Chapter Three

Stability

§ 25. The Function-Theoretic Center Problem

We begin with the definition of stability and instability. Let $\mathfrak{R}$ be a topological space whose points we denote by $\mathfrak{p}$, and let $\mathfrak{a}$ be a certain point in $\mathfrak{R}$. By a neighborhood here we will always mean a neighborhood of $\mathfrak{a}$ in $\mathfrak{R}$. Let $\mathfrak{p}_1 = S\mathfrak{p}$ be a topological mapping of a neighborhood $\mathfrak{U}_1$ onto a neighborhood $\mathfrak{V}_1$, whereby $\mathfrak{a} = S\mathfrak{a}$ is mapped onto itself. The inverse mapping $\mathfrak{p}_{-1} = S^{-1}\mathfrak{p}$ then carries $\mathfrak{V}_1$ onto $\mathfrak{U}_1$, and in general $\mathfrak{p}_n = S^n\mathfrak{p}$ $(n=0, \pm 1, \pm 2, \ldots)$ is a topological mapping of a neighborhood $\mathfrak{U}_n$ onto a neighborhood $\mathfrak{V}_n$, having $\mathfrak{a}$ as a fixed-point. For each point $\mathfrak{p} = \mathfrak{p}_0$ in the intersection $\mathfrak{U}_1 \cap \mathfrak{V}_1 = \mathfrak{W}$ we construct the successive images $\mathfrak{p}_{k+1} = S\mathfrak{p}_k$ $(k=0, 1, \ldots)$, as long as $\mathfrak{p}_k$ lies in $\mathfrak{U}_1$, and similarly $\mathfrak{p}_{-k-1} = S^{-1}\mathfrak{p}_{-k}$, as long as $\mathfrak{p}_{-k}$ lies in $\mathfrak{V}_1$. If the process terminates with a largest $k+1 = n$, then $\mathfrak{p}_0, \ldots, \mathfrak{p}_{n-1}$ all still lie in $\mathfrak{U}_1$, but $\mathfrak{p}_n$ no longer does; similarly for the negative indices. In this way, to each $\mathfrak{p}$ in $\mathfrak{W}$ there is associated a sequence of image points $\mathfrak{p}_k$ $(k = \ldots, -1, 0, 1, \ldots)$, which is finite, infinite on one side, or infinite on both sides.

The mapping S is said to be stable at the fixed-point $\mathfrak{a}$ if for each neighborhood $\mathfrak{U} \subset \mathfrak{W}$ there exists a neighborhood $\mathfrak{V} \subset \mathfrak{U}$ of $\mathfrak{a}$ whose images $S^n\mathfrak{V}$ $(n = \pm 1, \pm 2, \ldots)$ all lie in $\mathfrak{U}$. Instability on the other hand, is defined not as the logical negation of stability, but in terms of the following stronger requirement. The mapping S is said to be unstable at the fixed-point $\mathfrak{a}$ if there exists a neighborhood $\mathfrak{U} \subset \mathfrak{W}$ such that for each point $\mathfrak{p} \neq \mathfrak{a}$ in $\mathfrak{U}$ at least one image point $\mathfrak{p}_n$ lies outside $\mathfrak{U}$.

Let us restate the above definition in another form. A point set $\mathfrak{M} \subset \mathfrak{W}$ is said to be invariant under the mapping S if $\mathfrak{M} = S\mathfrak{M}$. The fixed-point $\mathfrak{a}$ is, of course, trivially an invariant point set. We now show that S is stable if and only if each neighborhood $\mathfrak{U}$ contains an invariant neighborhood $\mathfrak{V}$. If for each neighborhood $\mathfrak{U}$ there exists a neighborhood $\mathfrak{V} = S\mathfrak{V} \subset \mathfrak{U}$, then certainly $\mathfrak{V}$ has the necessary property required in the definition of stability, and consequently S is stable. Conversely, under the assumption that S is stable, for each neighborhood $\mathfrak{U} \subset \mathfrak{W}$ there exists a neighborhood $\mathfrak{Q} \subset \mathfrak{U}$ such that $S^n\mathfrak{Q} \subset \mathfrak{U}$ $(n=0, \pm 1, \pm 2, \ldots)$. The union

$\mathfrak{B} = \bigcup_n (S^n \mathfrak{Q})$ of all the $S^n \mathfrak{Q}$ is then invariant under S and is again a neighborhood, whereby the assertion is proved. Correspondingly, let us show that S is unstable if and only if there exists a neighborhood $\mathfrak{U}$ that contains no invariant subsets other than the fixed-point $\mathfrak{a}$. Indeed, if such a neighborhood $\mathfrak{U}$ exists then certainly the intersection $\mathfrak{U} \cap \mathfrak{W}$ has the same property, and we may therefore assume that $\mathfrak{U} \subset \mathfrak{W}$. If $\mathfrak{p}$ is then any point $\neq \mathfrak{a}$ in $\mathfrak{U}$, the images $\mathfrak{p}_n$ cannot all lie in $\mathfrak{U}$ or else $\mathfrak{M} = \bigcup_n \mathfrak{p}_n$ would be an invariant subset of $\mathfrak{U}$ that contains a point $\neq \mathfrak{a}$. Consequently S is unstable. Conversely, if S is unstable, there exists a neighborhood $\mathfrak{U} \subset \mathfrak{W}$ such that for each $\mathfrak{p} \neq \mathfrak{a}$ in $\mathfrak{U}$ at least one image $\mathfrak{p}_n$ does not lie in $\mathfrak{U}$. If $\mathfrak{p}$ is now any point of an invariant subset $\mathfrak{M} = S\mathfrak{M}$ of $\mathfrak{U}$, all the images $\mathfrak{p}_n$ of $\mathfrak{p}$ must lie in $\mathfrak{M}$ and therefore certainly in $\mathfrak{U}$, from which it follows that $\mathfrak{p} = \mathfrak{a}$. This again proves the assertion.

