

Chapter 2

The Periodic Sturm–Liouville Equations

One-dimensional crystals are the simplest crystals. Historically, much of our current fundamental understanding of electronic structures of crystals was obtained through the analysis of one-dimensional crystals [1–3]. Among the most well-known examples are the Kronig–Penney model [4], Kramers’ general analysis of the band structure of one-dimensional crystals [5], Tamm’s surface states [6], and so forth. A clear understanding of electronic states in one-dimensional finite crystals is the basis for further understandings of the electronic states in low-dimensional systems and finite crystals. For this purpose, we need to have a clear understanding of solutions of the Schrödinger equations for one-dimensional crystals.

The Schrödinger differential equation for a one-dimensional crystal can be written as¹:

$$-y''(x) + [v(x) - \lambda]y(x) = 0, \quad (2.1)$$

where $v(x + a) = v(x)$ is the periodic potential. Eastham’s book [7] provided a comprehensive and in-depth mathematical theory of a general class of periodic differential equations, where Eq. (2.1) is a specific and simple form. The theory in his book provided the major mathematics basis of the first edition of this book.

The relevant mathematical theory has made significant progress [8–10] since Eastham’s book was published in 1975. Zettl [9] pointed out that in the Eastham’s book “strong smoothness and positivity restrictions are placed on the coefficients. However, many, but not all, of the proofs given there are valid under much less severe restrictions on the coefficients.” The modern theory of periodic Sturm–Liouville equations [8–10] with less severe restrictions can treat more general problems.

In this chapter, we study the basic theory of the periodic Sturm–Liouville equations [8–10], to prepare for investigations of electronic states in semi-infinite and finite

¹In this book, a prime on a function denotes differentiation with respect to the variable of the function. If the function has two or more variables, a prime on the function denotes differentiation with respect to the variable x .

one-dimensional crystals in next two chapters and relevant problems in Appendices. We begin with a brief review of some elementary knowledge of the theory of the second-order linear homogeneous ordinary differential equations. Then, we present two basic Sturm Theorems on zeros of solutions of these equations. In the theory of boundary value problems for ordinary differential equations, the existence and locations of zeros of solutions of such equations are often of central importance. In the major part of this chapter, we learn the basic theory of the periodic Sturm–Liouville equations and the zeros of their solutions. Based on the mathematical theory and theorems in this chapter, a general theoretical formalism for investigations of the existence and properties of surface states in a semi-infinite one-dimensional crystal can be developed, and general results on electronic states in ideal one-dimensional finite crystals can be rigorously proven. For our purpose, we work on to understand how the mathematical results in Eastham’s book [7] can be extended by the modern theory of periodic Sturm–Liouville equations [8–10], by following the steps in Chap. 2 of the first edition.

We are interested in the Sturm–Liouville equations with periodic coefficients [8–10]:

$$[p(x)y'(x)]' + [\lambda w(x) - q(x)]y(x) = 0, \quad (2.2)$$

where $p(x) > 0$, $w(x) > 0$ and $p(x)$, $q(x)$, $w(x)$ are *piecewise continuous* real periodic functions with period a :

$$p(x + a) = p(x), \quad q(x + a) = q(x), \quad w(x + a) = w(x).$$

The one-dimensional Schrödinger equation (2.1) corresponds to a specific and simple form of Eq. (2.2) where $p(x) = w(x) = 1$ and $q(x) = v(x)$. The basic theory of periodic Sturm–Liouville equations summarized in this chapter is somewhat more general and advanced than what we need in treating Eq. (2.1) for electronic states in one-dimensional crystals.² However, such a more general and up-to-date theory can treat more general one-dimensional problems in physics, including the one-dimensional photonic crystals and phononic crystals in Appendices. Readers who are interested in a more complete and general mathematical theory are recommended to read original books [7–10]. Readers who are not interested in the proofs of relevant theorems may skip those parts of this chapter.

2.1 Elementary Theory and Two Basic Sturm Theorems

We begin with a class of second-order linear homogeneous ordinary differential equations:

$$[(p(x)y'(x))]' + q(x)y(x) = 0, \quad -\infty < x < +\infty. \quad (2.3)$$

²As the mathematical basis of the first edition of the book, in [7] it was assumed that $p(x)$ is real-valued, *continuous* and nowhere zero, and $p'(x)$ is *piecewise continuous*.

Here, $p(x)$ and $q(x)$ are piecewise continuous real finite functions.

$p(x)y'(x)$ is called the quasi-derivative of $y(x)$ [8–11], distinguished from the classical derivative $y'(x)$. In this chapter, the quasi-derivative plays the roles of the classic derivative in the first edition of the book. This is the essential difference between the modern theory of periodic Sturm–Liouville equations [8–10] summarized in this chapter and the theory presented in the Eastham’s book [7]. In equations on many physical problems, the classical derivative $y'(x)$ may not exist in some cases, but the quasi-derivative $p(x)y'(x)$ exists and is continuous.³ The applications of quasi-derivatives [8–11] significantly extends the ranges of the problems treatable by the modern Sturm–Liouville theory.

Equation (2.3) can be written in a matrix form [8–11]:

$$\begin{pmatrix} y(x) \\ p(x)y'(x) \end{pmatrix}' = \begin{pmatrix} 0 & \frac{1}{p(x)} \\ -q(x) & 0 \end{pmatrix} \begin{pmatrix} y(x) \\ p(x)y'(x) \end{pmatrix}, \quad -\infty < x < +\infty. \quad (2.4)$$

Equation (2.4) is a simple special case of a more general first-order linear homogeneous differential equation of matrices:

$$Y' = AY \quad (2.5)$$

with

$$Y = \begin{pmatrix} y(x) \\ p(x)y'(x) \end{pmatrix}, \quad A = \begin{pmatrix} 0 & \frac{1}{p(x)} \\ -q(x) & 0 \end{pmatrix}, \quad -\infty < x < +\infty. \quad (2.6)$$

For physical problems interested in this book, the functions $p(x)$ and $q(x)$ are as stated at the beginning of this section. Mathematicians might be more interested in properties of a more general Eq. (2.5) and relevant equations. The elements of the matrix A might often be more general.

³Suppose x_i is an isolated point where the function $p(x)$ is not continuous. By integrating (2.3) from $x_i - \delta$ to $x_i + \delta$, where δ is an infinitesimal positive number, we obtain

$$\int_{x_i-\delta}^{x_i+\delta} [p(x)y'(x)]' dx = - \int_{x_i-\delta}^{x_i+\delta} q(x)y(x) dx.$$

Since δ is an infinitesimal positive number, we have

$$\int_{x_i-\delta}^{x_i+\delta} q(x)y(x) dx = 0. \quad \delta \rightarrow 0$$

Thus

$$[p(x)y'(x)]_{x_i-0} = [p(x)y'(x)]_{x_i+0}.$$

That is, the quasi-derivative $p(x)y'(x)$ is continuous at x_i despite the fact that the classic derivative y' does not exist at x_i : $y'_{(x_i-0)} \neq y'_{(x_i+0)}$.

Properties of solutions of Eq. (2.3) or Eq. (2.4) can be obtained from the properties of solutions of the more general Eq. (2.5) [8–12].

1. Two linear-independent solutions.

Any nontrivial solution Y of Eq. (2.4) can be written as a linear combination of two linearly independent solutions Y_1 and Y_2 of Eq. (2.4):

$$Y = c_1 Y_1 + c_2 Y_2, \quad (2.7)$$

or alternatively,

$$\begin{pmatrix} y(x) \\ p(x)y'(x) \end{pmatrix} = c_1 \begin{pmatrix} y_1(x) \\ p(x)y_1'(x) \end{pmatrix} + c_2 \begin{pmatrix} y_2(x) \\ p(x)y_2'(x) \end{pmatrix}. \quad (2.8)$$

Here, c_1 and c_2 are two independent constants.

2. The fundamental matrix and the Wronskian.

Let Y_1 and Y_2 be two linearly independent solutions of Eq. (2.4). The following matrix Φ is called a fundamental matrix of Eq. (2.4):

$$\Phi = \begin{pmatrix} y_1(x) & y_2(x) \\ p(x)y_1'(x) & p(x)y_2'(x) \end{pmatrix}. \quad (2.9)$$

The Wronskian $W(y_1, y_2)$ of two solutions $y_1(x)$ and $y_2(x)$ of Eq. (2.3) is defined as

$$W(y_1, y_2) = y_1(x) p(x)y_2'(x) - p(x)y_1'(x) y_2(x). \quad (2.10)$$

The Wronskian $W(y_1, y_2)$ of two distinct solutions $y_1(x)$ and $y_2(x)$ of Eq. (2.3) is a constant:

$$[W(y_1, y_2)]' = [y_1(x) p(x)y_2'(x) - p(x)y_1'(x) y_2(x)]' = 0. \quad (2.11)$$

A necessary and sufficient condition that two solutions y_1 and y_2 of Eq. (2.3) are linearly independent is that

$$W(y_1, y_2) = y_1(x) p(x)y_2'(x) - p(x)y_1'(x) y_2(x) \neq 0.$$

3. The variation of parameters formula.

The nonhomogeneous differential equation

$$[p(x)z'(x)]' + q(x)z(x) = F \quad (2.12)$$

can be written in a matrix form:

$$\begin{pmatrix} z(x) \\ p(x)z'(x) \end{pmatrix}' = \begin{pmatrix} 0 & \frac{1}{p(x)} \\ -q(x) & 0 \end{pmatrix} \begin{pmatrix} z(x) \\ p(x)z'(x) \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix}. \quad (2.13)$$

The nonhomogeneous differential equation (2.13) can be solved as

$$\begin{pmatrix} z(x) \\ p(x)z'(x) \end{pmatrix} = \Phi(x) \int^x \Phi^{-1}(t) \begin{pmatrix} 0 \\ F(t) \end{pmatrix} dt. \quad (2.14)$$

Here Φ is a fundamental matrix of Eq. (2.4).

Let $y_1(x)$ and $y_2(x)$ be two linearly independent solutions of Eqs. (2.3), (2.14) can be more explicitly written as

$$\begin{pmatrix} z(x) \\ p(x)z'(x) \end{pmatrix} = \begin{pmatrix} y_1(x) & y_2(x) \\ p(x)y_1'(x) & p(x)y_2'(x) \end{pmatrix} \begin{pmatrix} -\int^x \frac{1}{W(t)} y_2(t) F(t) dt \\ \int^x \frac{1}{W(t)} y_1(t) F(t) dt \end{pmatrix}, \quad (2.15)$$

or

$$z(x) = -\int^x \frac{F(t)y_2(t)}{W(t)} dt y_1(x) + \int^x \frac{F(t)y_1(t)}{W(t)} dt y_2(x), \quad (2.16)$$

and

$$p(x)z'(x) = -\int^x \frac{F(t)y_2(t)}{W(t)} dt p(x)y_1'(x) + \int^x \frac{F(t)y_1(t)}{W(t)} dt p(x)y_2'(x), \quad (2.17)$$

here $W(t) = [W(y_1, y_2)]_t = y_1(t) p(t)y_2'(t) - p(t)y_1'(t) y_2(t)$ is defined in (2.11).

