Argument Principle

The argument principle helps us determine the number of zeroes of a holomorphic function on the domain enclosed by a curve in its domain of definition. To introduce it, let us start out with the well-known situation of estimating the number of zeroes for a real-valued function $y = f(x)$ of a real variable x, for example, $f(x) = \sum_{j=0}^{n} a_j x^j$ is a polynomial of degree n with real coefficients.

For the case of a real-valued function $y = f(x)$ of a real variable x, the tool at our disposal is the intermediate value theorem which says that for a real-valued continuous function $f(x)$ on a closed interval $[a, b]$, any value between $f(a)$ and $f(b)$ will be assumed by f at some point of $[a, b]$, Thus, if a continuous function $f(x)$ on $[a, b]$ has different signs at the two end-points a and b, then there must be a zero of $f(x)$ on the interval $[a, b]$. This tells us that the "sense" defined by the functional value $f(x)$ at the boundary $\{a, b\}$ of the domain $[a, b]$ can give us some information on the number of roots of f in the domain $[a, b]$. The "direction" or "sense" defined by a nonzero real number α means the direction of the vector pointing from the origin to α. A positive number defines the direction from left to right. A negative number defines the direction from right to left. This relation between the number of zeroes of the function on an interval and the information on the directions defined the functional values on the boundary of the interval is what interests us.

Now consider the case of a holomorphic function $w = f(z)$ on a domain Ω in $\mathbb{C}$. Let C be a piecewise continuously differentiable curve in Ω whose enclosure D is completely inside Ω. We would like to ask the question how to get some information on the number of zeroes of f on D from the "directions" defined by its functional values $f(z)$ on the boundary C of the domain D. The "direction" defined by $f(z)$ means the direction of the vector pointing from the origin to $f(z)$. It is measured by $\arg f(z)$. For the real variables case we can get some useful information if the "directions" defined by the functional values at the two points in the boundary are different, that is, there is a *change* in the "direction" defined by the functional values at the end-points. For the case of a holomorphic function $w = f(z)$ we also look for a *change* in the direction defined by the functional value $f(z)$ when z is on the boundary C of D. The way to get a number from the change of the direction defined by boundary functional values is to ask how many complete revolutions (with the sign from being counterclockwise or clockwise taken into account) the change of direction along the boundary has made. This tells us how many zeroes are precisely inside the enclosure D of the boundary C. This is known as the *argument principle*, because the direction defined by $f(z)$ is precisely the argument $\arg f(z)$ of $f(z)$.

Theorem (Argument Principle). Let $w = f(z)$ be a holomorphic function on a domain Ω in $\mathbb{C}$. Let C be a closed piecewise smooth simple curve (it i.e., not self-intersecting) in Ω so that the domain D enclosed by C is contained in Ω. Assume that $f(z)$ is nowhere zero on C. Then the number of zeroes (with multiplicities counted) of $f(z)$ in D is equal to $\frac{1}{2\pi}$ times the change of the argument $\arg f(z)$ as z goes around C once in the counterclockwise sense.

Proof. Let the roots of $f(z)$ in D be $a_1, \cdots, a_k$ with multiplicities $\nu_1, \cdots, \nu_j$ respectively. Let

$$g(z) = \frac{f(z)}{(z-a_1)^{\nu_1}(z-a_2)^{\nu_2}\cdots(z-a_k)^{\nu_k}}.$$

By Cauchy's theorem applied to the logarithmic derivative

$$\frac{d}{dz}\log g(z) = \frac{g'(z)}{g(z)}$$

of $g(z)$ on D with boundary C, we get

$$(*) \qquad \int_C \left(\frac{d}{dz}\log g(z)\right) dz = 0.$$

From the definition of $g(z)$ we have

$$(\dagger) \qquad \frac{d}{dz}\log g(z) = \frac{d}{dz}\log f(z) - \frac{\nu_1}{z-a_1} - \frac{\nu_2}{z-a_2} + \cdots - \frac{\nu_k}{z-a_k}.$$

We now interrupt our proof to give an analytical definition of the change of the argument along a curve.

