Commun. Math. Phys. 132, 103–118 (1990)

Communications in
Mathematical
Physics

On the Measure of the Spectrum for the Almost Mathieu Operator*

J. Avron[1], P. H. M. v. Mouche[2], and B. Simon[3]

[1] Department of Physics, Technion, Haifa, Israel
[2] Landbouwhogschool Wageningen, Vakgroep Wiskunde, Wageningen, The Netherlands
[3] Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

Dedicated to Res Jost and Arthur Wightman

Abstract. We obtain partial results on the conjecture that for the almost Mathieu operator at irrational frequency, α, the measure of the spectrum, $S(\alpha,\lambda,\theta)=|4-2|\lambda||$. For $|\lambda|\neq 2$ we show that if α_n is rational and $\alpha_n\to\alpha$ irrational, then $S_+(\alpha_n,\lambda,\theta)\to|4-2|\lambda||$.

1. Introduction

In this paper we will discuss the almost Mathieu operator, also called Harper's equation. This is the operator, $h_{\alpha,\lambda,\theta}$ on $l^2(\mathbb{Z})$ defined by

$$h_{\alpha,\lambda,\theta}=h_0+v\,,\qquad (h_0u)(n)=u(n+1)+u(n-1)\,,$$
$$(vu)(n)=\lambda\cos(2\pi\alpha n+\theta)u(n)\,,$$

where λ, α, θ are real parameters. This is the simplest of almost periodic Jacobi matrices and there has been considerable literature studying it [1, 2, 4–6, 14, 17, 18].

We will be interested in $S(\alpha,\lambda,\theta)$, the Lebesgue measure of the spectrum $\sigma(h_{\alpha,\lambda,\theta})$. It is a fundamental result (e.g. [2]) that for α irrational, S is independent of θ for α, λ fixed but this is not true if α is rational. In that case we define $S_\pm(\alpha,\lambda)$ to be the Lebesgue measure of $\sigma_\pm(\alpha,\lambda)$ where

$$\sigma_-(\alpha,\lambda)=\bigcap_\theta\sigma(\alpha,\lambda,\theta)\,,\qquad \sigma_+(\alpha,\lambda)=\bigcup_\theta\sigma(\alpha,\lambda,\theta)\,.$$

As explained in [2], $\bigcup_\theta\sigma(\alpha,\lambda,\theta)$ is the more natural object in that it has a set theoretic continuity in α.

* Research partially supported by U.S. NSF grant number DMS-8801918 and by BSF under grant number 88-00325

We are interested here in a conjecture that goes back at least to Aubry and Andre [1] that

$$S(\alpha,\lambda,\theta)=|4-2|\lambda|| \qquad \alpha \text{ irrational}. \tag{1.1}$$

By symmetry we can suppose $\lambda \geqq 0$ which we henceforth do. Thouless [17] has proven the following lower bound:

$$S_+(\alpha,\lambda)\geqq(4-2\lambda) \qquad \alpha \text{ rational};\ \lambda\geqq 0,$$

and he argued that therefore

$$S(\alpha,\lambda,\theta)\geqq(4-2\lambda) \qquad \alpha \text{ irrational};\ \lambda\geqq 0.$$

While a proof of this was not given, we will see it is not hard to prove from the rational case. In a subsequent work, Thouless [18] presented the result that

$$\lim_{q\to\infty} qS\left(\frac{1}{q},2,0\right)=32\beta(2)/\pi,$$

where $\beta(2)$ is Catalan's constant. In that paper, the issue of gap edges ordering that we discuss in Sect. 3 is also discussed. We will prove

Theorem 1. *For α rational and $0<\lambda<2$:*

$$S_-(\alpha,\lambda)=2|2-\lambda|.$$

For $\lambda\geqq 2$: $S_-=0$.

As for S_+, we will prove that

Theorem 2. *For p,q relatively prime and $0\leqq\lambda<2$,*

$$S_+(\alpha,\lambda)\leqq S_-(\alpha,\lambda)+4\pi\left(\frac{\lambda}{2}\right)^{q/2}.$$

We recall Andre-Aubry duality [1, 2] which implies that

$$S_+(\alpha,\lambda)=\frac{\lambda}{2}S_+\left(\alpha,\frac{4}{\lambda}\right) \qquad \alpha \text{ rational},\ \lambda>0,$$

$$S(\alpha,\lambda,\theta)=\frac{\lambda}{2}S\left(\alpha,\frac{4}{\lambda},\theta\right) \qquad \alpha \text{ irrational},\ \lambda>0.$$

This implies with Theorem 2 that if p_n, q_n are relatively prime and $q_n\to\infty$, then

$$\lim_{n\to\infty} S_+\left(\frac{p_n}{q_n},\lambda\right)\to|4-2\lambda| \qquad \lambda\neq 2,\ \lambda\geqq 0.$$

This strongly supports the conjecture (1.1) but as we will explain we have not succeeded in proving it.

In Sect. 2, we prove a result on the Mathieu operator that can be considered the continuum analog of Theorem 1 [or of the conjecture (1.1)]. In Sect. 3, we reduce the proof of Theorem 1 to a result on the ordering in terms of symmetry of levels of certain operators. In Sect. 4, we prove a result on degenerate perturbation theory which we use in Sect. 5 to prove the theorem on the ordering of level α. In Sect. 6 we prove Theorem 2. Finally, in Sect. 7, we discuss the problems with extending the result to prove (1.1).

2. The Mathieu Equation: A Warmup

As a warmup for the main theorem, we want to prove the following theorem about the Mathieu equation:

Theorem 3. *Let* $H_\lambda = -\frac{d^2}{dx^2} + \lambda\cos(x)$ *on* $L^2(-\infty,\infty)$. *Let* $e_0(\lambda) = \inf\sigma(H_\lambda)$ *and let* $(\mu_{n,-}(\lambda), \mu_{n,+}(\lambda))$ *be the* n^{th} *gap in the spectrum. Then*

$$e_0 + \sum_{n=1}^{\infty} (\mu_{n,+} - \mu_{n,-}) = 2|\lambda|. \tag{2.1}$$

Proof. Let $\mu_{D,n}$ be the n^{th} Dirichlet eigenvalue for the Mathieu operator on $(0,2\pi)$. There is a general sum rule that for general $-\frac{d^2}{dx^2} + V(x)$ (with a period 2π) [13] =

$$\mathrm{e}_0 + \sum_{n=1}^{\infty} [\mu_{n,+} + \mu_{n,-} - 2\mu_{D,n}] = 2V(0). \tag{2.2}$$

