On linear functional equations

Friedrich Riesz (in Kolozsvár)

The present work deals with the inverse problem for a certain class of linear transformations of continuous functions, along with an application to FREDHOLM's integral equation. In this, we are less concerned with new results than with a test of a extremely simple method. All is based on several theorems, developed in §1 and concerning linear functional manifolds [linear spaces of functions], which derive almost immediately from the definition of uniform convergence. The most important proofs are a kind of finiteness proof which show that certain processes cannot be continued indefinitely but must necessarily stop. The most important concept used therein is the concept, introduced into general set theory by Mr FRÉCHET, of a compact set (here, more specifically, compact sequence) which has been of great use in various branches of Analysis. This concept permits a particularly simple and fortunate formulation of the definition of a completely continuous transformation which imitates in essence the similar concept formulation of Mr HILBERT for functions of infinitely many variables.

The restriction made in this work to continuous functions doesn't really matter. The reader familiar with the recent investigations into various function spaces will recognize at once the general applicability of the method; he will also notice that some of these, among them the collection of square-integrable functions and the infinite-dimensional HILBERT space, permit simplifications, while the seemingly simpler case treated here may be considered a touchstone for the general applicability.

§1. Definitions and propositions["auxiliary theorems"].

In the following, we consider the collection of all functions $f(x)$ defined on the interval $a \leq x \leq b$ and continuous there. Hence the variable x is assumed to be real, while function values may be complex. But I want to stress at once that our developments are also directly valid for the smaller collection of all real functions.

For the sake of brevity, we will call the collection considered a *functional space*. In addition, we call the maximal value of $|f(x)|$ the *norm of* $f(x)$ and denote it $\|f\|$; the quantity $\|f\|$ is thus positive in general and vanishes only when $f(x)$ is identically zero. Further, we have the relations

$$\|cf(x)\| = |c|\|f(x)\|; \quad \|f_1 + f_2\| \leq \|f_1\| + \|f_2\|.$$

By the *distance* of the functions f_1, f_2 we mean the norm $\|f_1 - f_2\| = \|f_2 - f_1\|$ of its difference. With that, the uniform convergence of the function sequence $\{f_n\}$ to the limit function f is equivalent with the distance $\|f - f_n\|$ converging to 0. A necessary and sufficient condition for the uniform convergence of a sequence $\{f_n\}$ is, according to the so-called general convergence principle, the relation $\|f_m - f_n\| \to 0$ for $m \to \infty$, $n \to \infty$. In particular, a sequence $\{f_n\}$ for which all distances $\|f_m - f_n\|$ $(m \neq n)$ have a nonzero, hence strictly positive, infimum cannot converge uniformly.

In the following, we are going to be concerned with the inverse problem for *linear transformations*. A transformation T which associates to each element f of our functional space a uniquely determined element $T[f]$ is going to be called *linear* if it is *distributive* and *bounded*. The transformation is called distributive if for all f

$$T[cf] = cT[f]; \quad T[f_1 + f_2] = T[f_1] + T[f_2].$$

The transformation is called bounded when there is a constant M such that, for all f,

$$\|T[f]\| \leq M\|f\|.$$

It follows at once from the definition that T maps every bounded function sequence $\{f_n\}$, i.e., any sequence for which every $\|f_n\|$ lies below some bound, again into such a one. Also, it follows from

$$\|T[f] - T[f_n]\| = \|T[f - f_n]\| \leq M\|f - f_n\|$$

that each uniformly convergent sequence is carried into such a one, and that the limiting functions correspond to each other, in short, that T is continuous.

The notations cT, $T_1 + T_2$, T_1T_2, T^n are so immediate that there is no need to explain them further. It is also immediate that the new transformation derived in this way, as sum, product, or power of those, also are linear.

We denote by E the *identity transformation* that associates each function to itself. We are now going to be concerned with the inversion of transformations of the type $B = E - A$, where E is the identity transformation but A belongs to a special type, namely it is *completely continuous*. In order to introduce the concept of complete continuity and also capture it in the right way, we must first discuss another concept, that of a *compact sequence*.

