Inequalities for the perimeter of an ellipse

G.J.O. Jameson, *Math. Gazette* **98** (2014)

The perimeter of the ellipse $x^2/a^2 + y^2/b^2 = 1$ is $4J(a,b)$, where $J(a,b)$ is the "elliptic integral"

$$J(a,b) = \int_0^{\pi/2} (a^2 \cos^2\theta + b^2 \sin^2\theta)^{1/2}\, d\theta. \tag{1}$$

This integral is interesting in its own right, quite apart from its application to the ellipse. It is often considered together with the companion integral

$$I(a,b) = \int_0^{\pi/2} \frac{1}{(a^2 \cos^2\theta + b^2 \sin^2\theta)^{1/2}}\, d\theta. \tag{2}$$

Of course, we may as well assume that a and b are non-negative. We record first some elementary facts about these integrals:

(E1) $J(a,a) = \int_0^{\pi/2} a\, d\theta = \frac{\pi a}{2}, \qquad I(a,a) = \frac{\pi}{2a}$;

(E2) $J(a,0) = \int_0^{\pi/2} a\cos\theta\, d\theta = a, \qquad I(a,0)$ is undefined;

(E3) for $c > 0$, $J(ca,cb) = cJ(a,b)$ and $I(ca,cb) = \frac{1}{c} I(a,b)$;

(E4) $J(b,a) = J(a,b)$ and $I(b,a) = I(a,b)$ (substitute $\theta = \frac{\pi}{2} - \phi$);

(E5) $J(a,b)$ increases with a and with b, and $I(a,b)$ decreases.

In general, neither integral is amenable to evaluation by simply writing down an anti-derivative. However, both can be evaluated in terms of the *arithmetic-geometric mean* $M(a,b)$ of a and b. This is the common limit of the sequences (a_n) and (b_n) defined by the iteration $a_0 = a,\ b_0 = b$ and

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n), \qquad b_{n+1} = (a_n b_n)^{1/2}.$$

Write $a_n^2 - b_n^2 = c_n^2$ and $S = \sum_{n=0}^{\infty} 2^{n-1} c_n^2$. It turns out that

$$I(a,b) = \frac{\pi}{2M(a,b)}, \tag{3}$$

$$J(a,b) = (a^2 - S) I(a,b). \tag{4}$$

Moreover, the iteration converges very rapidly, so these identities amount to an effective way to calculate the integrals. Identity (3) is a famous theorem of Gauss, dating from 1799. Identity (4) is the basis of the very efficient Gauss-Brent-Salamin algorithm for calculation of π. We will not enter into any details here; for a highly readable account, see [Lo].

While (4), in principle, is an exact evaluation of $J(a,b)$, it certainly does not convey a quick and transparent indication of its magnitude in terms of a and b. This purpose is better achieved by inequalities comparing $J(a,b)$ with simple expressions like $a+b$ and $(a^2+b^2)^{1/2}$. One lower bound, $\frac{\pi}{4}(a+b)$, was given in [LS], as a solution to a *Gazette* problem. However, bounds of this sort are accurate for some values of b/a and less accurate for others, so for a really satisfactory estimation of $J(a,b)$, more than one lower bound (and equally, more than one upper bound) is needed. Here we will describe three different ways to derive such bounds directly from (1). I hope that some readers will share my view that the methods themselves are as interesting as the conclusions.

First, we mention a pair of bounds that follow at once from elementary facts (E1) and (E5): if $a \geq b$, then $J(b,b) \leq J(a,b) \leq J(a,a)$, so that

$$\frac{\pi}{2} b \leq J(a,b) \leq \frac{\pi}{2} a. \tag{5}$$

Geometrically, this is saying that the perimeter of the ellipse lies between those of the inscribed and circumscribed circles.

Another pair of bounds that seem geometrically obvious is

$$(a^2+b^2)^{1/2} \leq J(a,b) \leq a+b, \tag{6}$$

since this says that the curve of the quarter-ellipse is longer than the straight-line path between the same points, but shorter than two sides of the rectangle. However, it is instructive to see how these inequalities can be proved analytically. For the right-hand inequality, this is very easy: since $(x^2+y^2)^{1/2} \leq x+y$ for positive x, y, we have $(a^2\cos^2\theta + b^2\sin^2\theta)^{1/2} \leq a\cos\theta + b\sin\theta$ for each θ. Integration on $[0,\pi/2]$ gives $J(a,b) \leq a+b$.

We will improve this inequality below, as well as giving an analytic proof of the left-hand inequality in (6). We will also establish inequalities in the reverse direction to both sides of (6), with suitable constants inserted.

