Geometric interpretation of Fourier series

A. Eremenko

January 31, 2021

1. Let me recall some linear algebra. In linear algebra we study *vector spaces*; a vector space is a collection of objects which can be added and multiplied by numbers so that the usual rules hold. The numbers must be elements of a filed. By default we use complex numbers, so our vector spaces are complex vector spaces. Examples of vector spaces are: $\mathbf{C}^n$, it consists of complex column vectors of size n; this is a complex vector space, and $\mathbf{R}^n$, a similar real vector space.

If vectors $v_1, v_2, \ldots$ span the space, then every element x of this space can be written as a linear combination

$$x = c_1 v_1 + c_2 v_2 + \ldots. \tag{1}$$

To find the coefficients c_n of this linear combination one had to solve a system of linear equations.

A complex vector space can be equipped with an *Hermitian (dot) product.* This is an operation which to every pair of vectors x, y puts into correspondence a number (x, y) so that the following three conditions (i)-(iii) are satisfied.

(i) $(x, y) = \overline{(y, x)}$ (here $\overline{(,)}$ is the complex conjugation!)

(ii) $(c_1 x_1 + c_2 x_2, y) = c_1(x_1, y) + c_2(x_2, y)$

From (i) and (ii) follows that

$$(x, c_2 y_1 + c_2 y_2) = \overline{c_1}(x, y_1) + \overline{c_2}(x, y_2).$$

They express this by saying that the Hermitian product is linear with respect to the first argument, and *anti-linear* with respect to the second argument[1].

(iii) $(x, x) \geq 0$ and $(x, x) = 0$ *only if* $x = 0$.

An example of the dot product on $\mathbf{C}^n$ is

$$(x, y) = x^T \overline{y} = x_1 \overline{y_1} + x_2 \overline{y_2} + \ldots + x_n \overline{y_n},$$

which is called the *standard Hermitian product on* $\mathbf{C}^n$. Please, verify that this example indeed has all properties (i)-(iii).

On a real vector space, the dot product is defined similarly, except in this case the conjugation bars do nothing and can be omitted both in the definition and in the example.

It is important to understand that there can be many different dot products on the same vector space. For example, we can take any sequence of positive numbers $a_1, a_2, \ldots, a_n$ and define a dot product in $\mathbf{C}^n$ by the formula

$$(x, y) = a_1 x_1 \overline{y_1} + a_2 x_2 \overline{y_2} + \ldots + a_n x_n \overline{y_n}.$$

More generally, we can take any *positive definite Hermitian matrix* A and define the dot product by the formula $(x, y) = x^T A \overline{y}$. A matrix is called Hermitian if

$$A^T = \overline{A} \quad \text{that is} \quad a_{j,i} = \overline{a_{i,j}} \quad \text{for all} \quad i, j, \tag{2}$$

which is equivalent to $(Ax, y) = (x, Ay)$ for the standard Hermitian product and all x, y. A real matrix is Hermitian if and only if it is symmetric.

The length of a vector a. k. a. *the norm* is defined by the formula $\|x\| = \sqrt{(x, x)}$ (non-negative square root!)

A system of vectors $v_1, v_2, \ldots$ is called *orthogonal* if

$$(v_k, v_j) = 0 \quad \text{for every} \quad k \neq j.$$

An orthogonal system is called *orthonormal* if in addition all vectors of the system have unit length. In other words, an orthogonal system satisfies

$$(v_k, v_j) = \begin{cases} 0, & k \neq j, \\ 1, & k = j. \end{cases} \tag{3}$$

[1]In some books, especially in physics, they use the opposite convention, linear with respect to the second argument and anti-linear with respect to the first. I use the convention from our textbook.

Every orthogonal system of non-zero vectors is linearly independent. This means that every vector in the span of this system has a *unique* representation in the form (1). Indeed, multiplying both sides of (1) on v_n we obtain

$$(x, v_n) = c_n(v_n, v_n),$$

because all other summands give zero, and since $v_n \neq 0$ we have $(v_n, v_n) \neq 0$, by property (iii) of the dot product, so we can divide and obtain

$$c_n = \frac{(x, v_n)}{(v_n, v_n)}. \tag{4}$$

For an orthonormal system, these formulas simplify to $c_n = (x, v_n)$. So orthogonal systems have a great advantage: to obtain a representation of the form (1) of any vector x in the span of such system, one has a simple formula (4) instead of having to solve a system of linear equations.