Since all quantities in (2.1) are even in λ (on account of translation of x to $x+\pi$), we can suppose $\lambda > 0$. Since $\cos(x)$ is even, one of $\mu_{n,+}, \mu_{n,-}$ is $\mu_{D,n}$ and the other is the Neumann eigenvalue $\mu_{N,n}$. (We count so the Dirichlet eigenvalues start at $n=1$ but the Neumann at $n=0$.) Thus (2.2) becomes

$$e_0 + \sum_{n=1}^{\infty} (\mu_{N,n} - \mu_{D,n}) = 2\lambda. \tag{2.3}$$

Suppose that we prove that for all $\lambda > 0$,

$$\mu_{N,n} - \mu_{D,n} > 0. \tag{2.4}$$

Then for $n = 1, 2, \ldots,$

$$\mu_{N,n} = \mu_{n,+}; \qquad \mu_{D,n} = \mu_{n,-},$$

and (2.3) is precisely (2.1).

It is a standard fact [15] that for all $\lambda \neq 0$, $\mu_{n,+}(\lambda) \neq \mu_{n,-}(\lambda)$ (special to the Mathieu equation). Thus it suffices to prove (2.4) for λ large.

The gap edges $\mu_{n,\pm}$ are well known [15] to be precisely the eigenvalues with periodic and antiperiodic boundary conditions for $\frac{-d^2}{dx^2} + V(x)$ on $L^2(0,2\pi)$. Equivalently, these are the periodic B.C. eigenvalues for the operator on $L^2(-2\pi,2\pi)$. If one looks at $\cos x$ on $(-2\pi,2\pi)$, this is a classic double well problem with minima at $\pm\pi$ and reflection symmetry about $x=0$. In terms of eigenfunctions being even or odd under that symmetry, it is known that in the λ large region the ordering is [10, 16]: E, O, E, O, ... Since even means Neumann on $(0,2\pi)$ and odd means Dirichlet, we see that for each n and λ large $\mu_{n,D} < \mu_{n,N}$ (recall the numbering conventions E, O, E, O, ... means $\mu_{N,0} < \mu_{D,1} < \mu_{N,1} < \mu_{D,2} < \ldots$). □

Equation (2.1) can be reinterpreted in a way that shows why it is a warmup. Consider the sets $\sigma(H_\lambda)\backslash\sigma(H_0)$ and $\sigma(H_0)\backslash\sigma(H_\lambda)$. The first is $[e_0, 0)$ with the negative gaps in $\sigma(H_\lambda)$ removed. The second is the union of the positive gaps. Thus, with $|\cdot|$ = Lebesgue measure:

$$|\sigma(H_\lambda)\backslash\sigma(H_0)| - |\sigma(H_0)\backslash\sigma(H_\lambda)| = -e_0 - \sum_n (\mu_{n,+} - \mu_{n,-}),$$

so (2.1) says that

$$|\sigma(H_\lambda)\backslash\sigma(H_0)| - |\sigma(H_0)\backslash\sigma(H_\lambda)| = -2|\lambda|. \tag{2.5}$$

For finite measure sets $|A\backslash B| - |B\backslash A| = |A| - |B|$, so that *formally* (2.5) says

$$|\sigma(H_\lambda)| = |\sigma(H_0)| - 2|\lambda| \quad \text{(formal!)},$$

which is the continuum analog of Theorem 1 where $|\sigma(h_0)| = 4$. □

Theorem 3 has the following amusing consequence:

Proposition 4. *Let* $H_{\lambda,\omega} = -\frac{d^2}{dx^2} + \lambda\cos(\omega x)$, *and let* $|G|(\lambda,\omega)$ *be the total measure of its gaps. Then*

$$|G|(\lambda,\omega) = \begin{cases} 3\lambda + \omega\sqrt{\lambda/2} + 0(e^{-c/\omega}), & 0<\omega\ll 1 \\ 2\lambda + 0(\lambda^2/\omega^2), & \omega \gg 1, \end{cases}$$

in particular $\lim\limits_{\omega\to 0} |G|(\lambda,(\omega)) = 3\lambda$, *and* $\lim\limits_{\omega\to\infty} |G|(\lambda,\omega) = 2\lambda$.

Remark. This is interesting because $H_{\lambda,0}$ is just the shifted Laplacian which, of course, has no gaps in the spectrum. So, the limit of the gap measure (as $\omega\to 0$) is larger than the gap measure of the limit. The misbehavior of the measure is related to the difficulties we have encountered in proving the Aubry-Andre conjecture (see Sect. 7).

Proof. By scaling

$$H_{\lambda,\omega} \cong \omega^2 H_{\lambda/\omega^2,1},$$

from which it follows using (2.1), that

$$\begin{aligned} |G|(\lambda,\omega) &= \omega^2 |G|(\lambda,\omega^2,1) \\ &= \omega^2\left(\frac{2\lambda}{\omega^2} - e_0(\lambda/\omega^2)\right) \\ &= 2\lambda - \omega^2 e_0(\lambda/\omega^2), \end{aligned}$$

since

$$e_0(\mu) = \begin{cases} 0(\mu^2), & \mu \ll 1 \\ -\mu + \sqrt{\mu/2} + O(e^{-\mu}), & \mu \gg 1, \end{cases}$$

the result follows. □

3. Proof of Theorem 1 up to Level Ordering

In this section we prove Theorem 1 assuming some facts about level ordering which we will not prove until Sect. 5.

Let $\alpha = p/q$ with p, q relatively prime. In finding the spectrum of $h_{p/q,\lambda,\theta}$, a key role is played by the discriminant of the problem. Recall that if $v(n)$ has period q, one defines the transfer matrix

$$T(E) = \begin{pmatrix} E - v(0) & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} E - v(1) & -1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} E - v(q-1) & -1 \\ 1 & 0 \end{pmatrix}$$

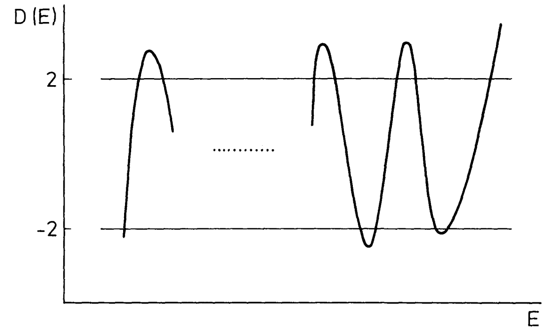

Fig. 1

and the discriminant

$$D(E) = \mathrm{Tr}(T(E)).$$

For $h_{p/q,\lambda,\theta}$ we will use $D_{p/q}(E,\lambda,\theta)$. If one follows $D(E)$ as a function of E from E near $+\infty$ downwards it is large for E large and then falls monotonically to below -2 (could be equal to -2). It then turns around and crosses -2 in arriving monotonically to $+2$, etc. There are q regions where it passes monotonically from 2 to -2 [15]. Schematics are shown in Fig. 1 (for q odd).

