On the history of the isomorphism problem of dynamical systems with special regard to von Neumann's contribution

Miklós Rédei · Charlotte Werndl

Abstract   This article reviews some major episodes in the history of the spatial iso-morphism problem of dynamical systems theory (ergodic theory). In particular, by analysing, both systematically and in historical context, a hitherto unpublished letter written in 1941 by John von Neumann to Stanislaw Ulam, this article clarifies von Neumann's contribution to discovering the relationship between spatial isomorphism and spectral isomorphism. The main message of the article is that von Neumann's argument described in his letter to Ulam is the very first proof that spatial isomor-phism and spectral isomorphism are not equivalent because spectral isomorphism is weaker than spatial isomorphism: von Neumann shows that spectrally isomorphic ergodic dynamical systems with mixed spectra need not be spatially isomorphic.

1 Introduction

The aim of this article is to review some major episodes in the history of the spacial iso-morphism problem of dynamical systems. In particular, by publishing and analyzing (both systematically and in historical context) a hitherto unpublished letter[1] written

---

[1] In Sect. 3 we reproduce von Neumann's letter in full. Doing so, we follow von Neumann's nota-tion and terminology closely, and we keep his notation and terminology throughout the article when we review subsequent developments about the isomorphism problem although some of that notation and terminology has become outdated. One instance of such non-standard usage of words is 'spacial isomorphism' instead of 'spatial isomorphism'.

Communicated by : Jeremy Gray.

M. Rédei (✉) · C. Werndl
Department of Philosophy, Logic and Scientific Method, London School of Economics,
Houghton Street, London WC2A 2AE, UK
e-mail: m.redei@lse.ac.uk

C. Werndl
e-mail: c.s.werndl@lse.ac.uk

⚛ Springer

in 1941 by John von Neumann to Stanislaw Ulam, this article aims to clarify von Neumann's contribution to discovering the relationship between spacial isomorphism and spectral isomorphism. The main message of the article is that von Neumann's argument described (in a sketchy way) in his letter to Ulam is the very first proof that spectral isomorphism and spacial isomorphism are *not* equivalent because spectral isomorphism is weaker than spacial isomorphism: von Neumann shows that spectrally isomorphic ergodic dynamical systems with mixed spectra need not be spacially isomorphic.

Dynamical systems are mathematical models of systems whose time evolution is deterministic, and these include physical systems described by classical mechanics and classical statistical mechanics. The origins of dynamical systems, especially the beginnings of ergodic theory, can be traced back to Boltzmann's work on the foundations of classical statistical mechanics in the second half of the nineteenth century (von Plato 1991; Szász 1996).

Modern ergodic theory developed in the early 1930s; the articles of von Neumann (1932a,b,c) and Koopman and Neumann (1932) contributed substantially to the birth of this new discipline. More specifically, the isomorphism problem—the problem of classifying the equivalence classes of dynamical systems with respect to a natural notion of spacial isomorphism—was first formulated in von Neumann (1932c).

The development that gave a decisive boost to dynamical systems theory, and which motivated in particular von Neumann to turn his attention to the study of the ergodic properties of dynamical systems, was the emergence of what has become called 'the Koopman formalism': dynamical systems theory in terms of functional analysis. Koopman (1931) observed that to each dynamical system there corresponds a group of unitary operators on the associated L_2 Hilbert space. Hence the properties of dynamical systems can be analyzed in terms of Hilbert space operator theory. The Koopman formalism proved to be very powerful and successful. Many stochastic properties of dynamical system (such as ergodicity and different types of mixing) could be fully characterized in terms of the Koopman operators associated with the dynamical system. This success raised the hope that the isomorphism problem could also be solved with this technique. Von Neumann (1932c) introduced a very natural notion of spectral isomorphism between Koopman representatives of dynamical systems, and he proved that for a restricted class of ergodic dynamical systems (those with pure point spectrum) spectral isomorphism is equivalent to spacial isomorphism. He also conjectured that this is true in general. If this were true, it would have shown that the technique of the Koopman formalism is exhaustive in the sense of being strong enough to describe every probabilistic aspect of dynamical systems.

It turned out, however, that the situation is much more complicated, and a long and intricate history of research on the relation of spectral isomorphism to spacial isomorphism ensued after von Neumann's 1932 article. A key event in this history is von Neumann's proof in his 1941 letter to Ulam that spacial isomorphism is stronger than spectral isomorphism. Von Neumann never published his proof but his result was referred to in important later publications on the problem. Given the significance of the isomorphism problem, von Neumann's letter is an important historical document. Also, understanding von Neumann's proof is helpful in putting later developments into perspective. As it will be seen, von Neumann's presentation of the proof

🍂 Springer

in his letter is sketchy, and we will fill in the details in order to make the proof understandable.

By giving a review of some major episodes in the history of the isomorphism problem we also hope to contribute to the extremely meager literature on the history of the isomorphism problem: while the literature on the history of ergodic theory is rich and there are a number of articles discussing specifically von Neumann's contribution to it (see Mackey 1990; Ornstein 1990; Zund 2002 and the references therein), we are not aware of any historically oriented article devoted exclusively to the isomorphism problem, although some articles on the history of ergodic theory mention it (e.g., Mackey 1990; Lo Bello 1983).

The structure of the article is the following. Section 2 recalls some definitions, including the Koopman formalism, defines the spacial isomorphism problem and states the main result on the problem that had been known before 1941—von Neumann's 1932 result. In Sect. 3 von Neumann's letter is reproduced without any comments on its content. Section 4 gives a detailed reconstruction of von Neumann's proof, filling in the (sometimes very large) gaps. Section 5 reviews how von Neumann's result was referred to in the mathematics literature. Section 6 is a collection of important more recent results on the isomorphism problem. The article is concluded with Sect. 7 where we highlight the historical significance of von Neumann's letter.

2 The isomorphism problem and von Neumann's 1932 result

First, we recall some definitions that will be needed for the ensuing discussion, using notation that is compatible with that in von Neumann's letter. $(\phi, \Sigma_\phi, \mu, S)$ is a (discrete)[2] *dynamical system* if ϕ^3 is a set (called the phase space), Σ_ϕ is a σ-algebra of subsets of ϕ, μ is a measure on Σ_ϕ with $\mu(\phi) = 1$, and $S : \phi \to \phi$ is a bijective measurable function such that also S^{-1} is measurable.[4]

The definition of spacial isomorphism captures the idea that dynamical systems are equivalent from a probabilistic perspective.

---

[2] We only focus on discrete-time systems because von Neumann's letter and all the relevant results are about them. (The only exception are the notions and results in von Neumann (1932c) discussed in this and the next section. They are formulated for continuous-time systems but can easily be transferred to discrete-time systems.)

[3] 'ϕ' is the notation von Neumann uses in his letter for the phase space (cf. Sect. 3). This differs from modern usage (where 'ϕ' is usually reserved for functions).

[4] We also assume, as is standard in ergodic theory, that (ϕ, Σ_ϕ, μ) is a Lebesgue space (cf. Petersen 1983). (ϕ, Σ_ϕ, μ) is called a *Lebesgue space* if either ϕ is countable or there is a measure space (K, Σ_K, ν), where $K = [a, b] \subseteq \mathbb{R}$ is an interval, Σ_K is the Lebesgue σ-algebra and ν the Lebesgue measure, such that the following holds: There is a countable set $W \subseteq \phi$, there is a $K_0 \subseteq K$ with $\nu(K_0) = 1$, there is a $\phi_0 \subseteq \phi$ with $\mu(\phi_0) = 1$ and there is a bijective function $c : \phi_0 \setminus W \to K_0$ such that

 (i) $c(A) \in \Sigma_K$ for all $A \in \Sigma_\phi$, $A \subseteq \phi_0 \setminus W$,
 (ii) $c^{-1}(B) \in \Sigma_\phi$ for all $B \in \Sigma_K$, $B \subseteq K_0$,
 (iii) $\nu(c(A)) = \mu(A)$ for all $A \in \Sigma_\phi$, $A \subseteq \phi_0 \setminus W$.

⚪ Springer

Definition 1 Two dynamical systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are called *spacially isomorphic* if there are measurable sets $\phi_0 \subseteq \phi$ and $\psi_0 \subseteq \psi$ with the properties 1. and 2. below and a bijection $c : \phi_0 \to \psi_0$ satisfying 3.(i)-(iii):

1. $S(\phi_0) \subseteq \phi_0$, $T(\psi_0) \subseteq \psi_0$;
2. $\mu(\phi \backslash \phi_0) = 0$, $\nu(\psi \backslash \psi_0) = 0$;
3. (i) c and c^{-1} are measurable with respect to
     $\Sigma_{\phi_0} := \{A \cap \phi_0 \mid A \in \Sigma_\phi\}$ and $\Sigma_{\psi_0} := \{B \cap \psi_0 \mid B \in \Sigma_\psi\}$,
   (ii) $\nu(c(A)) = \mu(A)$ for all $A \in \Sigma_{\phi_0}$,
   (iii) $c(S(x)) = T(c(x))$ for all $x \in \phi_0$.