Suppose that $v_1, v_2, \ldots$ is an orthogonal system, a vector x is given by (1), and let us compute the norm of x. We have, using (3):

$$\begin{aligned} \|x\|^2 &= (x, x) = \left(\sum_n c_n v_n\right)\left(\sum_n c_n v_n\right) \\ &= \sum_{m,n} c_m \overline{c_n}(v_m, v_n) = \sum_n |c_n|^2 \|v_n\|^2, \end{aligned} \tag{5}$$

again because $(v_m, v_n) = 0$ when $m \neq n$. So the squared norm of x equals to the sum of the squared norms of the summands in (1). This is the Pythagorean theorem!

2. In this course we consider vector spaces of infinite dimension. Some examples of such spaces are:

a) The space of all polynomials on the real line, or on some interval.

b) The space of all continuous functions on an interval, it is denoted by $C[a, b]$. The interval can be finite or infinite.

I recall that all functions that we consider are complex-valued by default. Letter C stands for "continuous".

c) The space $C^k[a, b]$ which consists of all functions having continuous derivatives up to order k on an interval, and the space C^∞ of infinitely differentiable functions. (An interval can be open, closed or semi-open. Including an end-point in the notation means that derivatives must exist and be continuous at this point as well).

d) The space $L^2(a,b)$ of "square-integrable" functions, the precise definition will be discussed later, but roughly speaking, it consists of those functions defined on (a,b) for which the integral

$$\int_a^b |f(x)|^2 dx < \infty$$

is finite. Again, the interval can be finite or infinite.

e) The space of square summable sequences ℓ^2 consists of all complex sequences $(c_n)_{n=-\infty}^{\infty}$ such that

$$\sum_{-\infty}^{\infty} |c_n|^2 < \infty. \tag{6}$$

All these are vector spaces: any linear combination of two functions (or sequences) with the stated property also possesses this property. Now we consider dot products.

The analog of the standard Hermitian product on functions defined on some interval $[a,b]$ is

$$(f,g) = \int_a^b f(x)\overline{g(x)}dx. \tag{7}$$

It is an Hermitian product on all spaces (a)-(c). Verification of properties (i), (ii) of an Hermitian product causes no difficulty. More subtle is property (iii): we have to show that

$$\int_a^b |f(x)|^2 dx = 0 \quad \text{implies} \quad f = 0. \tag{8}$$

The last statement means that $f(x) = 0$ *for all* $x \in [a,b]$. In general, this is not true for functions: take a function which is equal to 0 at all points except one. This function is not the zero-function, but its integral is 0. What saves our examples is that all functions in a)-c) are *continuous*. For continuous functions f, statement (8) is true.

A simple and important property which every Hermitian product possesses is the Cauchy–Schwarz inequality:

$$|(x,y)| \leq \|x\| \, \|y\|;$$

for the dot product defined in (7) this means

$$\left|\int_a^b f(x)\overline{g(x)}dx\right| \le \sqrt{\int_a^b |f(x)|^2dx}\sqrt{\int_a^b |g(x)|^2dx}. \tag{9}$$

The length of the vector associated with the dot product (7) is called the *square norm*:

$$\|f\|_2 = \sqrt{\int_a^b |f(x)|^2dx}.$$

Subscript 2 is used it to distinguish it from other norms discussed below. Some of them are defined using dot products, some are not.