Notion of the Change of Argument Along a Curve. Let Γ be a curve (not necessarily closed and not necessarily simple) in a domain G in *mathbbC* which is piecewise smooth and is parametrized by $t \to \varphi(t)$ for $\alpha \leq t \leq \beta$.

Let $F(z)$ be any smooth function on G which is nowhere zero on Γ. We define the change of argument of $F(z)$ along Γ by

$$\operatorname{Im}\left(\int_{t=\alpha}^{\beta} \frac{d(F\circ\varphi)}{F\circ\varphi}\right)$$

and we denote it by $\Delta_\Gamma \arg F$. Since $d\log(F\circ\varphi) = \frac{d(F\circ\varphi)}{F\circ\varphi}$, by the fundamental theorem of calculus and because the imaginary part of $\log(F\circ\varphi)$ is the same as $\arg(F\circ\varphi)\Delta_\Gamma F$, it follows that $\Delta_\Gamma F$ is precisely the change of the argument of F along Γ.

We now return to the proof of the argument principle. Putting together $(*)$, $(\dagger)$, and $(\flat)$ and using

$$\int_C \frac{dz}{z-a_j} = 2\pi i,$$

we get

$$0 = \Delta_C \log f - 2\pi i\left(\nu_1 + \nu_2 + \cdots + \nu_k\right),$$

or

$$\frac{1}{2\pi}\Delta_C \log f = \nu_1 + \nu_2 + \cdots + \nu_k.$$

Now $\log|f| = \operatorname{Re}\log f$ is a well-defined continuous function on C. Thus $\Delta_C \log|f| = 0$ and

$$\Delta_C \log f = \Delta_C\left(\log|f| + i\arg f\right) = \Delta_C \arg f$$

and

$$\nu_1 + \nu_2 + \cdots + \nu_k = \frac{1}{2\pi}\Delta_C \arg f.$$

Q.E.D.

Remark on Argument Principle for Meromorphic Function on Multi-Connected Domains. The above proof works also when D is defined by several curves $C_1, \cdots, C_m$ and f is meromorphic on Ω. More precisely, let D a bounded domain in $\mathbb{C}$ which is contained in Ω and let the boundary ∂D of D in $\mathbb{C}$ be composed of several piecewise smooth simple closed curves $C_1, \cdots, C_m$ in Ω. Let f be a meromorphic function in Ω. Then $\frac{1}{2\pi}\Delta_{\partial D}\arg f$ equals the number of zeroes of f minus the number of poles inside D with multiplicities counted.

The only modification in the above proof which is needed is to replace C by ∂D and to define

$$g(z)=\frac{f(z)\,(z-b_1)^{\lambda_1}\,(z-b_2)^{\lambda_2}\cdots(z-b_\ell)^{\lambda_\ell}}{(z-a_1)^{\nu_1}\,(z-a_2)^{\nu_2}\cdots(z-a_k)^{\nu_k}},$$

where $b_1,\cdots,b_\ell$ are the poles of $f(z)$ in D with multiplicities $\lambda_1,\cdots,\lambda_\ell$ respectively.

Example. Let $f(z)=z^4+z^3+4z^2+2z+3$. We are going to verify that there is no zero of $f(z)$ in the first quadrant by applying the argument principle to the part of the first quadrant which is inside the disk of radius R centered at the origin and letting $R\to 0$. We will calculate the change of $\arg f$

(i) along the real axis from the origin to $+\infty$,

(ii) around the arc

$$C_R:=\{\,|z|=R,\ \mathrm{Re}\,z\geq 0,\ \mathrm{Im}\,z\geq 0\,\}$$

in the counterclockwise sense, and

(iii) along the imaginary axis from $\infty\, i$ to the origin.

For the computation in (i) we observe that $f(z)>0$ on $[0,\infty)\subset\mathbb{R}$ so that $\arg f(z)$ can be continuously set to 0 when z goes along the real axis from the origin to $+\infty$. Hence change of $\arg f$ along the real axis from the origin to $+\infty$ is zero.