A mapping S that is not unstable thus has the property that each neighborhood contains an invariant point set with $\mathfrak{a}$ as a proper subset, while for a stable mapping S each neighborhood actually contains an invariant neighborhood. Consequently a stable mapping is necessarily not unstable, but a mapping that is not stable need not be unstable. A mapping S is said to be mixed at a fixed-point $\mathfrak{a}$ if it is neither stable nor unstable there. That there actually exist mixed mappings is seen by the simple example of the affine mapping $x_1 = x + y$, $y_1 = y$ in the (x, y)-plane, which has each point of the abscissa axis as a fixed-point. A bounded set is invariant under this mapping if and only if it lies on the abscissa axis. Since for arbitrary $r > 0$ the disk $x^2 + y^2 < r^2$ contains no invariant neighborhood of $(x, y) = (0, 0)$ but contains the invariant interval $-r < x < r$, $y = 0$, at the origin this mapping is neither stable nor unstable.

We carry over the definition of stability and instability to systems of differential equations

$$\dot{x}_k = f_k(x) \qquad (k = 1, \ldots, m). \tag{1}$$

Let $x = \xi^*$ be an equilibrium solution, so that $f_k(\xi^*) = 0$, and assume that a Lipschitz condition holds in a neighborhood of $x = \xi^*$. We again denote by $x(t, \xi)$ the solution to (1) with initial values $x_k = \xi_k$ at $t = 0$. Passage from ξ to $x(t, \xi)$ then defines for each fixed t a topological mapping S_t in a neighborhood of the fixed-point $x = \xi^*$. The definitions of stability and instability of the system (1) at the given equilibrium point are then obtained by taking for $\mathfrak{a}, \mathfrak{p}, S^n$, and $\mathfrak{p}_n = S^n \mathfrak{p}$ $(n = 0, \pm 1, \ldots)$ in the previous definitions the corresponding quantities ξ^*, ξ, S_t, and $\xi_t = x(t, \xi)$ as t varies over the reals. By introducing the modification that only positive values of t are permitted, one may speak also of stability or instability with respect to future time. This notion has, of course, significance in problems dealing with mechanics. Also the definition of the mixed case carries over in an obvious way.

Before turning to problems relating to stability for differential equations, we will look at the particular case when S is a conformal mapping in the plane. Already here some of the characteristic difficulties show up, although they can still be overcome by the available methods of analysis. Without loss of generality the fixed-point may be taken as the origin of the complex z-plane. The conformal mapping is then given by a power series

$$z_1 = f(z) = \lambda z + a_2 z^2 + a_3 z^3 + \cdots \qquad (\lambda \neq 0) \tag{2}$$

with complex coefficients, which converges in a neighborhood of $z = 0$. We wish to investigate when this mapping is stable, unstable, or mixed at $z = 0$. Assume first that S is stable. The circle of convergence $\mathfrak{K}$ for the series (2) then contains an invariant neighborhood $\mathfrak{V} = S\mathfrak{V}$ of the origin. This neighborhood may not be connected, but it does contain a connected invariant neighborhood; indeed, if $\mathfrak{L}$ is an open disk in $\mathfrak{V}$ containing the origin, the union of all the images $S^n \mathfrak{L}$ $(n = 0, \pm 1, \ldots)$ has the desired property. We may therefore assume that $\mathfrak{V}$ is already connected. Our aim here is to find an invariant neighborhood in $\mathfrak{K}$ that can be mapped conformally onto the unit disk. This can be achieved in, say, one of the following two ways. Perhaps $\mathfrak{V}$ is not simply connected. Then one adds to $\mathfrak{V}$ all points that lie in the interior of any simple closed curve $\mathfrak{C}$ contained in $\mathfrak{V}$. The resulting set $\mathfrak{U}$ is then again a connected neighborhood within $\mathfrak{K}$, and is easily seen to be simply connected. Because of the invariance of $\mathfrak{V}$, together with $\mathfrak{C}$ also $S\mathfrak{C}$ belongs to $\mathfrak{V}$, from which it follows that $\mathfrak{U}$ is invariant. Now, by the Riemann mapping theorem, $\mathfrak{U}$ can be mapped conformally onto a disk $|\zeta| < \varrho$ so that $z = 0$ goes into $\zeta = 0$ and the derivative z_ζ at $\zeta = 0$ has the value 1. Let

$$z = \phi(\zeta) = \zeta + b_2 \zeta^2 + \cdots \qquad (|\zeta| < \varrho) \tag{3}$$

be the inverse conformal mapping, whereby the series converges certainly in the circle $|\zeta| < \varrho$. We denote the mapping (3) by C and consider $T = C^{-1} S C$. Since the region $\mathfrak{U}$ was invariant under S, the disk $|\zeta| < \varrho$ is evidently invariant under the conformal mapping T, which has the center $\zeta = 0$ as a fixed-point. It follows from a well-known theorem in function theory that T is a linear mapping of the form

$$\zeta_1 = \mu \zeta \qquad (|\mu| = 1), \tag{4}$$

that is, a rotation about the origin. One can also arrive at (4) as follows, without constructing the set $\mathfrak{U}$. One constructs for $\mathfrak{V}$ the universal covering surface $\tilde{\mathfrak{V}}$, which by definition is simply connected. It has more than one boundary point, since this is already true of $\mathfrak{B}$. The conformal mapping S can now be extended to $\tilde{\mathfrak{B}}$ so that $S\tilde{\mathfrak{B}} = \tilde{\mathfrak{B}}$ and a fixed-point lies over the point $z=0$ of $\mathfrak{B}$. By the uniformization theorem one can then again map $\tilde{\mathfrak{B}}$ conformally onto a circle in the ζ-plane and set up the expression (3), where z now varies over the covering surface as ζ runs over $|\zeta| < \varrho$. The subsequent conclusion then follows as before.