For the space ℓ^2 defined in example e), the Cauchy-Schwarz inequality says

$$\sum_{-\infty}^{\infty} c_n\overline{d_n} \le \sqrt{\sum_{-\infty}^{\infty}|c_n|^2}\sqrt{\sum_{-\infty}^{\infty}|d_n|^2}.$$

Example. Consider the system of functions $\phi_n(x) = e^{inx}$, $-\infty < n < \infty$ defined on the interval $-\pi \le x \le \pi$. This system is orthogonal with respect to the dot product (7) with $[a,b] = (-\pi,\pi)$:

$$(\phi_m,\phi_n) = \int_{-\pi}^{\pi} e^{imx}e^{-inx}dx = \begin{cases} 0, & m \ne n, \\ 2\pi, & m = n. \end{cases} \tag{10}$$

So the system

$$\frac{1}{\sqrt{2\pi}}e^{inx}, \quad -\infty < n < \infty$$

is orthonormal in spaces of functions in $[-\pi,\pi]$ (examples (b), (c), (d)).

Suppose that a function f belongs to the span of (ϕ_n), that is

$$f(x) = \sum_n c_n e^{inx}, \quad -\pi \le x \le \pi.$$

To find c_m we dot-multiply both sides on $\phi_m(x) = e^{imx}$ and use the orthogonality relation (10). We obtain formula (4) for c_n which coincides with Fourier formula:

$$c_n = \frac{(f,\phi_n)}{(\phi_n,\phi_n)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx.$$

So *Fourier expansion is similar to expansion of a vector in an orthogonal system.*

There is one important difference: a linear combination in linear algebra is by definition a *finite* sum, while Fourier expansion is in general *infinite.* The reason is that spaces of functions are infinite dimensional. It is true that *every* vector space has a basis, so that each vector is a finite linear combination of the basis vectors, but such a basis is so enormous that it is useless in most infinite-dimensional spaces.

Example: the space of all polynomials has a basis consisting of all powers of x. But the space of all continuous functions *has no useful basis.*

Instead of a basis we will use a generalization which is called a *complete orthogonal system.* An orthogonal system (ϕ_n) is called complete in a space V with a Hermitian product, if for every vector f in V:

$$(f, \phi_n) = 0 \quad \text{for all} \quad n \quad \text{implies} \quad f = 0. \tag{11}$$

In other words, there is no vector orthogonal to every vector of our system, except the zero vector. In a finite-dimensional space, this is *equivalent* to the definition of an orthogonal basis.

The main property of complete systems is that every vector f can be expanded in an (infinite) series

$$f = \sum_n c_n \phi_n, \tag{12}$$

and this expansion is unique, and the coefficients are given by “Fourier formulas”

$$c_n = (f, \phi_n)/(\phi_n, \phi_n).$$

Convergence of the series means by definition that

$$\left\| x - \sum_{k=0}^{n} c_k \phi_k \right\|_2 \to 0 \quad \text{as} \quad n \to \infty, \tag{13}$$

where the norm $\|.\|_2$ is defined using the Hermitian product.

Convergence of any sequence $f_n \to f$ means that $\|f - f_n\|_2 \to 0$ as $n \to \infty$. For the standard dot product (7) this is called L^2*-convergence.*

Notice an important fact: applying the Pythagorean theorem to the expansion (12) to we obtain

$$\sum_n |c_n|^2 (\phi_n, \phi_n) = \|f\|^2.$$

In particular, for the system $\phi_n(x) = e^{nx}$, this means

$$(2\pi) \sum_{n=-\infty}^{\infty} |c_n|^2 = \|f\|^2 = \int_a^b |f(x)|^2 dx. \tag{14}$$

This is called the *Parseval identity.* It implies that the sequence of Fourier coefficients of a function of the space L^2 (example d)) belongs to the space ℓ^2 defined in example e) above.

In general, one can consider several types of convergence for functions. The most important types for us are:

a) Pointwise convergence: we say $f_n \to f$ pointvise, if $f_n(x) \to f(x)$ for every x from the common domain of definition. This type of convergence is not convenient, mainly because convergence of functions does not imply convergence of integrals.