Our first method uses the *Cauchy-Schwarz inequality*, which states: *for non-negative numbers* a_r, b_r $(1 \leq r \leq n)$, *we have*

$$\sum_{r=1}^{n} a_r b_r \leq \left(\sum_{r=1}^{n} a_r^2\right)^{1/2} \left(\sum_{r=1}^{n} b_r^2\right)^{1/2}.$$

This can be stated very concisely in vector notation: it says $\boldsymbol{a}.\boldsymbol{b} \leq |\boldsymbol{a}|.|\boldsymbol{b}|$, where the length of the n-dimensional vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ is $|\boldsymbol{a}| = (\sum_{r=1}^{n} a_r^2)^{1/2}$, and $\boldsymbol{a}.\boldsymbol{b}$ means $\sum_{r=1}^{n} a_r b_r$. The proof is short and neat, so we repeat it: writing $\boldsymbol{a}.\boldsymbol{b} = S$, we have, for any λ,

$$0 \leq \sum_{r=1}^{n} (a_r - \lambda b_r)^2 = |\boldsymbol{a}|^2 - 2\lambda S + \lambda^2 |\boldsymbol{b}|^2.$$

With λ chosen to be $S/|\boldsymbol{b}|^2$, this says $|\boldsymbol{a}|^2 - S^2/|\boldsymbol{b}|^2 \geq 0$, hence $S \leq |\boldsymbol{a}|.|\boldsymbol{b}|$.

There is a corresponding statement for integrals, proved in an analogous way. We write $\int_a^b f$ (this is perfectly adequate notation!) for $\int_a^b f(x)\,dx$. The inequality states: *if f and g are non-negative, integrable functions on $[a,b]$, then*

$$\int_a^b fg \leq \left(\int_a^b f^2\right)^{1/2} \left(\int_a^b g^2\right)^{1/2}.$$

The Cauchy-Schwarz inequality enables us to give a rather elegant analytic proof of the left-hand inequality in (6), as follows:

PROPOSITION 1. *We have*

$$J(a,b) \geq (a^2+b^2)^{1/2}. \tag{7}$$

Proof. By the Cauchy-Schwarz inequality,

$$a^2\cos\theta + b^2\sin\theta = a(a\cos\theta) + b(b\sin\theta) \leq (a^2+b^2)^{1/2}(a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}.$$

Integrating on $[0,\pi/2]$, we obtain

$$a^2+b^2 \leq (a^2+b^2)^{1/2} J(a,b),$$

hence (7). □

Note that (7) is exact when $b=0$, since $J(a,0)=a$, but not when $b=a$. A variation of the reasoning gives a second lower bound which is exact when $b=a$ but not when $b=0$, and which also improves upon the one in (5). This bound was established in [LS], by essentially the same method (though without explicit mention of the Cauchy-Schwarz inequality).

PROPOSITION 2. *We have*

$$J(a,b) \geq \frac{\pi}{4}(a+b). \tag{8}$$

Equality holds when $a=b$.

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned} a\cos^2\theta + b\sin^2\theta &= (a\cos\theta)\cos\theta + (b\sin\theta)\sin\theta \\ &\leq (a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}(\cos^2\theta+\sin^2\theta)^{1/2} \\ &= (a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}. \end{aligned}$$

Integration gives (8), since $\int_0^{\pi/2}\cos^2\theta\,d\theta = \int_0^{\pi/2}\sin^2\theta\,d\theta = \frac{\pi}{4}$. Equality holds when $a=b$, by (E1). □

Note that this says that the perimeter of the whole ellipse is at least $\pi(a+b)$. By the inequality of the means, this is not less than $2\pi(ab)^{1/2}$, still with equality occurring when $a=b$. As observed in [LS], this means that among ellipses with a given area, the one with the smallest perimeter is the circle. Equally, among ellipses with a given perimeter, the circle is the one with the largest area.

An upper bound for $J(a,b)$ is provided by the Cauchy-Schwarz inequality for integrals:

PROPOSITION 3. *We have*

$$J(a,b) \le \frac{\pi}{2\sqrt{2}}(a^2+b^2)^{1/2}. \tag{9}$$

Equality holds when $a=b$.

Proof. Apply the Cauchy-Schwarz inequality for integrals with f written as $f.1$ and both sides squared. We obtain

$$J(a,b)^2 \le \left(\int_0^{\pi/2} 1\right) \int_0^{\pi/2} (a^2\cos^2\theta + b^2\sin^2\theta)\, d\theta = \frac{\pi}{2}\,\frac{\pi}{4}\,(a^2+b^2). \qquad \square$$

Since $\pi/(2\sqrt{2}) \approx 1.1107$, the pair of bounds in (7) and (9) shows that $J(a,b)$ is actually modelled fairly well by $(a^2+b^2)^{1/2}$.