There are two basic theorems on zeros of solutions of the differential equation (2.3).

Theorem 2.1 (Sturm Separation Theorem [Theorem 2.6.2 in [9]]) *Let y_1 and y_2 be two linearly independent real solutions of (2.3); then there is exact one zero of y_2 between two consecutive zeros of y_1 .*

Proof Suppose α and β are two consecutive zeros of y_1 ,

$$y_1(\alpha) = y_1(\beta) = 0; \quad (2.18)$$

then it can be proven that there is, at least, one zero of y_2 in (α, β) .

Without losing generality, we may assume that $y_1(x) > 0$ in (α, β) . Then we have

$$y_1(\alpha + \delta) - y_1(\alpha) = \int_{\alpha}^{\alpha+\delta} \frac{p(x)y_1'(x)}{p(x)} dx > 0 \quad (2.19)$$

for a small $\delta > 0$. Since $p(\alpha)y_1'(\alpha) \neq 0$ and $p(x)y_1'(x)$ is continuous, (2.19) indicates that

$$p(\alpha)y_1'(\alpha) > 0. \quad (2.20)$$

Similarly, we have

$$p(\beta)y_1'(\beta) < 0. \quad (2.21)$$

From (2.11) we have $[W(y_1, y_2)]_{\alpha} = [W(y_1, y_2)]_{\beta} \neq 0$, thus

$$y_1(\alpha) p(\alpha) y_2'(\alpha) - p(\alpha) y_1'(\alpha) y_2(\alpha) - y_1(\beta) p(\beta) y_2'(\beta) + p(\beta) y_1'(\beta) y_2(\beta) = 0.$$

By (2.18) we further obtain

$$-p(\alpha) y_1'(\alpha) y_2(\alpha) + p(\beta) y_1'(\beta) y_2(\beta) = 0. \quad (2.22)$$

Since y_2 and y_1 are linearly independent, neither $y_2(\alpha)$ nor $y_2(\beta)$ is zero. Equation (2.22) can be true only when $y_2(\alpha)$ and $y_2(\beta)$ have different signs. Since $y_2(x)$ is continuous, there must be at least one zero of $y_2(x)$ in (α, β) .

However, if there are more than one zeros of y_2 in (α, β) , then according to what we have just proven, there is at least one extra zero of y_1 between two zeros of y_2 in (α, β) . This is contradictory to the supposition that α and β are two consecutive zeros of y_1 . Therefore, there is always one and only one zero of y_2 between two consecutive zeros of y_1 . Similarly, there is always one and only one zero of y_1 between two consecutive zeros of y_2 . The zeros of two linearly independent real solutions y_1 and y_2 of (2.3) are distributed alternatively and separated from each other. $\square$

Theorem 2.2 (Sturm Comparison Theorem [Theorem 2.6.3 in [9]]) *Suppose in two differential equations*

$$(p_1 y_1')' + q_1 y_1 = 0, \quad (2.23)$$

and

$$(p_2 y_2')' + q_2 y_2 = 0, \quad (2.24)$$

where

$$0 < p_2 \leq p_1 \quad \text{and} \quad q_2 \geq q_1 \quad (2.25)$$

are true, and α and β are two zeros of a nontrivial real solution y_1 of the first equation (2.23), then there is, at least, one zero of any nontrivial real solution y_2 of (2.24) in $[\alpha, \beta]$.

Proof Suppose that this is not true — that y_2 is not zero anywhere in $[\alpha, \beta]$, we may assume $y_2 > 0$ in $[\alpha, \beta]$. Without loss of generality, we may assume that α and β are two consecutive zeros of y_1 :

$$y_1(\alpha) = y_1(\beta) = 0, \quad (2.26)$$

and $y_1 > 0$ in (α, β) , then we can have

$$\left[\frac{y_1}{y_2} (p_1 y_1' y_2 - p_2 y_2' y_1) \right]' = (q_2 - q_1) y_1^2 + (p_1 - p_2) y_1'^2 + p_2 \frac{(y_1 y_2' - y_2 y_1')^2}{y_2^2} \quad (2.27)$$

in $[\alpha, \beta]$. By doing an integration of (2.27) from α to β and note that the integration of the left side is zero due to (2.26), we obtain that

$$\int_{\alpha}^{\beta} (q_2 - q_1) y_1^2 dx + \int_{\alpha}^{\beta} (p_1 - p_2) y_1'^2 dx = - \int_{\alpha}^{\beta} p_2 \frac{(y_1 y_2' - y_2 y_1')^2}{y_2^2} dx.$$

This equation can be valid only when $q_1 = q_2$, $p_1 = p_2$ and y_1, y_2 are linearly dependent. Thus, the theorem is proven. $\square$

2.2 The Floquet Theory

Now we consider the solutions of an Eq. (2.3) where $p(x)$ and $q(x)$ are real periodic functions with the same period a :

$$[p(x)y']' + q(x)y = 0, \quad p(x+a) = p(x), \quad q(x+a) = q(x). \quad (2.28)$$

Here, a is a nonzero real constant.

Theorem 2.3 (Theorem 2.7.1 in [9]) *There exist at least one nonzero constant ρ and one nontrivial solution $y(x)$ of (2.28) such that*

$$y(x+a) = \rho y(x), \quad (2.29)$$

Proof We can choose two linearly independent solutions $\eta_1(x)$ and $\eta_2(x)$ of (2.28) according to

$$\eta_1(0) = 1, \quad p(0)\eta_1'(0) = 0; \quad \eta_2(0) = 0, \quad p(0)\eta_2'(0) = 1. \quad (2.30)$$

These solutions are usually called normalized solutions of (2.28) [13].

Since the corresponding $\eta_1(x+a)$ and $\eta_2(x+a)$ are also two linearly independent nontrivial solutions of (2.28), we can write $\eta_1(x+a)$ and $\eta_2(x+a)$ as linearly combinations of $\eta_1(x)$ and $\eta_2(x)$:

$$\begin{aligned} \eta_1(x+a) &= A_{11}\eta_1(x) + A_{12}\eta_2(x), \\ \eta_2(x+a) &= A_{21}\eta_1(x) + A_{22}\eta_2(x), \end{aligned} \quad (2.31)$$

where A_{ij} ($1 \leq i, j \leq 2$) are four constants. From (2.30) and (2.31), we obtain that

$$A_{11} = \eta_1(a), \quad A_{21} = \eta_2(a), \quad A_{12} = p(a)\eta_1'(a), \quad A_{22} = p(a)\eta_2'(a). \quad (2.32)$$

Any nontrivial solution $y(x)$ of (2.28) can be written as

$$y(x) = c_1\eta_1(x) + c_2\eta_2(x),$$

where c_i are constants. If there is a nonzero ρ that makes

$$\begin{aligned} (A_{11} - \rho)c_1 + A_{21}c_2 &= 0, \\ A_{12}c_1 + (A_{22} - \rho)c_2 &= 0 \end{aligned} \quad (2.33)$$

true, then (2.28) has a nontrivial solution of the form (2.29). The requirement that c_i in (2.33) are not both zero leads to the condition

$$\rho^2 - [\eta_1(a) + p(a)\eta_2'(a)]\rho + 1 = 0. \quad (2.34)$$

Here, $[W(\eta_1, \eta_2)]_a = [W(\eta_1, \eta_2)]_0$ thus $\eta_1(a)p(a)\eta_2'(a) - p(a)\eta_1'(a)\eta_2(a) = 1$ was used. The quadratic equation (2.34) is called the characteristic equation associated with (2.28) [13]. Equation (2.34) for ρ has, at least, one nonzero root since it has a nonzero constant term. $\square$

Equation (2.28) may have one nontrivial solution of the form (2.29) or two linearly independent nontrivial solutions of the form (2.29), depending on whether the matrix $A = (A_{ij})$ in (2.32) has only one eigenvector or two linearly independent eigenvectors.

Theorem 2.4 (Theorem 2.7.2 in [9]) *There exist linearly independent solutions $y_1(x)$ and $y_2(x)$ of (2.28) such that either*

(i)

$$\begin{aligned} y_1(x) &= e^{h_1 x} p_1(x), \\ y_2(x) &= e^{h_2 x} p_2(x), \end{aligned} \quad (2.35)$$

here h_1 and h_2 are constants, not necessarily distinct, $p_i(x)$, $i = 1, 2$ are periodic with period a , or

(ii)

$$\begin{aligned} y_1(x) &= e^{hx} p_1(x), \\ y_2(x) &= e^{hx} [x p_1(x) + p_2(x)], \end{aligned} \quad (2.36)$$

here h is a constant and $p_i(x)$, $i = 1, 2$ are periodic with period a .

Proof The characteristic equation (2.34) may have either two distinct roots or a repeated root.

1. If the characteristic equation (2.34) has two distinct roots ρ_1 and ρ_2 , then there are two linearly independent nontrivial solutions of $y_1(x)$ and $y_2(x)$ of (2.28) such that

$$y_i(x + a) = \rho_i y_i(x), \quad i = 1, 2.$$

We can define h_1 and h_2 so that

$$e^{ah_i} = \rho_i \quad (2.37)$$

and then two functions $p_i(x)$ by

$$p_i(x) = e^{-h_i x} y_i(x).$$

It is easy to see that $p_1(x)$ and $p_2(x)$ are periodic functions with period a :

$$p_i(x + a) = e^{-h_i(x+a)} \rho_i y_i(x) = p_i(x).$$

Thus, (2.28) has two linearly independent nontrivial solutions:

$$y_1(x) = e^{h_1 x} p_1(x); \quad y_2(x) = e^{h_2 x} p_2(x). \quad (2.38)$$

2. Now, we consider the case that the characteristic equation (2.34) has a repeated root ρ . Define h by

$$e^{ah} = \rho. \quad (2.39)$$

According to Theorem 2.3, (2.28) has a nontrivial solution of the form (2.29):

$$y_1(x + a) = \rho y_1(x).$$

Suppose $Y_2(x)$ is any solution of (2.28) that is linearly independent of $y_1(x)$. Since $Y_2(x + a)$ is also a nontrivial solution of (2.28) we can write

$$Y_2(x + a) = c_1 y_1(x) + c_2 Y_2(x), \quad (2.40)$$

here c_1 and c_2 are constants. Since

$$[W(y_1, Y_2)]_{x+a} = \rho c_2 [W(y_1, Y_2)]_x$$

and $[W(y_1, Y_2)]_x$ does not depend on x , therefore,

$$\rho c_2 = 1 = \rho^2,$$

the second equality is due to that the constant term in (2.34) is equal to 1. Thus,

$$c_2 = \rho.$$

Equation (2.40) can be written as

$$Y_2(x + a) = c_1 y_1(x) + \rho Y_2(x). \quad (2.41)$$

There could be two different cases:

- 2.1. $c_1 = 0$.

