23 Separation of Variables – General Consideration –

Separation of variables is probably the only systematic way to solve linear PDEs. Its essence is the construction of the problem-adapted orthogonal function system. We have already studied the method in **18** when the ordinary Fourier expansion is applicable. The principles have been exhausted there. Here the general features of the method are outlined with a summary of prerequisites and limitations. Practically, if the reader wishes to solve a PDE boundary value problem, consult a collection of worked-out problems. We should not forget that if we need an exact method, it is a sure sign of our ignorance about the problem.

Key words: special function, eigenvalue problem.

Summary:
(1) Practically, the method works only when the domain has a special shape. Possible shapes are best seen in 'style books,' that is, books collecting worked-out problems. If the reader cannot find any good example in them, it may be wise to give up exact solutions (→**23.2**).
(2) The essence of separation is the problem-adapted Fourier-type expansion; consequently, in order to justify the method we need almost all the machinery of functional analysis (→**23.3**).

23.1 Separation of variables: general idea. All our time-dependent linear problems (→**1**) have the following form:

$$L_t\psi(\boldsymbol{x},t) = Q\psi(\boldsymbol{x},t), \tag{23.1}$$

where the operator L_t acts only on the functions of time, and Q on the functions of space coordinates. The time and space coordinates can be separated trivially as

$$L_t\psi_1(t) = \mu\psi_1(t), \tag{23.2}$$
$$Q\psi_2(\boldsymbol{x}) = \mu\psi_2(\boldsymbol{x}). \tag{23.3}$$

Since the first equation is an ODE, it is easy to obtain its general solution. If Q has a 'good' property, we can generalize the eigenvalue

expansion method for a finite dimensional vector space. Formally (23.1) can be transformed into

$$L_t\langle\varphi_\mu(\boldsymbol{x})|\psi(\boldsymbol{x},t)\rangle = \mu\langle\varphi_\mu(\boldsymbol{x})|\psi(\boldsymbol{x},t)\rangle, \tag{23.4}$$

where $\varphi_\mu(\boldsymbol{x})$ is the eigenfunction of the operator Q ($Q\varphi_\mu(\boldsymbol{x}) = \mu\varphi_\mu(\boldsymbol{x})$) and (consistently with the notation in **20.21**)

$$\langle\varphi_\mu(\boldsymbol{x})|\psi(\boldsymbol{x},t)\rangle \equiv \int_D dx\overline{\varphi_\mu(\boldsymbol{x})}\psi(\boldsymbol{x},t). \tag{23.5}$$

This is an analogue of the integral to compute the Fourier coefficients (→**20.14**, **20.24**). The final solution is formally given by

$$\psi(\boldsymbol{x},t) = \sum_\mu \langle\varphi_\mu(\boldsymbol{x})|\psi(\boldsymbol{x},t)\rangle\varphi_\mu(\boldsymbol{x}), \tag{23.6}$$

where the summation is over all the eigenvalues. Hence, the key problem of the separation of variables is to find a problem-adapted generalized Fourier expansion.

23.2 Practical procedure via separation of variables. As we have seen in **18** boundary conditions make the separation procedure more complicated than stated above (→**26B**, **27B**). We will see an illustration in **23.9**. A practical procedure to solve a PDE with inhomogeneous boundary conditions by separation of variables can be summarized as follows:
(A) If the domain shape is not regular (roughly speaking, if the boundary does not consist of part of planes and conic surfaces), forget about exact analytic methods (→**23.4**).
(B) If the domain is'well-shaped,' then consult a typical problem source book of the boundary-value problem. For example, the lecturer find N. N. Lebedev, I. P. Skalskaya and Y. S. Ufliand, *Worked Problems in Applied Mathematics* (Dover 1965) very useful.[327] If the reader cannot find any similar problem, unless she wishes to be an expert of special functions, it is wise for her to give up analytical methods to obtain exact solutions.
(C) If the reader insists on analytical solutions:
(1) Decompose the problem into the problems with inhomogeneous boundary conditions only in one coordinate direction with the aid of superposition principle exactly as we did in **18.2**. The remaining coordinate directions become (generalized) eigenvalue problems.
(2) The (generalized) eigenfunctions of the separated homogeneous problems dictate the form of the solution. (This is the step of constructing the problem-adapted generalized Fourier expansion scheme.)