The spectrum of $h_0 + v$ is the inverse image under D of the interval $[-2, 2]$. The band edges are the points where $D(E) = \pm 2$. $D(E) = 2$ are the eigenvalues of the operator $h_0 + v$ with periodic boundary conditions and $D(E) = -2$ are the eigenvalues of the same operator with antiperiodic boundary conditions.

As with so much else in the study of the almost Mathieu equation, our analysis depends on a remarkable formula of Chambers [6] and Butler-Brown [5] giving the θ dependence of $D_{p/q}(E,\lambda,\theta)$. Let $\Delta_{p/q}(E,\lambda) \equiv D_{p/q}(E,\lambda,\theta=\pi/2q)$.

Proposition 3.1. *If p, q are relatively prime:*

$$D_{p/q}(E,\lambda,\theta) = \Delta_{p/q}(E,\lambda) - 2\left(\frac{\lambda}{2}\right)^q \cos(q\theta). \tag{3.1}$$

Sketch. For the reader's convenience, here is a sketch of the proof. Imagine writing out $\cos(2\pi\alpha j + \theta)$ and multiplying the matrices defining T. It is clearly a sum of terms whose θ dependence is $e^{im\theta}$, $m = -q, -q+1, --, q-1, q$. By cyclicity of the trace, D is invariant under adding $2\pi p/q$ to θ. Since p is relatively prime to q, it must be invariant under adding $2\pi/q$ to θ so that D must have a Fourier expansion $e^{im\theta}$ with m divisible by q. It follows that only $m = 0$, $\pm q$ are present. It is easy to compute the $\pm q$ terms and (3.1) holds. $\square$

Henceforth we use $\lambda \geqq 0$ for convenience. As a direct consequence of this we have [4].

Corollary 3.2. *$\sigma_+(\alpha,\lambda)$ is the inverse image under Δ of the interval*

$$\left[-2-2\left(\frac{\lambda}{2}\right)^q, 2+2\left(\frac{\lambda}{2}\right)^q\right].$$

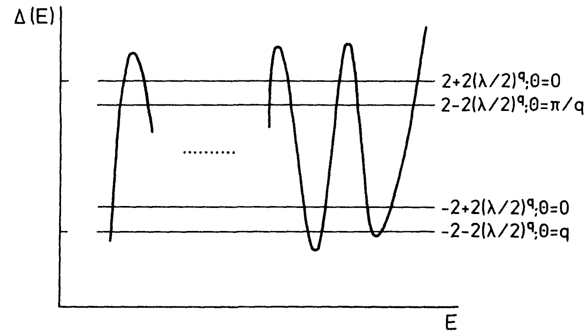

Fig. 2

If $\lambda > 2$, $\sigma_-(\alpha, \lambda)$ is empty, if $\lambda = 2$, $\sigma_-(\alpha, \lambda)$ is a discrete set, and if $\lambda < 2$, $\sigma_-(\alpha, \lambda)$ is the inverse image of $\left[-2+2\left(\frac{\lambda}{2}\right)^q, 2-2\left(\frac{\lambda}{2}\right)^q\right]$*; see Fig. 2.*

To study S_-, particular relevance is associated to the cases where $\Delta = \pm 2 \mp 2\left(\frac{\lambda}{2}\right)^q$ which gives the edges of S_-. The case $\Delta = -2+2\left(\frac{\lambda}{2}\right)^q$ corresponds to $\theta = 0$ and antiperiodic boundary conditions. That is if we take the sites $n = 0, \ldots, q-1$

$$\begin{aligned} (h_0)_{ij} &= 0, \qquad |i-j| \neq 1, q-1 \\ &= 1, \qquad |i-j| = 1 \\ &= -1, \qquad |i-j| = q-1 . \end{aligned}$$

(Note: The only pairs with $|i-j| = q-1$ are $i=0, j=q-1$ and its symmetric pair.) If q is odd, we can take the points

$$n = -\frac{q-1}{2}, \quad -\frac{q-1}{2}+1, \ldots, \frac{q-1}{2}$$

and still take h_0 with $h_0 = -1$ for the $\left(-\frac{(q-1)}{2}, \frac{(q-1)}{2}\right)$ coupling. If q is even we take $n = -\frac{q}{2}+1, \ldots, +\frac{q}{2}-1, \frac{q}{2}$ and the $\left(\frac{q}{2}, -\frac{q}{2}+1\right)$ coupling negative. We define the symmetry:

$$\begin{aligned} (Ru)(n) &= u(-n), \qquad q \text{ odd or } \left(q \text{ even and } n \neq \frac{q}{2}\right), \\ &= -u(n), \qquad q \text{ even}, \ n = \frac{q}{2}. \end{aligned}$$

With this strange R, it is easy to see that both h_0 and v are invariant under the symmetry R, that is, they commute with R. The minus in $-u\left(\frac{q}{2}\right)$ is needed to turn $(h_0)_{q/2-1, q/2}$ into $(h_0)_{-q/2+1, q/2}$. Since h is invariant, we can classify all its eigenvalues as even or odd. In Sect. 5 we will prove:

Theorem 4a. *For q odd, the order of levels for $\theta=0$ with antiperiodic boundary conditions is*

$$E\ O\ E\ O \dots E.$$

For q even, $\theta=0$, and the same boundary conditions, it is

$$O\ E\ O\ E \dots E.$$

Suppose now that q is odd. Then $\frac{\pi}{q}=\pi-\left(\frac{q-1}{2}\right)\frac{2\pi}{q}$ so that $\theta=\frac{\pi}{q}$ and $\theta=\pi$ are translates and we may as well take $\theta=\pi$, i.e. $-\lambda\cos(2\pi\alpha n)$. This potential is obviously invariant under $(Ru)(n)=u(-n)$ if we take

$$n=-\left(\frac{q-1}{2}\right),\dots,0,\dots,\left(\frac{q-1}{2}\right).$$

If q is even, $\theta=\frac{\pi}{q}$ is equivalent to taking n half-integral with the potential $\cos(2\pi\alpha n)$ with

$$n=-\left(\frac{q-1}{2}\right),\dots,-\frac{1}{2},\frac{1}{2},\dots,\left(\frac{q-1}{2}\right),$$

and again we have invariance. In Sect. 5, we will also prove.