Equation (2.41) becomes

$$Y_2(x + a) = \rho Y_2(x).$$

We can choose $y_2(x) = Y_2(x)$. Thus, (2.28) has two linearly independent solutions $y_1(x)$ and $y_2(x)$ and

$$y_1(x + a) = e^{ah} y_1(x), \quad y_2(x + a) = e^{ah} y_2(x).$$

The first part of the theorem is proven. This case corresponds to the case that the matrix $A = (A_{ij})$ has one repeated eigenvalue ρ but two linearly independent eigenvectors. Consequently, (2.28) may have two linearly independent nontrivial solutions of the form (2.29).

2.2. $c_1 \neq 0$.

Define

$$p_1(x) = e^{-hx} y_1(x), \quad p_2(x) = (a\rho/c_1) e^{-hx} Y_2(x) - x p_1(x);$$

then we have

$$p_1(x + a) = e^{-h(x+a)} y_1(x + a) = p_1(x)$$

and

$$\begin{aligned} p_2(x + a) &= (a\rho/c_1) e^{-h(x+a)} Y_2(x + a) - (x + a) p_1(x + a) \\ &= (a\rho/c_1) e^{-hx} Y_2(x) - x p_1(x) = p_2(x). \end{aligned}$$

Thus, $p_1(x)$ and $p_2(x)$ are periodic functions. Since

$$y_1(x) = e^{hx} p_1(x), \quad Y_2(x) = (c_1/a\rho) e^{hx} [x p_1(x) + p_2(x)],$$

we may choose

$$y_2(x) = (a\rho/c_1) Y_2(x).$$

Thus,

$$y_2(x) = e^{hx} [x p_1(x) + p_2(x)]$$

and the part (ii) of the theorem is proven. □

The part (i) of the theorem corresponds to the cases where Eq. (2.28) has two linearly independent nontrivial solutions of the form (2.29); the part (ii) of the theorem corresponds to the cases where Eq. (2.28) has only one nontrivial solution of the form (2.29).

2.3 Discriminant and Linearly Independent Solutions

From the last section, we see that the linearly independent solutions of Eq. (2.28) are determined by the roots ρ of the characteristic equation (2.34), which are determined by a real number

$$D = \eta_1(a) + p(a)\eta_2'(a). \quad (2.42)$$

This real number D determines the forms of roots ρ of the characteristic equation (2.34) and thus the forms of two linearly independent solutions of (2.28). This real number is called the discriminant of Eq. (2.28).⁴

There can be five different cases.

A. $-2 < D < 2$.

In this case, the two roots ρ_1 and ρ_2 of the characteristic equation (2.34) are two distinct nonreal numbers. They are complex conjugates of each other and have moduli equal to unity. h_i in (2.35) can be chosen as imaginary numbers $\pm ik$, where $0 < k < \pi/a$. Equation (2.28) has two linearly independent solutions,

$$\begin{aligned} y_1(x) &= e^{ikx} p_1(x), \\ y_2(x) &= e^{-ikx} p_2(x). \end{aligned} \quad (2.43)$$

Here $p_i(x)$, $i = 1, 2$ are periodic functions with period a . k in (2.43) is related to the discriminant D by

$$\cos ka = \frac{1}{2}D. \quad (2.44)$$

B. $D = 2$.

There are two possible subcases:

B.1. $\eta_2(a)$ and $p(a)\eta_1'(a)$ are not both zero.

In this subcase, not all elements of the matrix $(A - I\rho)$ (I is the unit matrix) are zero. The matrix $A = (A_{ij})$ has only one independent eigenvector.

Equation (2.28) can have two linearly independent solutions as

$$\begin{aligned} y_1(x) &= p_1(x), \\ y_2(x) &= x p_1(x) + p_2(x). \end{aligned} \quad (2.45)$$

Here $p_i(x)$, $i = 1, 2$ are periodic functions with period a .

B.2. $\eta_2(a) = p(a)\eta_1'(a) = 0$.

⁴It is easy to see that as the trace of matrix A_{ij} in (2.32), the discriminant D of Eq. (2.28) does not depend on how the origin 0 is chosen in (2.30).

In this subcase, $\eta_1(a) = p(a)\eta_2'(a) = 1$. All elements of the matrix $(A - I\rho)$ are zero. The matrix $A = (A_{ij})$ has two linearly independent eigenvectors.

Equation (2.28) can have two linearly independent solutions:

$$\begin{aligned} y_1(x) &= p_1(x), \\ y_2(x) &= p_2(x). \end{aligned} \quad (2.46)$$

Here $p_i(x)$, $i = 1, 2$ are periodic functions with period a .

C. $D > 2$.

In this case, the roots ρ_1 and ρ_2 of the characteristic equation (2.34) are two distinct positive real numbers that are not equal to unity. h_i in (2.35) can be chosen as real numbers $\pm\beta$. Equation (2.28) can have two linearly independent solutions:

$$\begin{aligned} y_1(x) &= e^{\beta x} p_1(x), \\ y_2(x) &= e^{-\beta x} p_2(x), \end{aligned} \quad (2.47)$$

and $p_i(x)$, $i = 1, 2$ are periodic functions with period a . β in (2.47) is a positive real number related to the discriminant D by

$$\cosh \beta a = \frac{1}{2} D. \quad (2.48)$$

D. $D = -2$.

There are two possible subcases:

D.1. $\eta_2(a)$ and $p(a)\eta_1'(a)$ are not both zero.

Equation (2.28) can have two linearly independent solutions as

$$\begin{aligned} y_1(x) &= s_1(x), \\ y_2(x) &= x s_1(x) + s_2(x). \end{aligned} \quad (2.49)$$

Here $s_i(x)$, $i = 1, 2$ are semi-periodic functions with semi-period a : $s_i(x + a) = -s_i(x)$.

D.2. $\eta_2(a) = p(a)\eta_1'(a) = 0$.

Equation (2.28) can have two linearly independent solutions:

$$\begin{aligned} y_1(x) &= s_1(x), \\ y_2(x) &= s_2(x). \end{aligned} \quad (2.50)$$

Here $s_i(x)$, $i = 1, 2$ are semi-periodic functions with semi-period a .

E. $D < -2$.

In this case, the roots ρ_1 and ρ_2 of the characteristic equation (2.34) are two distinct negative real numbers that are not equal to -1 . h_i in (2.35) can be chosen as complex numbers $\pm(\beta + i\pi/a)$. Equation (2.28) can have two linearly independent solutions as

$$\begin{aligned} y_1(x) &= e^{\beta x} s_1(x), \\ y_2(x) &= e^{-\beta x} s_2(x), \end{aligned} \quad (2.51)$$

and $s_i(x)$, $i = 1, 2$ are semi-periodic functions with semi-period a . β in (2.51) is a positive real number related to the discriminant D by

$$\cosh \beta a = -\frac{1}{2}D. \quad (2.52)$$

2.4 The Spectral Theory

Now, we consider a periodic Sturm–Liouville equation (2.2),

$$[p(x)y'(x)]' + [\lambda w(x) - q(x)]y(x) = 0,$$

where $p(x) > 0$, $w(x) > 0$ and $p(x)$, $q(x)$, $w(x)$ are piecewise continuous real periodic functions with period a :

$$p(x+a) = p(x), \quad q(x+a) = q(x), \quad w(x+a) = w(x).$$

The normalized solutions $\eta_i(x, \lambda)$, $i = 1, 2$ of (2.2) are defined as

$$\eta_1(0, \lambda) = 1, \quad p(0)\eta_1'(0, \lambda) = 0; \quad \eta_2(0, \lambda) = 0, \quad p(0)\eta_2'(0, \lambda) = 1, \quad (2.53)$$

and the discriminant of (2.2) is⁵

$$D(\lambda) = \eta_1(a, \lambda) + p(a)\eta_2'(a, \lambda). \quad (2.54)$$

The two linearly independent solutions of (2.2) are determined by $D(\lambda)$ in (2.54). To understand the properties of solutions of Eq. (2.2), we need to know how $D(\lambda)$ changes as λ changes. For this purpose, we first give two definitions.

⁵For the Schrödinger equation (2.1) of a one-dimensional crystal, the discriminant is $D(\lambda) = \eta_1(a, \lambda) + \eta_2'(a, \lambda)$ since $p(x) = 1$.

2.4.1 Two Eigenvalue Problems

We consider the solutions of Eq. (2.2) under the conditions

$$y(a) = y(0), \quad p(a)y'(a) = p(0)y'(0). \quad (2.55)$$

The corresponding eigenvalues are denoted by λ_n and can be ordered according to

$$\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots,$$

and the eigenfunctions can be chosen as to be real-valued and denoted as $\zeta_n(x)$. $\zeta_n(x)$ can be further required to form an orthonormal set over $[0, a]$:

$$\int_0^a \zeta_m(x) \zeta_n(x) \, dx = \delta_{m,n}.$$

$\zeta_n(x)$ can be extended by (2.55) to the whole of $(-\infty, +\infty)$ as continuous and piecewise quasi-differentiable functions⁶ with period a . Thus, λ_n are the values of λ for which Eq. (2.2) has a nontrivial solution with period a .

Similarly, we can also consider the solutions of (2.2) under the conditions

$$y(a) = -y(0), \quad p(a)y'(a) = -p(0)y'(0). \quad (2.56)$$

The corresponding eigenvalues are denoted by μ_n and can be ordered according to

$$\mu_0 \leq \mu_1 \leq \mu_2 \leq \cdots.$$

The corresponding eigenfunctions can be chosen to be real-valued and denoted as $\xi_n(x)$. $\xi_n(x)$ can be further required to form an orthonormal set over $[0, a]$:

$$\int_0^a \xi_m(x) \xi_n(x) \, dx = \delta_{m,n}.$$

$\xi_n(x)$ can be extended by (2.56) to the whole of $(-\infty, +\infty)$ as continuous and piecewise quasi-differentiable functions with semi-period a . Thus, μ_n are the values of λ for which Eq. (2.2) has a nontrivial solution with semi-period a .

⁶Each $\zeta_n(x)$ and its quasi-derivative $p(x)\zeta'_n(x)$ are continuous.

