are small enough so that $f'(z) \neq 0$ for $z \in D$, this will be a univalent function mapping D conformally onto an arc-of-circle domain A . We now face a numerical problem: choosing these N parameters so that $\mathcal{A}(f)$ is maximal. There are two natural approaches to this: either maximize $\mathcal{A}(f)$ directly, or make use of the characterization (3) to enforce the constant modulus condition along a circular arc. We have tried both.

Let f be our solution to the omitted area problem using the above construction. From the construction it is not clear if f is univalent. Certainly the maps f_1 and f_3 are univalent inside their domains of definition. As a sum of an analytic function in C and a polynomial, f_2 is certainly analytic in C . Hence f is analytic on the open unit disk. Let $\exp(i\theta)$ trace the unit circle in the positive direction with θ strictly increasing. Then by construction, $f(\exp(i\theta))$ traces the boundary of A once in the positive direction. If the image of $\exp(i[\theta_1, 2\pi - \theta_1])$ under f is a simple curve, then by the argument principle, f is univalent. We check whether this last condition holds by discretizing the curve.

As just mentioned, we have implemented two methods for finding the extremal function for the omitted area problem. In the first method we set up an optimization problem with parameters $d_5, d_7, \dots, d_{2N-1}, \theta_2, a_0$. We iterate to maximize $\mathcal{A}(f)$. At every iteration, given all the parameters, coefficients d_1 and d_3 are determined so that the linear equations (7) and (8) are satisfied.

In the second method we search for functions satisfying the constant-modulus condition. The set of unknowns is slightly enlarged by the addition of a parameter M representing the value of $|f'|$ on $\exp(i[\theta_2, \pi])$. We set up an over-determined system and minimize the following expression

$$\sum_{k=1}^L \left((|f'(e^{i\gamma_k})| - M)^2 \right) \quad (15)$$

for $\theta_2 < \gamma_k < \pi$ with $L = 50N$. The points γ_k are chosen so as to be denser near θ_2 .

Both methods work, and their results agree. The second method proves to be more efficient. Its results suggest that the correct value of A^* is 0.2385813248π to ten digits. The convergence to this number using the two different methods is shown in Figure 4. Results of the two methods are shown in Tables 1 and 2.

5 The optimal function f^*

The error is decreasing exponentially, which suggests that the function $p(z)$ which we approximate by a truncated Taylor series, is analytic in the closed unit disk. Indeed the coefficients d_j suggest that the radius of convergence of the Taylor expansion is about 1.6. The two methods converge at the same

N	omitted area (π)
2	0.23830377896845
3	0.23854060029864
4	0.23857325048507
5	0.23857972028678
6	0.23858098546805
7	0.23858125225852
8	0.23858130840494

Table 1: Results for Method 1.

N	omitted area (π)	$\max_{\theta \in [\theta_2, \pi]} f'(e^{i\theta}) - M$
2	0.23719466292381	$5.6e - 2$
3	0.23838494717525	$2.5e - 2$
4	0.23854792113829	$1.2e - 2$
5	0.23857485970016	$5.6e - 3$
6	0.23858001311453	$2.7e - 3$
7	0.23858103791058	$1.3e - 3$
8	0.23858126165609	$6.7e - 4$
9	0.23858130994730	$3.4e - 4$
10	0.23858132142118	$1.7e - 4$
11	0.23858132397435	$8.9e - 5$
12	0.23858132463176	$4.5e - 5$
13	0.23858132477544	$2.4e - 5$
14	0.23858132481692	$1.2e - 5$

Table 2: Results for Method 2. The final value for the omitted area seems to be correct to 10 digits.