Spacial isomorphism was first defined in von Neumann(1932c). Von Neumann's (1932c) article also was the first that called for a *classification of dynamical systems up to spacial isomorphism* – this is the *spacial isomorphism problem*. The spacial isomorphism problem is widely regarded as one of the most important problems, if not *the* most important problem, of ergodic theory (cf. Halmos 1956, 96; Halmos 1961, 75; Petersen 1983, 4; Rohlin 1960, 1). This is understandable because ergodic theory is the theory of the probabilistic behaviour of dynamical systems; it is thus a crucial question which systems are equivalent from the probabilistic perspective.

There are two main ways to approach this problem. First, one tries to find *spacial invariants* (properties which are the same for spacially isomorphic dynamical systems) which show that certain kinds of dynamical systems are not spacially isomorphic. Second, one aims to find a sufficiently large number of invariants which, taken together, provide sufficient conditions for systems to be isomorphic.

Next we briefly describe the Koopman formalism, which will lead to the notion of spectral isomorphism of dynamical systems. Given a dynamical system $(\phi, \Sigma_\phi, \mu, S)$ consider the Hilbert space $L_2(\phi)$ of complex-valued square integrable functions on ϕ. The operator

$$U : L_2(\phi) \to L_2(\phi), \quad U(f) = f(S(x))$$

is called the *Koopman operator* of $(\phi, \Sigma_\phi, \mu, S)$. Koopman (1931) showed that the Koopman operator is *unitary* (cf. Arnold and Avez 1968). The spectrum of the Koopman operator U is called the *spectrum* of the dynamical system.

Definition 2 Two dynamical systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are called *spectrally (or unitary) isomorphic* if their Koopman operators U and V are unitarily equivalent, i.e. if there is a unitary operator W between $L_2(\phi)$ and $L_2(\psi)$ such that $U = W^*VW$.

Spectral invariants are properties which are the same for spectrally isomorphic dynamical systems. Obviously, the spectrum of a dynamical system is a spectral invariant: two dynamical systems cannot be spectrally isomorphic if they have different spectrum. However, having the same spectrum does not entail that they are spectrally isomorphic (Halmos 1951, 75).

What is the relation of spectral and spacial invariants? Are they identical? If yes, then spacial isomorphism is equivalent to spectral isomorphism. These questions were

raised by von Neumann (1932c). An affirmative answer to the question about the equivalence of spectral and spacial isomorphism would mean that the Koopman formalism can deal with all the probabilistic questions about dynamical systems (cf. von Neumann 1932c; Halmos 1957).

The main result on this problem before von Neumann's 1941 letter to Ulam was presented in von Neumann's (1932c) seminal article. The article starts by listing several properties of dynamical systems which are spacial invariants but are also spectral invariants and hence can be characterized completely in terms of the Koopman representatives of dynamical systems. We recall here two of the most prominent of these properties and their spectral characterizations.

A dynamical system $(\phi, \Sigma_\phi, \mu, S)$ is called (cf. Petersen 1983)

Ergodic: (or *metrically transitive*): if it does not have non-trivial S-invariant measurable sets, i.e. if there is no set $A \in \Sigma_\phi, 0 < \mu(A) < 1$, such that $S(A) = A$.

Spectral characterisation of ergodicity: 1 is a simple proper value[5] of the Koopman operator U.

Weakly mixing: if for all $A, B \in \Sigma_\phi$ we have

$$\lim_{t \to \infty} \frac{1}{t} \sum_{i=0}^{t-1} |\mu(S^i(A) \cap B) - \mu(A)\mu(B)| = 0.$$

Spectral characterisation of weak mixing: 1 is the only proper value of the Koopman operator and this proper value is simple.

The focus of von Neumann 1932c then turned to the question about the relationship between spacial and spectral isomorphism. It is obvious that if $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are spacially isomorphic, then they are spectrally isomorphic: if c is the spacial isomorphism, then W defined by

$$W(f(\phi)) = f(c^{-1}(\psi))$$

sets up the spectral equivalence of the dynamical systems. This implies that spectral invariants, such as the spectrum of a system, are also spacial invariants. Thus the question to answer was whether the converse is true.

To tackle this question, von Neumann (1932c) distinguished between several types of spectra of dynamical systems. A dynamical system $(\phi, \Sigma_\phi, \mu, S)$ is said to have *pure point spectrum* if the proper functions of the Koopman operator U form a basis of $L_2(\phi)$. A dynamical system has *pure continuous spectrum* if the only proper functions of its Koopman operator are $\delta 1, \delta \in \mathbb{C}$. A dynamical system has *mixed spectrum* if it has neither a pure point spectrum nor a pure continuous spectrum.

---

[5] 'Proper value' and 'proper function' are also called 'eigenvalue' and 'eigenfunction', but we follow here (and throughout the article) von Neumann's terminology in his letter (see Sect. 3). 'Proper value' and 'proper function' also is the terminology in the literature we will refer to (e.g. Halmos 1949, 1956). A proper value δ is simple (also called 'non-degenerate') if the proper functions belonging to δ span a one-dimensional subspace in $L_2(\phi)$.

Von Neumann then proved the following theorem.[6]

Theorem 1 *Let $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ be ergodic dynamical systems with pure point spectrum. Then the dynamical systems are spacially isomorphic if, and only if, they are spectrally isomorphic (von Neumann 1932c).*

It is easy to see that spectrally isomorphic systems have the same proper values. Thus Theorem 1 tells us that ergodic systems with pure point spectrum are spacially isomorphic precisely when they have the same proper values.

Theorem 1 was the main result on the isomorphism problem before von Neumann' 1941 letter. The theorem shows that for ergodic systems with pure point spectrum the Koopman formalism can deal with all probabilistic questions about dynamical systems. To the best of our knowledge, nothing was known about the other cases of a mixed spectrum and continuous spectrum before the 1941 letter. Yet in these early days of ergodic theory it was sometimes conjectured that, in general, dynamical systems are spacially isomorphic just in case they are spectrally isomorphic (cf. Halmos 1961, 77). Even von Neumann (1932b), when discussing systems with pure continuous spectrum, states that results similar to Theorem 1 can be expected in this case.

Let us now turn to the next step in the history of the isomorphism problem, viz. von Neumann's letter.

3 Von Neumann's 1941 letter: the text

The letter below was written by John von Neumann to Stanislaw Ulam (the first page of the original is shown in Fig.1). Von Neumann and Ulam were not only colleagues but friends as well. This explains the informal tone and chatty style of the letter and the references to personal matters. Ulam and von Neumann corresponded on a number of issues (both scientific and non-scientific) over many years. A selection of von Neumann's letters to Ulam was published in Rédei (2005). One of these, not fully dated but most likely written in 1939, touches upon the isomorphism problem briefly (Rédei 2005, 252–253): Ulam's attention is called to the article van Neumann (1932c), and in particular to the theorem that systems with pure point spectrum are spacially isomorphic if, and only if, they are spectrally isomorphic (Theorem 1). This letter suggests that von Neumann did not yet know in 1939 that spacial and spectral isomorphism do not always go together and, in particular, that he had not yet discovered the result described in the 1941 letter.

To the best of our knowledge, the 1941 letter has not been published. The original, hand-written version is in the Ulam articles, in the Archives of the American Philosophical Society, Philadelphia, U.S.A. The language of the letter is English. We reproduce it below without comments and keeping the original spelling. The content of it will be analyzed in detail in the next section.

---

[6] For a modern proof, see Halmos (1957, 46–50). This result is also discussed by the more historically oriented articles Halmos (1949), Halmos (1957), Mackey (1974, 197), Mackey (1990), Reed and Simon (1980, Sect. VII.4), Rohlin (1967, 2 and 5) and Weiss (1972, 672–673).
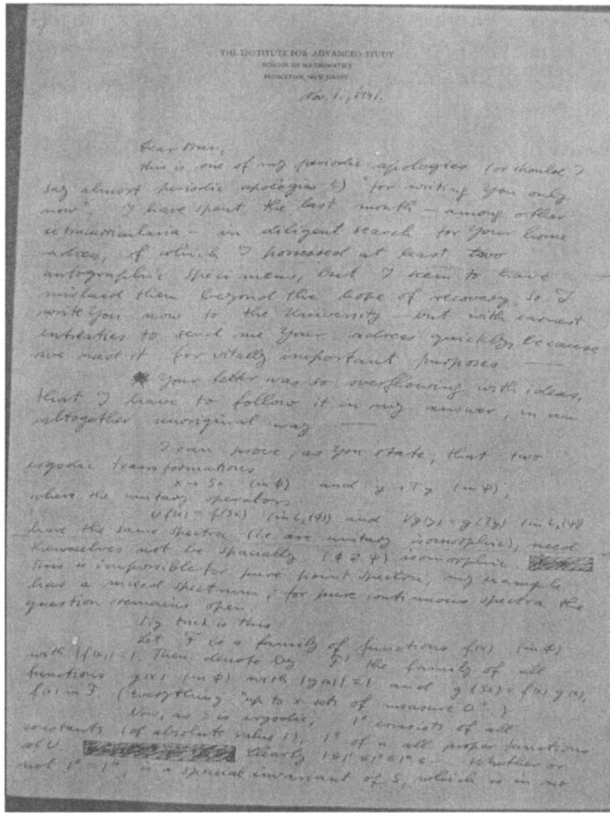
Fig. 1 First page of von Neumann's 1941 letter

November 1, 1941.