Following Fréchet, a sequence $\{f_n\}$ is called compact if each of its subsequences contains a further uniformly convergent subsequence. In particular, each uniformly convergent sequence is compact but not conversely since, e.g., through the interweaving of two uniformly convergent sequences with different limit functions one also obtains a compact sequence.

A necessary and sufficient condition for a sequence to be compact has already been given long ago by Arzelà[1]. We won't make any use of this for the time being and merely give a condition whose absence indicates in a given case that a given sequence is *not* compact. This condition is that, for each compact sequence $\{f_n\}$, the infimum of the distances $\|f_m - f_n\|$ $(m \neq n)$ must be zero since the sequence contains uniformly convergent [sub]sequences.

A further property of compact sequences of importance here is that each compact sequence is also bounded. For, in the contrary case, it would have to contain a sequence with norms monotonely growing to infinity all of whose subsequences would have the same property hence could not be uniformly convergent. *On the other hand, not every bounded sequence needs to be compact:* E.g., for $0 \leq x \leq 1$, the sequence $f_n(x) = x^n$ is bounded but not compact since it and all its subsequences converge to a function that is discontinuous at $x = 1$.

The fact just stressed, namely that a sequence can be bounded without being compact, provides the basis for what is special about the *completely continuous* linear transformations compared to general ones. For, as we already said, each linear transformation carries bounded sequences to bounded sequences, uniformly convergent ones to uniformly convergent ones and thus also compact ones to compact ones. We now define: a linear transformation is to be called *completely continuous* when it carries each *bounded* sequence to a *compact* one.

Simplest examples of completely continuous transformations are: $T[f] = f(a)$ which carries each function $f(x)$ to the constant function $= f(a)$; also, $T[f] = f(a) + f(b)x$ or, more generally, $T[f] = f(a_1)g_1(x) + \cdots + f(a_m)g_m(x)$, where $a_1, \ldots, a_m, g_1, \ldots, g_m$ are given points of the interval, resp. given continuous functions. Further examples are provided by the integral

$$T[f] = \int_a^x f(x)\,\mathrm{d}x$$

and, more generally, the integral

$$K[f] = \int_a^b k(x,y)f(y)\,\mathrm{d}y,$$

with which we are going to be concerned in the application to the Fredholm integral equation of the more general results to be obtained. The simplest example of a *not* completely continuous transformation offers the identity transformation E which carries each sequence, hence also each bounded but not compact one, to itself.

It follows immediately from the definition that *the product T_1T_2 is certainly completely continuous when at least one factor is completely continuous.* Since, further, multiplication by a constant or the termwise addition produces compact sequences from compact sequences, it follows that, along with T, T_1, T_2, also cT and $T_1 + T_2$ are completely continuous.

[1] C. Arzelà, "Sulle funzioni di linee", Memorie d. R. Accad. d. Scienze di Bologna, serie 5, t. V (1895), p. 225–244.

We have to explain one more concept which is basic for what is to follow, namely the concept of the *linear manifold*. By this, we mean any manifold of elements of our functional space that satisfies the following conditions: 1) with f, f_1, f_2 it also contains cf, $f_1 + f_2$; 2) are the elements of a uniformly convergent sequence f_n contained in it, then it also contains the limit function f. Examples of linear manifolds are provided by the functional space itself, also, in order to mention at once the other extreme, the manifold consisting of the sole function $f = 0$. Further, as follows directly from the definition, each arbitrary set of functions determines also two linear manifolds, namely 1) the collection of all linear combinations and their limit functions (in the sense of uniform convergence), 2) the collection of all continuous functions for which the product integral with any element of the set is zero.

We want to establish some theorems concerning linear manifolds that derive almost immediately from the definitions and which will serve us as lemmas in the considerations to follow.

Proposition 1. *If L is an arbitrary linear manifold and g is a function not belonging to it, then there is a function f_1 in L such that, for all functions f in L, there holds the inequality*

$$\|g - f\| \geq \frac{1}{2}\|g - f_1\|.$$

Proof: Since the function g does not belong to the manifold L, the infimum d of the distances $\|g - f\|$ is different from zero; for, in the contrary case, L would contain a sequence converging uniformly to g, hence also g. We now choose f_1 so that $\|g - f_1\| \leq 2d$; since, on the other hand, the distance $\|g - f\| \geq d$ for all f, our inequality follows.