[327] This is an accompanying workbook of N. N. Lebedev, *Special Functions & Their Applications* (Dover, 1965), which is an excellent book.

(3) Fix the expansion coefficients with the aid of the inhomogeneous boundary conditions and the orthogonality of the eigenfunctions as in **18**. See **23.9** for an illustration

23.3 What do we need to justify and implement our procedure? Here, we summarize the requirements.
(1) When can we justify the expansion (23.6)? To answer this question, we need a rudimentary knowledge of Hilbert space (→**20**) and the operators on it (→**34**). After a suitable preparation we can generalize Fourier expansion and integral transformations (→**34B.6**).
(2) We must be able to find explicitly the eigenfunctions of Q defined on a linear space satisfying the auxiliary conditions. We use the method of separation of variables to reduce the problems to lower dimensional (hopefully 1-space) problems. Therefore, we need methods to solve linear ODEs (→**24**) and associated eigenvalue problems (the Sturm-Liouville problems **35**).

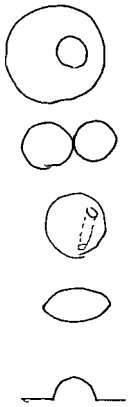
Discussion: Fourier expansion of multivariable functions: addendum to separation of variables.
We have claimed that the key element of the justification of the separation of variables is the (generalized) Fourier expansion of the function in terms of the 'equation adapted' orthonormal basis.[328] Generally, we have several variables and we need multiple Fourier expansion. Then a natural question is whether the totality of the tensor products constructed from ON bases for individual coordinates is indeed an ON basis. The answer is in a certain sense affirmative, but somewhat delicate.
(1) The (generalized) Fourier expansion of $f(x_1, \cdots, x_n)$ is well defined if f is integrable thanks to Fubini's theorem (→**20.15**). That is, the value of the Fourier coefficients do not depend on the order of expansion.
(2) To reconstruct the original function from the Fourier coefficients, we can apply the individual inverse transforms successively. This is allowed, but the Fourier coefficients may not be integrable, so to interpret the inverse transform as an n-tuple integral (not as n successive one dimensional integrals) is delicate[329] and some extra condition on f is generally required.[330]

23.4 What problems can we solve by hand?. To have an analytic solutions, we must be able to solve the eigenvalue problem by hand. To this end almost always separation of variables is mandatory. As is mentioned in **23.2**(A), this requires not only a special form of the operator,

[328]The reader might say any ON basis will do for our purpose. If we need not worry about the (termwise) differentiability of the Fourier sum, then this is indeed the case. However, we are solving differential equations, so that we must be sensitive about the uniformity of the convergence of the resultant Fourier series (→**17.12-13**).
[329]That is, we must in general inverse transform in the reverse order of the operation used in the calculation of the coefficients.
[330]See Kolmogorov and Fomin, second ed. Chapter 8, Section 4. Perhaps not available in English.

but also a special shape of the domain.[331] Therefore, problems we can solve analytically are very limited even for the Laplace equation. For situations frequently encountered in practice (e.g., the Laplacian in a ball) eigenfunctions of separated operators are well known and called *special functions*. In short, we can solve by hand only very standard PDE under very standard auxiliary conditions. That is why the advice in (B) of **23.2** "see a style book" is practical.

Exercise.
(A) Specify appropriate curvilinear coordinates to solve the following problems (if the problems are separable at all):
(1) From a solid ball of radius a, another ball of radius $b(< a)$ which is completely inside the first ball is removed. Temperatures of inside and outside surfaces are given. Find the steady temperature distribution in the solid.
(2) There are two osculating identical conducting balls. Compute the electric field when the balls are maintained at V with respect to the infinity.
(3) A cylindrical hole of radius r is made through a solid conducting ball of radius $R(> r)$ slightly off the center. Find the electric field when the solid has the total charge Q.
(4) A lens-shaped conductor is maintained at voltage V relative to infinity. Assume that the surfaces are with the same radius of curvature R and the thickness of the lens is $2d$, where $d < R$.
(5) A conducting plane has a semicylindrical boss of radius a. The plane is maintained with the electric potential V. Find the electric field in the space.
(B) Two identical conducting spheres of radius a are placed with the separation of $2l$ between the centers. Both the spheres are maintained at voltage V relative to infinity. Find the electrostatic potential due to these spheres.