For the computation in (ii), we use the parametrization $\theta\mapsto Re^{i\theta}$ for $0\leq\theta\leq\frac{\pi}{2}$. We have

$$\begin{aligned}\frac{1}{2\pi}\Delta_{C_R}\arg f&=\frac{1}{2\pi}\Delta_{C_R}\left(R^4e^{4i\theta}+R^3e^{3i\theta}+4R^2e^{2i\theta}+2Re^{i\theta}+3\right)\\&=\frac{1}{2\pi}\Delta_{C_R}\left(e^{4i\theta}+\frac{1}{R}e^{3i\theta}+\frac{4}{R^2}e^{2i\theta}+\frac{2}{R^3}Re^{i\theta}+\frac{3}{R^4}\right)\\&\to\frac{1}{2\pi}\Delta_{C_R}\left(e^{4i\theta}\right)=1\quad\text{as } R\to\infty.\end{aligned}$$

For computation of (iii) we restrict $f(z)=z^4+z^3+4z^2+2z+3$ to the imaginary axis $z=iy$ and get

$$f(iy)=y^4-iy^3-4y^2+2iy+3=\left(y^4-4y^2+3\right)+i\left(-y^3+2y\right).$$

We look at the graph of the real part $y^4 - 4y^2 + 3$ on $0 \leq y < \infty$. We are only interested in where the graph is positive, zero, or negative. The roots of $y^4 - 4y^2 + 3 = 0$ are given by the formula for the roots of a quadratic polynomial

$$y^2 = \frac{4 \pm \sqrt{4^2 - 4 \cdot 3}}{2} = \frac{4 \pm \sqrt{16 - 12}}{2} = 1 \text{ or } 3.$$

We can also get this right away by factoring $y^4 - 4y^2 + 3 = (y^2 - 3)(y^2 - 1)$. Thus on $0 \leq y < \infty$ the roots of $y^4 - 4y^2 + 3 = 0$ are $y = 1$ and $y = \sqrt{3}$. At $y = 0$ the value of $\operatorname{Re} f(iy) = y^4 - 4y^2 + 3 = 3$ is positive. So we know that

$$\operatorname{Re} f(iy) = y^4 - 4y^2 + 3 \begin{cases} > 0 & \text{on } [0, 1) \\ = 0 & \text{at } 1 \\ < 0 & \text{on } (1, \sqrt{3}) \\ = 0 & \text{at } \sqrt{3} \\ > 0 & \text{on } (\sqrt{3}, \infty) \end{cases}$$

We now look at the graph of the imaginary part $y^4 - 4y^2 + 3$ on $0 \leq y < \infty$. Again we are only interested in where the graph is positive, zero, or negative. The roots of $-y^3 + 2y = 0$ are $y = 0$ and $y = \pm\sqrt{2}$. Thus on $0 \leq y < \infty$ the roots of $-y^3 + 2y = 0$ are $y = 0$ and $y = \sqrt{2}$. For very large value of y the value of $-y^3 + 2y$ is negative. So we know that

$$\operatorname{Im} f(iy) = -y^3 + 2y \begin{cases} = 0 & \text{at } 0 \\ > 0 & \text{on } (0, \sqrt{2}) \\ = 0 & \text{at } \sqrt{2} \\ < 0 & \text{on } (\sqrt{2}, \infty) \end{cases}$$

We now put together the information on the sign changes of the real and imaginary part of $f(iy)$ for $0 \leq y < \infty$.

at 0	on $(0, 1)$	at 1	on $(1, \sqrt{2})$	at $\sqrt{2}$	on $(\sqrt{2}, \sqrt{3})$	at $\sqrt{3}$	on $(\sqrt{3}, \infty)$
$\operatorname{Re} > 0$	$\operatorname{Re} > 0$	$\operatorname{Re} = 0$	$\operatorname{Re} < 0$	$\operatorname{Re} < 0$	$\operatorname{Re} < 0$	$\operatorname{Re} = 0$	$\operatorname{Re} > 0$
$\operatorname{Im} = 0$	$\operatorname{Im} > 0$	$\operatorname{Im} > 0$	$\operatorname{Im} > 0$	$\operatorname{Im} = 0$	$\operatorname{Im} < 0$	$\operatorname{Im} < 0$	$\operatorname{Im} < 0$

Let us trace the locus of the point $f(iy)$ as $z = iy$ goes from the origin along the imaginary axis all the way to infinity.

(i) The point $f(iy)$ starts out on the positive real axis (real part > 0 and imaginary part $= 0$).