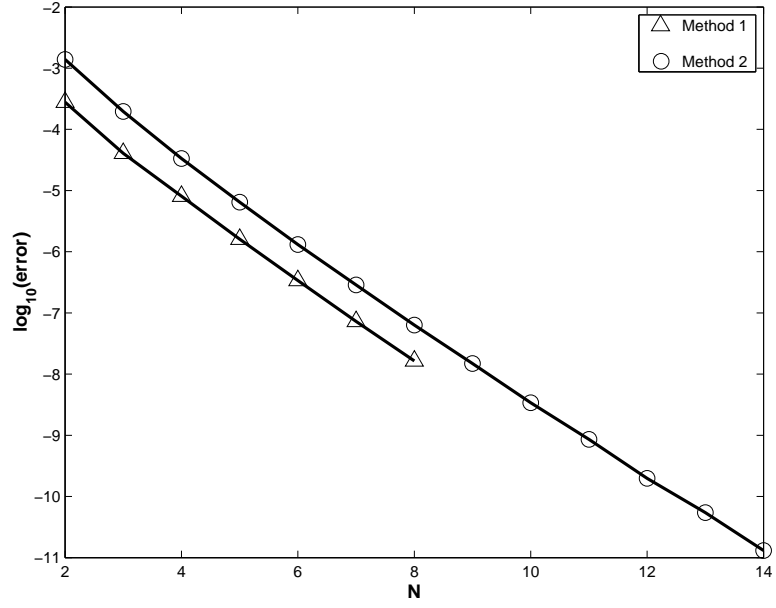


Figure 4: Convergence of the omitted area to the number 0.2385813248π using the two methods. These data points were calculated as the difference between the computed area for each N and the number 0.2385813248300π .

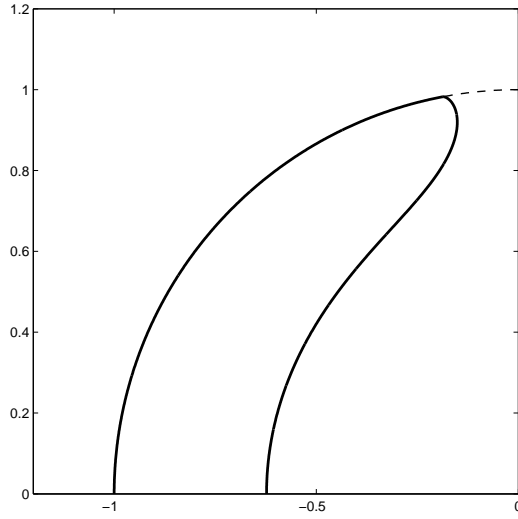


Figure 5: The omitted area domain for the extremal function f^* .

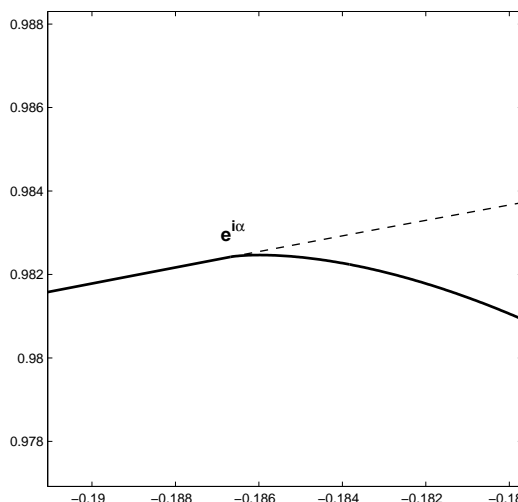


Figure 6: Closeup near the point $e^{i\alpha} = f^*(e^{i\theta_2})$, showing continuity of the derivative there.

rate, but for any fixed N , Method 1 gives a slightly better result. This is to be expected since at any stage the first method is finding the maximum area whereas the second method is finding the best approximation to the extremal function. Yet the optimization problem in the first method is much more difficult to solve. The final result took 7 hours to compute in MATLAB on a Pentium III 800 MHz processor, whereas all the data obtained using the second method can be computed on the same machine within an hour.