Discussion: Lamè's problem.
The most general case we can solve with the aid of separation of variables is the *confocal rectangular parallelepiped* whose surfaces are made of confocal quadratic surfaces given by

$$\frac{x^2}{s-e_1}+\frac{y^2}{s-e_2}+\frac{z^2}{s-e_3}=1. \tag{23.7}$$

The necessary special functions are called Lamè functions which are not studied very well.

23.5 What is a special function? The word 'special function' is used to denote collectively (1) Γ-function ($\rightarrow$**9**) and related functions

[331] We must be able to employ the standard orthogonal curvilinear coordinates. For example, for 3d Schrödinger equation, complete separation of variables is possible only when the potential function V has the following form:

$$V = h_1V(q^1) + h_2V(q^2) + h_3V(q^3),$$

where h's are the ones given in **2D.3** (H. P. Robertson, Math. Ann. **98**, 749 (1928); L. P. Eisenstein, Ann. Math. **35**, 284 (1934)).

like polygamma functions,[332] (2) functions described by indefinite integrals of elementary functions like the probability integral (→**25.11**), (3) elliptic functions, (4) solutions of second order ODE obtained by separating variables, and (5) solutions to special ODE like Painlevè equations.[333] Solutions to the second order linear ODE with 3 regular singular points (special functions of hypergeometric type) or with 1 irregular singular point resulting from the merging of two regular singular points (→**24B.2**) in the former (special functions of confluent type) are called *classical special functions.*

23.6 Are the analytic solutions useful? It is not easy to say yes. Often the obtained solutions are series solutions in terms of special functions. Since special functions are mere symbols, one must look up tables or use, e.g., Mathematica or Maple (even trigonometric functions are no exceptions; we need a table or a pocket calculator). Hence, if she wants a detailed behavior of the solution, a lot of numerical work is needed anyway. One might say that in order to know qualitative or asymptotic behaviors of a solution, analytic forms are useful. This is true. However, to require a complete solution in order to get qualitative or asymptotic behaviors does not sound elegant.

It should be clearly recognized that necessity of full analytic solution is a clear sign of the sad fact that we do not understand the problem.

23.7 Importance of qualitative understanding. It is important to know how to solve the problems by hand: what special functions are suitable, how they behave qualitatively, etc. To teach these has been the main objective of the conventional math-phys courses.[334] However, for most scientists (esp. pure scientists) to juggle tons of special functions is not important at all.[335] It is much more important to acquire the sense or feeling of correct physics and mathematics so that we will not be outsmarted by computers, or not to be drowned in the flood of numbers. The reader must be able to walk, but in order to go to the Pacific coast she need not retrace the Oregon Trail on foot!

---

[332] Polygamma functions: the nth-derivative of $\log\Gamma(z)$ is called the (n+1)-Gamma function. In particular, $n = 1$ is called digamma function, $n = 2$ is called trigamma function, etc.

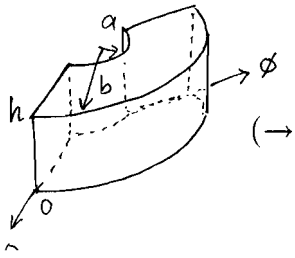
[333] See E L Ince, *Ordinary Differential Equations* (Dover, 1956; original 1926) Chapter 14.

[334] See, e.g., H. W. Wyld, *Mathematical Methods for Physics* (Benjamin, 1976).

[335] Perhaps more than 50 years ago there were one-year courses solely devoted to trigonometrics in universities (remember that the universities in those days were not remedial schools of the high school education). This sounds absurd now. To realize that some topics are unimportant is an important progress.

23.8 Use of symbol manipulation programs. Many standard analytic methods, e.g., the series expansion method (→**24B**), are best implemented with the aid of mathematics softwares like Mathematica or Maple. Special functions are available in the standard mathematics softwares. For example, with Mathematica, if the reader types in `BesselJ[n,z]`, then she gets $J_n(z)$ (→**27A.1**). Hence, we need not be extremely familiar with special functions, although we should know their general features. Most analytical calculations can be mechanized, so it is probably wiser to practice the use of these programs than to experience lengthy practice sessions of analytical methods.