Dear Stan,

this is one of my periodic apologies (or should I say almost periodic apologies?) "for writing you only now". I have spent the last month—among other extracurricularia—in diligent search for your home adress, of which I possessed at least two autographic specimens, but I seem to have mislaid them beyond the hope of recovery. So I write you now to the University—but with earnest entreaties to send me your adress quickly, because we need it for vitally important purposes. ——

Your letter was so overflowing with ideas, that I have to follow it in my answer, in an altogether unoriginal way. ——

I can prove, as you state, that two ergodic transformations

$$x \rightarrow Sx \ (\text{in } \phi) \quad \text{and} \quad y \rightarrow Ty \ (\text{in } \psi), \tag{1}$$

where the unitary operators

$$Uf(x) = f(Sx) \ (\text{in } L_2(\phi)) \quad \text{and} \quad Vg(y) = g(Ty) \ (\text{in } L_2(\psi)) \tag{2}$$

have the same spectra (i.e. are spectrally isomorphic), need themselves not be spacially
($\phi \leftrightarrow \psi$) isomorphic. This is impossible for pure point spectra, my example has a
mixed spectrum, for pure continuous spectra the question remains open.

My trick is this:

Let F be a family of functions $f(x)$ (in ϕ) with $|f(x)| = 1$. Then denote by F' the
family of all functions $g(x)$ (in ϕ) with $|g(x)| = 1$ and $g(Sx) = f(x)g(x)$, $f(x)$ in
F. (Everything "up to x-sets of measure 0".)

Now,[7] as S is ergodic, $1'$ consists of all constants (of absolute value 1), $1''$ of all
proper functions of U. Clearly, $1 \subseteq 1' \subseteq 1'' \subseteq 1''' \subseteq \dots$. Whether or not $1'' = 1'''$, is
a spacial invariant of S, which is in no obvious way determined by its spectrum. But:

(α) For a pure point spectrum it must be determined by the spectrum, since then the
spectrum is known to determine S up to a spacial isomorphism.

(β) For a pure continuous spectrum it is equally worthless, since then $1' = 1''$, hence
$1'' = 1'''$.

Thus the chance for this invariant arises when there is a mixed spectrum.

Now let ϕ be the space of all u, v (both mod1), with Lebesgue measure. Choose a
fixed, irrational γ, and define

$$S(u, v) = (u + \gamma, v + u). \tag{3}$$

(This example was constructed sometime ago by Halmos, for a different purpose.)

It is easy to show, that S has this spectrum

I) A simple point spectrum $e^{2\pi i k \gamma}$, $k = 0, \pm 1, \pm 2, \dots$.

II) An everywhere infinitely multiple, Lebesgue measure spectrum, covering all of
$|\lambda| = 1$

$1''$ consists precisely of all $e^{2\pi i k u}$, $k = 0, \pm 1, \pm 2, \dots$. $1'''$ contains $e^{2\pi i v}$. Hence
$1'' \neq 1'''$.

Consider on the other hand any ergodic transformation

$$w \to Rw \quad (\text{in } \Omega) \tag{4}$$

which has this spectrum:

I') A simple point spectrum 1.

II') Same as II) above.

(E.g. the Hopfe one: $w = (x_n \mid n = 0, \pm 1, \pm 2, \dots)$, $Rw = (x_{n+1} \mid n = 0, \pm 1,
\pm 2, \dots)$—each x_n-s range being ± 1.)

Define

$$T(u, w) = (u + \gamma, Rw). \tag{5}$$

---

[7] In this paragraph, and throughout von Neumann's letter, 1 stands for the constant function on ϕ, taking
the value $1 \in \mathbb{R}$. This is slightly confusing because the symbol refers both to a function and the real number
'one'; however, the context will always make clear what 1 stands for. Also note another slight abuse of
notation in this paragraph and in the letter: The previous paragraph defines F' for a *set* F of functions.
However, von Neumann shortens $\{1\}'$ to $1'$ in case where $F = \{1\}$, and the same holds for higher primes.
So, for instance, $\{\{1\}'\}'$ is denoted by $1''$ etc.

⌖ Springer

Then it is easy to show that T has the same spectrum I)+II) as S. And a simple computation gives $1''' = 1'' = $ set of all $e^{2\pi i k u}$, $k = 0, \pm 1, \pm 2, \ldots$, i.e. $1'' = 1'''$. Thus, S and T are spacially non-isomorphic.

Re Kakutani: K. has decided to stay on. He has an I.A.S. stipend for 1941/42. If the war should force him to remain beyond Sept. 1942, we will do everything humanly possible for him. If he could be elected to the Harvard Society of Fellows, that would be super-excellent. Did you find out from Henderson whether there is any chance of this? I will be too glad to do anything, to write to anybody, etc.—if there is a chance of achieving something. ——

Norberts resignation from the Academy: I have no idea whether he has resigned or not. I remember that he used to talk about it 2–3–4 years ago. I think it is nonsense. ——

Many thanks for the invitation to talk at Madison, including the Mammon. I have no way to foresee the future in this slightly opaque century—but I hope that I may get Midwest again before too long—but I don't know for the moment when or how. ——

Many congratulations to your + Oxtoby early article, which appeared so late in the "Annals". It is really a pleasure to read it. ——

My respectful homage to Françoise, God knows that my similarity to Riquet á la Houppe (if any—which I doubt) is rigorously exterior. I am a miserable sinner, and I never rescue maidens, except for base and egoistical motives. But her high—and utterly undeserved—consideration thrills me to the core of my being. ——

I am also highly pleased by the Russian campaign, although I fear that it will be like Old Times—i.e. 1914–1917—when the german wore Russia down after 2–3 years. But they will probably not be able to get rid of some kind of war in that direction—more or less like honorable Sino-Japanese incident.

I suppose that Schickelgruber is licked alright, but it will be a long and bloody affair. And as to the US: You know what the court physician said to the German prince when his wife bore him a daughter: "Majestät werden sich nochmal bemühen müssen." ——

Klari hopes for an epistle from you. It seem to be a difficult piece of accounting-letters. She has had a rather unpleasant eye-trouble lately—probably over-strain—and it is still not quite over. ——

I am writing a book on games and economics with the economist O. Morgenstern. Whadayasay.

Also a lot whatnots on operator theory, for which I have a monopoly in production as well as in consumption. So that's that.

Please remind me to Françoise.
With the best from both of us
hoping to hear from you soon
as ever John.

4 Von Neumann's 1941 letter: explanation of the proof

In the letter von Neumann shows that spectrally isomorphic dynamical systems with a mixed spectrum need not be spacially isomorphic. He starts with the following construction (see paragraphs 4–5). Consider an ergodic dynamical system $(\phi, \Sigma_\phi, \mu, S)$. Recall that $L_2(\phi)$ is the set of all complex-valued square-integrable functions on ϕ.

🅐 Springer

Given a family F of functions $f \in L_2(\phi)$ whose absolute value is always 1 (i.e. $|f(x)| = 1$ for all $x \in \phi$), let F' be the family of functions $g \in L_2(\phi)$ whose absolute value is always 1 (i.e. $|g(x)| = 1$ for all $x \in \phi$) where there is an $f \in F$ such that

$$g(Sx) = f(x)g(x). \tag{6}$$

Let 1 denote the function $f(x) = 1$ for all $x \in \phi$, and consider $1'$ (see footnote 7). For ergodic dynamical systems the only *invariant functions* (i.e. functions g with $g(S(x)) = g(x)$) are the constants (Arnold and Avez 1968). Hence $1'$ is the set of all constant functions of absolute value 1. Thus $1''$ is the set of all proper functions of U of absolute value 1. Clearly, for any sets G, H, if $G \subseteq H$, then $G' \subseteq H'$. Let us call the functions in $1''$, $1'''$, $1''''$ ... *generalised proper functions* with *generalised proper values* $1'$, $1''$, $1'''$ ....[8] Because $1 \subseteq 1'$, $1 \subseteq 1' \subseteq 1'' \subseteq 1'''$ ....

Von Neumann states (see paragraphs 3 and 5) that whether $1'' = 1'''$ is in no obvious way determined by the spectral properties. What is clear is that for spacially isomorphic dynamical systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ either for both $1'' = 1'''$ or for both $1'' \neq 1'''$. This is so because for spacially isomorphic systems their unitary operators are related via $U = W^* V W$ where $W(f) = f(c^{-1}(y))$. Therefore, $U(g) = fg$ if, and only if, $V(W(g)) = W(fg) = W(f)W(g)$, and hence g is a generalised proper function of $(\phi, \Sigma_\phi, \mu, S)$ if, and only if, $W(g)$ is a generalised proper function of $(\psi, \Sigma_\psi, \nu, T)$.

Von Neumann's idea now was *to find two dynamical systems which are spectrally isomorphic but where for one system $1'' = 1'''$ and for the other $1'' \neq 1'''$, implying that they are not spacially isomorphic*. Note that this strategy will not work for ergodic systems with pure point spectrum because for those systems spacial and spectral isomorphism are equivalent. Spectrally isomorphic systems have the same proper values; hence this strategy will also *not* work for systems with *pure continuous spectrum* because their only proper functions are $\delta 1$, $\delta \in \mathbb{C}$, and thus $1' = 1'' = 1'''$. Hence von Neumann focused on systems with *mixed spectrum* and constructed two spectrally isomorphic dynamical systems having mixed spectrum such that for one $1'' = 1'''$ holds, and for the other $1'' \neq 1'''$ holds.