We now describe a second strategy, based on the triangle inequality for vectors. Note that, in the notation for the length of vectors,

$$J(a,b) = \int_0^{\pi/2} |(a\cos\theta, b\sin\theta)|\, d\theta.$$

The familiar triangle inequality for vectors says that $|\boldsymbol{u}_1+\boldsymbol{u}_2| \le |\boldsymbol{u}_1| + |\boldsymbol{u}_2|$. This is geometrically obvious in the case $n=2$, and in general it follows easily from the Cauchy-Schwarz inequality. We show that $J(a,b)$ also satisfies the triangle inequality, in the following sense:

PROPOSITION 4. *Let $\boldsymbol{u}_j = (a_j, b_j)$ $(j=1,2)$, where a_j, $b_j \ge 0$. Then*

$$J(\boldsymbol{u}_1+\boldsymbol{u}_2) \le J(\boldsymbol{u}_1) + J(\boldsymbol{u}_2). \tag{10}$$

Proof. By the ordinary triangle inequality, we have for each θ

$$|[(a_1+a_2)\cos\theta, (b_1+b_2)\sin\theta]| \le |(a_1\cos\theta, b_1\sin\theta)| + |(a_2\cos\theta, b_2\sin\theta)|.$$

Integrating on $[0, \frac{\pi}{2}]$, we deduce that

$$J(a_1+a_2, b_1+b_2) \le J(a_1,b_1) + J(a_2,b_2),$$

which equates to (10). $\square$

Note. Since $J(a,b) = J(|a|,|b|)$, (10) actually holds without the condition that the components are non-negative. In the usual parlance, $J(a,b)$ is a *norm* on the vector space $\mathbb{R}^2$, and we are comparing it with other norms like $|a|+|b|$ and $(a^2+b^2)^{1/2}$.

By suitable choices of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, we can read off various inequalities for $J(a,b)$. Firstly, $J(a,b) \le J(a,0) + J(0,b) = a+b$, as seen in (6). Secondly:

Second proof of (8). Since $(a+b, a+b) = (a,b) + (b,a)$, we have

$$(a+b)\frac{\pi}{2} = J(a+b, a+b) \le J(a,b) + J(b,a) = 2J(a,b). \qquad \square$$

Thirdly, we can derive an upper bound that simultaneously improves on those in (5) and (6), and is exact at both ends, at the cost of being unsymmetrical:

PROPOSITION 5. *For $0 \le b \le a$, we have*

$$J(a,b) \le a + \left(\frac{\pi}{2} - 1\right) b. \tag{11}$$

Equality holds when $b = 0$ and when $b = a$.

Proof. Since $(a,b) = (a-b, 0) + (b,b)$,

$$J(a,b) \le J(a-b,0) + J(b,b) = (a-b) + \frac{\pi}{2}b. \qquad \square$$

This bound is more natural than one might at first think. For fixed a, it represents the linear function of b that agrees with $J(a,b)$ at $b=0$ and $b=a$. In fact, (11) reflects the fact that $J(a,b)$ is a *convex* function of b. Recall that a function is "convex" if it lies below the straight-line chords between pairs of points of its graph. This means that if $b_1 < b_2$ and $b_\lambda = (1-\lambda)b_1 + \lambda b_2$, where $0 < \lambda < 1$, then

$$f(b_\lambda) \le (1-\lambda) f(b_1) + \lambda f(b_2).$$

By Proposition 4 and (E3), this holds with $f(b) = J(a,b)$, since $(a, b_\lambda) = (1-\lambda)(a,b_1) + \lambda(a,b_2)$.

Comparison of the bounds, and a question about the derivative

Our two lower bounds for $J(a,b)$ are $m_1 = \frac{\pi}{4}(a+b)$ (exact when $b=a$) and $m_2 = (a^2+b^2)^{1/2}$ (exact when $b=0$). To explore how they compare, first fix $a=1$. The two bounds coincide when $1 + b^2 = \frac{\pi^2}{16}(1+b)^2$. The solutions of this quadratic are b_1 and $1/b_1$, where $b_1 \approx 0.34823$. The better (i.e. larger) lower bound is m_2 when $0 < b < b_1$ and when

$b > 1/b_1$. For general a, apply this to b/a to conclude that m_2 is the larger one when either b/a or a/b is less than b_1.