Example: functions $f_n(x) = (n+1)x^n$ converge to zero on $(0,1)$, as $n \to \infty$, but their integrals

$$\int_0^1 f_n(x)dx = \int_0^1 (n+1)x^n dx = 1$$

do not converge to 0 as $n \to \infty$.

b) Uniform convergence: we say that $f_n \to f$ uniformly on X if

$$\sup_{x \in X} |f_n(x) - f(x)| \to 0, \quad n \to \infty.$$

With this type of convergence one can pass to the limit under the integral sign. But this type of convergence is too restrictive for Fourier analysis. The main reason is that uniform limit of continuous functions is continuous. So there is no hope for uniform convergence of Fourier expansion even for simplest discontinuous functions, like step functions, which are useful both in theory and in the applications.

c) L^2-convergence we discussed above. This is very convenient for Fourier analysis, except for one difficulty which we address now.

3. Space L^2. One has to define correctly the space of functions we deal with. Our first requirement is that integrals in the definitions of the dot product and the of the norm must exist. Second requirement is that implication (11) must hold. On the other hand, we do not want to exclude discontinuous functions. Moreover, we want the L^2 limits of functions of our space to belong to our space.

There are two ways to define the space with required properties. Both are somewhat abstract, and require a generalization of the very notion of "function".

a) Consider all functions on a given interval I (which can be finite or infinite) with the property that

$$\int_I |f(x)|^2 dx < \infty.$$

Then the dot product is well-defined in view of inequality (9). To deal with (11), we define small sets which do not influence integrals. We say that a set $E \subset I$ is *negligible* if *for every* $\delta > 0$ *it can be covered by a sequence of open intervals of lengths* $\delta_j > 0$*, such that*

$$\sum_{j=0}^{\infty} \delta_j < \delta.$$

Intuitive meaning: such sets E have zero length; they can be covered by intervals whose sum of lengths is arbitrarily small. For example, every finite set is negligible, and every countable set is negligible. For example, the set of all rational numbers is negligible. There are also uncountable negligible sets.

Then we *identify* any two functions f and g which differ on a negligible set. We say in this case that $f(x) = g(x)$ almost everywhere. Then it is true that $f \geq 0$ and

$$\int f(x) dx = 0$$

imply that $f(x) = 0$ almost everywhere. Then our space L^2 consists not of functions themselves but of *classes of functions*, where functions belong to the same class if they coincide almost everywhere.

For example, the *Dirichlet's function* which equals 1 at all rational points and 0 at all irrational points is identified with the zero function, and represents the zero vector of the space L^2.

This generalization of the notion of function seems very natural from the point of view of engineers and physicists: we really never measure a value of a function at a point. Every measuring device only measures an average value at nearby points.

So a value of a function at a point really makes sense only for *continuous functions* when these averages tend to the value at a point when we average over smaller and smaller intervals.

This definitions also requires a generalization of the notion of integral. Indeed, the integral in the sense of Riemann (which is taught in undergraduate courses) does not exist for such functions as Dirichlet's function. The required generalization is called *Lebesgue's integral.* Lebesgue's integral "does not feel" any change of the function on a negligible set.

b) The second method is using the procedure of *completion.* To explain what is this, we first consider how real numbers are defined.

Suppose we begin with rational numbers. We want to define real numbers as limits of rational numbers. To each convergent sequence of rational numbers we want to put into correspondence a new object and call it a real number. But how can we say that a sequence is convergent, without mentioning its limit? (Which is not defined yet).

A way to do this was found by Cauchy. Let (a_n) be a sequence. It is called a *Cauchy sequence* if $|a_m - a_n| \to 0$ when $m, n \to \infty$. More precisely: for every $\epsilon > 0$ there exists N such that whenever $m > N$ and $n > N$ we have $|a_m - a_n| < \epsilon$.

Two Cauchy sequences (a_n) and (b_n) are called *equivalent* if $|a_n - b_n| \to 0$.

Now real numbers are defined simply as *classes of equivalent Cauchy sequences.* In other words, each Cauchy sequence is declared to define a real number. And two equivalent Cauchy sequences define the same real number. This is the definition of real numbers, proposed by Cantor. With this definition, one has to explain what the sum and products of real numbers are, and also the limit of sequences of real numbers. All this can be done in terms of Cauchy sequences.