Consider (see paragraph 6) the dynamical system where $\phi = [0, 1) \times [0, 1)$, Σ_ϕ is the Lebesgue σ-algebra, μ is the Lebesgue measure and

$$S(u, v) = (u + \gamma \ (\text{mod} \ 1), v + u \ (\text{mod} \ 1)) \tag{7}$$

for a fixed irrational γ. It is shown for this system that $1'' \neq 1'''$.

---

[8] In the literature they are called like this (Abramov 1964, Halmos 1949, Halmos 1956, Rohlin 1960). They have been first introduced when proving the result of von Neumann's letter (see Sect. 5). This terminology is motivated as follows: $1''$ are proper functions of absolute value 1 with proper values $1'$ of absolute value 1, and the generalised proper functions and proper values are obtained by allowing f in Eq. 6 to be not just the constant functions but any function in $1''$ or $1'''$, etc. From unitarity it follows that the absolute value of any proper value of the Koopman operator is 1. It suffices to consider proper functions of absolute value 1 because for ergodic systems the absolute value of any proper function g is constant, i.e. $|g(x)| = c$ for a constant c for all $x \in \phi$ (Arnold and Avez 1968, 26).

⚛ Springer

It is well known (see paragraph 7) that the functions

$$g_{k,m}(u, v) = e^{2\pi i k u} e^{2\pi i v m} \qquad k, m \in \mathbb{Z}$$

form an orthonormal basis of $L_2(\phi)$ (Arnold and Avez 1968; Halmos 1956). We have

$$U(g_{k,m}) = e^{2\pi i k \gamma} g_{k+m,m};$$

hence $g_{k,0}, k \in \mathbb{Z}$, are proper functions. Let H_0 be the closed subspace in $L_2(\phi)$ generated by $\{g_{k,0} : k \in \mathbb{Z}\}$. The functions $h_k := g_{k,0}, k \in \mathbb{Z}$, form an orthonormal basis of H_0. Thus U restricted to H_0 has pure point spectrum with proper functions h_k and proper values $e^{2\pi i k \gamma}$—this is point I) referred to by von Neumann in the letter.

Now by applying U, the functions $g_{k,m}$ for $m \neq 0$ get permuted among themselves and multiplied by constant factors. We get rid of these constant factors by setting

$$f_{k,m} = a_{k,m} g_{k,m}, \quad \text{with } a_{k,m} = (\sqrt[8m]{e^{2\pi i \gamma k}})^{(2n-m)^2}, \quad k, m \in \mathbb{Z}, \ m \neq 0. \qquad (8)$$

A simple calculation yields $U(f_{k,m}) = f_{k+m,m}$. Since $|a_{k,m}| = 1$, the functions $f_{k,m}$ are still an orthonormal set. We can label the functions $f_{k,m}$ such that we obtain a sequence of functions $h_{i,j}, i, j \in \mathbb{Z}$, with $U(h_{i,j}) = h_{i+1,j}$. The subspace generated by $h_{i,j}$ is the orthogonal complement $H_0^\perp$ of H_0 in $L_2(\phi)$, and $h_{i,j}$ is an orthonormal basis of $H_0^\perp$.

An operator O is said to have an *infinitely multiple homogeneous Lebesgue measure spectrum* if there exists an orthonormal basis $b_{i,j}, i, j \in \mathbb{Z}$, with $O(b_{i,j}) = b_{i+1,j}$ (Arnold and Avez 1968, pp. 28–30; Cornfeld et al. 1982, Appendix 2). Consequently, U acting on $H_0^\perp$ has an infinitely multiple homogeneous Lebesgue measure spectrum—this is point II) referred to by von Neumann.

Let us show that there are no other proper functions of absolute value 1 besides $\delta g_{k,0}, |\delta| = 1$. Suppose that $g(u, v)$ is a proper function with proper value λ. Because $g_{k,m}$ is an orthonormal basis, we can expand g as

$$g(u, v) = \sum_{k,m} \chi_{k,m} g_{k,m}(u, v), \quad \chi_{k,m} \in \mathbb{C}.$$

Then

$$Ug = \sum_{k,m} e^{2\pi i (k-m)\gamma} \chi_{k-m,m} g_{k,m} = \lambda \sum_{k,m} \chi_{k,m} g_{k,m}. \qquad (9)$$

Thus $|\chi_{k-m,m}| = |\lambda \chi_{k,m}|$ for all k, m. It is well known (e.g. Halmos Halmos 1951, 20) that if κ, κ_j are elements in an Hilbert space with κ_j pairwise orthogonal, then $\sum_j \kappa_j = \kappa$ if, and only if, $\sum_j \|\kappa_j\|^2$ converges. Hence $\sum_{k,m} |\chi_{k,m}|^2$ converges, implying that $\chi_{k,m} = 0$ whenever $m \neq 0$. Therefore, $1''$ – the set of proper functions – consists of $\delta e^{2\pi i k u}, |\delta| = 1, k \in \mathbb{Z}$. Because for $g = e^{2\pi i v}$ we have

$$g(S(u, v)) = e^{2\pi i (u+v)} = e^{2\pi i u} g;$$

$1'''$ contains $e^{2\pi i v}$. Consequently, $1'' \neq 1'''$.

Consider (see paragraph 8) a dynamical system $(\Omega, \Sigma_\Omega, \rho, R)$ with Koopman operator X where

  (i) $\delta 1$, $|\delta| = 1$, $\delta \in \mathbb{C}$, are the only proper functions with absolute value 1;
  (ii) X acting on $H_1^\perp$, where H_1 is the subspace generated by the function 1, has infinitely multiple homogeneous Lebesgue measure spectrum.

For instance, let Ω be the set of all bi-infinite sequences $w = (\ldots w_{-1}, w_0, w_1 \ldots)$ with w_i either -1 or 1. Let Σ_Ω be the σ-algebra generated by the semi-algebra of sets

$$C_{i_1 \ldots i_n}^{e_1 \ldots e_n} = \{w \in \Omega \mid w_{i_1} = e_1, \ldots, w_{i_n} = e_n, \ i_j \in \mathbb{Z}, \ i_1 < \ldots < i_n, \ e_j \in \{-1, 1\}, 1 \le j \le n\}.$$
(10)

Let ρ be the unique extension to a measure on Ω of the pre-measure defined by

$$\bar{\rho}\left(C_{i_1 \ldots i_n}^{e_1 \ldots e_n}\right) = p_{e_1} \ldots p_{e_n},$$

where p_{-1} and p_1 are real numbers with $p_{-1} + p_1 = 1, 0 < p_{-1}, p_1 < 1$. Finally, let R be defined by

$$R(\ldots w_{-1}, w_0, w_1 \ldots) = (\ldots w_0, w_1, w_2 \ldots).$$

The dynamical system $(\Omega, \Sigma_\Omega, \rho, R)$ is called a *Bernoulli shift* on the symbols -1 and 1 (cf. Werndl 2009a). It is a standard result that it satisfies (i) and (ii) (Arnold and Avez 1968, 29–32).

Consider (see paragraphs 8–9) the dynamical system $(\psi, \Sigma_\psi, \nu, T)$ where $\psi = [0, 1) \times \Omega$, Σ_ψ denotes the product σ-algebra, ν is the product measure and

$$T(u, w) = (u + \gamma \ (\text{mod } 1), \ R(w)).$$
(11)

Let V be the Koopman operator of $(\psi, \Sigma_\psi, \nu, T)$. Let $d_{k,m}(w), k, m \in \mathbb{Z}$, be the orthonormal set of functions with $X(d_{k,m}) = d_{k+1,m}$ and which, together with 1, form an orthonormal basis of $L_2(\Omega)$ (because of (ii) above, such a set exists). Define

$$\bar{h}_k(u, w) = e^{2\pi i k u} \quad \text{and} \quad p_{l,k,m}(u, w) = e^{2\pi i l u} d_{k,m}, \quad k, l, m \in \mathbb{Z}.$$

$\bar{h}_k$ and $p_{l,k,m}$ form an orthonormal basis of $L_2(\psi)$ because $d_{k,m}$ and 1 form an orthonormal basis of $L_2(\Omega)$ and $e^{2\pi i k u}$ form an orthonormal basis of $L_2([0, 1))$, and for the spaces of square integrable functions $L_2(Y)$ and $L_2(Z)$ with orthonormal bases $y_i(u)$ and $z_j(w)$, $y_i(u)z_j(w)$ form an orthonormal basis of $L_2(Y \times Z)$ (Alicki and Fannes 2001, 47). Clearly, $V(\bar{h}_k) = e^{2\pi i k \gamma} \bar{h}_k$. Thus $\bar{h}_k$ are proper functions with proper values $e^{2\pi i k \gamma}$, and V restricted to the closed subspace F_0 generated by the $\bar{h}_k$ has pure point spectrum (this is point I) referred to by von Neumann). One easily verifies that $V(p_{l,k,m}) = e^{2\pi i l \gamma} p_{l,k+1,m}$. We get rid of the constant factors by setting $t_{l,k,m} = e^{2\pi i l k \gamma} p_{l,k,m}$. Clearly, $V(t_{l,k,m}) = t_{l,k+1,m}$. The set of pairs (k, m) can be put into one-to-one correspondence with $\mathbb{Z}$. If $\bar{h}_{i,j}$ is used to denote $t_{l,k,m}$ whenever (l, m)

🐾 Springer

corresponds to j, $V(\bar{h}_{i,j}) = \bar{h}_{i+1,j}$. Consequently, V acting on $F_0^\perp$ has an infinitely multiple homogeneous Lebesgue measure spectrum (this is point II) referred to by von Neumann).