Meanwhile, for $b \le a$, our upper bounds are $M_1 = a + (\frac{\pi}{2} - 1)b$ (exact at both ends) and $M_2 = \frac{\pi}{2\sqrt{2}}(a^2 + b^2)^{1/2}$ (exact when $b = a$). When $a = 1$, the bounds coincide when $[1 + (\frac{\pi}{2} - 1)b]^2 = \frac{\pi^2}{8}(1 + b)^2$. One solution is $b = 1$, and the other is $b_2 \approx 0.25741$. The better (i.e. smaller) upper bound is M_1 when $\frac{b}{a} < b_2$.

Fixing $a = 1$, write $J_1(b)$ for $J(1, b)$. Do we now have enough information to give an accurate sketch of $J_1(b)$ as a function of b on the interval $[0, 1]$? The reader may care to draw a sketch of our four bounds on this interval, and it will be seen that between them they confine $J_1(b)$ to a fairly narrow box. But one ingredient for the sketch is still missing: it would be desirable to know the derivative, at least at the end-points. At $b = 1$, we do know: $J_1(b)$ is sandwiched between $m_1 = \frac{\pi}{4}(1 + b)$ and $M_2 = \frac{\pi}{2\sqrt{2}}(1 + b^2)^{1/2}$, and both have gradient $\pi/4$ there, hence $J_1'(1) = \pi/4$. Of course, this means that both m_1 and M_2 give good approximations to $J_1(b)$ for b close to 1. Also, m_1 is the tangent to the graph at $b = 1$, and the fact that the function is above the tangent again reflects convexity.

However, at $b = 0$, $J_1(b)$ is sandwiched between $(1 + b^2)^{1/2}$ (with gradient 0) and $1 + (\frac{\pi}{2} - 1)b$ (with gradient $\frac{\pi}{2} - 1$), which tells us nothing about $J_1'(0)$. We will now settle this problem. Contrary to what one might expect, we will do so by deploying some corresponding estimations for $I(a, b)$.

Inequalities for $I(a, b)$

A pair of bounds for $I(a, b)$ follows at once from identity (3): since $(ab)^{1/2} \le M(a, b) \le \frac{1}{2}(a + b)$, we have

$$\frac{\pi}{a + b} \le I(a, b) \le \frac{\pi}{2(ab)^{1/2}}. \qquad (12)$$

The right-hand inequality can also be proved directly by the Cauchy-Schwarz inequality for integrals, most easily with $I(a, b)$ expressed in the equivalent form

$$I(a, b) = \int_0^\infty \frac{1}{(x^2 + a^2)^{1/2}(x^2 + b^2)^{1/2}}\, dx,$$

which equates to (2) by the substitution $x = b \tan \theta$.

The left-hand inequality in (12) does no justice to the fact that $I(a, b)$ actually tends to infinity as $b \to 0^+$. In fact, there is a theorem describing $I(1, b)$ very well for small b: a simple version of it states

$$\log \frac{4}{b} < I(1, b) < \log \frac{4}{b} + \tfrac{1}{2}b. \qquad (13)$$

The proof, which is not very hard, can be seen in [Jam].

The integral $L(a,b)$, and a closer estimation for small b

Integrals of the following type form a very neat link between $J(a,b)$ and $I(a,b)$ (here we are following the account in [Lo]). For $a > 0$ and $b \geq 0$, define

$$L(a,b) = \int_0^{\pi/2} \frac{\cos^2\theta}{(a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}}\, d\theta. \tag{14}$$

The first thing to note is that $L(b,a) \neq L(a,b)$. In fact, the substitution $\theta = \frac{\pi}{2} - \phi$ gives

$$L(b,a) = \int_0^{\pi/2} \frac{\sin^2\theta}{(a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}}\, d\theta. \tag{15}$$

From (14) and (15), we deduce at once (for $a,\ b > 0$) the following attractive pair of identities:

$$L(a,b) + L(b,a) = I(a,b), \tag{16}$$

$$a^2 L(a,b) + b^2 L(b,a) = J(a,b). \tag{17}$$

By eliminating $L(b,a)$ in (16) and (17), we deduce

$$(a^2 - b^2)L(a,b) = J(a,b) - b^2 I(a,b). \tag{18}$$

It is clear that $L(a,b)$ decreases with a and b, also $L(a,0) = 1/a$ and $L(a,a) = \pi/(4a)$, so for $0 \leq b \leq a$, $L(a,b)$ lies between these two values. Note that $L(0,b)$ is not defined.

By differentiation under the integral sign, we have $\frac{\partial}{\partial b} J(a,b) = bL(b,a)$. To complete the symmetry, we mention that also $\frac{\partial}{\partial b} I(a,b) = -\frac{1}{b} L(a,b)$ (but this only becomes apparent after applying the substitution $x = b\tan\theta$ to both integrals).