The relation $T = C^{-1}SC$ can be expressed in the form $CT = SC$, whereupon (2), (3), (4) combine into the identity $\phi(\mu\zeta) = f(\phi(\zeta))$. This is the Schröder functional equation [1]. By comparison of the linear terms it follows that $\lambda = \mu$. Denoting the two conformal mappings determined in the previous paragraph by C_1 and C_2, one sees from $C_1^{-1}SC_1 = T = C_2^{-1}SC_2$ that $C_1^{-1}C_2 = C_0$ commutes with T. If λ is not a root of unity, by inserting the respective power series into the relation $C_0 T = T C_0$ one finds that C_0 is the identity mapping, so that $C_1 = C_2$. In particular, this implies that $\mathfrak{B} = \tilde{\mathfrak{B}} = \mathfrak{U}$ is simply connected, although this will not be used subsequently.

In view of (4) one has $|\lambda| = 1$, which therefore is a necessary condition for stability of S. We will now show that S is stable if and only if $|\lambda| = 1$ and the Schröder functional equation

$$\phi(\lambda\zeta) = f(\phi(\zeta)) \tag{5}$$

has a convergent power series solution $\phi(\zeta) = \zeta + \cdots$. The necessity of this condition is precisely what was shown in the preceding argument. Conversely, if there exists a convergent solution $\phi(\zeta)$ to (5) with $|\lambda| = 1$, then the substitution $z = \phi(\zeta)$, $z_1 = \phi(\zeta_1)$ transforms the given mapping $z_1 = f(z)$ into the rotation $\zeta_1 = \lambda\zeta$ which trivially is stable, since as invariant neighborhoods one can take all circles in the ζ-plane with center at the origin $\zeta = 0$. Because $\phi(\zeta)$, as well as its inverse series, converges in a sufficiently small neighborhood of the origin, it follows that also the given mapping $S = CTC^{-1}$ is stable. This proves the assertion. The name "center problem" is derived from the fact that in case of stability the family of concentric circles about the origin in the ζ-plane gives rise to the invariant neighborhoods of $z = 0$.

To investigate whether the mapping S is stable it is therefore enough to discuss whether Schröder's functional equation can be solved by a convergent power series $\phi(\zeta) = \zeta + \cdots$. Setting up $\phi(\zeta)$ as a series with undetermined coefficients, we first seek a solution to (5) in terms of a formal power series. Under the assumption that λ is not a root of unity, comparison of coefficients will give rise to exactly one solution, which we will call the Schröder series. Let $n \geqq 2$, and assume that the coefficients b_k $(1 < k < n)$ in (3) have already been determined so that both sides of (5) agree in terms of order $k < n$. For $n = 2$ the assumption is valid.

Expressing (5) in the form

$$\phi(\lambda\zeta)-\lambda\phi(\zeta)=f(\phi(\zeta))-\lambda\phi(\zeta)$$

we have

$$\sum_{l=2}^{\infty}(\lambda^l-\lambda)\,b_l\zeta^l=\sum_{l=2}^{\infty}a_l\,\phi^l(\zeta)\,,\tag{6}$$

and consequently the coefficient of ζ^n on the right is a polynomial in the a_l $(l=2,\ldots,n)$ and the already known b_k $(k=2,\ldots,n-1)$ with integral coefficients, while the corresponding coefficient on the left side of (6) is equal to $(\lambda^n-\lambda)\,b_n$. Since λ is not a root of unity and $\neq 0$ we have $\lambda^n-\lambda\neq 0$ $(n=2,3,\ldots)$, and consequently b_n is uniquely determined. In this way one obtains recursively the coefficients of the Schröder series $\phi(\zeta)=\zeta+b_2\zeta^2+\cdots$ which formally satisfies the Schröder functional equation (5).

Before investigating convergence of the above series $\phi(\zeta)$, we will consider the case when λ is a root of unity. Let $\lambda^n=1$ $(n>0)$, where also $n=1$ is admitted. If S is stable, then $T=C^{-1}SC$ again has the normal form $\zeta_1=\lambda\zeta$, and $T^k=C^{-1}S^kC$ is the mapping $\zeta_1=\lambda^k\zeta$. Consequently T^n is the identity mapping E, and therefore also $S^n=E$. Conversely, if $S^n=E$ and $\mathfrak{U}$ is a neighborhood of $z=0$ within the circle of convergence $\mathfrak{K}$ of $f(z)$, one selects any sufficiently small neighborhood $\mathfrak{V}$ of $z=0$ for which the n images $S^k\mathfrak{V}$ $(k=0,\ldots,n-1)$ are still completely contained in $\mathfrak{U}$. Because $S^n\mathfrak{V}=\mathfrak{V}$, the union of the $S^k\mathfrak{V}$ is an invariant neighborhood within $\mathfrak{U}$, and it follows that S is stable. Thus, in the case $\lambda^n=1$ $(n>0)$ the mapping S is stable if and only if $S^n=E$. As an example we consider the mapping

$$z_1=\frac{z}{1-z}=z+z^2+\cdots,\quad \lambda=1\,,$$

for which S^n is given by

$$z_n=\frac{z}{1-nz}\qquad (n=\pm 1,\pm 2,\ldots)\,,$$

and consequently is never the identity. Because $S\neq E$ and $\lambda=1$, this mapping is necessarily not stable. This can also be seen directly by setting $z=1/n$, where the natural number n can be arbitrarily large. On the other hand, if one sets $z=ir, 0<r<1$, then $|z_n|<r$ and the totality of images of z together with z form an invariant set within the circle $|z|\leqq r$. This shows that S is not unstable, and is therefore mixed. It is not known, however, whether it may happen that λ is a root of unity and S is unstable. From now on we will assume throughout that λ is not a root of unity.