Proposition 2. *If one of the two linear manifolds L_1, L_2, say L_2, is a proper part of L_1, i.e., if L_2 is contained in L_1 without being identical, then there exists in L_1 a function g_1 such that, on the one hand*

$$\|g_1\| = 1,$$

on the other hand, for all elements f of L_2,

$$\|g_1 - f\| \geq \frac{1}{2}.$$

Proof: By assumption, L_1 contains at least one element g that doesn't belong to L_2. By proposition 1, there then is in L_2 an element f_2 so that, for all f in L_2 there holds the inequality

$$\frac{\|g - f\|}{\|g - f_2\|} \geq \frac{1}{2}.$$

We set

$$g_1 = \frac{g - f_2}{\|g - f_2\|};$$

then we have $\|g_1\| = 1$, further, g_1, as a linear combination of g and f_2, is in L_1 and, finally,

$$\|g_1 - f\| = \|\frac{g - f_2}{\|g - f_2\|} - f\| = \frac{\|g - f_2 - \|g - f_2\|f\|}{\|g - f_2\|} = \frac{\|g - f_3\|}{\|g - f_2\|},$$

where the function $f_3 = f_2 + \|g - f_2\|f$, being a linear combination of f_2 and f, is in L_2; therefore we also have

$$\|g_1 - f\| = \frac{\|g - f_3\|}{\|g - f_2\|} \geq \frac{1}{2}.$$

In both propositions, the number $\frac{1}{2}$ can evidently be replaced by an arbitrary positive number < 1. On the other hand, in general, one cannot replace it by 1 itself. E.g., if we take for L_1 the collection of all functions for which $g(a) = 0$, but for L_2 the one for which in addition also its integral over (a, b) vanishes. If now there were a function g_1 in L_1 such that $\|g_1\| = 1$ and, for all $f \in L_2$, the distance $\|g_1 - f\| \geq 1$, then this function g_1 would also have to have the additional extremal property that it maximizes the absolute value of the integral of g over all g with $\|g\| \leq 1$. For, if there were a function g_2 in L_1 for which $\|g_2\| \leq 1$ and the integral greater than for g_1, then the equation

$$\int_a^b g_1(x)\,\mathrm{d}s - \xi \int_a^b g_2(x)\,\mathrm{d}s = 0$$

would provide a number ξ for which $|\xi| < 1$ and, on the other hand, the function $f = g_1 - \xi g_2$ would belong to L_2. But then, $g_1 - f = \xi g_2 \leq |\xi| < 1$, contrary to our assumption. Therefore, the absolute value of the integral reaches its maximum at g_1. But now, because of the condition $\|g\| \leq 1$, this maximum is certainly $\leq b - a$, on the other hand one can come arbitrarily close to this value $b - a$ by functions g that are almost everywhere equal to 1 and only near a approach the value 0 continuously. Therefore, the integral of g_1 is, in absolute value, equal to $b - a$, i.e., equal to the length of the interval of integration. But that, because of the continuity of g_1 and since $\|g_1\| = 1$, would be possible only if everywhere $|g_1| = 1$ which contradicts the assumption $g_1(a) = 0$.

The example just discussed shows that, in Proposition 2, the number $\frac{1}{2}$ cannot, in general, be replaced by 1. A corresponding example for Proposition 1 is obtained by choosing for L the manifold L_2 just used, and choosing for g the function $x - a$ or any arbitrary function from L_1 that doesn't also belong to L_2. In one particular case, though, one may use in both theorems the number 1, namely when L, respectively L_2 is finite-dimensional. By this we mean the case where all elements of the manifold are linear combinations of a *finite* number of them. It is sufficient to rewrite correspondingly only the first proposition.