Theorem 4b. *For q odd, $\theta=\pi$, periodic boundary conditions the ordering of levels is*

$$E\ O\ E\ O \dots E.$$

For q even, $\theta=\frac{\pi}{q}$, periodic boundary conditions, it is

$$O\ E\ O\ E \dots E.$$

The next step in the proof of Theorem 1 concerns the difference between traces over the even and odd spaces. For each of these Hamiltonians, H, let $\Gamma(H)=\operatorname{Tr}(H\restriction\mathscr{H}_e)-\operatorname{Tr}(H\restriction\mathscr{H}_0)$, where $\mathscr{H}_e$ (respectively $\mathscr{H}_0$) is the subset of states on which $R=+1$ (respectively -1), then:

Proposition 3.3. *$\Gamma(H)$ has the following values:*
(a) *q odd; $\theta=\pi$, periodic* B.C. $\Gamma(H)=-\lambda+2$,
(b) *q odd; $\theta=0$, antiperiodic* B.C. $\Gamma(H)=\lambda-2$,
(c) *q even; $\theta=\frac{\pi}{q}$, periodic* B.C. $\Gamma(H)=4$,
(d) *q even; $\theta=0$, antiperiodic* B.C. $\Gamma(H)=2\lambda$.

Proof. (a) A basis for the even states is

$$\{\delta_0\}\cup\left\{\frac{\delta_j+\delta_{-j}}{\sqrt{2}}\right\}_{j=1}^{(q-1)/2}$$

and for the odd states $\left\{\frac{\delta_j-\delta_{-j}}{\sqrt{2}}\right\}_{j=1}^{(q-1)/2}$. The terms from the potential $v=-\lambda\cos(2\pi\alpha n)$ cancel exactly for $j=1,\dots,(q-1)/2$ but $(\delta_0,v\delta_0)=-\lambda$ contributes. h_0 has a diagonal matrix element because $h_0\delta_{(q-1)/2}=\delta_{-(q-1)/2}+\delta_{(q-3)/2}$. This diagonal matrix element is 1 on the even space and -1 on the odd so $\Gamma=-\lambda+1-(-1)=2-\lambda$.

(b) The basis is the same as in (a) but now $v=\lambda\cos(2\pi\alpha n)$ so $(\delta_0, v\delta_0)=\lambda$. Because of the antiperiodic boundary conditions the diagonal matrix elements of h_0 have opposite signs so $\Gamma=\lambda-1-(+1)=\lambda-2$.

(c) As noted above, the reflection is natural in terms of a basis δ_j; $j=\pm\frac{1}{2}, \pm\frac{3}{2}, \ldots, \pm\frac{q-1}{2}$.

Then a basis for even states is $\left\{\frac{\delta_j+\delta_{-j}}{\sqrt{2}}\right\}_{j=1/2,3/2,\ldots,(q-1)/2}$ and for the odd states

$$\left\{\frac{\delta_j-\delta_{-j}}{\sqrt{2}}\right\}_{j=1/2,3/2,\ldots,(q-1)/2}.$$

The potential terms cancel exactly but h_0 has diagonal terms for $\frac{\delta_j\pm\delta_{-j}}{\sqrt{2}}$ with $j=\frac{1}{2}$, $\frac{(q+1)}{2}$. For the even states each such matrix element is 1 and for the odd states -1. Since $1+1-(-1-1)=4$, $\Gamma(H)=4$.

(d) A basis for the even states is $\{\delta_0\}\cup\left\{\frac{\delta_j+\delta_{-j}}{\sqrt{2}}\right\}_{j=1}^{(q-2)/2}$ and for the odd states

$$\{\delta_{q/2}\}\cup\left\{\frac{\delta_j-\delta_{-j}}{\sqrt{2}}\right\}_{j=1}^{(q-2)/2};$$

$\delta_{q/2}$ is odd as discussed above. h_0 has no diagonal matrix elements. The v matrix elements canceled except for $\delta_{q/2}$ and δ_0. Since $\lambda-(-\lambda)=2\lambda$, we have that $\Gamma(H)=2\lambda$.

Proof of Theorem 1. Consider first the case q odd. Then the ordering of levels gives the picture in Fig. 3.

S_- is the sum of the bands. Half the bands $\left[\text{actually } \left(\frac{q+1}{2}\right)\right]$ run from even antiperiodic up to even periodic while half the bands $\left[\text{actually } \left(\frac{q-1}{2}\right)\right]$ run from

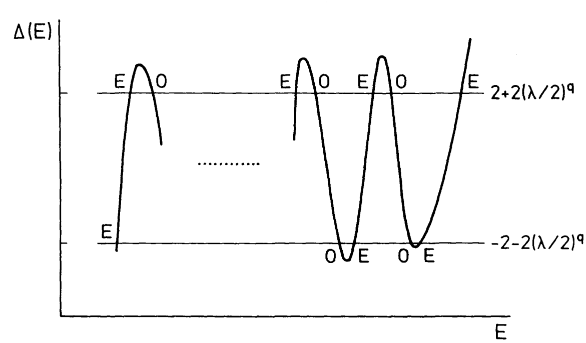

Fig. 3

odd and periodic up to odd and antiperiodic. Thus, by Proposition 3.3,

$$\begin{aligned} S_- &= (\text{even periodic}) - (\text{even anti}) + (\text{odd anti}) - (\text{odd periodic}) \\ &= \Gamma(H, \text{periodic}) - \Gamma(H, \text{antiperiodic}) \\ &= -\lambda + 2 - (\lambda - 2) = 4 - 2\lambda . \end{aligned}$$

The argument for q even is similar: Γ is positive at large negative E so the first band runs from odd to odd. But

$$\Gamma(H, \text{periodic}) - \Gamma(H, \text{antiperiodic}) = 4 - 2\lambda$$

still holds. □

4. Degenerate Perturbation Theory

In understanding where we will need the theorem below, think of R as reflection symmetry, A as the potential v, B as h_0, μ as λ^{-1} and φ_n as a renumbering of Kronecker delta functions.