So $J_1'(b) = bL(b,1)$, and in particular, $J_1'(1) = L(1,1) = \pi/4$, as we saw earlier. This expression for $J_1'(b)$ is undefined at $b = 0$, but we don't need it: we can easily determine $J_1'(0)$ using (18) and the elementary inequality (12).

PROPOSITION 6. *We have $J(1,b) \leq 1 + \frac{\pi}{2} b^{3/2}$ for $0 \leq b \leq 1$, hence $J_1'(0) = 0$.*

Proof. By (18),

$$J(1,b) = b^2 I(1,b) + (1 - b^2)L(1,b).$$

Now $L(1,b) \leq 1$ for $0 \leq b \leq 1$, so, with (12), we have $J(1,b) \leq 1 + b^2 I(1,b) \leq 1 + \frac{\pi}{2} b^{3/2}$. Hence

$$0 \leq \frac{J_1(b) - 1}{b} \leq \frac{\pi}{2} b^{1/2}.$$

This tends to 0 when $b \to 0^+$, which shows that $J_1'(0) = 0$. □

This might seem good enough. However, given (13), it will now cost us very little further work to derive corresponding bounds for $J(1,b)$, giving a highly accurate estimation for small b. We will express $J_1(b)$ as the integral of its derivative $bL(b,1)$, so we want bounds for $L(b,1)$.

LEMMA. *For* $0 < b < 1$,

$$\log \frac{1}{b} + c_1 \le L(b,1) \le \log \frac{1}{b} + c_2, \tag{19}$$

where $c_1 = \log 4 - 1$ *and* $c_2 = \log 4 + \frac{1}{2} - \pi/4$.

Proof. By (16), we have $L(b,1) = I(1,b) - L(1,b)$. Now $\frac{\pi}{4} \le L(1,b) \le 1$, and by (13),

$$\log \frac{1}{b} + \log 4 \le I(1,b) \le \log \frac{1}{b} + \log 4 + \tfrac{1}{2}$$

for $0 < b < 1$. Together, these inequalities give (19). □

PROPOSITION 7. *For* $0 < b < 1$,

$$J(1,b) = 1 + \tfrac{1}{2} b^2 \log \frac{1}{b} + r(b), \tag{20}$$

where $c_3 b^2 \le r(b) \le c_4 b^2$, *with* $c_3 = \log 2 - \frac{1}{4}$ *and* $c_4 = \frac{1}{2} + \log 2 - \frac{\pi}{8}$ *(so, in particular,* $0 < r(b) < b^2$*).*

Proof. We have

$$J_1(b) - 1 = \int_0^b J_1'(t)\,dt = \int_0^b tL(t,1)\,dt,$$

and by (19), $-t\log t + c_1 t \le tL(t,1) \le -t \log t + c_2 t$. Now

$$\int_0^b (-t\log t)\,dt = \left[-\tfrac{1}{2}t^2 \log t\right]_0^b + \int_0^b \tfrac{1}{2}t\,dt = \tfrac{1}{2}b^2 \log \frac{1}{b} + \tfrac{1}{4}b^2.$$

The stated bounds are found by adding, respectively, $\int_0^b c_1 t\,dt = \frac{1}{2}c_1 b^2$ and $\frac{1}{2}c_2 b^2$. □

So we now have a pair of bounds that differ by less than b^2. The previous lower bound $m_2 = (1+b^2)^{1/2}$ has the correct gradient 0 at $b = 0$, but the lower bound in (20) is obviously stronger for small b. In fact, it is greater than $1 + \frac{1}{2}b^2$ (hence greater than m_2) for all $b \le \frac{3}{4}$.

Since $J(a,1) = aJ(1,\frac{1}{a})$, we can derive the following estimation, effective for large a:

$$J(a,1) = a + \frac{\log a}{2a} + r_1(a),$$

where $c_3/a \le r_1(a) \le c_4/a$.

Finally, a fact that may seem rather surprising: J_1 does not have a second derivative at 0, because by the Lemma, $\frac{1}{b}J_1'(b) = L(b,1) \to \infty$ as $b \to 0^+$.

References

[Jam] G.J.O. Jameson, An approximation to the arithmetic-geometric mean, *Math. Gaz.* **98** (2014).

[Lo] Nick Lord, Recent calculations of π; the Gauss-Salamin algorithm, *Math. Gaz.* **76** (1992), 231–242.

[LS] Nick Lord and H.-J. Seiffert, Solution to Problem 76D, *Math. Gaz.* **77** (1993), 126.