Proposition 3. *If L is a linear manifold of finite dimension and g is a function not belonging to it, then there exists in L a function f^* such that, for all f in L, there holds the inequality*

$$\|g - f\| \geq \|g - f^*\|.$$

The proof of this assertion is based on the

Proposition 4. *When a sequence of elements of a linear manifold of finite dimension is bounded, then it is also compact.*

Proof of the Propositions 3. and 4.: By assumption, all elements of this manifold can be written

$$g = c_1 g_1 + c_2 g_2 + \cdots + c_k g_k.$$

We may assume that the functions $g_1, \ldots, g_k$ on which this representation is based are linearly independent; in the contrary case we would leave off the superfluous ones. To prove 4., it is now sufficient to prove that the assumption of a bound for

$$\|g\| = \|c_1 g_1 + c_2 g_2 + \cdots + c_k g_k\|$$

implies the existence of a corresponding bound for all $|c_i|$, i.e., that, for a bounded sequence of elements g, also the corresponding points $(c_1, \ldots, c_k)$ of k-dimensional space form a bounded sequence, and this immediately implies Proposition 4., by the Bolzano-Weierstrass Theorem.

Thus, it remains to show that the boundedness of $\|g\|$ implies also a bound for the $|c_i|$. The contrary assumption would imply the existence of a bounded sequence of functions g for which the corresponding sums $|c_1| + \cdots + |c_k|$ grow without bound. From this sequence we could then obtain, by dividing each function by the corresponding sum $|c_1| + \cdots + |c_k|$, a new sequence which converges uniformly to 0, and for each of its elements we had $|c_1| + \cdots + |c_k| = 1$. By the Bolzano-Weierstrass theorem, there would then be a subsequence for which the corresponding coefficients c_i converge to corresponding limit values c_i^*, and also $|c_1^*| + \cdots + |c_k^*| = 1$. But, since $c_1 \to c_1^*, \ldots, c_k \to c_k^*$ implies

$$c_1 g_1 + \cdots + c_k g_k \to c_1^* g_1 + \cdots + c_k^* g_k,$$

while, on the other hand, the whole sequence, hence also this subsequence, converges to 0, we would have to have

$$c_1^* g_1 + \cdots + c_k^* g_k = 0,$$

hence, because of the assumed linear independence of the functions $g_1, \ldots, g_k$, also $c_1^* = 0, \ldots, c_k^* = 0$; but this is contradicted by $|c_1^*| + \cdots + |c_k^*| = 1$.

Thus 4. is proved. Now, 3. follows from 4. by the following considerations. It is to be proved that $\|g - f\|$ actually takes on its infimum. Let $\{f_n\}$ be a sequence for which $\|g - f_n\|$ converges to the infimum d of $\|g - f\|$; then the sequence $\{g - f_n\}$ is certainly bounded and, because of $\|f_n\| \leq \|g\| + \|g - f_n\|$, so is the sequence $\{f_n\}$. By Proposition 4., the bounded sequence $\{f_n\}$ is therefore also compact. Thus, there is a uniformly convergent subsequence, and the limit function f^* of this subsequence has, because of $\|g - f^*\| \leq \|g - f_n\| + \|f_n - f^*\| \to d$ the desired property to provide a minimum for $\|g - f\|$.

Proposition 5. which we now establish is a complement to Proposition 4.; for it states that the compactness of all bounded sequences of elements of a linear manifold of *finite* dimension is *characteristic*.

Proposition 5. *If every bounded sequence of elements of a linear manifold is compact, then the manifold is finite-dimensional.*