(ii) The point $f(iy)$ is in the first quadrant when $0 < y < 1$ (real part > 0 and imaginary part > 0).

(iii) The point $f(iy)$ is on the upper imaginary axis when $y = 1$ (real part $= 0$ and imaginary part > 0).

(iv) The point $f(iy)$ is in the second quadrant when $1 < y < \sqrt{2}$ (real part < 0 and imaginary part > 0).

(v) The point $f(iy)$ is on the negative real axis when $y = \sqrt{2}$ (real part < 0 and imaginary part $= 0$).

(vi) The point $f(iy)$ is in the third quadrant when $\sqrt{2} < y < \sqrt{3}$ (real part < 0 and imaginary part < 0).

(vii) The point $f(iy)$ is on the lower imaginary axis when $y = \sqrt{3}$ (real part $= 0$ and imaginary part < 0).

(viii) The point $f(iy)$ is in the fourth quadrant when $\sqrt{3} < y < \infty$ (real part > 0 and imaginary part < 0).

(ix) Finally

$$\lim_{y\to\infty} \frac{\operatorname{Im} f(iy)}{\operatorname{Re} f(iy)} = \lim_{y\to\infty} \frac{-y^3 + 2y}{y^4 - 4y^2 + 3} = 0$$

so that the point $f(iy)$ reaches the positive real axis as $y \to \infty$.

This means that $\arg f(iy)$ increases precisely by 2π as y goes from 0 to ∞. When the direction is reversed and y decreases from ∞ to 0, the change of $\arg f(iy)$ is precisely -2π. This precisely cancels

$$\lim_{R\to\infty} \frac{1}{2\pi} \Delta_{C_R} \arg f = 1$$

and we conclude from the argument principle that the number of roots of $f(z) = 0$ in the first quadrant is zero.

Because the coefficients of $f(z) = z^4 + z^3 + 4z^2 + 2z + 3$ are real, there is symmetric with respect to the real axis, to determine the number of roots in each of the other three quadrants, it suffices to determine whether there are any roots on the negative real axis. We use the grouping of the first three terms and another grouping of the first two terms to show that there are no roots of $f(z)$ on the negative real axis. These two ways of grouping gives

$$\begin{aligned} f(-x) &= x^4 - x^3 + 4x^2 - 2x + 3 = \left(x^4 - x^3 + 4x^2\right) + (-2x + 3) \\ &= \left(x^4 + x^2(1-x) + 3x^2\right) + (2(1-x) + 1) > 0 \end{aligned}$$

for $0 < x < 1$ and

$$\begin{aligned} f(-x) &= x^4 - x^3 + 4x^2 - 2x + 3 = \left(x^4 - x^3\right) + \left(4x^2 - 2x + 3\right) \\ &= \left(x^3(x-1)\right) + \left(2x(x-1) + 2x^2 + 3\right) > 0 \end{aligned}$$

for $x > 1$. Thus we can conclude that there are no zeroes in the fourth quadrant. There are two roots each in the second and the third quadrant.

Interpretation of the Argument Principle in Terms of the law of Gauss and the Streamline Function. Let us consider the case of a holomorphic function $f(z)$ on a domain D in $\mathbb{C}$ which is nowhere zero on the boundary C of D. Suppose we have an electrostatic potential of the form $-\log|f(z)|$. Then the flux which goes out of a closed curve C is given by

$$\int_C \vec{n} \cdot \nabla \log|f(z)|$$

where $\vec{n}$ is the unit normal vector of C pointing out, because the electric field is given by $-\nabla(-\log|f(z)|) = \nabla \log|f(z)|$. Fix a point P_0 on C. We can choose the coordinate z of $\mathbb{C}$ so that

(i) P_0 is the origin of this coordinate system z,

(ii) the unit vector in the direction of the positive real axis is $\vec{n}$ at P_0, and

(iii) the unit vector in the direction pointing up along the imaginary axis is the unit tangent vector $\vec{t}$ of C at P_0.