2.4.2 The Function $D(\lambda)$

The following theorem describes how $D(\lambda)$ changes as λ changes regarding the eigenvalues λ_n and μ_n defined by the two eigenvalue problems (2.55) and (2.56)⁷:

Theorem 2.5 (Extended Theorem 2.3.1 in [7])

(i) The numbers λ_n and μ_n occur in the order

$$\lambda_0 < \mu_0 \leq \mu_1 < \lambda_1 \leq \lambda_2 < \mu_2 \leq \mu_3 < \lambda_3 \leq \lambda_4 < \cdots. \quad (2.57)$$

As λ increases from $-\infty$ to $+\infty$, $D(\lambda)$ changes as the following, where $m = 0, 1, 2, \dots$:

- (ii) In the interval $(-\infty, \lambda_0)$, $D(\lambda) > 2$.
- (iii) In the intervals $[\lambda_{2m}, \mu_{2m}]$, $D(\lambda)$ decreases from $+2$ to -2 .
- (iv) In the intervals (μ_{2m}, μ_{2m+1}) , $D(\lambda) < -2$.
- (v) In the intervals $[\mu_{2m+1}, \lambda_{2m+1}]$, $D(\lambda)$ increases from -2 to $+2$.
- (vi) In the intervals $(\lambda_{2m+1}, \lambda_{2m+2})$, $D(\lambda) > 2$.

Proof This theorem can be proven in several steps.

(1) There exists a Λ such that for all $\lambda < \Lambda$, $D(\lambda) > 2$.

We can choose a Λ so that for all x in $(-\infty, +\infty)$,

$$[q(x) - \Lambda w(x)] > 0$$

is true.

Suppose $y(x)$ is any nontrivial solution of (2.2) for which $y(0) \geq 0$ and $p(0)y'(0) \geq 0$; then there is always an interval $(0, \Delta)$ in which $y(x) > 0$.

For all $\lambda \leq \Lambda$, in any interval $(0, X)$ for which $y(x) > 0$ we have

$$[p(x)y'(x)]' = [q(x) - \lambda w(x)]y(x) > 0;$$

thus, in the interval $(0, X)$, we have $p(x)y'(x) > 0$ and $y(x)$ is increasing in $(0, X)$. Therefore, $y(x)$ has no zero $x = x_0$ in $(0, +\infty)$ and both $y(x)$ and $p(x)y'(x)$ are increasing.

Since both $\eta_1(x, \lambda)$ and $\eta_2(x, \lambda)$ defined in (2.53) satisfy

$$\eta_1(0, \lambda) \geq 0, \quad p(0)\eta_1'(0, \lambda) \geq 0; \quad \eta_2(0, \lambda) \geq 0, \quad p(0)\eta_2'(0, \lambda) \geq 0,$$

⁷Relevant contents can also be found in Theorem 12.7 in [8], Theorem 4.8.1 in [9], Theorems 1.6.1 and 2.4.2 in [10], Theorem 5.33 in [11], and also in References [14, 15].

both $\eta_1(x, \lambda)$, $p(x)\eta_1'(x, \lambda)$ and $\eta_2(x, \lambda)$, $p(x)\eta_2'(x, \lambda)$ are increasing in $(0, +\infty)$ for all $\lambda \leq \Lambda$. In particular, we have

$$\eta_1(a, \lambda) > \eta_1(0, \lambda) = 1; \quad p(a)\eta_2'(a, \lambda) > p(0)\eta_2'(0, \lambda) = 1.$$

Thus, for all $\lambda \leq \Lambda$, we have $D(\lambda) > 2$.

However, as λ increases, $[p(x)y'(x)]'/y(x) = [q(x) - \lambda w(x)]$ will become negative and, consequently, $D(\lambda)$ will decrease as λ increases.

(2) For any λ such that $|D(\lambda)| < 2$, $D'(\lambda)$ is not zero.

Differentiating (2.2) with respect to λ with $y(x) = \eta_1(x, \lambda)$, we obtain

$$\frac{d}{dx} \left[p(x) \frac{d}{dx} \left[\frac{\partial \eta_1(x, \lambda)}{\partial \lambda} \right] \right] + [\lambda w(x) - q(x)] \frac{\partial \eta_1(x, \lambda)}{\partial \lambda} = -w(x)\eta_1(x, \lambda). \quad (2.58)$$

By using the variation of parameters formula (2.16), we solve $\partial \eta_1(x, \lambda)/\partial \lambda$ from (2.58) with the initial condition $\frac{\partial \eta_1(0, \lambda)}{\partial \lambda} = 0$ (from (2.53)) and obtain that

$$\frac{\partial \eta_1(x, \lambda)}{\partial \lambda} = \int_0^x [\eta_1(x, \lambda)\eta_2(t, \lambda) - \eta_2(x, \lambda)\eta_1(t, \lambda)]w(t)\eta_1(t, \lambda) dt, \quad (2.59)$$

by noting that $W[\eta_1(t, \lambda), \eta_2(t, \lambda)] = 1$.

Similarly by differentiating (2.2) with respect to λ with $y(x) = \eta_2(x, \lambda)$, we obtain

$$\frac{d}{dx} \left[p(x) \frac{d}{dx} \left[\frac{\partial \eta_2(x, \lambda)}{\partial \lambda} \right] \right] + [\lambda w(x) - q(x)] \frac{\partial \eta_2(x, \lambda)}{\partial \lambda} = -w(x)\eta_2(x, \lambda). \quad (2.60)$$

By using the variation of parameters formula (2.16), we solve $\frac{\partial \eta_2(x, \lambda)}{\partial \lambda}$ from (2.60) with the initial condition that $\frac{\partial \eta_2(0, \lambda)}{\partial \lambda} = 0$ (from (2.53)) and obtain that

$$\frac{\partial \eta_2(x, \lambda)}{\partial \lambda} = \int_0^x [\eta_1(x, \lambda)\eta_2(t, \lambda) - \eta_2(x, \lambda)\eta_1(t, \lambda)]w(t)\eta_2(t, \lambda) dt, \quad (2.61)$$

By using the variation of parameters formula (2.17), we solve $\frac{\partial p(x)\eta_2'(x, \lambda)}{\partial \lambda}$ from (2.60) with the initial condition $\frac{\partial p(0)\eta_2'(0, \lambda)}{\partial \lambda} = 0$ (from (2.53)) and obtain that

$$\begin{aligned} \frac{\partial p(x)\eta_2'(x, \lambda)}{\partial \lambda} &= \int_0^x [p(x)\eta_1'(x, \lambda)\eta_2(t, \lambda) \\ &\quad - p(x)\eta_2'(x, \lambda)\eta_1(t, \lambda)]w(t)\eta_2(t, \lambda) dt. \end{aligned} \quad (2.62)$$

Combining (2.62) and (2.59) and putting $x = a$, we obtain that

$$D'(\lambda) = \int_0^a [p\eta_1'\eta_2^2(t, \lambda) + (\eta_1 - p\eta_2')\eta_1(t, \lambda)\eta_2(t, \lambda) - \eta_2\eta_1^2(t, \lambda)] w(t) dt, \quad (2.63)$$

and thus

$$\begin{aligned} 4\eta_2 D'(\lambda) = & - \int_0^a [2\eta_2\eta_1(t, \lambda) - (\eta_1 - p\eta_2')\eta_2(t, \lambda)]^2 w(t) dt \\ & - [4 - D^2(\lambda)] \int_0^a \eta_2^2(t, \lambda) w(t) dt. \end{aligned} \quad (2.64)$$

In (2.63) and (2.64) we have written $\eta_i = \eta_i(a, \lambda)$ and $p\eta_i' = p(a)\eta_i'(a, \lambda)$ for brevity.

If $|D(\lambda)| < 2$, from (2.64) we have $\eta_2 D'(\lambda) < 0$; thus, $D'(\lambda) \neq 0$. Therefore, only in the regions of λ in which $|D(\lambda)| \geq 2$ can $D'(\lambda) = 0$ be true.

(3) At a zero λ_n of $D(\lambda) - 2 = 0$, if and only if

$$\eta_2(a, \lambda_n) = p(a)\eta_1'(a, \lambda_n) = 0, \quad (2.65)$$

$D'(\lambda_n) = 0$ is true. Further, if $D'(\lambda_n) = 0$, then $D''(\lambda_n) < 0$.

(3a) Equation (2.65) gives that $\eta_1(a, \lambda_n)p(a)\eta_2'(a, \lambda_n) = 1$ by $W[\eta_1(a, \lambda_n), \eta_2(a, \lambda_n)] = 1$. It further gives

$$\eta_1(a, \lambda_n) = p(a)\eta_2'(a, \lambda_n) = 1 \quad (2.66)$$

by $D(\lambda_n) = 2$. Equation (2.63) then gives that $D'(\lambda_n) = 0$. According to **B.2** in Sect. 2.3, this case corresponds to the case that $D(\lambda) - 2$ has a double zero at $\lambda = \lambda_n$.

(3b) On the other hand, if $D'(\lambda_n) = 0$, then the first integrand on the right of (2.64) must be identically zero since $D(\lambda_n) = 2$. Since $\eta_1(t, \lambda)$ and $\eta_2(t, \lambda)$ are linearly independent, $\eta_2(a, \lambda_n) = 0$ and $\eta_1(a, \lambda_n) = p(a)\eta_2'(a, \lambda_n)$ must be true. From (2.63), we obtain that $p(a)\eta_1'(a, \lambda_n) = 0$.

(3c) To further prove $D''(\lambda_n) < 0$ when $D'(\lambda_n) = 0$, we differentiate (2.58) with respect to λ and obtain

$$\begin{aligned} \frac{d}{dx} \left[\frac{\partial^2 p(x)\eta_1'(x, \lambda)}{\partial \lambda^2} \right] + [\lambda w(x) - q(x)] \frac{\partial^2 \eta_1(x, \lambda)}{\partial \lambda^2} \\ = -2w(x) \frac{\partial}{\partial \lambda} \eta_1(x, \lambda). \end{aligned} \quad (2.67)$$

Applying the variation of parameters formula (2.16) to solve $\partial^2 \eta_1(x, \lambda) / \partial \lambda^2$ from (2.67) with the initial condition $\frac{\partial^2 \eta_1(0, \lambda)}{\partial \lambda^2} = 0$ (from (2.53)) we obtain that

$$\frac{\partial^2 \eta_1(x, \lambda)}{\partial \lambda^2} = 2 \int_0^x [\eta_1(x, \lambda) \eta_2(t, \lambda) - \eta_2(x, \lambda) \eta_1(t, \lambda)] w(t) \frac{\partial}{\partial \lambda} \eta_1(t, \lambda) dt \quad (2.68)$$

by noting that $W[\eta_1(t, \lambda), \eta_2(t, \lambda)] = 1$.