We next investigate convergence of the formally constructed Schröder series $\phi(z)$ for the case $|\lambda| \neq 1$. This can be readily accomplished by the usual method of majorants. From the convergence of the series (2) there exists a positive number a such that $|a_{n+1}| < a^n$ $(n=1,2,\ldots)$. If az_1, az are introduced as variables in place of z_1, z in the transformation (2), one obtains again a conformal mapping of the form (2) with the same value for λ, but for which now

$$|a_{n+1}| < 1 \qquad (n=1,2,\ldots). \tag{7}$$

For the investigation of convergence we may therefore assume (7) at the outset. Moreover, since $|\lambda| \neq 1$, there exists a positive constant c such that

$$|\lambda^{n+1} - \lambda| > c > 0 \qquad (n=1,2,\ldots). \tag{8}$$

If the coefficients b_{n+1} in the Schröder series are determined by the recursive procedure associated with (6), it follows from (7), (8) that the formal solution $\Phi(\zeta) = \zeta + c_2\zeta^2 + \cdots$ to the functional equation

$$c(\Phi - \zeta) = \sum_{l=2}^{\infty} \Phi^l \tag{9}$$

is a majorant for $\phi(\zeta)$. On the other hand, the series

$$\zeta = \Phi - c^{-1} \sum_{l=2}^{\infty} \Phi^l,$$

which converges for $|\Phi| < 1$, has an inverse that converges in a neighborhood of $\zeta = 0$. This completes the convergence proof. As in § 17, one can also readily obtain from this a lower bound for the radius of convergence. We already know from $|\lambda| \neq 1$ that the mapping S is not stable. Because of the convergence of C just proved, we can construct the normal form $C^{-1}SC = T$, and it is immediately evident that the mapping $\zeta_1 = \lambda\zeta$ is actually unstable. Indeed, if one considers any point $\zeta \neq 0$ in an arbitrary bounded neighborhood $\mathfrak{U}$ of $\zeta = 0$, then because $|\lambda| \neq 1$, for n sufficiently large, positive or negative, the point $\zeta_n = \lambda^n\zeta$ will no longer lie in $\mathfrak{U}$. The instability of T implies that of $S = CTC^{-1}$, and consequently for $|\lambda| \neq 1$ the mapping S is necessarily unstable. This can also be shown directly without use of the normal form T.

For future discussion we may restrict ourselves to the case where λ is in absolute value 1, and is not a root of unity. In this case the investigation of convergence of the Schröder series requires finer estimates, to which we now turn. We first show that the set of λ for which there exists a convergent power series $f(z) = \lambda z + \cdots$ whose Schröder series diverges form a dense set on the unit circle $|\lambda| = 1$ [2]. For this divergence proof it will be enough to consider only series $f(z)$ whose coefficients $a_n\,(n=2,3,\ldots)$ are always equal to $\pm\dfrac{1}{n!}$, with the choice in sign determined recursively. In particular, such $f(z)$ converge everywhere. We turn once more to the determination of the b_n from equation (6). By comparison of coefficients one obtains for each $n>1$ the expression $(\lambda^n-\lambda)b_n-a_n$ as a polynomial in the a_k, b_k with $1<k<n$, and it is therefore obviously possible to choose $a_n=\pm\dfrac{1}{n!}$ recursively in such a way that

$$(10)\qquad |b_n|\geqq\frac{1}{n!}|\lambda^n-\lambda|^{-1}=\frac{1}{n!}|\lambda^{n-1}-1|^{-1}\qquad(n=2,3,\ldots).$$

Suppose now that for a given λ the inequality

$$(11)\qquad |\lambda^n-1|<(n!)^{-2}$$

is satisfied for infinitely many natural numbers n, and let $f(z)$ be a power series whose coefficients $a_2, a_3, \ldots$ have been determined in the above manner. Then on the one hand the series in z is everywhere convergent, while on the other hand the corresponding Schröder series $\phi(\zeta)$ diverges for each $\zeta\neq0$, since by (10), (11) the general term $b_n\zeta^n$ does not even tend to 0. The mapping $z_1=f(z)=\lambda z+\cdots$ therefore is not stable. However, it is not known whether it is mixed or unstable.

It remains to show that there is a dense set of values λ on the unit circle that are not roots of unity and that satisfy the inequality (11) for infinitely many n. If one sets $\lambda=e^{2\pi i\alpha}$ $(0\leqq\alpha<1)$ and for each natural number n chooses the integer m so that

$$(12)\qquad -\tfrac{1}{2}\leqq n\alpha-m<\tfrac{1}{2},$$

then

$$|\lambda^n-1|=|e^{2\pi i n\alpha}-1|=|e^{\pi i n\alpha}-e^{-\pi i n\alpha}|$$
$$=2|\sin(\pi n\alpha)|=2\sin(\pi|n\alpha-m|).$$

Because $|n\alpha-m|=\vartheta\leqq\frac{1}{2}$, it follows that $2\vartheta\leqq\sin(\pi\vartheta)\leqq\pi\vartheta$ and therefore

$$(13)\qquad 4\vartheta\leqq|\lambda^n-1|\leqq2\pi\vartheta\leqq7\vartheta.$$

Consequently it is enough to construct a set of irrational numbers α, dense in the interval $0\leqq\alpha<1$, for which the inequalities

$$(14)\qquad |n\alpha-m|<\frac{1}{7(n!)^2},\qquad n>0$$

have infinitely many integral solutions n, m. This can be readily accomplished as follows, using the representation of real numbers in terms of simple continued fractions. As is well known, to each irrational number α in the interval $0<\alpha<1$ one can associate a sequence of natural numbers $r_1, r_2, \ldots$ so that the sequence of fractions $\frac{p_k}{q_k}$ $(k=0,1,\ldots)$, recursively defined according to the prescription

$$(15)\quad \begin{cases} p_0=0, \quad q_0=1, \quad p_1=1, \quad q_1=r_1, \\ p_k=r_k p_{k-1}+p_{k-2}, \quad q_k=r_k q_{k-1}+q_{k-2} \quad (k=2,3,\ldots), \end{cases}$$

converges to α. The numbers $r_1, r_2, \ldots$ are uniquely determined by α and are known as the partial quotients of α. Moreover from the theory of continued fractions one obtains the inequality

$$(16)\qquad |q_k\alpha-p_k|<\frac{1}{q_{k+1}}<\frac{1}{r_{k+1}q_k}\leqq\frac{1}{r_{k+1}}\qquad (k=1,2,\ldots).$$

Conversely, corresponding to each such sequence $r_1, r_2, \ldots$ there is an irrational number α in the interval $0<\alpha<1$ with these prescribed partial quotients in its continued fraction representation.