Theorem 4.1. *Suppose that A and B are finite self-adjoint real matrices and R a unitary matrix which obeys $R^2 = 1$ and*

$$RAR^{-1} = A, \qquad RBR^{-1} = B.$$

Suppose that E_0 is a doubly degenerate eigenvalue of A for which there are orthogonal eigenvectors φ_0, φ_1 so that

$$R\varphi_0 = \varphi_1 .$$

Let $\varphi_2, \ldots, \varphi_n$ be a labelling of other eigenvectors of A to yield a complete set and let $A\varphi_m = E_m\varphi_m$. Let $E_\pm(\mu)$ denote the eigenvalues of $A + \mu B$ whose eigenvectors approach $\varphi_\pm = \frac{1}{\sqrt{2}}(\varphi_0 \pm \varphi_1)$ as $\mu \to 0$. Because of the symmetry, R, we know $\varphi_\pm$ are the limiting vectors. Suppose that for all $l < p$ and all $\varphi_{n_1}, \ldots, \varphi_{n_{l-1}}$ we have that

$$(\varphi_1, B\varphi_{n_1})(\varphi_{n_1}, B\varphi_{n_2}) \ldots (\varphi_{n_{l-1}}, B\varphi_0) = 0 .$$

Then, as $\mu \downarrow 0$:

$$E_+ - E_- = 2\gamma\mu^p + O(\mu^{p+1}),$$

where

$$\gamma = \sum_{\substack{\varphi_{n_1}, \ldots, \varphi_{n_{p-1}} \\ E(\varphi_{n_j}) \neq E_0}} \frac{(\varphi_1, B\varphi_{n_1}) \ldots (\varphi_{n_{p-1}}, B\varphi_0)}{(E_0 - E_{n_1}) \ldots (E_0 - E_{n_{p-1}})} .$$

Proof. We begin standard eigenvalue perturbation theory [12, 15], namely we consider the projection $P(\mu)$ onto the eigenspaces for $E_\pm$. Then, for μ small:

$$P(\mu) = \frac{1}{2\pi i} \oint \frac{dz}{z - (A + \mu B)} .$$

Clearly

$$\begin{aligned} E_{\pm} &= (\varphi_{\pm}, (A+\mu B)P\varphi_{\pm})/(\varphi_{\pm}, P\varphi_{\pm}) \\ &= E_0 + \mu(\varphi_{\pm}, BP\varphi_{\pm})/(\varphi_{\pm}, P\varphi_{\pm}). \end{aligned} \tag{4.1}$$

By expanding $[z-(A+\mu B)]^{-1}$ in a geometric series with remainder in the usual way [12, 15], one sees that

$$(\varphi_0, P\varphi_1) = O(\mu^p), \tag{4.2a}$$

$$(\varphi_0, BP\varphi_1) = \mu^{p-1}\gamma + O(\mu^p). \tag{4.2b}$$

For (4.1), we look at

$$(\varphi_{\pm}, P\varphi_{\pm}) = (\varphi_0, P\varphi_0) \pm O(\mu^p),$$

$$(\varphi_{\pm}, BP\varphi_{\pm}) = (\varphi_0, BP\varphi_0) \pm \gamma\mu^{p-1} + O(\mu^p),$$

so (4.1)–(4.2) implies

$$E_{\pm} = E_0 + \mu[(\varphi_0, BP\varphi_0)/(\varphi_0, P\varphi_0)] \pm \gamma\mu^p/(\varphi_0, P\varphi_0) + O(\mu^{p+1}).$$

Since $(\varphi_0, P\varphi_0) = 1 + O(\mu)$, the theorem is proven. □

5. The Ordering of Levels

We want to prove Theorems 4a, b (from Sect. 3) using Theorem 4.1. The levels in Theorem 4a, b are non-degenerate for $\lambda \neq 0, \infty$. This is because the gaps can only close *in principle* for the case where $\Delta = \pm\left(2+2\left(\frac{\lambda}{2}\right)^q\right)$ not for the case $\Delta = \pm\left(2-2\left(\frac{\lambda}{2}\right)^q\right)$ of interest to us. In fact [9, 14], (except for the middle gap if q is even) the gaps don't even close for $\Delta = \pm\left(2+2\left(\frac{\lambda}{2}\right)^q\right)$ but we will not need this subtle theorem.

We write $h_0 + \lambda v = \lambda(v + \lambda^{-1}h_0)$ and think of h_0 as a perturbation of v. Levels of v will be degenerate and symmetric so we can apply Theorem 4.1.

Case 1: $\theta = 0$, antiperiodic B.C., q odd. The potential is

$$v(n) = \cos(2\pi pn/q).$$

The top level with $n=0$ is non-degenerate and even. The others are degenerate due to $n \to -n$ symmetry and the ordering of levels (Theorem 4a) says that they split even odd, i.e. the quantity γ of Theorem 4.1 is negative. h_0, the perturbation, links only neighboring levels. δ_m and δ_{-m} will be linked first via one of these chains going through $n=0$ or $n = \pm\left(\frac{q-1}{2}\right)$. For the $n=0$ chains all h_0 matrix elements are positive. All energy denominators come in pairs except for $(E_m - E_n = 0)^{-1}$ which is negative because $E_{n=0} = 1$ is the largest eigenvalue, of v. If the chain goes through $\delta_{\pm(q-1)/2}$ all energy denominators come in pairs but one h_0 matrix element

$$(\delta_{(q-1)/2}, h_0\delta_{-(q-1)/2}) = -1$$

because of the antiperiodic boundary conditions. So γ is always negative and we have the claimed $E\ O\ E\ O \dots E\ O\ E$ ordering.

Remark. We see that the $\theta=0$, periodic B.C. splitting (not one we need!) is complicated. $\gamma<0$ for chains going through $n=0$, while for chains through $n=\pm(q-1)/2$ we have $\gamma>0$.

Case 2: $\theta=0$, antiperiodic B.C., q even. The potential is now

$$v(n)=\cos(2\pi pn/q).$$

$n=0$ and $n=q/2$ are non-degenerate with $n=0$ even and $n=\frac{q}{2}$ odd as explained in Sect. 3. $n=q/2$ is the bottom level, $n=0$ the top. All other levels are degenerate and we want to show that $\gamma<0$ so the spliting is even below odd and we get $O\ E\ O\ E\ O \dots E$. Again chains can go through $n=0$ or $n=q/2$. The ones through δ_0 have all h_0 matrix elements positive and all energy denominators paired except for $(E_m-E_{n=0})^{-1}$ which is negative. For chains through $n=q/2$, one matrix element is negative. The energy denominators are all paired except for $(E_m-E_{n=q/2})^{-1}$ which is positive. So $\gamma<0$.