We differentiate (2.60) with respect to λ and obtain

$$\begin{aligned} \frac{d}{dx} \left[\frac{\partial^2 p(x) \eta'_2(x, \lambda)}{\partial \lambda^2} \right] + [\lambda w(x) - q(x)] \frac{\partial^2 \eta_2(x, \lambda)}{\partial \lambda^2} \\ = -2w(x) \frac{\partial}{\partial \lambda} \eta_2(x, \lambda). \end{aligned} \quad (2.69)$$

By applying the variation of parameters formula (2.17) again to solve $\frac{\partial^2 p(x) \eta'_2(x, \lambda)}{\partial \lambda^2}$ from (2.69) with the initial condition $\frac{\partial^2 p(0) \eta'_2(0, \lambda)}{\partial \lambda^2} = 0$ (from (2.53)), we obtain

$$\begin{aligned} \frac{\partial^2 p(x) \eta'_2(x, \lambda)}{\partial \lambda^2} = 2 \int_0^x [p(x) \eta'_1(x, \lambda) \eta_2(t, \lambda) \\ - p(x) \eta'_2(x, \lambda) \eta_1(t, \lambda)] w(t) \frac{\partial}{\partial \lambda} \eta_2(t, \lambda) dt. \end{aligned} \quad (2.70)$$

Therefore, by combining (2.68) and (2.70) and noting that when $D'(\lambda_n) = 0$, (2.65) and (2.66) are true, we obtain that

$$\begin{aligned} D''(\lambda_n) = 2 \int_0^a \left[\eta_2(t, \lambda_n) \frac{\partial \eta_1(t, \lambda)}{\partial \lambda} - \eta_1(t, \lambda_n) \frac{\partial \eta_2(t, \lambda)}{\partial \lambda} \right]_{\lambda=\lambda_n} w(t) dt \\ = -2 \int_0^a w(t) dt \int_0^t [\eta_1(t, \lambda_n) \eta_2(\tau, \lambda_n) - \eta_2(t, \lambda_n) \eta_1(\tau, \lambda_n)]^2 w(\tau) d\tau. \end{aligned} \quad (2.71)$$

Equations (2.59) and (2.61) were used in obtaining the second equality. The right side of (2.71) is less than zero since the integrand in the double integral is positive.

- (4) It can be similarly proven that there is a corresponding result to (3) for the zeros μ_n of $D(\lambda) + 2$: If and only if

$$\eta_2(a, \mu_n) = p(a) \eta'_1(a, \mu_n) = 0, \quad (2.72)$$

$D'(\mu_n) = 0$ is true. Further, $D''(\mu_n) > 0$ when $D'(\mu_n) = 0$. This case corresponds to that $D(\lambda) + 2$ has a double zero at $\lambda = \mu_n$.

- (5) Therefore, except cases in (3) or (4), only in the regions of λ in which $D(\lambda) < -2$ or $D(\lambda) > 2$ can $D'(\lambda) = 0$ be true. The $D(\lambda)$ – λ curve can change direction only in such regions.

- (6) From above results of (1)–(5), we can discuss the behavior of $D(\lambda)$ as λ increases from $-\infty$ to $+\infty$. When λ is a large negative real number, $D(\lambda) > 2$ by (1). As λ increases from $-\infty$, we have $D(\lambda) > 2$ until λ reaches the first zero λ_0 of $D(\lambda) - 2$. Since λ_0 is not a maximum of $D(\lambda)$, $D''(\lambda_0) \neq 0$; thus, $D'(\lambda_0) \neq 0$ by (3). The $D(\lambda)$ – λ curve intersects the line $D = 2$ at $\lambda = \lambda_0$; thus, to the immediate right of λ_0 , we have $D(\lambda) < 2$. Then by (2), as λ increases from λ_0 , $D(\lambda)$ decreases until λ reaches the first zero μ_0 of $D(\lambda) + 2$. Thus, in the interval $(-\infty, \lambda_0)$, $D(\lambda) > 2$, and in the interval $[\lambda_0, \mu_0]$, $D(\lambda)$ decreases from $+2$ to -2 .

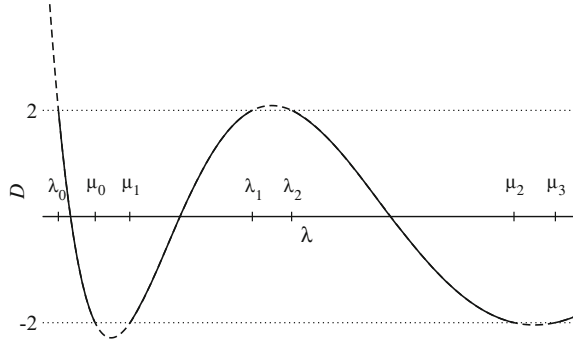
In general, μ_0 will be a simple zero of $D(\lambda) + 2$, so the $D(\lambda)$ – λ curve intersects the line $D = -2$ at $\lambda = \mu_0$, and to the immediate right of μ_0 $D(\lambda) < -2$. As λ increase, $D(\lambda) < -2$ will remain true until λ reaches the second zero μ_1 of $D(\lambda) + 2$, since, by (5), the $D(\lambda)$ – λ curve can change direction in a region where $D(\lambda) < -2$. Since μ_1 is not a minimum of $D(\lambda)$, μ_1 is a simple zero of $D(\lambda) + 2$. The $D(\lambda)$ – λ curve intersects the line $D = -2$ again at $\lambda = \mu_1$. To the immediate right of μ_1 , we have $D(\lambda) > -2$; then according to (2), as λ increases from μ_1 , $D(\lambda)$ increases until λ reaches the next zero λ_1 of $D(\lambda) - 2$. Thus, in the interval (μ_0, μ_1) , $D(\lambda) < -2$, and in the interval $[\mu_1, \lambda_1]$, $D(\lambda)$ increases from -2 to $+2$.

In general, λ_1 will be a simple zero of $D(\lambda) - 2$, so the $D(\lambda)$ – λ curve intersects the line $D = 2$ at $\lambda = \lambda_1$, and to the immediate right of λ_1 , we have $D(\lambda) > 2$. As λ increase, $D(\lambda) > 2$ will remain to be true until λ reaches the third zero λ_2 of $D(\lambda) - 2$, since, by (5), the $D(\lambda)$ – λ curve can change direction in a region where $D(\lambda) > 2$. The argument we used starting from $\lambda = \lambda_0$ can be repeated starting from $\lambda = \lambda_2$, and can be repeated again and again as λ increases to $+\infty$.

Now all parts of the theorem have been proven except when $D(\lambda) \pm 2$ has double zeros. If, for example, $D(\lambda) - 2$ has a double zero at a specific $\lambda = \lambda_{2m+1}$ (i.e., $\lambda_{2m+2} = \lambda_{2m+1}$). From **B.2** in Sect. 2.3, this can happen only when $\eta_2(a, \lambda_{2m+1}) = \eta'_1(a, \lambda_{2m+1}) = 0$; therefore, $D'(\lambda_{2m+1}) = 0$ and $D''(\lambda_{2m+1}) < 0$ is true by (3). Consequently, to the immediate right of $\lambda = \lambda_{2m+1} = \lambda_{2m+2}$ we have $D(\lambda) < 2$. In such a case, the $D(\lambda)$ – λ curve merely *touches* the line $D = 2$ at $\lambda = \lambda_{2m+1} = \lambda_{2m+2}$ rather than intersects the line $D = 2$ twice at $\lambda = \lambda_{2m+1}$ and at $\lambda = \lambda_{2m+2}$. The previous analysis of $D(\lambda)$ can repeatedly continue again. The cases where $D(\lambda) + 2$ has double zeros can be similarly analyzed by using (4). $\square$

Therefore, in general, when λ increases from $-\infty$ to $+\infty$, the discriminant $D(\lambda)$ of Eq. (2.2) as defined in (2.54) changes, as shown typically in Fig. 2.1 [1, 5, 8, 9, 11, 14, 15]. The permitted eigenvalue bands of (2.2) are in the ranges of λ for which $-2 \leq D(\lambda) \leq 2$ (solid lines). No eigenmode exists in the ranges of λ for which $D(\lambda) > 2$ or $D(\lambda) < -2$ (dashed lines).

Fig. 2.1 A typical $D(\lambda)$ – λ curve. The permitted eigenvalue bands of Eq. (2.2) are in the ranges of λ for which $-2 \leq D(\lambda) \leq 2$ (solid lines). No eigenvalue exists in the ranges of λ for which $D(\lambda) > 2$ or $D(\lambda) < -2$ (dashed lines)



2.5 Band Structure of Eigenvalues

The permitted band structure of a periodic Sturm–Liouville equation (2.2) has some especially simple and general properties [1, 2, 5, 8, 9, 11, 15].

Only for λ in the range $[\lambda_0, +\infty)$ can eigenmodes exist in a crystal with translational invariance. There are five different cases:

- A. λ in an interval (λ_{2m}, μ_{2m}) or $(\mu_{2m+1}, \lambda_{2m+1})$, where $m = 0, 1, 2, \dots, -2 < D(\lambda) < 2$. Two linearly independent solutions can be chosen as

$$\begin{aligned} y_1(x, \lambda) &= e^{ik(\lambda)x} p_1(x, \lambda), \\ y_2(x, \lambda) &= e^{-ik(\lambda)x} p_2(x, \lambda), \end{aligned} \quad (2.73)$$

and $p_i(x, \lambda)$ are periodic functions depending on λ . They are the two well-known one-dimensional Bloch states $\phi_n(k, x)$ and $\phi_n(-k, x)$ with wave vector k or $-k$. Their corresponding energies can be written as $\varepsilon_n(k)$ and $\varepsilon_n(-k)$, with $\varepsilon_n(-k) = \varepsilon_n(k)$. Therefore, the intervals (λ_{2m}, μ_{2m}) and $(\mu_{2m+1}, \lambda_{2m+1})$ correspond to the inside of permitted energy bands of Eq. (2.2).

In each permitted band, where $0 < k(\lambda) < \pi/a$ is determined by (2.51):

$$\cos ka = \frac{1}{2} D(\lambda). \quad (2.74)$$

It has been shown that $\varepsilon_n(k)$ is always a monotonic function of k inside each permitted energy band [2]. Here we give another simple proof.

Suppose this is not true, there is an energy band in which $\varepsilon_n(k)$ is not a monotonic function of k . Then there must be at least one λ inside the energy band for which there are at least two distinct k_1 and k_2 in $(0, \frac{\pi}{a})$ for which $\varepsilon_n(k_1) = \varepsilon_n(k_2) = \lambda$. That means (2.2) has at least four linearly independent solutions for such a λ : two with $k = k_1$ and two with $k = k_2$ in (2.73). This is contradictory to that a periodic Sturm–Liouville equation such as (2.2) can only have two linearly independent solutions.

B. At $\lambda = \lambda_n$, $D(\lambda) = 2$.

B.1. In most cases, λ_n is a simple zero of $D(\lambda) - 2$. Equation (2.2) has two linearly independent solutions with forms as

$$\begin{aligned} y_1(x, \lambda) &= p_1(x, \lambda_n), \\ y_2(x, \lambda) &= x p_1(x, \lambda_n) + p_2(x, \lambda_n), \end{aligned} \quad (2.75)$$

and $p_i(x, \lambda)$ are periodic functions depending on λ . In crystals with translational invariance, only the periodic function solution y_1 is permitted. λ_n corresponding to a band-edge eigenvalue at $k = 0$, $\varepsilon_n(0)$. For $n > 0$, Case **B.1** corresponds to the cases where there is a nonzero band gap between $\varepsilon_{2m+1}(0)$ and $\varepsilon_{2m+2}(0)$.