Since $\log f(z)$ is holomorphic and its real part is $\log|f(z)|$ and its imaginary part is $\arg f(z)$, by the Cauchy-Riemann equations

$$\vec{n} \cdot \nabla \log|f(z)| = \frac{\partial}{\partial \vec{n}} |f(z)| = \frac{\partial}{\partial \vec{t}} \arg f(z)$$

so that by the fundamental theorem of calculus

$$\int_C \vec{n} \cdot \nabla \log|f(z)| = \int_C \frac{\partial}{\partial \vec{t}} \arg f(z) = \Delta_C \arg f.$$

Thus we conclude that the change of the argument of a holomorphic function f along a closed curve is equal to the flux across C of the electric field whose electrostatic potential is $-\log|f(z)|$. By Gauss's law such a flux across C is equal (up to a normalizing constant which we just assume to be 1) to the total electric point charge inside (if all the electric charge inside assumed to be in the form of point charges). We can interpret a point charge as a zero of $f(z)$ as follows. For the model case of

(i) D equal to the disk $|z-a|<r$,

(ii) $C =$ equal to the circle $|z-a|=r$,

(iii) the holomorphic function $f(z)$ equal to $z-a$,

we have

$$\begin{aligned}\int_{|z-a|=r} \vec{n}\cdot\nabla\log|z-a| &= \int_{\theta=0}^{2\pi}\left(\frac{\partial}{\partial r}\log|z-a|\right) r d\theta \\ &= \int_{\theta=0}^{2\pi}\left(\frac{\partial}{\partial r}\log r\right) r d\theta = 2\pi,\end{aligned}$$

where $z-a=re^{i\theta}$. For the general case where the zeroes of $f(z)$ in D are $a_1,\cdots,a_k$ with multiplicities $\nu_1,\cdots,\nu_k$ respectively, we can write

$$f(z)=g(z)\,(z-a_1)^{\nu_1}\,(z-a_2)^{\nu_2}\cdots(z-a_k)^{\nu_k}$$

with $g(z)$ holomorphic on D up to C without any zero. Let D_j is a disk in D with boundary C_j also in D. Then

$$\int_C \vec{n}\cdot\nabla\log|f(z)| = \int_C \vec{n}\cdot\nabla\log|g(z)| + \sum_{j=1}^k \nu_j \int_C \vec{n}\cdot\nabla\log|z-a_j|.$$

The divergence theorem gives

$$\int_C \vec{n}\cdot\nabla\log|g(z)| = \int_D \Delta\log|g(z)| = 0,$$

because $\log|g(z)|$ is harmonic on D up to the boundary C. When we apply the divergence theorem to $\log|z-a_j|$ on $D-D_j$ we get

$$\int_C \vec{n}\cdot\nabla\log|z-a_j| - \int_{C_j} \vec{n}\cdot\nabla\log|z-a_j| = \int_{D-D_j} \Delta\log|z-a_j| = 0,$$

because $z - a_j$ is nowhere zero on $D - D_j$ and as a consequence $\log|z - a_j|$ is harmonic on $D - D_j$. From the model case we have

$$\int_{C_j} \vec{n} \cdot \nabla \log|z - a_j| = 2\pi.$$

Thus

$$\int_C \vec{n} \cdot \nabla \log|f(z)| = \sum_{j=1}^{k} \nu_j \int_C \vec{n} \cdot \nabla \log|z - a_j| = 2\pi \sum_{j=1}^{k} \nu_j$$

and we have the argument principle as a consequence of Gauss's law and the Cauchy-Riemann equations.

A moment ago we showed by using the Cauchy-Riemann equations that the flux of the electric field with electrostatic potential $-\log|f(z)|$ across a closed curve C is equal to the change of $\arg f(z)$ along C. The argument $\arg f(z)$ is the imaginary part of $\log f(z)$ whose real part is $\log|f(z)|$. Thus $\arg f(z)$ is the complex conjugate of $\log|f(z)|$. Flux lines (streamlines) are given by the condition that the streamline function $\arg f(z)$ is equal to a constant. This reasoning can be applied to any curve (not necessarily closed) to give us the conclusion that the change of the argument $\arg f(z)$ along Γ which is equal to the difference of the values of the flux line function (streamline function) $\arg f(z)$ at the two end-points of the curve Γ when the electrostatic potential of the electric field is $-\log|f(z)|$.