Let β now be an arbitrary irrational number in the interval $0<\beta<1$ with $s_1, s_2, \ldots$ the partial quotients in its continued fraction expansion. For l an arbitrary fixed natural number one defines

$$(17)\qquad r_k=s_k \quad (0<k\leqq l), \qquad r_{k+1}=7(q_k!)^2 \quad (k\geqq l),$$

where $q_0, q_1, \ldots, q_k$ are again recursively determined in accordance with (15). For the continued fraction α with partial quotients $r_1, r_2, \ldots$ we have inequality (16), and since the first l partial quotients in the continued fractions of α and β agree, also $|q_l\beta-p_l|<q_l^{-1}$. It follows that

$$|\alpha-\beta|\leqq\left|\alpha-\frac{p_l}{q_l}\right|+\left|\beta-\frac{p_l}{q_l}\right|<2q_l^{-2}\leqq 2l^{-2},$$

while on the other hand, by (16), (17), the infinitely many pairs $n=q_k$, $m=p_k$ $(k=l, l+1, \ldots)$ satisfy (14). Since l can be chosen arbitrarily large, the corresponding numbers $\alpha=\alpha_l$ accumulate at β, and β being arbitrary, we have shown that the set of values α in question form a dense set in the unit interval.

Let Λ denote the set of values $\lambda=e^{2\pi i\alpha}$ on the unit circle such that for each power series $f(z)=\lambda z+a_2z^2+\cdots$ convergent in a neighborhood of $z=0$ the corresponding solution $\phi(\zeta)=\zeta+b_2\zeta^2+\cdots$ to the Schröder functional equation converges in a neighborhood of $\zeta=0$. We will now prove that Λ, as a subset of the unit circle, has linear Lebesgue measure 2π, or, equivalently, that the set A of the corresponding values α in the unit interval has Lebesgue measure 1. This will show that the set of irrational α for which there is at least one convergent series $f(z)$, with leading coefficient $\lambda = e^{2\pi i\alpha}$, whose Schröder series $\phi(\zeta)$ diverges, is a set of measure zero. In particular, this says that generally the mapping S is stable, provided that the necessary condition $|\lambda| = 1$ is satisfied.

For two given positive numbers ε, μ we consider the set $B(\varepsilon, \mu)$ of all numbers α in the unit interval E for which the inequalities

$$|n\alpha - m| < \varepsilon n^{-\mu}, \qquad n > 0 \tag{18}$$

have at least one integral solution n, m. Obviously

$$B(\varepsilon', \mu') \subset B(\varepsilon, \mu) \qquad (\varepsilon' \leqq \varepsilon, \mu \leqq \mu').$$

If one permits k to range over all natural numbers and considers the intersection

$$B = \bigcap_k B(k^{-1}, 2) \tag{19}$$

of all the $B(k^{-1}, 2)$, then certainly

$$B \subset B(\varepsilon, 2) \tag{20}$$

for each ε. We denote the Lebesgue measure of a measurable set Γ by $m(\Gamma)$ and estimate the measure of $B(\varepsilon, 2)$ from above. By (18), this set is a countable union of intervals, and therefore measurable, while by (19) then also B is measurable. For each solution n, m of (18) we have

$$-\varepsilon < m < n + \varepsilon \tag{21}$$

whenever α is in E, while on the other hand for given n, m the interval for α defined by (18) has length $2\varepsilon n^{-\mu-1}$. Noting that for each fixed natural number n the number of integers m satisfying (21) is smaller than $n + 2\varepsilon + 1$, and keeping in mind (20), we finally have

$$m(B(\varepsilon, 2)) \leqq \sum_{n=1}^{\infty} 2\varepsilon(n + 2\varepsilon + 1)n^{-3} < 4\varepsilon(\varepsilon + 1) \sum_{n=1}^{\infty} n^{-2}$$

$$m(B) < \frac{2\pi^2}{3} \varepsilon(\varepsilon + 1),$$

and since ε can be arbitrarily small, it follows that $m(B) = 0$. If Δ denotes the set of all α in E for which (18) has a solution for each choice of ε, μ, then by (19) the set Δ is contained in B, so that certainly $m(\Delta) = 0$. The complementary set $\Gamma = E - \Delta$ thus has $m(\Gamma) = 1$, and Γ is characterized by the property that for each number α in Γ there exist two positive numbers ε, μ such that for all natural numbers n and integers m we have

$$|n\alpha - m| > \varepsilon n^{-\mu}. \tag{22}$$

We will show in the next section that for all α in Γ the Schröder series of any convergent series $f(z)=\lambda z+\cdots$, with $\lambda=e^{2\pi i\alpha}$, is also convergent. In that case, by definition, we have $A\supset\Gamma$, so that also $m(A)=1$, as asserted.

§ 26. The Convergence Proof

In the previous section we saw that Schröder's functional equation has a unique formal power series solution $\phi(\zeta)=\zeta+\cdots$. To prove convergence of this series we will begin with a new construction of ϕ using an iteration process that converges rapidly enough to be insensitive to the effect of the small divisors λ^n-1.

Using the notation of the previous section, we consider an α in Γ, which by (25; 22) is irrational. Consequently $\lambda=e^{2\pi i\alpha}$ is not a root of unity, and we set

$$\varrho_n=|\lambda^n-1|^{-1}\qquad (n=1,2,\ldots).$$

With m as in (25; 12), we obtain from (25; 13), (25; 22) the estimate

$$\varrho_n\leqq\frac{1}{4}|n\alpha-m|^{-1}\leqq\frac{n^\mu}{4\varepsilon}=\frac{c_0 n^\mu}{\mu!},\tag{1}$$

where μ is a natural number and ε,μ may depend on α. Here $c_0=\mu!/4\varepsilon$ and c_1,c_2,c_3 will denote positive constants that depend only on α.