Recall that h_k and $h_{i,j}$ with $U(h_{i,j}) = h_{i+1,j}$ form an orthonormal basis of $L_2(\phi)$ and that $\bar{h}_k$ and $\bar{h}_{i,j}$ with $V(\bar{h}_{i,j}) = \bar{h}_{i+1,j}$ form an orthonormal basis of $L_2(\psi)$. Clearly, the operator defined by

$$W\left(\sum_k \alpha_k h_k + \sum_{i,j} \beta_{i,j} h_{i,j}\right) := \sum_k \alpha_k \bar{h}_k + \sum_{i,j} \beta_{i,j} \bar{h}_{i,j}, \quad \alpha_i, \beta_{i,j} \in \mathbb{C}, \quad (12)$$

is unitary. Hence $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are spectrally isomorphic.

Now (see paragraph 9) we already know that for $(\phi, \Sigma_\phi, \mu, S)$ it holds that $1'' \neq 1'''$. It remains to show that for $(\psi, \Sigma_\psi, \nu, T)$ we have $1'' = 1'''$, implying that $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are *not* spacially isomorphic. For $(\psi, \Sigma_\psi, \nu, T)$ the set $1''$ consists of all proper functions of absolute value 1. Because spectrally isomorphic systems have the same proper values, the proper functions are $\delta e^{2\pi i k u}$, $|\delta| = 1$. If $g \in 1'''$, then $g(u + \gamma, R(w)) = \delta e^{2\pi i k u} g(u, w)$ for some $k \in \mathbb{Z}$. Let us expand $g(u, w)$ in terms of the $d_{k,m}$:

$$g(u, w) = n_0(u) + \sum_{k,m} n_{k,m}(u) d_{k,m}(w).$$

Then

$$g(u + \gamma, Rw) = n_0(u + \gamma) + \sum_{k,m} n_{k-1,m}(u + \gamma) d_{k,m}(w) \quad \text{and}$$

$$\delta e^{2\pi i k u} g(u, w) = \delta e^{2\pi i k u} n_0(u) + \sum_{k,m} \delta e^{2\pi i k u} n_{k,m}(u) d_{k,m}(w). \quad (13)$$

Consequently,

$$n_{k-1,m}(u + \gamma) = \delta e^{2\pi i k u} n_{k,m}(u). \quad (14)$$

Taking the norms of both sides of (14) yields $\left\|n_{k-1,m}\right\| = \left\|n_{k,m}\right\|$. Recall that if κ, κ_j are elements in an Hilbert space with κ_j pairwise orthogonal, then $\sum_j \kappa_j = \kappa$ if, and only if, $\sum_j \left\|\kappa_j\right\|^2$ converges. Hence $\sum_{k,m} \left\|n_{k,m}\right\|^2$ converges and because $\left\|n_{k-1,m}\right\| = \left\|n_{k,m}\right\|$, $n_{k,m} = 0$ for all $k, m \in \mathbb{Z}$ and $g(u, w) = n_0(u)$. Equation 13 imply that

$$n_0(u + \gamma) = \delta e^{2\pi i k u} n_0(u).$$

It is well known that $\{\vartheta_l(u) := e^{2\pi i l u} : l \in \mathbb{Z}\}$, form an orthonormal basis of $L_2((0, 1])$. Hence $n_0(u) = \sum_l o_l \vartheta_l(u), o_l \in \mathbb{C}$. Then

$$n_0(u + \gamma) = \sum_l o_l e^{2\pi i l \gamma} \vartheta_l(u) \quad \text{and} \quad \delta e^{2\pi i k u} n_0(u) = \sum_l \delta o_{l-k} \vartheta_l(u). \quad (15)$$

Consequently, $|o_{l-k}| = |o_l|$ for all l. This implies that $k = 0$ because $\sum_l |o_l|^2$ converges and the o_l cannot be all 0. But for $k = 0$, g is a proper function of V. Thus $1'' = 1'''$.

5 Von Neumann's 1941 letter: literature referring to the result

To summarise, *the proof in von Neumann's 1941 letter shows that for dynamical systems with mixed spectrum spectral isomorphism does not imply spacial isomorphism.* The letter is the first document with a proof that spectral isomorphism does not imply spacial isomorphism. One would have expected that von Neumann went on to publish this result. However, this never happened, and we do not know why. In order to place the letter in historical context and to later assess in Sect. 7 the importance of it, we need to discuss the extant mathematical literature that refers to the result of the letter.

Two of Halmos' publications are relevant here. First, Halmos (1949), in a survey of recent advances in ergodic theory, remarks that for mixed spectrum systems it can be shown that spectral isomorphism does not imply spacial isomorphism (the remark is three paragraphs long). To quote: "This construction has not been published so far—it is the result of joint work by von Neumann and myself" (Halmos 1949, 1025). No proof is presented but a few details are given, which make clear that Halmos refers to the result of the letter. Namely, he states that the dynamical systems which are spectrally but not spacially isomorphic are the systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ mentioned in the letter (see Sect. 4). Moreover, he states that the proof is based on the newly introduced notions of generalised proper values and generalised proper functions. Halmos' (1949) remark is important for three reasons: it gives a few details of the proof, it shows that the result in the letter was joint work by von Neumann and Halmos and that they intended to publish it, and it suggests that generalised proper values and generalised proper functions were first introduced by von Neumann and Halmos.

Second, the only publication we have found where a similar proof of the result of the letter is given is relatively late in 1956 in Halmos' 'Lectures on Ergodic Theory' (in the Chapter 'Generalized Proper Values'). There are no remarks to the effect that a similar construction was used by von Neumann and Halmos to first show that spectral isomorphism does not imply spacial isomorphism. (Halmos might not have felt the need to include such commentary in lecture notes.) As already stated, Halmos' proof is similar to the one in the letter, but there are also differences, most importantly the following: first, the requirements on $R(w)$ differ. Von Neumann requires that $(\Omega, \Sigma_\Omega, \rho, R)$ has (i) pure continuous spectrum and that (ii) the Koopman operator acting on $H_1^\perp$, where H_1 is the subspace generated by the function 1, has infinitely multiple homogenous Lebesgue measure spectrum. The example he gives is the two-shift on the symbols 1 and -1. Halmos (1956) requires $R(w)$ to be more specific and different from the two-shift, namely to be a mixing system on the unit circle such that (i) and (ii) hold.[9] Second, von Neumann calculates that $1''' \neq 1''$ for $S(u, v)$ and that $1''' = 1''$ for $T(u, w)$. Halmos instead focuses on the least positive integer, not ruling

---

[9] Clearly, this difference is inconsequential because all that is needed for the proof are (i) and (ii).

out infinity, such that 1 with n dashes equals 1 with $n + 1$ dashes, and he calculates that this number is 3 for $S(u, v)$ and 2 for $T(u, w)$. Third, Halmos explicitly constructs the spectral isomorphism. Von Neumann only shows that the spectra of $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, v, T)$ are the same, and he seems to have inferred from this that they are spectrally isomorphic, which is easy to show in this case. (Yet in general it is not true that systems with the same spectrum are spectrally isomorphic—see Halmos 1951, 75.) Finally, unsurprisingly, Halmos's proof is relatively detailed whereas von Neumann's proof in the letter is sketchy.

Another publication referring to the result of the letter is Anzai (1951). Among other things, this article proves the same result as the letter, viz. that spectrally isomorphic dynamical systems with mixed spectrum need not be spacially isomorphic. However, both the non-isomorphic dynamical systems and the methods used to prove the result are entirely different from the ones in letter. Interestingly, in the introduction one finds the following acknowledgements:

> The author is much indebted to Professor Kakutani for his kind discussions [...]. Further he taught the author that Professor J. von Neumann had proved the following theorem: The ergodic transformation $T(x, y) \rightarrow (x + \gamma, x + y)$ on the torus is spectrally isomorphic to the direct product transformation of the translation $x \rightarrow x + \gamma$ on the circle and the shift-transformation on the infinite dimensional torus 2, though these transformations are not spacially isomorphic to each other. This fact has been the stimulation in obtaining the results of Sect. 6. (Anzai 1951, 84)

Clearly, Anzai is here referring to the result of the letter. His quote makes clear that von Neumann had told Kakutani about that result. Von Neumann's letter contains a paragraph about Kakutani (see the paragraph beginning with 'Re Kakutani'), telling us that Kakutani has decided to stay in the US and that if the war continues, they will try to help him as much as they can.