B.2. In some special cases, λ_n ($n > 0$) is a double zero of $D(\lambda) - 2$: $\lambda_{2m+1} = \lambda_{2m+2}$. Equation (2.2) has two linearly independent solutions with forms as

$$\begin{aligned} y_1(x, \lambda) &= p_1(x, \lambda_{2m+1}), \\ y_2(x, \lambda) &= p_2(x, \lambda_{2m+1}). \end{aligned} \quad (2.76)$$

Their corresponding eigenvalues are $\varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0) = \lambda_{2m+1}$, and there is a zero band gap between $\varepsilon_{2m+1}(0)$ and $\varepsilon_{2m+2}(0)$.

C. λ in an interval $(\lambda_{2m+1}, \lambda_{2m+2})$, $D(\lambda) > 2$. In such a case λ is inside a band gap at $k = 0$ of Eq. (2.2). The two linearly independent solutions of Eq. (2.2) can be written as

$$\begin{aligned} y_1(x, \lambda) &= e^{\beta(\lambda)x} p_1(x, \lambda), \\ y_2(x, \lambda) &= e^{-\beta(\lambda)x} p_2(x, \lambda). \end{aligned} \quad (2.77)$$

Here, $\beta(\lambda) > 0$ is determined by (2.48):

$$\cosh \beta a = \frac{1}{2} D(\lambda) \quad (2.78)$$

and $p_i(x, \lambda)$, $i = 1, 2$ are periodic functions. These forbidden solutions in crystals with translational invariance might play a significant role in the physics of one-dimensional semi-infinite crystals and crystals of finite length.

D. At $\lambda = \mu_n$, $D(\lambda) = -2$.

D.1. In most cases, μ_n is a simple zero of $D(\lambda) + 2$. Equation (2.2) has two linearly independent solutions with forms as

$$\begin{aligned} y_1(x, \lambda) &= s_1(x, \mu_n), \\ y_2(x, \lambda) &= x s_1(x, \mu_n) + s_2(x, \mu_n), \end{aligned} \quad (2.79)$$

and $s_i(x, \lambda)$ are semi-periodic functions depending on λ . In crystals of infinite size, only the semi-periodic function solution y_1 is permitted. μ_n

corresponds to a band-edge eigenvalue $\varepsilon_n(\frac{\pi}{a})$ at $k = \frac{\pi}{a}$. Case **D.1** corresponds to the cases that there is a nonzero band gap between $\varepsilon_{2m}(\frac{\pi}{a})$ and $\varepsilon_{2m+1}(\frac{\pi}{a})$.

- D.2.** In some special cases where μ_n is a double zero of $D(\lambda) + 2$: $\mu_{2m} = \mu_{2m+1}$. Equation (2.2) has two linearly independent solutions as

$$\begin{aligned} y_1(x, \lambda) &= s_1(x, \mu_{2m}), \\ y_2(x, \lambda) &= s_2(x, \mu_{2m}). \end{aligned} \quad (2.80)$$

Their corresponding eigenvalues are $\varepsilon_{2m}(\frac{\pi}{a}) = \varepsilon_{2m+1}(\frac{\pi}{a}) = \mu_{2m}$, and there is a zero band gap between $\varepsilon_{2m}(\frac{\pi}{a})$ and $\varepsilon_{2m+1}(\frac{\pi}{a})$.

- E.** λ in an interval (μ_{2m}, μ_{2m+1}) , $D(\lambda) < -2$. In such a case λ is inside a band gap at $k = \pi/a$ of Eq. (2.2). The two linearly independent solutions can be written as

$$\begin{aligned} y_1(x, \lambda) &= e^{\beta(\lambda)x} s_1(x, \lambda), \\ y_2(x, \lambda) &= e^{-\beta(\lambda)x} s_2(x, \lambda). \end{aligned} \quad (2.81)$$

Here, $\beta(\lambda) > 0$ is determined by (2.52):

$$\cosh \beta a = -\frac{1}{2} D(\lambda), \quad (2.82)$$

and $s_i(x, \lambda)$ are semi-periodic functions. These forbidden solutions in crystals with translational invariance might play a significant role in the physics of one-dimensional semi-infinite crystals and crystals of finite length.

Therefore, in above cases **A**, **B**, and **D**, the permitted eigen solutions can exist as solutions of the periodic Sturm–Liouville equation (2.2). By combining our discussions in these three cases, we see that in the permitted band $\varepsilon_n(k)$ and permitted eigen solutions $\phi_n(k, x)$, the wave vector k in $\varepsilon_n(k)$ and $\phi_n(k, x)$ is limited in the Brillouin zone,

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a}, \quad (2.83)$$

and that λ_n , μ_n , and $\zeta_n(x)$, $\xi_n(x)$ defined in Sect. 2.4.1 are the band-edge energies and band-edge wave functions:

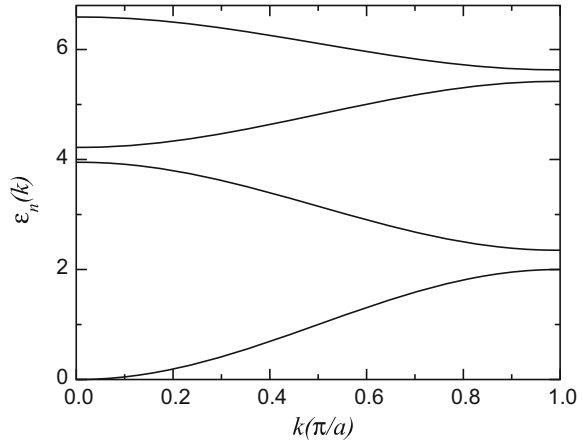
$$\lambda_n = \varepsilon_n(0), \quad \zeta_n(x) = \phi_n(0, x), \quad (2.84)$$

and

$$\mu_n = \varepsilon_n\left(\frac{\pi}{a}\right), \quad \xi_n(x) = \phi_n\left(\frac{\pi}{a}, x\right). \quad (2.85)$$

Therefore, (2.57) can be rewritten as

Fig. 2.2 A typical permitted band structure determined by (2.74) as solutions of a periodic Sturm–Liouville equation (2.2)



$$\begin{aligned} \varepsilon_0(0) < \varepsilon_0\left(\frac{\pi}{a}\right) &\leq \varepsilon_1\left(\frac{\pi}{a}\right) < \varepsilon_1(0) \leq \varepsilon_2(0) \\ &< \varepsilon_2\left(\frac{\pi}{a}\right) \leq \varepsilon_3\left(\frac{\pi}{a}\right) < \varepsilon_3(0) \leq \varepsilon_4(0) < \cdots \end{aligned} \quad (2.86)$$

A typical permitted band structure determined by (2.74) as solutions of a Sturm–Liouville equation Eq. (2.2) is shown in Fig. 2.2.

From the figure, we can see and understand that the permitted band structure as solutions of a periodic Sturm–Liouville equation has some very simple and general properties, see, for example [2, 15]:

1. In each permitted band, $\varepsilon_n(-k) = \varepsilon_n(k)$ and $\varepsilon_n(k)$ ($k > 0$) is a monotonic function of k .
2. The minimum and the maximum of each permitted band are always located either at $k = 0$ or $k = \frac{\pi}{a}$.
3. There is no permitted band crossing or permitted band overlap.
4. Each permitted band and each band gap occur alternatively.
5. The band gaps are between $\varepsilon_{2m}(\frac{\pi}{a})$ and $\varepsilon_{2m+1}(\frac{\pi}{a})$ or between $\varepsilon_{2m+1}(0)$ and $\varepsilon_{2m+2}(0)$.

A crystal described by a one-dimensional Schrödinger equation (2.1) can also be considered as a one-dimensional free space—a crystal with a zero potential (empty lattice)—reconstructed by a perturbation of the crystal potential $v(x)$. The crystal with an empty lattice has a free-electron-like energy spectrum $\lambda(k) = ck^2$ (c is a proportional constant) without any band gap in an extended Brillouin zone scheme. In the reduced Brillouin zone scheme, the energy spectrum $\lambda(k) = ck^2$ is folded at the Brillouin zone boundary, *continuously and simply touches at $k = 0$ and $k = \frac{\pi}{a}$ with no band crossings or band overlaps*. Finite Fourier components of the crystal potential $v(x)$ —no matter how small it is—*will open band gaps at $k = 0$ or $k = \frac{\pi}{a}$, with no band crossing or band overlap*, see, for example [2, 15]. Such a picture would be helpful for future understanding of the significant differences between

a one-dimensional crystal and a multi-dimensional crystal that we will discuss in Chap. 5.⁸

The theory of periodic Sturm–Liouville equations discussed so far has provided a general understanding of the permitted band structure of a one-dimensional crystal with translational invariance. In the next two chapters and in the appendices, we will treat states/modes in one-dimensional semi-infinite crystals or ideal finite crystals, in which the translational invariance is truncated. We will meet the complex band structure of periodic Sturm–Liouville equations. Similarly, the complex band structure of an Eq. (2.2) is also completely and analytically obtained from the discriminant $D(\lambda)$ (2.54) of the equation, as shown in Eqs. (2.78) and (2.82). The following several theorems on the zeros of solutions of Eq. (2.2) play a significant role in helping us to understand the states/modes in those systems.⁹

2.6 Zeros of Solutions

Now, we consider the solutions of (2.2) under the condition

$$y(a) = y(0) = 0. \quad (2.87)$$

The eigenvalues are denoted by Λ_n , and the corresponding eigenfunctions are denoted by $\Psi_n(x)$.

For physics problems in one-dimensional phononic crystals or photonic crystals, correspondingly there is an eigenvalue problem of (2.2) under the condition

$$p(a)y'(a) = p(0)y'(0) = 0, \quad (2.88)$$

the eigenvalues can be denoted by ν_n , and the corresponding eigenfunction can be written as $\Phi_n(x)$.

Theorem 2.6 (Extended Theorem 3.1.1 in [7], Theorem 13.10 in [8]) *For $m = 0, 1, 2, \dots$, we have*

$$\varepsilon_{2m}\left(\frac{\pi}{a}\right) \leq \Lambda_{2m} \leq \varepsilon_{2m+1}\left(\frac{\pi}{a}\right); \quad \varepsilon_{2m+1}(0) \leq \Lambda_{2m+1} \leq \varepsilon_{2m+2}(0). \quad (2.89)$$

and

$$\varepsilon_{2m}\left(\frac{\pi}{a}\right) \leq \nu_{2m+1} \leq \varepsilon_{2m+1}\left(\frac{\pi}{a}\right); \quad \varepsilon_{2m+1}(0) \leq \nu_{2m+2} \leq \varepsilon_{2m+2}(0), \quad (2.90)$$

$$\nu_0 \leq \varepsilon_0(0).$$

⁸These properties of the permitted bands are closely related to that a periodic Sturm–Liouville equation can have no more than two independent solutions.