Proof: In the contrary case, the manifold would contain a sequence $\{g_n\}$ all of whose elements are linearly independent of the remaining ones, i.e., none can be written as a linear combination of its predecessors. Denoting by L_k the collection of all linear combinations of $g_1, \ldots, g_k$, then it is certain that g_{k+1} is not contained in L_k. On the other hand, L_k is linear manifold; for, on the one hand it contains all linear combinations of its elements, on the other hand, as we made clear in the proof of Proposition 4., the condition $\|c_1 g_1 + \cdots + c_k g_k\| \to 0$ also implies $c_1 \to 0, \ldots, c_k \to 0$, consequently, the uniform convergence of a sequence $\{g^{(n)} = c_1^{(n)} g_1 + \cdots + c_k^{(n)} g_k\}$ to a limit function g^* implies the convergence of the coefficients $c_i^{(n)}$ to corresponding limit values c_i^*, hence $g^* = c_1^* g_1 + \cdots + c_k^* g_k$, hence it belongs to the manifold. Since, further, L_k is a proper subset of L_{k+1}, there is, by Proposition 2., a function f_k such that $\|f_k\| = 1$ while its distance from every function in L_k is at least $\frac{1}{2}$.[1] The functions f_k form, because of $\|f_k\| = 1$, a bounded sequence. On the other hand, for $i \neq k$, the distance $\|f_i - f_k\| \geq \frac{1}{2}$ since either f_i belongs to the manifold L_k or f_k belongs to the manifold L_i. Thus, the infimum of the distances $\|f_i - f_k\|$ $(i \neq k)$ is different from zero and the *bounded* sequence $\{f_k\}$ is, thus, not compact.

Proposition 6. *If the linear manifolds L_1 and L_2 have no common element other than $f = 0$ and if at least one of them is finite-dimensional, then there exists a constant C so that for every element f in L_1 and every element g in L_2 there holds*

$$\|f\| + \|g\| \leq C\|f + g\|.$$

Proof: In the contrary case, there would exist sequences $\{f_n\}$ and $\{g_n\}$ such that $\|f_n\| + \|g_n\| > n\|f_n + g_n\|$, and we can assume without loss of generality that $\|f_n\| + \|g_n\| = 1$ since this can always be achieved by dividing both f_n and g_n by $\|f_n\| + \|g_n\|$. Now assume that, e.g., L_1 is finite-dimensional; then, by Proposition 4., the bounded sequence $\{f_n\}$ is also compact; there is therefore a uniformly convergent subsequence $f^{(n)} \to f^*$. Since also $\|f^{(n)} + g^{(n)}\| < \frac{1}{n} \to 0$, hence uniformly $f^{(n)} + g^{(n)} \to 0$, so also uniformly $g^{(n)} \to -f^*$. This implies, on the one hand, $\|f^{(n)}\| \to \|f^*\|$,[2] $\|g^{(n)}\| \to \|-f^*\| = \|f^*\|$, and so, because of $\|f^{(nj)}\| + \|g^{(n)}\| = 1$, also $2\|f^*\| = 1$; on the other hand, f^*, as limit function of $\{f^{(n)}\}$ resp. of $\{-g^{(n)}\}$, would have to belong to both manifolds L_1, L_2 and therefore would have to vanish everywhere, which contradicts the relation $2\|f^*\| = 1$ just proved.

§2. The inversion of the linear transformation.

[The rest of the paper deals with the inversion of a map $B = E - A$ with A completely continuous, i.e., what we now call a *compact perturbation of the identity*, bringing for the first time an abstract discussion and proof of all the basic facts now rightly thought classical, including (in order) the finite dimensionality of the kernel of B, the uniform finite dimensionality of the kernels of its powers, the fact that B onto implies B 1-1 in which case B is bounded below, the closedness of the range of B, the existence of a finite n for which the space is the direct sum of $\ker B^n$ and $\operatorname{ran} B^n$, the fact that B onto implies B 1-1; etc. etc.

The original article, entitled 'Über lineare Funktionalgleichungen', appeared in *Acta Math.* **41** (1918), 71–98.]

translated by C. de Boor

[1] Since L_k is finite-dimensional, we could replace $\frac{1}{2}$ by 1; but this deeper fact doesn't matter here; the corresponding Proposition 3. is going to be used only later.

[2] One obtains the limit equation $\|f^{(n)}\| \to \|f^*\|$ for every uniformly convergent sequence $f^{(n)} \to f^*$ most simply from the two inequalities $\|f^*\| \leq \|f^* - f^{(n)}\| + \|f^{(n)}\|$, $\|f^{(n)}\| \leq \|f^* - f^{(n)}\| + \|f^*\|$ and the limit equation $\|f^* - f^{(n)}\| \to 0$.