Abramov (1962), a article in Russian published in English two years later (Abramov 1964), contains a proof of the result that totally ergodic[10] dynamical systems are spacially isomorphic if, and only if, their generalised proper values and generalised proper functions are equivalent (see Sect. 6 for more on this). In the introduction he refers to Halmos (1949) and remarks that generalised proper values and generalised proper functions "were introduced by von Neumann and Halmos [...], who proved with these concepts the existence of spectrally equivalent but metrically nonisomorphic automorphisms with mixed spectrum" (Abramov 1964, 37). Clearly, he refers here to the result in the letter but no further details about the proof are given and there is no indication that he knew more than what was announced in Halmos (1949).

Finally, Rohlin (1960) briefly mentions generalised proper values and generalised proper functions and states that in this way one can show that spectrally isomorphic systems need not be spacially isomorphic. However, no further details are given and there is no indication who proved the result. Because generalised proper values and

---

[10] A dynamical system $(\phi, \Sigma_\phi, \mu, S)$ is totally ergodic if $(\phi, \Sigma_\phi, \mu, S^t)$ is ergodic for all $t \in \mathbb{Z} \setminus \{0\}$.

functions are mentioned, it seems likely that Rohlin refers to the result of the letter, but it remains unclear what he knew about it.

Let us now report the main results on the isomorphism problem after the publication of the letter.

6 The main results after Von Neumann's 1941 letter

From 1942 until today a large number of various articles on the spacial isomorphism problem have been published, and much research is still being done on this problem nowadays. Thus we can only provide a summary of the most important results on the isomorphism problem after the publication of the letter. Our discussion will be chronological, and we will particularly focus on those results that are relevant to assess the importance of the letter later in Sect. 7.

In 1942, shortly after the letter had been written, von Neumann and Halmos published a article with a further contribution to the isomorphism problem. Their main result is about the same class of systems as Theorem 1, viz. ergodic systems with pure point spectrum.

Theorem 2 *Every ergodic dynamical system* $(\phi, \Sigma_\phi, \mu, S)$ *with pure point spectrum is spacially isomorphic to a rotation on a compact Abelian group (von Neumann and Halmos 1942).*[11]

The importance of this theorem is that it provides a normal form for the class of ergodic systems with pure point spectrum, and thus it can be used to answer many questions about this class. An interesting corollary should also be mentioned:

Corollary 1 *Every ergodic dynamical system* $(\phi, \Sigma_\phi, \mu, S)$ *with pure point spectrum is spacially isomorphic to its inverse, i.e. to* $(\phi, \Sigma_\phi, \mu, S^{-1})$ *(von Neumann and Halmos 1942).*

The next important contribution to the isomorphism problem was Anzai (1951). As already mentioned in Sect. 5, Anzai (1951) proves the same result as von Neumann's letter, viz. that there are spectrally isomorphic dynamical systems with mixed spectrum which are not spacially isomorphic. However, the non-isomorphic dynamical systems as well as the construction of the proof are very different from the ones in the letter.

Up to now none of the results were on systems with pure continuous spectrum. Systems with pure continuous spectrum are generic among all dynamical systems (comeagre in the strong neighbourhood topology) (Halmos 1944). Therefore, the outstanding open problem concerning the isomorphism problem was to classify systems with pure continuous spectrum, and, in particular, to find out whether for these systems

---

[11] For an accessible proof, see Halmos (Halmos 1957, 46–50). The more historically oriented articles Halmos (1949), Halmos (1957), Mackey (1974, 197–198) and Weiss (1972, 672–673) also discuss this result.

Springer

spectral isomorphism implies spacial isomorphism. Kolmogorov eventually made progress on this question. According to Sinai (1989), *in 1957 in a seminar in Russia Kolmogorov first presented examples of systems with pure continuous spectrum which are spectrally isomorphic but not spacially isomorphic.* This construction was then later published (Kolmogorov 1958, 1986; Sinai 1959). Let us outline Kolmogorov's proof.[12]

As a first ingredient, motivated by his work on information theory, Kolmogorov introduced a new spacial invariant—nowadays called the Kolmogorov-Sinai entropy. A *partition* $\alpha = \{\alpha_i \mid i = 1, \ldots, n\}$ of (ϕ, Σ_ϕ, μ) is a collection of non-empty, non-intersecting measurable sets that cover ϕ: $\alpha_i \cap \alpha_j = \emptyset$ for all $i \neq j$ and $\phi = \bigcup_{i=1}^n \alpha_i$. Dynamical systems and information theory can be connected as follows: each $x \in \phi$ produces, relative to a partition α (a coding), an infinite string of symbols $\ldots x_{-2}x_{-1}x_0x_1x_2 \ldots$ in an alphabet of k letters via the coding $x_j = \alpha_i$ if, and only if, $S^j(x) \in \alpha_i$, $j \in \mathbb{Z}$. Interpreting the system $(\phi, \Sigma_\phi, \mu, S)$ as the source, the output of the source are these strings $\ldots x_{-2}x_{-1}x_0x_1x_2 \ldots$. If the measure is interpreted as time-independent probability, then $H(\alpha, T) :=$

$$\lim_{n \to \infty} 1/n \sum_{i_j \in \{1, \ldots, k\}, 0 \leq j \leq n-1} -\mu(\alpha_{i_0} \cap S\alpha_{i_1} \ldots \cap S^{n-1}\alpha_{i_{n-1}}) \log(\mu(\alpha_{i_0} \ldots \cap S^{n-1}\alpha_{i_{n-1}})) \quad (16)$$

measures the average information which the system produces per step relative to α as time goes to infinity (Frigg and Werndl 2011; Petersen 1983, pp. 233–240; Werndl 2009b). Now

$$h(T) := \sup_\alpha \{H(\alpha, S)\} \quad (17)$$

is the *Kolmogorov-Sinai entropy* of $(\phi, \Sigma_\phi, \mu, S)$. It measures the highest average amount of information that the system can produce per step relative to a coding. It is easy to see that spacially isomorphic systems have the same Kolmogorov-Sinai entropy.

The second ingredient in the proof were Bernoulli shifts. In the modern framework of probability theory an independent process, i.e. a doubly-infinite sequence of independent rolls of an n-sided die where the probability of obtaining k is $p_k, k \in \bar{N} := \{N_1, \ldots, N_n\}$, with $\sum_{k=1}^n p_k = 1$, is modeled as follows. Let Ω be the set of all bi-infinite sequences $w = (\ldots w_{-1}, w_0, w_1 \ldots)$ with $w_i \in \bar{N}$, corresponding to the possible outcomes of an infinite sequence of independent trials. Let Σ_Ω be the set of

---

[12] This proof is standardly given to outline Kolmogorov's contribution (e.g. Halmos 1961; Rohlin 1967; Weiss 1972), but it is based on a definition of entropy introduced by Sinai (1959). According to Sinai (1989), this proof corresponds closely to the one Kolmogorov gave in this seminar. The proof Kolmogorov published—Kolmogorov (1958)—was different from the one in the seminar. Most importantly, its definition of entropy was based on a theorem that turned out to be wrong (Kolmogorov 1986), and so later Sinai (1959) definition was adopted.

🖄 Springer

all sets of infinite sequences to which probabilities can be assigned, and let μ be the probability measure on Σ_Ω.[13] Define the shift:

$$R : \Omega \to \Omega \quad R((\ldots w_{-1}, w_0, w_1 \ldots)) = (\ldots w_0, w_1, w_2 \ldots). \qquad (18)$$

$(\Omega, \Sigma_\Omega, \rho, R)$ is called a *Bernoulli shift with probabilities* $p_1, \ldots, p_n$.

A dynamical system $(\phi, \Sigma_\phi, \mu, S)$ is a *Lebesgue system* if there exists an orthonormal basis of $L_2(\phi)$ formed by the function $f(x) = 1$ for all $x \in \phi$, and the functions $f_{i,j}, i, j \in \mathbb{Z}$, such that $U(f_{i,j}) = f_{i,j+1}$ for all i, j (Arnold and Avez 1968, pp. 28–30).[14] Clearly, Lebesgue systems have pure continuous spectrum and it is easy to see that they are spectrally isomorphic.[15] A simple calculation shows that all Bernoulli shifts are Lebesgue systems, and hence Bernoulli shifts are spectrally isomorphic (Arnold and Avez 1968, pp. 30–31). However, it can be calculated that the Kolmogorov-Sinai entropy of the Bernoulli shift with probabilities $p_1, \ldots, p_n$ is $\sum_{i=1}^{n} p_i \log(p_i)$ and thus takes a continuum of different values. Consequently:[16]

Theorem 3 *There is a continuum of dynamical systems which are spectrally isomorphic but not spacially isomorphic.*

This result by Kolmogorov was hailed as the major breakthrough since von Neumann's earlier work and stimulated much further reserach (cf. Halmos 1961, pp. 75–77; Rohlin 1967, p. 45, Sinai 1989, pp. 834–835; Weiss 1972, pp. 674–676).