⁹Relevant contents can be found in [8, 10].

Proof Since $\Psi_n(x)$ is the eigenfunction corresponding to the n th eigenvalue under the condition (2.87), it has exactly n zeros in $(0, a)$. According to (2.53), $\eta_2(0, \lambda) = 0$; thus, each Λ_n is the solution of the equation

$$\eta_2(a, \Lambda_n) = 0, \quad (2.91)$$

and the corresponding eigenfunction

$$\Psi_n(x) = \eta_2(x, \Lambda_n).$$

Therefore, $\eta_2(x, \Lambda_n)$ has exactly n zeros in the interval $(0, a)$. According to (2.53), $p(0)\eta_2'(0, \Lambda_n) > 0$; thus,

$$p(a)\eta_2'(a, \Lambda_n) < 0 \ (n = \text{even}); \quad p(a)\eta_2'(a, \Lambda_n) > 0 \ (n = \text{odd}). \quad (2.92)$$

Since we have

$$\eta_1(a, \Lambda_n)p(a)\eta_2'(a, \Lambda_n) = 1,$$

from

$$\eta_1(a, \Lambda_n)p(a)\eta_2'(a, \Lambda_n) - p(a)\eta_1'(a, \Lambda_n)\eta_2(a, \Lambda_n) = 1$$

and (2.91), therefore

$$D(\Lambda_n) = \eta_1(a, \Lambda_n) + p(a)\eta_2'(a, \Lambda_n) = \frac{1}{p(a)\eta_2'(a, \Lambda_n)} + p(a)\eta_2'(a, \Lambda_n).$$

For $n = \text{even}$, we have

$$-D(\Lambda_{n=\text{even}}) = [|p(a)\eta_2'(a, \Lambda_{n=\text{even}})|^{-1/2} - |p(a)\eta_2'(a, \Lambda_{n=\text{even}})|^{1/2}]^2 + 2 \geq 2$$

and thus

$$D(\Lambda_{n=\text{even}}) \leq -2$$

from (2.92). Similarly

$$D(\Lambda_{n=\text{odd}}) \geq 2$$

from (2.92). Therefore, Λ_n and Λ_{n+1} are always in different band gaps, if we consider the special cases in which $\varepsilon_{2m}(\frac{\pi}{a}) = \varepsilon_{2m+1}(\frac{\pi}{a})$ or $\varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0)$ as a band gap with the gap size being zero.

Now, we consider two consecutive zeros $\varepsilon_{2m+1}(0)$ and $\varepsilon_{2m+2}(0)$ of $D(\lambda) - 2$, with either $D(\lambda) > 2$ between them, or $D'(\lambda) = 0$ thus $D(\lambda) - 2$ has a double zero at $\lambda = \varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0)$ (see Fig. 2.1).

In the special cases where $D'(\varepsilon_{2m+1}(0)) = 0$, $D(\lambda) - 2$ has repeated solutions $\varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0)$; then by (2.65) we always have $\eta_2(a, \varepsilon_{2m+1}(0)) = 0$ and thus (2.91) has one solution $\Lambda_n = \varepsilon_{2m+1}(0)$.

In most cases, $D(\lambda) > 2$ in $(\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0))$. According to (2.64), we have $\eta_2(a, \lambda)D'(\lambda) \leq 0$ at both $\varepsilon_{2m+1}(0)$ and $\varepsilon_{2m+2}(0)$. Since $D'(\varepsilon_{2m+1}(0)) > 0$ and $D'(\varepsilon_{2m+2}(0)) < 0$, we have $\eta_2(a, \varepsilon_{2m+1}(0)) \leq 0$ and $\eta_2(a, \varepsilon_{2m+2}(0)) \geq 0$. Thus, $\eta_2(a, \lambda)$ has at least one zero in $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$. Since Λ_n and Λ_{n+1} must be in different band gaps, there is no more than one Λ_n in $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$. Thus, there is one Λ_n ($n = \text{odd}$) in $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$.

The cases of two consecutive zeros $\varepsilon_{2m}(\frac{\pi}{a})$ and $\varepsilon_{2m+1}(\frac{\pi}{a})$ of $D(\lambda) + 2$ can be similarly considered; we will obtain the conclusion that there is one and only one Λ_n ($n = \text{even}$) in $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$.

Therefore, Λ_n starts occurring in $[\varepsilon_0(\frac{\pi}{a}), \varepsilon_1(\frac{\pi}{a})]$ and always occurs alternatively between $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ or $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$.

Equation (2.90) can be similarly proven. Each v_n is the solution of the equation

$$p(a)\eta'_1(a, v_n) = 0, \quad (2.93)$$

and

$$\Phi_n(x) = \eta_1(x, v_n),$$

where $\eta_i(x, \lambda)$ and $p(x)\eta'_i(x, \lambda)$ ($i = 1, 2$) are defined in (2.53). Therefore, $\eta_1(x, v_n)$ has exactly n zeros in the interval $[0, a)$ and exactly n zeros in the interval $(0, a)$. That means $\eta_1(a, v_0) > 0$ and $\eta_1(a, v_{n=\text{even}}) > 0$, $\eta_1(a, v_{n=\text{odd}}) < 0$.

We have

$$D(v_n) = \eta_1(a, v_n) + p(a)\eta'_2(a, v_n) = \eta_1(a, v_n) + [\eta_1(a, v_n)]^{-1},$$

since $\eta_1(a, v_n)p(a)\eta'_2(a, v_n) = 1$ by (2.93). Therefore if $n = \text{even}$, $\eta_1(a, v_n) > 0$, $D(v_n) > 2$; $n = \text{odd}$, $\eta_1(a, v_n) < 0$, $D(v_n) < -2$;

Therefore, v_n and v_{n+1} are always in different forbidden ranges, if we consider the special cases in which $\varepsilon_{2m}(\frac{\pi}{a}) = \varepsilon_{2m+1}(\frac{\pi}{a})$ or $\varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0)$ as a band gap with the gap size being zero. v_n starts to occur in $(-\infty, \varepsilon_0(0)]$ and then always occurs alternatively between $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ or $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$. $\square$

Theorem 2.7 (Extended Theorem 3.1.2 in [7], Theorem 2.5.1 in [10], Theorems 13.7 and 13.8 in [8]) *As solutions of Eq. (2.2),*

- (i) $\phi_0(0, x)$ has no zero in $[0, a]$.
- (ii) $\phi_{2m+1}(0, x)$ and $\phi_{2m+2}(0, x)$ have exactly $2m + 2$ zeros in $[0, a)$.
- (iii) $\phi_{2m}(\frac{\pi}{a}, x)$ and $\phi_{2m+1}(\frac{\pi}{a}, x)$ have exactly $2m + 1$ zeros in $[0, a)$.

Proof The Theorem can be proven by Theorem 2.2 and Eqs. (2.86) and (2.89).

(1) Since $\Psi_0(x)$ (i.e., $\eta_2(x, \Lambda_0)$) has no zero in $(0, a)$ and by (2.86) and (2.89) $\varepsilon_0(0) < \varepsilon_0(\frac{\pi}{a}) \leq \Lambda_0$, from Theorem 2.2, both $\phi_0(0, x)$ and $\phi_0(\frac{\pi}{a}, x)$ have no more than one zero in $[0, a)$. $\phi_0(0, x)$ as a periodic function can only have an even numbers of zeros in $[0, a)$, and $\phi_0(\frac{\pi}{a}, x)$ as a semi-periodic function can only have an odd number of zeros in $[0, a)$. Therefore, $\phi_0(0, x)$ must have no zero in $[0, a)$, and $\phi_0(\frac{\pi}{a}, x)$ must have exactly one zero in $[0, a)$.

(2) By (2.89) and (2.86) $\Lambda_{2m} \leq \varepsilon_{2m+1}(\frac{\pi}{a}) < \varepsilon_{2m+1}(0) \leq \Lambda_{2m+1}$, from Theorem 2.2 both $\phi_{2m+1}(\frac{\pi}{a}, x)$ and $\phi_{2m+1}(0, x)$ has at least $2m + 1$ but no more than $2m + 2$ zeros in $(0, a)$. $\phi_{2m+1}(\frac{\pi}{a}, x)$ as a semi-periodic function can only have an odd number of zeros in $[0, a)$, and $\phi_{2m+1}(0, x)$ as a periodic function can only have an even number of zeros in $[0, a)$. Thus $\phi_{2m+1}(\frac{\pi}{a}, x)$ must have exactly $2m + 1$ zeros in $[0, a)$ and $\phi_{2m+1}(0, x)$ must have exactly $2m + 2$ zeros in $[0, a)$.

(3) By (2.89) and (2.86) $\Lambda_{2m+1} \leq \varepsilon_{2m+2}(0) < \varepsilon_{2m+2}(\frac{\pi}{a}) \leq \Lambda_{2m+2}$, from Theorem 2.2 both $\phi_{2m+2}(0, x)$ and $\phi_{2m+2}(\frac{\pi}{a}, x)$ has at least $2m + 1$ but no more than $2m + 2$ zeros in $(0, a)$. $\phi_{2m+2}(0, x)$ as a periodic function can only have an even number of zeros in $[0, a)$, $\phi_{2m+2}(\frac{\pi}{a}, x)$ as a semi-periodic function can only have an odd number of zeros in $[0, a)$. Thus $\phi_{2m+2}(0, x)$ must have exactly $2m + 2$ zeros in $[0, a)$, $\phi_{2m+2}(\frac{\pi}{a}, x)$ must have exactly $2m + 1$ zeros in $[0, a)$.

The theorem is proven by combining items (1)–(3). $\square$

Now, we consider an eigenvalue problem of (2.2) in $[\tau, \tau + a]$ for a real number τ under the boundary condition

$$y(\tau) = y(\tau + a) = 0. \quad (2.94)$$

The corresponding eigenvalues can be written as $\Lambda_{\tau, n}$.

Correspondingly there is an eigenvalue problem of (2.2) under the condition

$$p(\tau + a)y'(\tau + a) = p(\tau)y'(\tau) = 0, \quad (2.95)$$

the eigenvalues can be denoted by $\nu_{\tau, n}$.

Theorem 2.8 (Extended Theorem 3.1.3 in [7], Theorem 13.10 in [8]) *As functions of τ , the ranges of $\Lambda_{\tau, 2m}$ are $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ and the ranges of $\Lambda_{\tau, 2m+1}$ are $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$;*

As functions of τ , the ranges of $\nu_{\tau, 2m+1}$ are $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ and the ranges of $\nu_{\tau, 2m+2}$ are $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$.