In the construction to follow we will obtain ϕ not by comparison of coefficients but by an iteration process consisting of a repeated substitution of variables. To describe this process, we denote the given transformation (25; 2) symbolically by S_0 and recall that our aim is to find a substitution C such that $T=C^{-1}S_0C$ is the linear transformation $\zeta_1=\lambda\zeta$. Rather than do this directly, we will first construct a substitution C_0 such that $C_0^{-1}S_0C_0=S_1$ is merely closer to the linear transformation T than was S_0. Then, starting with S_1, we will repeat this process to construct a substitution C_1 leading to a transformation

$$S_2=C_1^{-1}S_1C_1=C_1^{-1}C_0^{-1}S_0C_0C_1$$

that approximates T even more closely. Inductively, this process will lead to substitutions C_ν $(\nu=0,1,2,\ldots)$ and transformations

$$S_{\nu+1}=C_\nu^{-1}S_\nu C_\nu=B_\nu^{-1}S_0B_\nu,$$

with

$$B_\nu=C_0C_1\ldots C_\nu\tag{2}$$

converging to the desired substitution. For the success of this method it will be important that the composition of $C_0, C_1, \ldots, C_\nu$ is well defined, i.e. that the range of C_ν lies in the domain of definition of $C_{\nu-1}$, and that the sequence B_ν converges in some fixed neighborhood of the origin to an invertible substitution B of the form $z=\phi(\zeta)=\zeta+\cdots$, while $S_\nu \to T$ as $\nu \to \infty$.

If we assume the above statements to be true, it follows that

$$T = B^{-1} S_0 B,$$

so that B will be the desired substitution. The uniqueness of the formal power series for ϕ then assures us that B is represented by the Schröder series, which therefore must converge, since B is analytic near $\zeta=0$.

To carry out the above procedure, we begin with a transformation S expressed in the form

$$z_1 = f(z) = \lambda z + \hat{f}(z)$$

where $\hat{f}$ is a convergent power series starting with the quadratic term. The derivative $\hat{f}'$ is also analytic in some disk about the origin, and given $\delta > 0$ we can find $r > 0$ such that

$$|\hat{f}'| < \delta \quad \text{in} \quad |z| < r. \tag{3}$$

The substitution $z=\phi(\zeta)=\zeta+\hat{\phi}$ that linearizes the above transformation satisfies Schröder's functional equation (25; 5), which can be expressed in the form

$$\hat{\phi}(\lambda\zeta) - \lambda\hat{\phi}(\zeta) = \hat{f}(\phi(\zeta)).$$

Rather than solve this equation, we define the substitution $z=\zeta+\psi(\zeta)$, with ψ a power series beginning with the quadratic term, as the solution to the linear equation

$$\psi(\lambda\zeta) - \lambda\psi(\zeta) = \hat{f}(\zeta) = \sum_{k=2}^{\infty} a_k \zeta^k. \tag{4}$$

This substitution, which we denote by C, forms the basic step in the iteration process. Defining $S_+ = C^{-1} S C$ and expressing it as

$$\zeta_1 = g(\zeta) = \lambda\zeta + \hat{g}(\zeta), \tag{5}$$

we will show that, with δ, r suitably chosen, the function $\hat{g}$, which measures the deviation of S_+ from the linear transformation, is indeed smaller than the previous function $\hat{f}$.

To this end we choose constants δ, θ so that

$$0 < \theta < \tfrac{1}{5}, \quad c_0\delta < \theta^{\mu+2}, \quad 0 < \delta < \theta, \tag{6}$$

and then take $r>0$ sufficiently small so as to satisfy (3). First we estimate the solution ψ of (4) given by the series

$$\psi=\sum_{k=2}^{\infty}\frac{a_k}{\lambda^k-\lambda}\zeta^k .$$

Since f is analytic in $|z|<r$, by Cauchy's estimate we have

$$k|a_k|\leqq\frac{\delta}{r^{k-1}},$$

and in view of (1) the series ψ is seen to converge in $|\zeta|<r$. Indeed, in the somewhat smaller domain $|\zeta|<r(1-\theta)$ one obtains, using (1), the estimate

$$\begin{aligned}|\psi'|&\leqq\sum_{k=2}^{\infty}\frac{k|a_k|}{|\lambda^k-\lambda|}|\zeta|^{k-1}\leqq\frac{c_0\delta}{\mu!}\sum_{k=1}^{\infty}k^{\mu}(1-\theta)^k\\&<c_0\delta\sum_{k=0}^{\infty}\binom{k+\mu}{\mu}(1-\theta)^k=\frac{c_0\delta}{\theta^{\mu+1}},\end{aligned}$$

which together with (6) gives

$$|\psi'|<\theta\quad\text{in}\quad|\zeta|<r(1-\theta),\tag{7}$$

whereupon integration yields

$$|\psi|<\frac{c_0\delta}{\theta^{\mu+1}}r<\theta r\quad\text{in}\quad|\zeta|<r(1-\theta).\tag{8}$$