23.9 Case study of separation of variables: Laplace equation with Dirichlet condition. The purpose of this entry is to provide a show case with the aid of a fair difficult problem. The region is fan-shaped: $z \in [0,h]$, $\varphi \in [0,\phi]$ and $r \in [a,b]$:



$$\left[\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}+\frac{\partial^2}{\partial z^2}\right]\psi=0 \tag{23.8}$$

(→**2D.10**) with the boundary condition

$$\psi(r,\varphi,0) = f_0(r,\varphi),\quad \psi(a,\varphi,h)=f_h(r,\varphi), \tag{23.9}$$
$$\psi(r,0,z) = g_0(r,z),\quad \psi(r,\phi,z)=g_\phi(r,z), \tag{23.10}$$
$$\psi(a,\varphi,z) = h_a(\varphi,z),\quad \psi(b,\varphi,z)=h_b(\varphi,z). \tag{23.11}$$

First we perform the step (C)(1) of **23.2**. The separation procedure $\psi = R(r)\Phi(\varphi)Z(z)$ gives three distinct eigenvalue problems. The full solution is the superposition of the solutions to all the following three problems (1)-(3).
(1) With the boundary condition (r, φ homogeneous; z inhomogeneous):

$$\psi(r,\varphi,0) = f_0(r,\varphi),\quad \psi(r,\varphi,h)=f_h(r,\varphi), \tag{23.12}$$
$$\psi(r,0,z) = 0,\quad \psi(r,\varphi,z)=0, \tag{23.13}$$
$$\psi(a,\varphi,z) = 0,\quad \psi(b,\varphi,z)=0. \tag{23.14}$$

The separated equations are

$$\frac{d^2\Phi}{d\varphi^2} = -m^2\Phi, \tag{23.15}$$
$$\frac{d^2Z}{dz^2} = \alpha^2 Z, \tag{23.16}$$
$$\frac{1}{R}\left[\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right]-\frac{m^2}{r^2}+\alpha^2 = 0. \tag{23.17}$$

The eigenvalue problems are (23.15) and (23.17) with homogeneous Dirichlet boundary conditions ($\Phi(0) = \Phi(\phi) = 0$ and $R(a) = R(b) = 0$). The positivity of α^2 and m^2 follows from the negative definiteness of the operators.[336] The solution must have the following form:

$$\psi = \sum_{m,\alpha}(A_{m,\alpha}J_m(\alpha r)+B_{m,\alpha}N_m(\alpha r))(C_m \sin m\varphi+D_m \cos m\varphi)(E_{m,\alpha} \sinh \alpha z+F_{m,\alpha} \cosh \alpha z). \tag{23.18}$$

Here J_m is the Bessel function (→**27A.2-3**), and N_m is the Neumann function (→**27A.16**). m, C_m and D_m are fixed by the Dirichlet condition:

$$D_m = 0;\ C_m \sin m\phi + D_m \cos m\phi = 0. \tag{23.19}$$

We may choose $C_m = 1$ without any loss of generality. α, $A_{m,\alpha}$ and $B_{m,\alpha}$ are fixed by the Dirichlet condition

$$A_{m,\alpha}J_m(\alpha a) + B_{m,\alpha}N_m(\alpha a) = 0, \tag{23.20}$$
$$A_{m,\alpha}J_m(\alpha b) + B_{m,\alpha}N_m(\alpha b) = 0. \tag{23.21}$$

That is, $J_m(\alpha a)N_m(\alpha b) = J_m(\alpha b)N_m(\alpha a)$ fixes α. E and F are determined from the inhomogeneous boundary condition (23.14) with the aid of complete orthogonality (→**34B.5**) of the eigenfunctions constructed above (not easy or almost impossible bu hand for general a and b).
(2) With the boundary condition (r, z homogeneous; φ inhomogeneous)

$$\psi(r,\varphi,0) = 0;\ \psi(a,\varphi,h) = 0, \tag{23.22}$$
$$\psi(r,0,z) = g_0(r,z);\ \psi(r,\phi,z) = g_\phi(r,z), \tag{23.23}$$
$$\psi(a,\varphi,z) = 0;\ \psi(b,\varphi,z) = 0. \tag{23.24}$$