Proof Since $\varepsilon_n(0)$ are the values of λ for which the corresponding solutions $\phi_n(0, x)$ of (2.2) are periodic, and $\varepsilon_n(\frac{\pi}{a})$ are the values of λ for which the corresponding solutions $\phi_n(\frac{\pi}{a}, x)$ of (2.2) are semi-periodic, $\varepsilon_n(0)$ and $\varepsilon_n(\frac{\pi}{a})$ will remain unchanged if the basic interval in (2.55) and (2.56) is changed from $[0, a]$ to $[\tau, \tau + a]$. Consequently, the conclusions of Theorem 2.6 will remain unchanged if the basic interval is changed from $[0, a]$ to $[\tau, \tau + a]$. Therefore, from Theorem 2.6, we have

$$\varepsilon_{2m}\left(\frac{\pi}{a}\right) \leq \Lambda_{\tau, 2m} \leq \varepsilon_{2m+1}\left(\frac{\pi}{a}\right), \quad \varepsilon_{2m+1}(0) \leq \Lambda_{\tau, 2m+1} \leq \varepsilon_{2m+2}(0). \quad (2.96)$$

From part (iii) of Theorem 2.7, both $\phi_{2m}(\frac{\pi}{a}, x)$ and $\phi_{2m+1}(\frac{\pi}{a}, x)$ have exactly $2m + 1$ zeros in $[0, a)$. According to Theorem 2.2, the zeros of $\phi_{2m}(\frac{\pi}{a}, x)$ and $\phi_{2m+1}(\frac{\pi}{a}, x)$ are distributed alternatively. There is always one and only one zero of $\phi_{2m+1}(\frac{\pi}{a}, x)$ between two consecutive zeros of $\phi_{2m}(\frac{\pi}{a}, x)$, and there is always one and only one zero of $\phi_{2m}(\frac{\pi}{a}, x)$ between two consecutive zeros of $\phi_{2m+1}(\frac{\pi}{a}, x)$.

Suppose x_0 is any zero of $\phi_{2m}(\frac{\pi}{a}, x)$. Let $\tau = x_0$; then $\phi_{2m}(\frac{\pi}{a}, x)$ satisfies (2.94):

$$\phi_{2m}\left(\frac{\pi}{a}, \tau\right) = \phi_{2m}\left(\frac{\pi}{a}, \tau + a\right) = 0.$$

Again, from part (iii) of Theorem 2.7, $\phi_{2m}(\frac{\pi}{a}, x)$ has $2m$ zeros in the open interval $(x_0, x_0 + a)$; thus, $\phi_{2m}(\frac{\pi}{a}, x)$ is an eigenfunction of (2.2) under the boundary condition (2.94) corresponding to the eigenvalue $\Lambda_{x_0, 2m}$: $\Lambda_{x_0, 2m} = \varepsilon_{2m}(\frac{\pi}{a})$. Similarly, if x_1 is any zero of $\phi_{2m+1}(\frac{\pi}{a}, x)$, then $\phi_{2m+1}(\frac{\pi}{a}, x)$ is an eigenfunction of (2.2) under the boundary condition (2.94) corresponding to the eigenvalue $\Lambda_{x_1, 2m+1}$: $\Lambda_{x_1, 2m+1} = \varepsilon_{2m+1}(\frac{\pi}{a})$. Hence, as τ as the variable changes from $\tau = x_0$ to $\tau = x_1$, a zero of $\phi_{2m+1}(\frac{\pi}{a}, x)$ next to x_0 , as a function of τ , $\Lambda_{\tau, 2m}$ correspondingly and continuously changes from $\varepsilon_{2m}(\frac{\pi}{a})$ to $\varepsilon_{2m+1}(\frac{\pi}{a})$. Similarly, as τ as the variable changes from $\tau = x_1$ to $\tau = x_2$, the other zero of $\phi_{2m}(\frac{\pi}{a}, x)$ next to x_1 , as a function of τ , $\Lambda_{\tau, 2m}$ correspondingly and continuously changes back from $\varepsilon_{2m+1}(\frac{\pi}{a})$ to $\varepsilon_{2m}(\frac{\pi}{a})$. Therefore, as functions of τ , the ranges of $\Lambda_{\tau, 2m}$ are $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$.

Similarly, we can obtain that as functions of τ , the ranges of $\Lambda_{\tau, 2m+1}$ are $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$. The first part of the theorem is proven.

Similarly, we have

$$\varepsilon_{2m}\left(\frac{\pi}{a}\right) \leq \nu_{\tau, 2m+1} \leq \varepsilon_{2m+1}\left(\frac{\pi}{a}\right), \quad \varepsilon_{2m+1}(0) \leq \nu_{\tau, 2m+2} \leq \varepsilon_{2m+2}(0), \quad \nu_{\tau, 0} \leq \varepsilon_0(0). \quad (2.97)$$

If τ is not an isolated discontinuous point of $p(x)$, then the results for $\nu_{\tau, n}$ can be similarly obtained. Since, when $x = \tau$ is a turning point of the periodic or semi-periodic eigenfunction $\phi_n(k_g, x)$, $p(\tau + a)\phi'_n(k_g, \tau + a) = p(\tau)\phi'_n(k_g, \tau) = 0$ thus the equalities in Eq. (2.95) are attained.

In case if τ is an isolated discontinuous point of $p(x)$, then $\tau - \delta$ and $\tau + \delta$ —the real number δ can be as small as needed—are not an isolated discontinuous point of $p(x)$, hence the argument in the previous paragraph works.

Since $p(x)y'(x)$ is continuous at any isolated discontinuous point of $p(x)$, $\nu_{\tau, n}$ is also continuous at any isolated discontinuous point τ of $p(x)$. Thus the second part of the theorem is proven. $\square$

This theorem indicates that: *There is always one and only one eigenvalue $\Lambda_{\tau, n}$ of (2.2) under the boundary condition (2.94) in each gap $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ if $\varepsilon_{2m}(\frac{\pi}{a}) < \varepsilon_{2m+1}(\frac{\pi}{a})$ or $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$ if $\varepsilon_{2m+1}(0) < \varepsilon_{2m+2}(0)$. In some special cases when $\varepsilon_{2m}(\frac{\pi}{a}) = \varepsilon_{2m+1}(\frac{\pi}{a})$ or $\varepsilon_{2m+1}(0) = \varepsilon_{2m+2}(0)$, then we have $\Lambda_{\tau, 2m} = \varepsilon_{2m}(\frac{\pi}{a})$ or $\Lambda_{\tau, 2m+1} = \varepsilon_{2m+1}(0)$.*

A consequence of Theorem 2.8 is that in general a one-dimensional Bloch function $\phi_n(k, x)$ does not have a zero except $k = 0$ or $k = \frac{\pi}{a}$. Since if $\phi_n(k, x)$ has a zero at $x = x_0$, $\phi_n(k, x_0) = 0$, then we must have $\phi_n(k, x_0 + a) = 0$. According to Theorem 2.8, the corresponding eigenvalue $\Lambda_{\tau, n}$ must be in either $[\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a})]$ or $[\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0)]$. Since $(\varepsilon_{2m}(\frac{\pi}{a}), \varepsilon_{2m+1}(\frac{\pi}{a}))$ and $(\varepsilon_{2m+1}(0), \varepsilon_{2m+2}(0))$ are band gaps, only the Bloch functions at a band-edge $\phi_{n \neq 0}(0, x)$ or $\phi_n(\frac{\pi}{a}, x)$ may have a zero.

Theorem 2.9 (Extended Theorem 3.2.2 in [7], Theorem 2.5.2 in [10]) *Any nontrivial solution of (2.2) with $\lambda \leq \varepsilon_0(0)$ has at most one zero in $-\infty < x < +\infty$.*

Proof This theorem can be proven in two steps.

(1) From part (i) of Theorem 2.7, we know that $\phi_0(0, x)$, which is a nontrivial solution of (2.2) with $\lambda = \varepsilon_0(0)$, has no zero in $[0, a]$ and thus has no zero in $(-\infty, +\infty)$.

(2) If any nontrivial solution $y(x, \lambda)$ of (2.2) with $\lambda \leq \varepsilon_0(0)$ had more than one zeros in $(-\infty, +\infty)$, $\phi_0(0, x)$ would have at least one zero between two zeros of $y(x, \lambda)$ by Theorem 2.2. This is contradictory to (1). $\square$

The theory in this chapter could play a fundamental role in investigations of many physical problems on general one-dimensional crystals, including infinite crystals, semi-infinite crystals, and finite crystals. In the next two chapters, we will investigate electronic states in one-dimensional semi-infinite crystals and finite crystals. Theorems 2.7–2.9, especially the Theorem 2.8, play an essential role in the theory of electronic states in ideal one-dimensional crystals of finite length.

Based on the theory in this chapter, a corresponding theory for treating other one-dimensional crystals, including one-dimensional phononic crystals or photonic crystals will be presented in the Appendices.

References

1. F. Seitz, *The Modern Theory of Solids* (McGraw-Hill, New York, 1940)
2. H. Jones, *The Theory of Brillouin Zones and Electronic States in Crystals* (North-Holland, Amsterdam, 1960)
3. C. Kittel, *Introduction to Solid State Physics*, 7th edn. (Wiley, New York, 1996)
4. R.L. Kronig, W.G. Penney, Proc. Roy. Soc. Lond. Ser. A. **130**, 499 (1931)
5. H.A. Kramers, Physica **2**, 483 (1935)
6. I. Tamm, Phys. Z. Sowj. **1**, 733 (1932)
7. M.S.P. Eastham, *The Spectral Theory of Periodic Differential Equations* (Scottish Academic Press, Edinburgh, 1973). (and references therein)
8. J. Weidmann, *Spectral Theory of Ordinary Differential Operators* (Springer, Berlin, 1987)
9. A. Zettl, *Sturm-Liouville Theory* (American Mathematical Society, Providence, 2005) (In particular, p. 39)
10. B. Malcolm Brown, M.S.P. Eastham, K.M. Schmidt, *Periodic Differential Operators. Operator Theory: Advances and Applications*, vol. 230 (Springer, Heidelberg, 2013) (and references therein)
11. G. Teschl, *Ordinary Differential Equations and Dynamical Systems* (American Mathematical Society, Providence, 2012)
12. V.A. Yakubovich, V.M. Starzhinskii, *Linear Differential Equations with Periodic Coefficients*, vol. 1 (Chapter II) (Wiley, New York, 1975)
13. W. Magnus, S. Winkler, *Hill's Equation* (Interscience, New York, 1966)
14. W. Kohn, Phys. Rev. **115**, 809 (1959)
15. P.A. Kuchment, Bull. (New Ser.) Am. Math. Soc. **53**, 343 (2016)

Electronic States in Crystals of Finite Size

Quantum Confinement of Bloch Waves

Ren, S.Y.

2017, XVI, 283 p. 42 illus., Hardcover

ISBN: 978-981-10-4716-9