One can make a similar definition of the space L^2 on a given interval I. We start with continuous functions. If the interval I is infinite we also assume that our functions have bounded support, that is each of them is different from 0 only on a bounded interval. Then we define the distance between two functions as

$$\|f - g\|_2 = \sqrt{\int_I |f(x) - g(x)|^2 dx}.$$

Since our functions are continuous and different from zero only on a finite interval, this definition makes sense, even with the ordinary (Riemann's) definition of integral. Once we have a distance, we define Cauchy sequences exactly as above: a sequence (f_n) of functions is Cauchy if $\|f_m - f_n\|_2 \to 0$ as $m, n \to \infty$. And two Cauchy sequences are equivalent if $\|f_n - g_n\|_2 \to 0$.

Then the elements of $L^2(I)$ are defined as *equivalence classes of Cauchy sequences.* Again, addition and multiplication of elements of L^2 is defined using Cauchy sequences, and one verifies that we obtain a vector space. Then the integral of an element of L^2 is defined as the limit of integrals of f_n for some sequence (f_n) representing our element, and one shows that formula (7) defines a genuine dot product.

The crucial property of these completion spaces is *completeness*: every Cauchy sequence converges to an element of the space.

Examples. $f(x) = |x|^\alpha$ belongs to $L^2(-1,1)$ when $\alpha > -1/2$ and does not belong when $\alpha \leq -1/2$. Same about $L^2(0,1)$.

The function given by the same formula belongs to $L^2(1,\infty)$ if and only if $\alpha < -1/2$.

So the function defined by this formula *never* belongs to $L^2(0,+\infty)$ or to $L^2(-\infty,\infty)$.

Function $(\sin x)/x$ belongs to $L^2(-\infty,\infty)$. (And it does not matter how you define it at $x = 0$.)

The space L^2 permits to give a very satisfactory statement of Fourier correspondence.

Theorem of Riesz and Fisher *For every $f \in L^2(-\pi,\pi)$ the sequence*

$$c_n = \int_{-\pi}^{\pi} f(x) e^{-inx} dx, \quad -\infty < n < \infty$$

is defined, and belongs to ℓ^2 that is (6) holds. Conversely, for every sequence $(c_n) \in \ell^2$, the series

$$\sum_{n=-\infty}^{\infty} c_n e^{int}$$

converges in L^2 sense, and these two formulas define a 1*-to-*1 *correspondence between $L^2(-\pi,\pi)$ and ℓ^2. Moreover, this correspondence respects the norms (up to multiple 2π):*

$$\int_{-\pi}^{\pi} |f(x)|^2 dx = 2\pi \sum_{-\infty}^{\infty} |c_n|^2.$$

The last equality is called the Parseval identity. It says that Fourier transform respects distances (up to the factor 2π), and this also implies that it respects the dot product, and the angles between vectors. (The angle between vectors f and g is defined as the unique number in $\alpha \in [0, 1]$ such that

$$\cos \alpha = \frac{(f, g)}{\|f\| \|g\|}.$$

Such α exists in view of Cauchy–Schwarz inequality.

When Fourier coefficients and $\|f\|$ can be explicitly computed, we obtain from Parsevals's identity many interesting explicit sums of various series, like in the exercises on p. 37 of the book.

3. What are the properties of the system $(e^{inx})_{-\infty}^{\infty}$ and of the interval $(-\pi, \pi)$ which imply the Riesz–Fisher theorem? How to generalize it to other intervals and other systems of functions.

Let us go back to Linear Algebra. A basis in a vector space is a system of vectors which is linearly independent and spans the space. Every orthogonal system (of non-zero vectors) is linearly independent. How to check that an orthogonal system spans the space? One simple way is to check that there is no vector, other than zero vector which is orthogonal to each vector of our orthogonal system.

It turns out that this last property generalizes to our infinite expansions.

Definition. *A system of vectors ϕ_n in $L^2(I)$ is called* **complete** *if for every $f \in L^2(I)$ equations $(f, \phi_n) = 0$ for all n imply that $f = 0$.*

In other words, an orthogonal system is complete if only zero vector is orthogonal to all vectors of this system.

A vector space V with a dot product is called a *Hilbert space* if it is complete (=every Cauchy sequence has a limit).