Case 3: $\theta=\pi$, periodic B.C., q odd. The potential is

$$v(n)=-\cos(2\pi pn/q).$$

The bottom level is $n=0$ and is even. We claim $\gamma>0$ so the ordering of each of the other pairs is $O\ E$ and overall we have $E\ O\ E\ O\ E \dots E$. h_0 only has positive matrix elements. All energy denominators are paired except for $(E_m-E_{n=0})^{-1}$ which is positive so $\gamma>0$.

Case 4: $\theta=\frac{\pi}{q}$, periodic B.C., q even. As noted the basis is $\{\delta_n\}$ with

$$n=-\left(\frac{q-1}{2}\right),\dots,-\frac{1}{2},\frac{1}{2},\dots,\frac{(q-1)}{2}.$$

All states are paired, h_0 has only positive matrix elements and all energy denominators are paired so $\gamma>0$, odd is below even and the ordering is $O\ E\ O\ E \dots O\ E$ as claimed. $\square$

6. Proof of Theorem 2

By the analysis in Sect. 3, S_+/S_- is the inverse image under $\Delta_{p/q}$ of the intervals

$$\left(-2-2\left(\frac{\lambda}{2}\right)^q,-2+2\left(\frac{\lambda}{2}\right)^q\right)\cup\left(2-2\left(\frac{\lambda}{2}\right)^q,2+2\left(\frac{\lambda}{2}\right)^q\right).$$

S_+/S_- will be small because $\left(\frac{\lambda}{2}\right)^q$ is small. Consider one connected piece of the inverse image under $\Delta_{p/q}$ of

$$\left(2-2\left(\frac{\lambda}{2}\right)^q,2+2\left(\frac{\lambda}{2}\right)^q\right).$$

We want to think of this instead as the inverse image under $D_{p/q}(\cdot, \lambda, \theta=0)$ of $\left(2-4\left(\frac{\lambda}{2}\right)^q, 2\right)$ and in particular as a part of the spectrum of $h_{p/q,\lambda,\theta}=0$. Let $r(E) = \pi k(E)$ be the rotation number for this problem where k is the integrated density of states. We know that $r(E)$ is determined by

$$D_{p/q}(E, \lambda, \theta=0) = 2(-1)^q \cos(qr). \tag{6.1}$$

On the other hand, Deift-Simon [7] have proven that on the spectrum

$$\frac{dr}{dE} \geqq \frac{1}{2}. \tag{6.2}$$

Proof of Theorem 2. For simplicity, suppose that $D(E)$ is increasing on the piece of inverse image in question. Let E_0 be the point where $D(E)=2$ and $E_0-\delta E$ the point where $D(E)=2-4\left(\frac{\lambda}{2}\right)^q$. $D(E)=2\cos(qr)$ and for $|\delta z|<\pi$, $2\cos z \leqq 2-\frac{4}{\pi^2}(\delta z)^2$ near a point where $z=z_0+\delta z \cos(z_0)=1$. Thus

$$-4\left(\frac{\lambda}{2}\right)^q \leqq -[q(\delta r)]^2 \left(\frac{2}{\pi}\right)^2.$$

By (6.2), $\delta E \leqq 2\delta r$. Thus

$$\delta E \leqq (2\pi) q^{-1} \left(\frac{\lambda}{2}\right)^{q/2}.$$

There are $2q$ bands in S_+/S_- so Theorem 2 is proven. $\square$

7. The Irrational Case

In this final section, we want to make some remarks about the irrational case. We begin with a theorem about the continuity of gaps:

Proposition 7.1. *Let f be a C^1 function on the unit circle. Let $\sigma(\alpha, \theta)$ be the spectrum of*

$$h_0 + f(2\pi\alpha n + \theta) \equiv h(\alpha, \theta)$$

and let $\sigma(\alpha) = \bigcup_\theta \sigma(\alpha, \theta)$. There exists $C>0$ so that if $E \in \sigma(\alpha)$ and $|\alpha-\alpha'| \leqq C[\|f'\|_\infty]^{-1}$, then there is $E' \in \sigma(\alpha')$ with

$$|E-E'| \leqq 6\|f'\|_\infty^{1/2} |\alpha-\alpha'|^{1/2}.$$

Remarks. 1. One could decrease 6 somewhat with little effort.
2. Our method is related to that of [9] who used a sharp cutoff in place of our test function below and so obtain $|\alpha-\alpha'|^{1/3}$.
3. The result extends to functions with several frequencies.

Proof. The basic idea is to take an approximate eigenfunction for $h(\alpha, \theta)-E$ and cut it off over a distance L. The cutoff introduces error of order L^{-1} in the kinetic energy and the potential energy difference is of order $L|\alpha-\alpha'|\,\|f'\|_\infty$. The sum is optimized by the choice $L=O([|\alpha-\alpha'|\,\|f'\|_\infty]^{-1/2})$.

Explicitly given ε, find $0\neq\varphi_\varepsilon\in l^2(\mathbb{Z})$ and θ so that

$$\|(h(\alpha,\theta)-E)\varphi_\varepsilon\|\leqq\varepsilon\|\varphi_\varepsilon\|\,.$$

Let $\eta_{0,L}$ be the test function

$$\begin{aligned}\eta_{0,L}(n)&=(1-|n|/L), & |n|&\leqq L,\\ &=0, & |n|&\geqq L,\end{aligned}$$

and let

$$\eta_{j,L}(n)=\eta_{0,L}(n-j),$$

the test function centered at j. We want to show that for each L and some j

$$\|(h(\alpha,\theta)-E)\eta_{j,L}\varphi_\varepsilon\|\leqq[\varepsilon+O(L^{-1})]\,\|_{j,L}\varphi_\varepsilon\|\,.$$

Note first that

$$\sum_j[\eta_{j,L}(n)]^2=1+(L-1)(2L-1)/3L\equiv\alpha_L$$

is independent of n. Clearly:

$$\begin{aligned}\sum_j\|\eta_{j,L}(h(\alpha,\theta)-E)\varphi_\varepsilon\|^2&=\alpha_L\|(h(\alpha,\theta)-E)\varphi_\varepsilon\|^2\\ &\leqq\alpha_L\varepsilon^2\|\varphi_\varepsilon\|^2=\varepsilon^2\sum_j\|\eta_{jL}\varphi_\varepsilon\|^2\,.\end{aligned}$$