Another contribution which should be mentioned is Abramov's (1962) article in Russian which was published in English two years later (Abramov 1964). This article is important because it uses generalised proper values and generalised proper functions as introduced in von Neumann's letter (cf. Sect. 4) to provide sufficient conditions for spacial isomorphism. Let us explain. For a dynamical system $(\phi, \Sigma_\phi, \mu, S)$ denote the sequence of generalised proper functions $1'', 1''', 1'''' \ldots$ by $G_\phi^1, G_\phi^2, G_\phi^3 \ldots$ and the sequence of generalised proper values $1', 1'', 1''' \ldots$ by $H_\phi^1, H_\phi^2, H_\phi^3 \ldots$. It is easy to see that the G_ϕ^n and H_ϕ^n, $n \geq 1$, are groups under multiplication (Halmos 1956, 57). Generalising the notion of a pure point spectrum, Abramov (1962) introduced the definition that a dynamical system $(\phi, \Sigma_\phi, \mu, S)$ has *quasi-discrete spectrum* if $G_\phi = \cup_{n \geq 1} G_\phi^n$ forms a basis of $L_2(\phi)$.

---

[13] In detail: Σ_Ω is the σ-algebra generated by the cylinder-sets

$$C_{i_1 \ldots i_m}^{k_1 \ldots k_m} = \{w \in \Omega \mid w_{i_1} = k_1, \ldots, w_{i_m} = k_m, \ i_j \in \mathbb{Z}, \ i_1 < \ldots < i_m, \ k_j \in \bar{N}, \ 1 \leq j \leq m\}.$$

The sets have probability $\bar{\rho}(C_{i_1 \ldots i_m}^{k_1 \ldots k_m}) = p_{k_1} p_{k_2} \cdots p_{k_m}$ since the outcomes are independent. ρ is defined as the unique extension of $\bar{\rho}$ to a measure on Σ_Ω.

[14] $(\phi, \Sigma_\phi, \mu, S)$ is a Lebesgue system if, and only if, $\delta 1, |\delta| = 1$, are the only proper functions with absolute value 1, and (ii) S acting on $H_1^\perp$, where $H_1^\perp$ is the orthogonal complement of the subspace H_1 generated by function 1, has infinitely multiple homogeneous Lebesgue measure spectrum (cf. the discussion of paragraph 7 of von Neumann's letter).

[15] See Equation (12) for the construction of the spectral isomorphism.

[16] This result is also discussed by the more historically oriented articles and chapters Halmos (1961, pp. 75–77), LoBello (1983), Mackey (1974, p. 200), Reed and Simon(1980, p. 239), Rohlin (1967, p. 45) Sinai (1963, p. 68),Sinai (1989, pp. 834–835) and Weiss (1972, pp. 674–676).

Springer

As a next step, Abramov formalised the idea that dynamical systems have equivalent generalised proper values and generalised proper functions. Before we state the definition, note that the function $Q_\phi : H_\phi^{n+1} \to H_\phi^n$, $Q_\phi(f) = f(\phi(x))/f(x)$ is a group homomorphism. Now the generalised proper values and generalised proper functions of the dynamical systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ are *equivalent* if there exists a group isomorphism L of the group $H_\phi = \cup_{n \geq 1} H_\phi^n$ to the group $H_\psi = \cup_{n \geq 1} H_\psi^n$ such that

$$L(f) = f \quad \text{for all} \quad f \in H_\phi^1, \tag{19}$$

$$L(H_\phi^n) = H_\psi^n, \quad n \in \mathbb{N}, \tag{20}$$

$$Q_\phi = L^{-1} Q_\psi L. \tag{21}$$

The main result of Abramov (1962) can now be stated.

Theorem 4 *Totally ergodic dynamical systems $(\phi, \Sigma_\phi, \mu, S)$ and $(\psi, \Sigma_\psi, \nu, T)$ with quasi-discrete spectrum are spacially isomorphic if, and only if, the generalised proper values and generalised proper functions of the dynamical systems are equivalent.*

Recall von Neumann (1932c) result that ergodic dynamical systems are spacially isomorphic if, and only if, their proper values coincide (cf. Theorem 1). Theorem 4 is analogous to this result in the sense that systems with equivalent generalised proper values and generalised proper functions are spacially isomorphic. One can then also prove an analogue of Theorem 2, namely that every totally ergodic dynamical system of quasi-discrete spectrum is spacially isomorphic to an affine transformation on a compact connected Abelian group (Parry 1971). However, Theorem 4 is also very different from Theorem 1: spectrally isomorphic systems do always have equivalent generalised proper values and generalised proper functions; hence for systems with quasi-discrete spectrum spectral isomorphism does not imply spacial isomorphism.[17] The notions of generalised proper values and generalised proper functions introduced by von Neumann in the letter proved to be very fruitful, and Abramov (1962) work highlights this.

Another important contribution is Choksi (1965). Recall that von Neumann (1932) had proven that for *ergodic* systems with pure point spectrum spacial and spectral isomorphism are equivalent (Theorem 1). It was often believed, or at least hoped (e.g. von Neumann 1932c, p. 495), that this theorem could be extended to *nonergodic* dynamical systems by invoking von Neumann's ergodic decomposition theorem, saying that any dynamical system can be decomposed in ergodic parts. Choksi (1965) dashed this hope when he showed that for non-ergodic systems with pure point spectrum spectral isomorphism does not imply spacial isomorphism.

Finally, Ornstein's work needs to be mentioned. Since Kolmogorov's groundbreaking contribution to the spacial isomorphism problem in the late 1950s, it had been

---

[17] Totally ergodic systems with quasi-discrete spectrum have either pure point spectrum or a mixed spectrum consisting of the proper values and a infinitely multiple homogenous Lebesgue measure spectrum (cf. Sect. 4). All mixed-spectrum systems with the same proper values are spectrally isomorphic, but they need not be spacially isomorphic (Abramov 1962).

⚫ Springer

an open question whether, for certain systems, having the same Kolmogorov-Sinai entropy would be a sufficient condition for spacial isomorphism. In particular, the question arose whether Bernoulli shifts with the same Kolmogorov-Sinai entropy are spacially isomorphic. This question was answered in the positive by Ornstein's landmark work in the 1970s. In particular, Ornstein proved the following celebrated result (Ornstein 1970; see also Ornstein 1974).[18]

Theorem 5 *If a Bernoulli shift with probabilities $p_1, \ldots, p_n$ and a Bernoulli shift with probabilities $q_1, \ldots, q_m$ have the same Kolmogorov-Sinai entropy, i.e. $\sum_{i=1}^{n} p_i \log(p_i) = \sum_{j=1}^{m} q_j \log(q_j)$, then they are spacially isomorphic.*

Combined with Kolmogorov's result, this means that Bernoulli shifts are spacially isomorphic if, and only if, they have the same Kolmogorov-Sinai entropy. Ornstein developed many new techniques and so his work simulated much further research. Indeed, as already mentioned above, the isomorphism problem continues to be an active and lively research area. We have now presented the major developments on the isomorphism problem since von Neumann's 1941 letter. Now we are in a position to return to von Neumann's letter and to comment further on its historical importance.

7 Concluding remarks

We will conclude the article by highlighting what we regard as the three major reasons why the letter is historically important. *First, von Neumann's 1941 letter to Ulam is the earliest document containing a proof of the result that spectral isomorphism does not imply spacial isomorphism.* Clearly, the letter also contains the earliest proof showing that for systems with *mixed spectrum* spacial and spectral isomorphism do not always go together. Without the letter, the earliest document establishing that spectral isomorphism does not imply spacial isomorphism would be Anzai (1951). This led some like LoBello (1983) to claim, erroneously as we know now from von Neumann's letter, that it was Anzai (1951) who first established this result. From our discussion it is clear that instead Halmos and von Neumann should be credited with this. Furthermore, the letter is also the earliest document in which the notions of generalised proper values and generalised proper functions are introduced. As we have seen in Sect. 6 (see, in particular, the work by Abramov 1962), these notions proved fruitful: they were later employed in many articles to make progress on the isomorphism problem.

   Second, as discussed above, Halmos (1949) remarks that he and von Neumann have first shown that spectral isomorphism does not imply spacial isomorphism, and Anzai (1951) and Abramov (1962) briefly remark that von Neumann and Halmos proved that for mixed spectrum systems spectral isomorphism does not imply spacial ismorphism. However, without the letter, apart from Halmos (1949) comment that the proof relies on generalised proper values and proper functions, it remained unclear how the proof proceeded. To be sure, there exists one document containing a proof of the result of the letter based on generalised proper functions and proper values, viz. Halmos (1956),

---

[18] This results is also discussed by the more historically oriented articles Lo Bello (1983), Sinai (1989) and Weiss (1972).

and one could have guessed that this was von Neumann and Halmos' original proof. However, there is no commentary in Halmos (1956) explaining how his proof relates to the one that he and von Neumann had intended to publish. *Consequently, some uncertainty remained about the construction of the original proof. Von Neumann's letter removes this uncertainty and presents us with the original proof.*

Related to this, most of the extant articles on the history of ergodic theory do not address the question of *when* it was first shown and *who* first showed that spectral isomorphism does not imply spacial isomorphism (e.g. Mackey 1974, 1990; Reed and Simon 1980, Sect. VII.4; Sinai 1989; Weiss 1972). The presentation of the history of the isomorphism problem in some of these publications potentially leave one with the impression that it was Kolmogorov who first proved this result when he showed that systems with Lebesgue spectrum need not be spacially isomorphic (Kolmogorov's contribution was discussed in Sect. 6) (Reed and Simon 1980, Sect. VII.4; Weiss 1972). Indeed, as mentioned above, the only article we have found in which a correct answer to the "when and who?" question is given is the mathematics article Halmos (1949). It seems likely that a major reason why this question has often not been addressed is the following: Halmos' (1949), Anzai (1951) and Abramov (1962) remarks are very brief and left unclear how exactly the original proof proceeded, and hence this episode of the history of the ismorphism remained obscure. We hope that the publication of the letter and the story around it will make this important episode more widely known and more widely discussed.