Let ϕ_n be any orthogonal system (complete or not) in a Hilbert space V. Then for every $f \in V$ we can define Fourier coefficients,

$$c_n = \frac{(f, \phi_n)}{(\phi_n, \phi_n)}.$$

We always have *Bessel's inequality*:

$$\sum_n |c_n|^2 (\phi_n, \phi_n) \leq \|f\|^2.$$

The system (ϕ_n) is complete if and only if for every f we have equality in Bessel's inequality, that is (14). And equivalently, if for every f Fourier expansion holds

$$f = \sum_n c_n \phi_n.$$

Examples.

(i) System $(e^{inx})_{-\infty}^{\infty}$ is complete in $L^2(-\pi, \pi)$.

(ii) System $(1, \cos nx, \sin nx)_{n=1}^{\infty}$ is complete in the same space.

(iii) System $(\sin nx)_1^{\infty}$ is complete in $L^2(0, \pi)$.

(iv) System $(\cos nx)_0^{\infty}$ is complete in $L^2(0, \pi)$.

(v) System $(\cos nx)_1^{\infty}$ is not complete in $L^2(0, \pi)$. Indeed, function $f(x) = 1$ is orthogonal to all functions of this system. Its Fourier expansion with respect to this system is just 0, so it does not converge to f.

Statements (iii) and (iv) follow from (ii) by the even or odd extensions which were discussed earlier in this course.

In general if you have an orthogonal system (not containing zero functions) and you remove any function from it, the resulting system cannot be complete, since the removed function is orthogonal to all the rest.

For arbitrary complete system we have

Riesz and Fisher General theorem. *Let (ϕ_n) be an complete orthogonal system in L^2 which contains no zero vector. Suppose that $\|\phi_n\|$ are all equal. Then for every $f \in L^2$ the sequence*

$$c_n = \frac{(f, \phi_n)}{(\phi_n, \phi_n)} \tag{15}$$

is defined, and belongs to ℓ^2 that is (6) holds. Conversely, for every sequence $(c_n) \in \ell^2$, the series

$$\sum_n c_n \phi_n$$

converges in L^2 sense, and these two formulas define a 1*-to-*1 *correspondence between L^2 and ℓ^2. Moreover, this correspondence respects the norms:*

$$\int |f(x)|^2 dx = \sum_n |c_n|^2 \|\phi_n\|^2.$$

Sketch of the proof. For every n define

$$h_n = f - \sum_{k=0}^{n} c_n \phi_n.$$

Multiplying on ϕ_k with $k \leq n$ we obtain

$$(h_n, \phi_k) = (f, \phi_k) - c_n(\phi_k, \phi_k) = 0,$$

where we used orthogonality of ϕ_k and (15). Thus h_n is orthogonal to all ϕ_k for $k \leq n$. Now

$$f = \sum_{k=0}^{n} c_k \phi_k + h_n,$$

and we obtain

$$\begin{aligned} \|f\|^2 &= (f, f) = \left(\sum_k c_k \phi_k + h, \sum_j c_n \phi_j + h \right) \\ & \sum_{k,j} c_k \overline{c_j} (\phi_k, \phi_j) + \sum_k c_k (\phi_k, h) + \sum_j \overline{c_n} (h, \phi_j) + \|h_n\|^2 \end{aligned}$$

By orthogonality, in the first sum only terms with $j = k$ remain, and the second and third sums are zero in view of orthogonality of h_n to ϕ_k with $k \leq n$. So what remains is

$$\|f\|^2 = \sum_{k=1}^{n} |c_k|^2 \|\phi_k\|^2 + \|h_n\|^2,$$

which implies

$$\sum_k |c_k|^2 (\phi_k, \phi_k) \leq \|f\|^2.$$

This is called Bessel's inequality and it implies that the sequence (c_n) belongs to ℓ^2. Now for every sequence (c_n) in ℓ_2, the series $\sum_n c_n \phi_n$ and if we define

$$h = f - \sum_n c_n \phi_n,$$

then h is orthogonal to all functions ϕ_n. Since this system of functions is complete, $h = 0$, and the above computation of the norm gives the Parseval equality

$$\|f\|^2 = \sum_n |c_n|^2 \|\phi_n\|^2.$$