Since $\|u+v\|^2\leqq(1+\delta)\,\|v\|^2+(1+\delta^{-1})\,\|u\|^2$,

$$\sum_j\|(h-E)\eta_{j,L}\varphi_\varepsilon\|^2\leqq(1+\delta^{-1})\varepsilon^2\sum_j\|\eta_{j,L}\varphi_\varepsilon\|^2+(1+\delta)\sum_j\|[\eta_{j,L},h_0]\varphi_\varepsilon\|^2\,.$$

Now

$$\begin{aligned}[][\eta_{j,L},h_0]_{i,j\pm1}&=c_{i,j,L} & &\text{if}\quad |i|,|i\pm1|\leqq L\\ &=0 & &\text{otherwise}\end{aligned}$$

with each $c_{i,j,L}=\pm\frac{1}{L}$. It follows that

$$\sum_j\|[h_0,\eta_{j,L}]\varphi_\varepsilon\|^2\leqq L^{-2}\beta_L\|\varphi_\varepsilon\|^2$$

with $\beta_L\sim8L$ for L large. Since $\alpha_L\sim\frac{SL}{3}$ we see that

$$\sum_j\|(h-E)\eta_{j,L}\varphi_\varepsilon\|^2\leqq(1+\delta^{-1})\varepsilon^2\sum_j\|\eta_{j,L}\varphi_\varepsilon\|^2+(1+\delta)13L^{-2}\sum_j\|\eta_{j,L}\varphi_\varepsilon\|^2$$

if $L\geqq L_0(\delta)$. Thus for some $j,\eta_{j,L}\ \varphi_\varepsilon\neq0$ and:

$$\|(h(\alpha,\theta)-E)\eta_{j,L}\varphi_\varepsilon\|\leqq(\varepsilon^2(1+\delta^{-1})+(1+\delta)^2 12L^{-2})^{1/2}\|\eta_{j,L}\varphi_\varepsilon\|\,.$$

Next given α' near α, let θ' be such that

$$\alpha j+\theta=\alpha' j+\theta'\,.$$

Then on $\operatorname{supp}(\eta_{j,L}\varphi_\varepsilon)$

$$|f(2\pi\alpha n+\theta)-f(2\pi\alpha' n+\theta')|\leqq 2\pi\|f'\|_\infty|\alpha-\alpha'|L$$

so that

$$\|(h(\alpha',\theta')-E)\eta_{j,L}\varphi_\varepsilon\|\leqq c\|\eta_{j,L}\varphi_\varepsilon\|,$$

where

$$c=(\varepsilon^2(1+\delta^{-1})+(1+\delta)^2 12L^{-2})^{1/2}+(2\pi)\|f'\|_\infty|\alpha-\alpha'|L.$$

Let

$$\tilde{c}=\sqrt{12}L^{-1}+(2\pi)\|f'\|_\infty|\alpha-\alpha'|L,$$

whose minimum value is

$$2\sqrt[4]{12}\sqrt{2\pi}\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2}.$$

Take for $L=\sqrt[4]{12}[(2\pi)^{1/2}\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2}]^{-1}$. Since $2\sqrt[4]{12}\sqrt{2\pi}<6$ and ε can be taken arbitrarily small, the result is proven. □

Theorem 7.2. *Under the hypothesis and notation of Proposition 7.1, let* $E_\pm(\alpha) = \begin{smallmatrix}\sup\\ \inf\end{smallmatrix}\,\sigma(\alpha)$. *Then* $E_\pm$ *are Holder continuous of order* $\frac{1}{2}$ *and, indeed, for* $|\alpha-\alpha'|$ *small*

$$|E_\pm(\alpha)-E_\pm(\alpha')|\leqq 6\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2}.$$

Proof. Fix α. By Proposition 7.1 for $|\alpha-\alpha'|$ small,

$$\sigma(\alpha')\cap(-\infty,E_-(\alpha)+6\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2})\neq\emptyset,$$

so

$$E_-(\alpha')\leqq E_-(\alpha)+6\|f'\|^{1/2}|\alpha-\alpha'|^{1/2}.$$

Interchange α and α' to get the E_- result. The E_+ result is similar. □

Recall [3, 8, 11] that gaps in $\sigma(\alpha)$ are labelled by integer m with $k(E)=(m\alpha)$ in the gap. One definition of m is as follows. Fix E_0 in the gap. For each θ, there is a unique function $u_+(n,\theta)$ solving

$$(h(\alpha,\theta)-E_0)u_+=0 \qquad \text{(difference equation)}$$

with $u_+ l^2$ at $+\infty$. m is just the winding number of the vector $(u_+(0),u_+(1))$ in $\mathbb{R}^2$, i.e. as a map of S^1 to $\mathbb{R}^2\backslash\{0\}$. Let $E_\pm^m(\alpha)$ be the edges of this gap and $G_m(\alpha)=E_+^m(\alpha)-E_-^m(\alpha)$ its size.

Theorem 7.3. *Under the hypothesis of Proposition 7.2, if* $G_m(\alpha)>0$, *then for* $|\alpha-\alpha'|$ *small enough (how small only depends on* $G_m(\alpha)$ *and* $\|f_\alpha'\|^{1/2}$*), we have* $G_m(\alpha')>0$ *and*

$$|G_m(\alpha)-G_m(\alpha')|\leqq 12\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2}.$$

Proof. Since $h(\alpha,\theta)$ has no spectrum in $(E_-^m(\alpha),E_+^m(\alpha))$, we have that for $|\alpha'-\alpha|$ small, $h(\alpha',\theta)$ has no spectrum in

$$(E_-^m(\alpha)+6\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2},\ E_+^m(\alpha)-6\|f'\|_\infty^{1/2}|\alpha-\alpha'|^{1/2})$$

which is non-empty (for $|\alpha-\alpha'|$ small). Let E_0 be the middle of the gap. A simple continuity argument shows that the winding number on m at E_0 is constant on the interval from α to α'. Symmetry implies the result. $\square$

Now fix λ and let $\tilde{S}(\alpha)$ be the $S_+(\alpha,\lambda)$. Then

$$\tilde{S}(\alpha)=E_+(\alpha)-E_-(\alpha)-\sum_m G_m(\alpha). \tag{7.1}$$

The only bar to proving that $\tilde{S}(\alpha)$ is continuous, given the last two theorems, is the fact that the sum in (7.1) is infinite. If we obtain a summable bound on the individual terms, we could prove continuity in α. That this is not trivial is seen by:

Fact 1. $\tilde{S}(\alpha)$ is discontinuous at every rational α (at least for $0<\lambda<2$)! For given α rational, put $\alpha_n=p_n/q_n$ with $q_n\to\infty$ and $\alpha_n\to\alpha$. We have proven that $\tilde{S}(\alpha_n)\to 4-2\lambda$ $=|S_-(\alpha,\lambda)|<|S_+(\alpha,\lambda)|=\tilde{S}(\alpha)$.