Third, the letter highlights von Neumann's contribution to the isomorphism problem. Up to the present day, this contribution has usually been taken to be (i) the introduction of the notion of isomorphism and the formulation of the isomorphism problem (cf. von Neumann 1932c); (ii) the results in von Neumann (1932c), in particular, the proof that for ergodic dynamical systems with pure point spectrum spacial and spectral isomorphism are equivalent (cf. Theorem 1); and (iii) the results in von Neumann and Halmos (1942), in particular, the proof that any ergodic pure point spectrum system is spacially isomorphic to a rotation on a compact separable Abelian group (cf. Theorem 2) (cf. Halmos 1957; Sinai 1989; Weiss 1972). *The letter shows that this is not all: von Neumann, together with Halmos, was also the first to prove that spectrally isomorphic dynamical systems need not be spacially isomorphic.* Furthermore, the letter also highlights that von Neumann and Halmos were the first to introduce the important notions of generalised proper values and generalised proper functions. As discussed in Sect. 6 (see, in particular, the work by Abramov 1962), these notions proved very fruitful in later work on the isomorphism problem.

References

Abramov, L. 1962. Metric automorphism with quasi-discrete spectrum. *Izvestiya Rossiiskoi Akademii Nauk SSSR Seriya Matematicheskaya* 26: 513–530.

Ⓐ Springer

Abramov, L. 1964. Metric automorphisms with a quasi-discrete spectrum. *American Mathematical Society Translations (2)* 39: 37–56.

Alicki, R., and Fannes, M. 2001. *Quantum Dynamical Systems*. Oxford: Oxford University Press.

Anzai, H. 1951. Ergodic skew product transformations on the torus. *Osaka Mathematical Journal* 3: 83–99.

Arnold, V. and Avez, A. 1968. *Ergodic Problems of Classical Mechanics*. New York: W. A. Benjamin.

Choksi, J. 1965. Nonergodic transformations with discrete spectrum. *Illinois Journal of Mathematics* 9: 307–320.

Cornfeld, I., Fomin, S. and Sinai, Y. 1982. *Ergodic Theory*. Berlin: Springer.

Frigg, R. and Werndl, C. 2011. Entropy—a guide for the perplexed, *in* C. Beisbart and S. Hartmann, eds, Probabilities in Physics, Oxford: Oxford University Press (forthcoming).

Halmos, P. 1944. In general a measure-preserving transformation is mixing. *The Annals of Mathematics* 45: 786–792.

Halmos, P. 1949. Measurable transformations. *Bulletin of the American Mathematical Society* 55: 1015–1043

Halmos, P. 1951. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. New York: Chelsea Publishing Company.

Halmos, P. 1956. *Lectures on Ergodic Theory*. New York: Chelsea Publishing Company.

Halmos, P. 1957. Von Neumann on measure and ergodic theory. *Bulletin of the American Mathematical Society Translations* 64: 86–94.

Halmos, P. 1961. Recent progress in ergodic theory. *Bulletin of the American Mathematical Society* 67: 70–80.

Kolmogorov, A. 1958. A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces. *Doklady Akademii Nauk SSSR* 119: 861–864.

Kolmogorov, A. 1986. A new metric invariant for transitive dynamical systems and automorphisms of Lebesgue spaces. In *Proceedings of the Steklov Institute of Mathematics* 169: 97–102.

Koopman, B. 1931. Hamiltonian systems and transformations in Hilbert space. In: *Proceedings of the National Academy of Sciences of the United States of America* 17: 315–318.

Koopman, B., and von Neumann, J. 1932. Dynamical systems of continuous spectra. In: *Proceedings of the National Academy of Sciences of the United States of America* 18: 255–263. Reprinted in (Taub, A., ed. 1961a. *John von Neumann: Collected Works*, Vol. II. Operators, Ergodic Theory and Almost Periodic Functions in a Group. New York and Oxford: Pergamon Press) No. 14.

Lo Bello, A. 1983. On the origin and history of ergodic theory. *Bollettino di Storia delle Scienze Matematiche* 3: 73–75.

Mackey, G. 1974. Ergodic theory and its significance for statistical mechanics and probability theory. *Advances in Mathematics* 12: 178–268.

Mackey, G. 1990. Von Neumann and the early days of ergodic theory. In: *The Legacy of John von Neumann*, edited by J. Glimm, J. Impagliazzo and I. Singers. Vol. 50 of *Proceedings of Symposia in Pure Mathematics*, 25–38 (Providence: American Mathematical Society, 1990).

Ornstein, D. 1970. Bernoulli-shifts with the same entropy are isomorphic. *Advances in Mathematics* 4: 337–352.

Ornstein, D. 1974. *Ergodic Theory, Randomness, and Dynamical Systems*. New Haven and London: Yale University Press.

Ornstein, D. 1990. Von Neumann and ergodic theory. In: *The Legacy of John von Neumann*, edited by J. Glimm, J. Impagliazzo and I. Singers. Vol. 50 of *Proceedings of Symposia in Pure Mathematics*, 39–42 (Providence: American Mathematical Society, 1990).

Parry, W. 1971. Metric classification of ergodic nilflows and unipotent affines. *American Journal of Mathematics* 93: 819–828.

Petersen, K. 1983. *Ergodic Theory*, Cambridge: Cambridge University Press.

Rédei, M., ed. 2005. *John von Neumann: Selected Letters*, Vol. 27 of *History of Mathematics*. Rhode Island: American Mathematical Society and London Mathematical Society.

Reed, M. and Simon, B. 1980. *Methods of Modern Mathematical Physics I: Functional Analysis*. San Diego: Academic Press.

Rohlin, V. 1960. New progress in the theory of transformations with invariant measure. *Russian Mathematical Surveys* 15: 1–22.

Rohlin, V. 1967. Lectures on the entropy theory of measure-preserving transformations. *Russian Mathematical Surveys* 22: 1–52.

Sinai, Y. 1959. On the concept of entropy for dynamical systems. *Doklady Akademii Nauk SSSR* 124: 768–771.

Sinai, Y. 1963. Probabilistic ideas in ergodic theory. *American Mathematical Society Translations* 31: 62–84.

Sinai, Y. 1989. Kolmogorov's work on ergodic theory. *The Annals of Probability* 17: 833–839.

Szász, D. 1996. Boltzmann's ergodic hypothesis. A conjecture for centuries? *Studia Scientia Mathematica Hungaria* 31: 299–322.

Taub, A., ed. 1961a. *John von Neumann: Collected Works*, Vol. II. Operators, Ergodic Theory and Almost Periodic Functions in a Group, Pergamon Press, New York and Oxford.

Taub, A., ed. 1961b. *John von Neumann: Collected Works*, Vol. IV. Continuous Geometry and Other Topics, Pergamon Press, New York and Oxford.

von Neumann, J. 1932a. Applications of the ergodic hypothesis. *Proceedings of the National Academy of Sciences* 18: 263–266. Reprinted in (Taub, A., ed. 1961a. *John von Neumann: Collected Works*, Vol. II. Operators, Ergodic Theory and Almost Periodic Functions in a Group. New York and Oxford: Pergamon Press) No. 13.

von Neumann, J. 1932b. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 18: 70–82. Reprinted in (Taub, A., ed. 1961a. *John von Neumann: Collected Works*, Vol. II. Operators, Ergodic Theory and Almost Periodic Functions in a Group. New York and Oxford: Pergamon Press) No. 12.

von Neumann, J. 1932c. Zur Operatorenmethode in der klassischen Mechanik. *The Annals of Mathematics* 33: 587–642. Reprinted in (Taub, A., ed. 1961a. *John von Neumann: Collected Works*, Vol. II. Operators, Ergodic Theory and Almost Periodic Functions in a Group. New York and Oxford: Pergamon Press) No. 17.

von Neumann, J. and Halmos, P. 1942. Operator methods in classical mechanics II. *Annals of Mathematics* 43: 332–350. Reprinted in (Taub, A., ed. 1961b. *John von Neumann: Collected Works*, Vol. IV. Continuous Geometry and Other Topics. New York and Oxford: Pergamon Press) No. 23.

von Plato, J. 1991. Boltzmann's Ergodic Hypothesis. *Archive for History of Exact Sciences* 42: 71–89.

Weiss, B. 1972. The isomorphism problem in ergodic theory. *Bulletin of the American Mathematical Society* 78: 668–684.

Werndl, C. 2009a. Are deterministic descriptions and indeterministic descriptions observationally equivalent? *Studies in History and Philosophy of Modern Physics* 40: 232–242.

Werndl, C. 2009b. Justifying definitions in matemathics—going beyond Lakatos. *Philosophia Mathematica* 17: 313–340.

Zund, J. 2002. Georg David Birkhoff and John von Neumann: A question of priority and the ergodic theorems, 1931–1932. *Historia Mathematica* 29: 138–156.