The separated equations are

$$\frac{d^2\Phi}{d\varphi} = m^2\Phi, \tag{23.25}$$
$$\frac{d^2Z}{dz^2} = -\alpha^2 Z, \tag{23.26}$$
$$\frac{1}{R}\left[\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right] + \frac{m^2}{r^2} - \alpha^2 = 0. \tag{23.27}$$

Here the positivity of α^2 is obvious from the condition that (23.26) becomes an eigenvalue problem (it is is not elementary to see this →**34B.6**

[336] Intuitively speaking, the eigenfunctions must be oscillatory functions to satisfy the orthogonality condition. "Negative definiteness" of an operator L means $\langle f|L|f\rangle \leq 0$ for any ket $|f\rangle$. The Laplacian Δ is a typical example.

Discussion (B)). m^2 also must be positive so that (23.27) becomes an eigenvalue problem. Hence, we may assume

$$\psi = \sum_{m,\alpha}(A_{m,\alpha}I_{im}(\alpha r)+B_{m,\alpha}K_{im}(\alpha r))(C_m \sinh m\varphi+D_m \cosh m\varphi)(E_\alpha \sin \alpha z+F_\alpha \cos \alpha z), \tag{23.28}$$

where I and K are modified Bessel functions (→**27A.23**). Here α, E_α and F_α are fixed by the Dirichlet condition

$$F_\alpha = 0;\ E_\alpha \sin \alpha h + F_\alpha \cos \alpha h = 0. \tag{23.29}$$

$E_\alpha = 1$ is admissible. m, $A_{m,\alpha}$ and $B_{m,\alpha}$ are determined by the boundary conditions

$$A_{m,\alpha}I_{im}(\alpha a) + B_{m,\alpha}K_{im}(\alpha a) = 0, \tag{23.30}$$
$$A_{m,\alpha}I_{im}(\alpha b) + B_{m,\alpha}K_{im}(\alpha b) = 0. \tag{23.31}$$

That is, $I_{im}(\alpha a)K_{im}(\alpha b) = I_{im}(\alpha b)K_{im}(\alpha a)$ determines m. C and D are determined from the inhomogeneous boundary condition (23.23) with the aid of complete orthogonality of the eigenfunctions constructed above.[337]

(3) With the boundary condition (φ, z homogeneous; r inhomogeneous)

$$\psi(r,\varphi,0) = 0,\ \psi(r,\varphi,h) = 0, \tag{23.32}$$
$$\psi(r,0,z) = 0,\ ,\psi(r,\phi,z) = 0, \tag{23.33}$$
$$\psi(a,\varphi,z) = h_a(\varphi,z),\ \psi(b,\varphi,z) = h_b(\varphi,z). \tag{23.34}$$

The separated equations are[338]

$$\frac{d^2\Phi}{d\varphi} = -m^2\Phi, \tag{23.35}$$
$$\frac{d^2Z}{dz^2} = -\alpha^2 Z, \tag{23.36}$$
$$\frac{1}{R}\left[\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right]-\frac{m^2}{r^2}-\alpha^2 = 0. \tag{23.37}$$

The eigenvalue problems are easy ones: (23.35) and (23.36) with homogeneous Dirichlet conditions. We may thus assume

$$\psi = \sum_{m,\alpha}(A_{m,\alpha}I_m(\alpha r)+B_{m,\alpha}K_m(\alpha r))(C_m \sin m\theta+D_m \cos m\theta)(E_\alpha \sin \alpha z+F_\alpha \cos \alpha z). \tag{23.38}$$

[337] This problem is nontrivial, since we need modified Bessel functions of imaginary order. See N. N. Lebedev, *Special Functions & Their Applications* (Dover 1972) Section 6.5.

[338] In this case obviously m^2 and α^2 must be non-negative.

Here, I_m and K_m are modified Bessel functions (→**27A.23**). A and B must be fixed from the boundary condition (23.34).

23.10 Remarks to 23.9.
(1) If the region in the z-direction is not bounded, we need Fourier transformations; if the region is not bounded in the r-direction, we need the Fourier-Bessel(-Dini) transformation (→**27A.22**).
(2) The boundary condition in the φ direction may be periodic.
(3) The Neumann condition case is analogous.