It must be in this case that the total of $\sum G_m(\alpha_n)$ contributes to $\lim_{\alpha_n\to\alpha}(\sum G_m(\alpha_n))$ but not to $\sum G_m(\alpha)$ (!). We believe that $\tilde{S}(\alpha)$ is continuous at irrational α but have not found a proof. Let us try to explain why irrational α differs from rational α and explain how if one could prove Holder continuity of order $\chi>1/2$ uniformly in m, one could prove continuity at most irrational α.

Fact 2. Among all reals, the rationals are worst approximated by rationals (!). To be precise if α is real and $p_n/q_n\to\alpha$ (with p,q relatively prime) and $q_n\to\infty$, then

$$\varliminf_{n\to\infty} q_n\left|\alpha-\frac{p_n}{q_n}\right|>0,$$

for if $\alpha=p_0/q_0$ and $p_n/q_n\neq p_0/q_0$ $(q_n>q_0)$ then $|\alpha-p_n/q_n|\geqq 1/q_nq_0$ so that $\varliminf$ is larger than $1/q_0$. On the other hand *any* irrational α has a set of canonical rational approximations [19] p_n/q_n so that $q_{n+1}>q_n$ and

$$|\alpha-p_n/q_n|\leqq 1/q_nq_{n+1}<1/q_n^2.$$

Suppose we know that $|G_m(\alpha)-G_m(\alpha')|\leqq C|\alpha-\alpha'|^\chi$ with $\chi>1/2$. Since

$$\left|\frac{p_n}{q_n}-\frac{p_{n+1}}{q_{n+1}}\right|=1/q_nq_{n+1}$$

and for $k>n$, $\left|\frac{p_n}{q_n}-\frac{p_k}{q_k}\right|\leqq 1/q_nq_{n+1}$, we have that

$$|G_m(\alpha_n)-G_m(\alpha_k)|\leqq Cq_n^{-\chi}q_{n+1}^{-\chi},\qquad k\geqq n.$$

If $2|m|>q_n$, then $G_m(\alpha_n)=0$ so it follows that

$$|G_m(\alpha_k)|\leqq Cq_n^{-\chi}q_{n+1}^{-\chi},\qquad k\geqq n;\ |m|>q_n/2.$$

So for $q_n/2<|m|<q_{n+1}/2$:

$$|G_n(\alpha_k)|\leqq Cq_n^{-\chi}q_{n+1}^{-\chi}$$

for all α_k [since $G_n(\alpha_k)=0$ if $k\leqq n$]. Thus, to get continuity of $\tilde{S}(\alpha)$ as $\alpha_k\to\alpha$ we only need that

$$\sum(q_{n+1}-q_n)q_n^{-\chi}q_{n+1}^{-\chi}<\infty.$$

Since $q_n \geqq 2^{n/2}$, this follows if

$$q_{n+1} \leqq q_n^{\beta}$$

with $\beta < \frac{\chi}{1-\chi}$. Since $\frac{\chi}{1-\chi} > 1$, this holds for a.e. α. Alas we do not even know if the estimate is true with $\chi > 1/2$.

Note that upper semi-continuity of σ_+ is easy given continuity of G_n so that the lower bound on σ_+ follows from the rational case.

Acknowledgements. P. H. M. v. Mouche would like to thank B. Simon and D. Wales, and J. E. Avron would like to thank B. Simon and M. Cross for the hospitality of Caltech.

References

1. Aubry, S., Andre, G.: Analyticity breaking and Anderson localization in incommensurate lattices. Ann. Israel Phys. Soc. **3**, 133–164 (1980)
2. Avron, J., Simon, B.: Almost periodic Schrödinger operators. II. The integrated density of states. Duke Math. J. **50**, 369–391 (1983)
3. Bellisard, J., Lima, R., Testard, D.: Almost periodic Schrödinger operators. In: Mathematics and physics recent results, Vol. 1, pp. 1–64. Singapore: World Scientific 1985
4. Bellisard, J., Simon, B.: Cantor spectrum for the almost Mathieu equation. J. Funct. Anal. **48**, 408–419 (1982)
5. Butler, F., Brown, E.: Model calculations of magnetic band structure. Phys. Rev. **166**, 630–636 (1968)
6. Chambers, W.: Linear-network model for magnetic breakdown in two dimensions. Phys. Rev. **140**, A135–A143 (1965)
7. Deift, P., Simon, B.: Almost periodic Schrödinger operators. III. The absolutely continuous spectrum in one dimension. Commun. Math. Phys. **90**, 389–411 (1983)
8. Deylon, F., Souillard, B.: The rotation number for finite difference operators and its properties. Commun. Math. Phys. **89**, 415 (1983)
9. Elliot, G., Choi, M., Yui, N.: Gauss polynomials and the rotation algebra. Preprint
10. Harrell, E.: The band structure of a one dimensional periodic system in a scaling limit. Ann. Phys. **119**, 351–369 (1979)
11. Johnson, R., Moser, J.: The rotation number for almost periodic potentials. Commun. Math. Phys. **84**, 403 (1982)
12. Kato, T.: Perturbation Theory for Linear Operators, 2nd ed. Berlin, Heidelberg, New York: Springer 1980
13. McKean, H., van Moerbeke, P.: The spectrum of Hill's equation. J. Math. **30**, 214–274 (1975)
14. v. Mouche, P.M.H.: The coexistence problem for the discrete Mathieu operator. Commun. Math. Phys. **122**, 23–24 (1989)
15. Reed, M., Simon, B.: Methods of modern mathematical physics. IV. Analysis of operators. London: Academic Press 1978
16. Simon, B.: Semiclassical analysis of low lying eigenvalues. III. Width of the ground state and band in strongly coupled solids. Ann. Phys. **158**, 415–420 (1984)
17. Thouless, D.: Bandwidths for a quasiperiodic tight binding model. Phys. Rev. B **28**, 4272–4276 (1983)
18. Thouless, D.: Scaling for the discrete Mathieu equation. Commun. Math. Phys. (to appear)
19. Wall, H.S.: Analytic theory of continued fractions. New York: Van Nostrand 1948

Communicated by T. Spencer

Received December 4, 1989