This inequality shows that the substitution C maps the disk $|\zeta|<r(1-4\theta)$ into $|z|<r(1-3\theta)$, for by (8) we have

$$|z|\leqq|\zeta|+|\psi|<r(1-4\theta)+r\theta=r(1-3\theta).$$

Moreover, we claim that C^{-1} is defined in $|z|<r(1-2\theta)$ and maps this disk into $|\zeta|<r(1-\theta)$. To prove this, we have to show that for z in $|z|<r(1-2\theta)$ the equation

$$\zeta+\psi(\zeta)=z$$

has a unique solution in $|\zeta|<r(1-\theta)$. This follows, for example, from the explicit construction of ζ as $\lim_{n\to\infty}\zeta_n$, where $\zeta_0=0$ and

$$\zeta_{n+1}+\psi(\zeta_n)=z\qquad(n=0,1,2,\ldots).$$

Indeed, the above defines a sequence ζ_n of analytic functions of z which, by (7), satisfies

$$|\zeta_{n+1}-\zeta_n|=|\psi(\zeta_n)-\psi(\zeta_{n-1})|\leqq\theta|\zeta_n-\zeta_{n-1}|,$$

and hence

$$|\zeta_{n+1}-\zeta_n| \leqq \theta^n |\zeta_1-\zeta_0| = \theta^n |z|,$$

provided that $|\zeta_k| < r(1-\theta)$ $(k=0,1,\ldots,n)$. Thus, for $|z|<r(1-2\theta)$ we have

$$|\zeta_{n+1}| \leqq \sum_{k=1}^{n+1} |\zeta_k-\zeta_{k-1}| < (1-\theta)^{-1}|z| < \frac{1-2\theta}{1-\theta} r < (1-\theta) r,$$

which shows that the functions ζ_n are defined and analytic for $|z|<(1-2\theta)r$ $(n=0,1,2,\ldots)$. Since $\theta<1$, the sequence ζ_n converges for $|z|<(1-2\theta)r$ to $\zeta=\zeta(z)$, the desired inverse function of $\zeta+\psi(\zeta)$.

This allows us to define the transformation $S_+ = C^{-1}SC$ in the disk $|\zeta|<r(1-4\theta)$, since C maps this disk into $|z|<r(1-3\theta)$ which, in view of (3), (6), is mapped by S into

$$|z_1| \leqq |z| + |\hat{f}| < |z| + \delta r < r(1-2\theta)$$

and, finally, the latter is mapped by C^{-1} into the disk $|\zeta_1|<r(1-\theta)$. Consequently the function $\hat{g}$ in (5) is certainly analytic for $|\zeta|<r(1-4\theta)$, and we will estimate it in this region. To this end we express the relation $CS_+ = SC$ in the form

$$g(\zeta)+\psi(g) = f(\zeta+\psi)$$

or

$$\hat{g}(\zeta)+\psi(\lambda\zeta+\hat{g}) = \lambda\psi + \hat{f}(\zeta+\psi),$$

whereupon subtracting the defining relation (4) for ψ we obtain

$$\hat{g}(\zeta) = \psi(\lambda\zeta) - \psi(\lambda\zeta+\hat{g}) + \hat{f}(\zeta+\psi) - \hat{f}(\zeta).$$

We now use the mean value theorem to estimate $\gamma = \sup|\hat{g}(\zeta)|$ over $|\zeta|<r(1-4\theta)$, obtaining

$$\begin{aligned}\gamma &\leqq \sup|\psi'|\gamma + \sup|\hat{f}(\zeta+\psi)-\hat{f}(\zeta)| \\ &\leqq \theta\gamma + \sup|\hat{f}(\zeta+\psi)-\hat{f}(\zeta)|,\end{aligned}$$

and since $\theta<\frac{1}{5}$, from (3), (8) we have

$$\begin{aligned}\gamma &\leqq \frac{5}{4}\sup|\hat{f}(\zeta+\psi)-\hat{f}(\zeta)| \\ &\leqq \frac{5}{4}\delta\sup|\psi| < c_1\frac{\delta^2}{\theta^{\mu+1}} r.\end{aligned}$$

Finally, applying Cauchy's estimate to $|\hat{g}'|$ in a somewhat smaller domain, we get

(9) $$|\hat{g}'| < c_1 \frac{\delta^2}{\theta^{\mu+2}} \quad \text{for} \quad |\zeta| < r_+ = r(1-5\theta).$$

The essential feature of this estimate is the quadratic dependence of $|\hat{g}'|$ on the previous deviation δ, which gives rise to very fast convergence. We now iterate the above construction with $S_\nu, C_\nu, S_{\nu+1}$ in place of S, C, S_+, taking care that the quantities $r_\nu, \theta_\nu, \delta_\nu$ appearing in place of r, θ, δ, which now depend on ν $(\nu = 0, 1, 2, \ldots)$, are chosen so that (6), (9) hold at each step, and that the successive domains do not shrink to a point. This is achieved by making the initial choice of δ_0 sufficiently small and then setting

$$r_\nu = \frac{r}{2}(1+2^{-\nu}) \qquad (\nu = 0, 1, \ldots),$$

while defining θ_ν by the relation

$$\frac{r_{\nu+1}}{r_\nu} = 1 - 5\theta_\nu .$$

This last choice, which gives

$$5\theta_\nu = \frac{1}{2(2^\nu + 1)},$$

is motivated by the need to replace $|z| < r$ by $|\zeta| < r(1-5\theta)$ when passing from the estimate of S to that of S_+ in (9). The quantity $\delta_{\nu+1}$, which will estimate the deviation of $S_{\nu+1}$ from T, is now defined by

(10) $$\delta_{\nu+1} = \frac{c_1 \delta_\nu^2}{\theta_\nu^{\mu+2}} < c_2^{\nu+1} \delta_\nu^2 \qquad (\nu = 0, 1, \ldots).$$

Observing that the sequence $\eta_\nu = c_2^{\nu+2} \delta_\nu$ satisfies

$$0 < \eta_{\nu+1} < \eta_\nu^2$$

and therefore tends to zero faster than exponentially, provided that $\eta_0 < 1$,we see that for

$$\delta_0 < c_2^{-2}$$

the sequence δ_ν tends to zero as $\nu \to \infty$. One readily verifies that also (6) holds for $\delta = \delta_\nu, \theta = \theta_\nu$ $(\nu = 0, 1, 2, \ldots)$, provided that δ_0 is chosen sufficiently small.