Our methods converge to the same number. Indeed the best two approximations to the extremal function using the two methods are equal within an error of 5.7×10^{-5} along the arc $\exp(i[\theta_2, \pi])$. As a result of our computations we propose the following value of \mathcal{A}^* :

$$\mathcal{A}^* = 0.2385813248\pi.$$

We believe that this number is probably accurate to the full ten digits displayed. The extremal domain is plotted in Figures 5 and 6. We also show images of concentric circles around 0 and radial lines under f^* in Figure 8. Figure 7 plots $(f^*)'(z)$ near $z = e^{i\theta_2}$. The values of the parameters of our best approximation to the extremal function are the following: $M = 0.44074691$, $a_0 = 0.5787293$, $\theta_2 = 0.5575142$ and the coefficients d_j are shown in Table 3.

j	d_j
1	0.66674444254256
3	0.10151497157267
5	0.02603062412472
7	0.00831506726776
9	0.00289083802642
11	0.00112217881059
13	0.00044351035572
15	0.00019260405260
17	0.00008045133385
19	0.00003819142146
21	0.00001585332461
23	0.00000850626700
25	0.00000308999675
27	0.00000218932186

Table 3: The coefficients d_j for our best approximation to the extremal function f^* . Note that the even coefficients are 0.

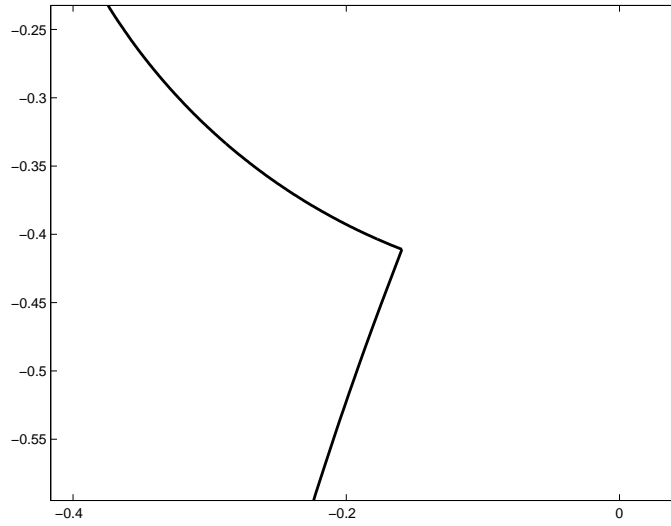


Figure 7: Plot of $(f^*)'(e^{i\theta})$ for θ near θ_2 .

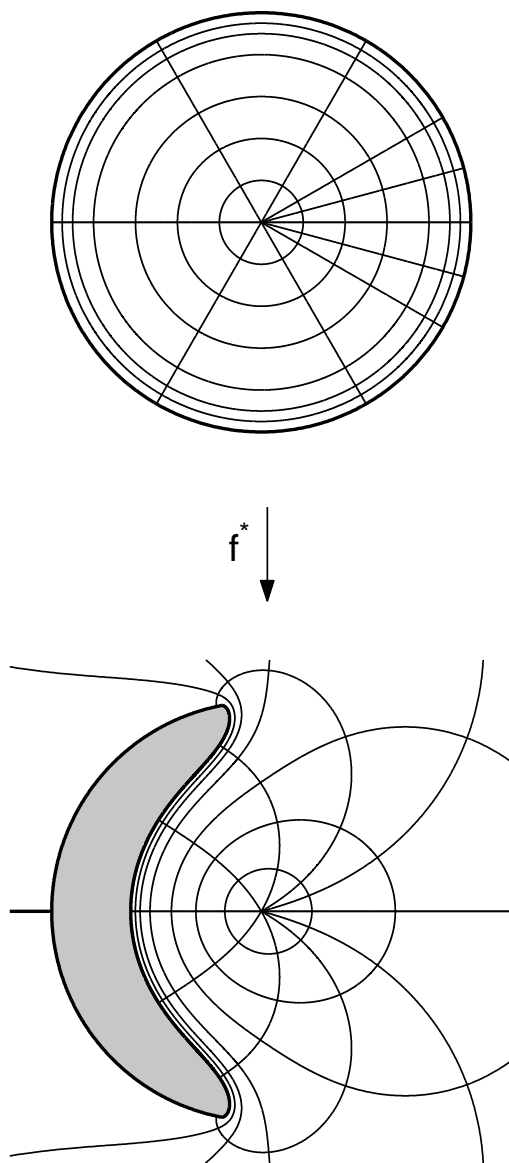


Figure 8: Images under f^* of circles of radius $r = .2, .4, .6, .8, .9, .95$ centred at 0, and of radial lines at angles $\theta = 0, \pm\pi/12, \pm\pi/6, \pm\pi/3, \pm2\pi/3, \pi$. In the lower plot, one can see from the uniform spacing of the curves near the boundary arc Γ that $(f^*)'$ has constant modulus there.