With the above choice of $r_\nu, \theta_\nu, \delta_\nu$, we proceed with our construction. We assume that S_ν is represented in the form

$$z_1 = f(z) = \lambda z + \hat{f}(z)$$

with

$$|\hat{f}'| < \delta_\nu \quad \text{in} \quad |z| < r_\nu ,$$

whereupon the previous considerations lead to a substitution $C = C_\nu$ transforming S_ν into $S_{\nu+1} = C_\nu^{-1} S_\nu C_\nu$, which we express in the form (5). By (9), (10) we have

$$|\hat{g}'| < \delta_{\nu+1} \quad \text{in} \quad |\zeta| < r_{\nu+1} ,$$

and we may therefore proceed inductively.

It is now easy to show that the sequence B_ν in (2) converges to the desired substitution in $|\zeta| < r/2$. First we recall that $C_{\nu+1}$ maps the disk of radius $r_{\nu+1}(1 - 4\theta_{\nu+1})$ into that of radius

$$r_{\nu+1}(1 - 3\theta_{\nu+1}) < r_{\nu+1} = r_\nu(1 - 5\theta_\nu) < r_\nu(1 - 4\theta_\nu) ,$$

so that $C_\nu C_{\nu+1}$ is defined for $|\zeta| < r_{\nu+1}(1 - 4\theta_{\nu+1})$. It follows by induction that B_ν is defined in $|\zeta| < r_\nu(1 - 4\theta_\nu)$, and since this radius is larger than

$$r_\nu(1 - 5\theta_\nu) = r_{\nu+1} > \frac{r}{2} ,$$

each substitution B_ν maps $|\zeta| < r/2$ into $|z| < r$. Moreover, since each of the factors $C_\varkappa$ in B_ν is a substitution of the form $z = \chi_\varkappa(\zeta) = \zeta + \psi_\varkappa$ with the identity as its linearized part, the same is true of each B_ν.

To show convergence of the sequence B_ν in $|\zeta| < r/2$, we express B_ν in the form $z = \beta_\nu(\zeta)$, with β_ν defined inductively by $\beta_0(\zeta) = \chi_0(\zeta)$ and

$$\beta_\nu(\zeta) = \beta_{\nu-1}(\chi_\nu(\zeta)) \qquad (\nu = 1, 2, \ldots) .$$

Using $\|\beta_\nu'\|$ to denote the maximum of $|\beta_\nu'|$ in $|\zeta| \leqq r_\nu(1 - \theta_\nu)$, we have in this domain

$$|\beta_\nu'| \leqq \|\beta_{\nu-1}'\| \, (1 + |\psi_\nu'|) ,$$

which in conjunction with (7) gives

$$\begin{aligned} \|\beta_\nu'\| &\leqq \|\beta_{\nu-1}'\| \, (1 + \theta_\nu) \\ &\leqq \prod_{\varkappa=0}^{\nu} (1 + \theta_\varkappa) \leqq \prod_{\varkappa=0}^{\infty} (1 + \theta_\varkappa) = c_3 , \end{aligned}$$

where the infinite product is readily seen to converge. This together with (8) shows that

$$\begin{aligned} |\beta_{\nu+1}(\zeta) - \beta_\nu(\zeta)| &= |\beta_\nu(\chi_{\nu+1}(\zeta)) - \beta_\nu(\zeta)| \\ &\leqq c_3 |\chi_{\nu+1} - \zeta| = c_3 |\psi_{\nu+1}| < c_3 \theta_{\nu+1} r , \end{aligned}$$

from which it follows that as $\nu \to \infty$ the sequence β_ν converges in $|\zeta| < r/2$ to an analytic function $\beta(\zeta)$. Since β_ν transforms S_0 into $S_{\nu+1}$, we have $B_\nu S_{\nu+1} = S_0 B_\nu$ or, writing $S_\varkappa$ in the form $z = f_\varkappa(\zeta)$ ($\varkappa = 0, 1, \ldots$),

$$\beta_\nu(f_{\nu+1}(\zeta)) = f_0(\beta_\nu(\zeta))$$

with $|f_{\nu+1}(\zeta) - \lambda\zeta| < \delta_{\mu+1} r \to 0$ as $\nu \to \infty$. Consequently, letting $\nu \to \infty$, we obtain

$$\beta(\lambda\zeta) = f_0(\beta(\zeta)),$$

so that β is a solution to Schröder's functional equation. Moreover, since uniform convergence of analytic functions implies convergence of derivatives in the interior, we have $\beta'(0) = 1, \beta(0) = 0$. Thus, the power series for β must agree with the unique formal expansion of ϕ. This completes the convergence proof.

The basic idea of using such a sequence of substitutions to prove convergence was introduced by A. N. Kolmogorov [1, 2] in a different context. The crucial point in this method is to find an iteration scheme in which the new error depends quadratically on the previous one, as in the case of Newton's method for finding roots of a function. This suffices to counteract efficiently the growth factors due to the small divisors $\lambda^n - 1$, as seen in (10), where the effect of the small divisors is reflected in the coefficient $c_2^{\nu+1}$. For a discussion of this method in the problem of transforming mappings into normal form we refer to [3].

In his original proof of this theorem, Siegel [4] actually succeeded by direct estimates of the coefficients, as in Cauchy's method of majorants. This, however, required more delicate estimates on the small divisors than (1) and, in particular, it was necessary to use the fact that the expressions $\lambda^n - 1$ are small for only relatively few integers n. On the other hand, for the theory of stability, to be developed in this chapter, it will be essential to have a quadratically convergent scheme, for which cruder estimates of the small divisors suffice.

§ 27. The Poincaré Center Problem

We consider a system of differential equations

$$\dot{x}_k = f_k(x) \qquad (k = 1, \ldots, m) \tag{1}$$

for which $x = 0$ is an equilibrium solution, whereby the functions $f_k(x)$ are convergent power series in a neighborhood of $x = 0$ with real coefficients and without constant term. If $x(t, \xi)$ denotes the solution to (1) with initial condition $x(0, \xi) = \xi$, the association of $x(t, \xi)$ to ξ for