References

- [1] Ahlfors, L.V., *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable*, Cambridge University Press, New York, 1997.
- [2] Barnard, R.W., *The omitted area problem for univalent functions*, Contemp. Math. **38** (1985) 53–60.
- [3] Barnard, R.W.; J.L. Lewis, *On the omitted area problem*, Michigan Math. J. **34** (1987), no.1 13–22.
- [4] Barnard, R.W.; K. Pearce, *Rounding corners of gearlike domains and the omitted area problem*, J. Comput. Appl. Math. **14** (1986), no.1-2 217–226.
- [5] de Branges, L., *A proof of the Bieberbach conjecture*, Acta Math. **154**. 137–152.
- [6] Driscoll, T.A., Algorithm 765: *A MATLAB toolbox for Schwarz-Christoffel mapping*, ACM Trans. Math. Softw. **22** (1996), 168–186. Software available at <http://www.math.udel.edu/~driscoll/SC/>
- [7] Goodman, A.W., *Note on regions omitted by univalent functions*, Bull. Amer. Math. Soc. **55** (1949) 363–369.
- [8] Goodman, A.W.; E. Reich, *On regions omitted by univalent functions II*, Canad. J. Math. **7** (1955) 83–88.
- [9] Hayman, W.K., *Multivalent Functions*, Cambridge University Press, Cambridge, 1994.
- [10] Henrici, P., *Applied and Computational Complex Analysis, vol. III*, Wiley, New York, 1986.
- [11] Jenkins, J.A., *On values omitted by univalent functions*, Amer. J. Math., **75** (1953), 406–408.
- [12] Lewis, J.L., *On the omitted area problem*, Indiana University Mathematics Journal, Vol 34, No. 3 (1985).

- [13] Markushevich, A. I., *Theory of Functions of a Complex Variable*, Chelsea Publishing Co., New York, 1985, vol. III.
- [14] Trefethen, L. N., *Spectral Methods in MATLAB*, SIAM, 2000.

LLOYD N. TREFETHEN, OXFORD UNIVERSITY COMPUTING
LABORATORY, WOLFSON BUILDING, PARKS ROAD, OXFORD
OX1 3QD, UK

E-mail address: LNT@comlab.ox.ac.uk

LEHEL BANJAI, OXFORD UNIVERSITY COMPUTING LABORA-
TORY, WOLFSON BUILDING, PARKS ROAD, OXFORD OX1
3QD, UK

E-mail address: lehelb@comlab.ox.ac.uk

A Code to compute the extremal function

```
function f = extremal(z)      % Evaluate f(z) for |z|<=1

if any(imag(z)<0) % Im z<0 handled by reflection
    L = imag(z)<0;
    f(~L) = extremal(z(~L));
    f(L) = conj(extremal(conj(z(L))));
    return
end

theta2 = .5575142;
a0 = .5787293;
d = [ 0.6667444425 0.1015149716 0.0260306241 0.0083150673 ...
      0.0028908380 0.0011221788 0.0004435104 0.0001926041 ...
      0.0000804513 0.0000381914 0.0000158533 0.0000085063 ...
      0.0000030900 0.0000021893 ];

p = kron(fliplr(d),[1 0]);
a1 = real(sin((pi*(1+a0*i))/2));
b = (4/(a1+1)-1)^2;
f0 = (1-z*exp(i*theta2))./(z-exp(i*theta2));
J = abs(abs(z)-1) <= 1e-15;
f0(J) = real(f0(J));
f1 = i*(sqrt(f0)-1)./(sqrt(f0)+1);
f2 = polyval(p,f1)+(2i/pi)*(log(f1+i)-log(f1-i))+3;
g1 = sin((pi*f2)/2);
g2 = (2*(g1+1)/(a1+1)-1).^2;
g3 = (1-1/b)*g2./(g2-1);
K = real(g1)<(a1-1)/2;
g4(K) = -sqrt(g3(K)); g4(~K) = sqrt(g3(~K));
f = (1+g4)./(1-g4);
```