Notes on waveguide localization by slow-wave regions (a.k.a. "total internal reflection")

Steven G. Johnson, Massachusetts Institute of Technology

Created 21 November 2014, updated November 22, 2014.

1 Introduction

A "waveguide" is a structure that allows waves to propagate along one (or more) direction(s) while confining them in one or more other "transverse" directions. These arise in many forms in many different wave equations, with a variety of mechanisms for the transverse confinement: silica optical fibers, sound waves propagating down hollow tubes, coaxial transmission lines, and so forth. One commonplace confinement mechanism is known as "total internal reflection" or "index guiding" in optics [1], where waves are trapped in a region ("core") of higher refractive index (e.g. the core of an optical fiber) surrounded by a "cladding" of a lower index. It turns out that this mechanism is not specific to optics, and the general principle is that waves can be trapped in a medium where the wave speed (phase velocity) is *slower* than in the surrounding medium. Moreover, it turns out that we can *prove* the existence of such localized solutions quite generally—even in cases that we cannot solve analytically—using a "variational" proof that is closely related to proofs of localization ("bound states") in quantum mechanics (i.e., the Schrödinger equation) [2–4]. In these notes, we demonstrate this approach in the context of a 2d scalar wave equation, but closely related proofs can be applied to other wave equations (e.g. Maxwell's equations [5, 6]).

2 Review of scalar waves

In these notes, we consider the scalar wave equation:

$$-\hat{A}u = c^2\nabla^2 u = \frac{\partial^2 u}{\partial t^2},$$

where the coefficient $c(\mathbf{x}) > 0$, which has units of speed, is some function of $\mathbf{x}$ and we have defined the linear operator $\hat{A} = -c^2\nabla^2$. It is straightforward to show that $\hat{A}$ is Hermitian (i.e. self-adjoint, $\hat{A} = \hat{A}^*$, where $*$ denotes the adjoint) and positive semidefinite under a variety of boundary conditions on

u (e.g. u or its normal derivative vanishes at the boundary of the domain) with the inner product $\langle u, v\rangle = \int \frac{\bar{u}v}{c^2}$ (where $\bar{u}$ denote complex conjugation), i.e. $\langle u, \hat{A}v\rangle = \langle \hat{A}u, v\rangle$ and $\langle u, \hat{A}u\rangle = \int |\nabla u|^2 \leq 0$ for all[1] u, v.

For reasonable $c(\mathbf{x})$ we will have a spectral theorem,[2] i.e. a complete basis of eigenfunctions for $\hat{A}$, and to understand the solution of the wave equation it is enough to understand the eigenfunctions[3] $u(\mathbf{x})$ satisfying $\hat{A}u = \lambda u$. It is easy to show that the self-adjointness means that λ will be real, and the positive-semidefiniteness means that $\lambda \geq 0$, so it is convenient to write $\lambda = \omega^2$ where ω is real with units of frequency. Plugging this back into the wave equation, we see that $-\hat{A}u = -\omega^2 u = \ddot{u}$ and hence each eigensolution describes oscillating modes $u(\mathbf{x})e^{\pm i\omega t}$ with (eigen)frequency ω. These are the "normal modes:" another consequence of self-adjointness is that eigensolutions are *orthogonal*: given $\hat{A}u_1 = \omega_1^2 u_1$ and $\hat{A}u_2 = \omega_2^2 u_2$ with $\omega_1^2 \neq \omega_2^2$, it is easy to show that $\langle u_1, u_2\rangle = 0$.

If c is a constant (homogeneous c), the eigensolutions yield the planewaves $u(\mathbf{x}, t) = e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ where $\omega(\mathbf{k}) = \pm c|\mathbf{k}|$, in which case c is the magnitude of both the phase velocity ($\omega/|\mathbf{k}|$) and the group velocity ($|\nabla_k \omega|$): it is the speed at which waves travel and information propagates. For the case of inhomogeneous c considered below, the solutions and hence the velocities are more complicated, but we can still in some sense think of smaller-c regions as being "slower" materials.

3 Waveguide eigenproblems

To define a waveguide, we will consider the 2d domain $\mathbb{R}^2$ formed by the xz plane[4] (or some subset $X \times \mathbb{R}$ for $X \subseteq \mathbb{R}$), where $c(x)$ depends only on x and not z. In this case, the translational symmetry implies[5] that will have *separable*

[1] Or rather, for all u, v in the appropriate Sobolev space for this problem, which basically means that we restrict ourselves to functions where these integrals are defined. In these notes, I won't worry about pinning down the precise function spaces.

[2] For finite-dimensional Hermitian operators, you always have a basis of eigenvectors (diagonalizability). For infinite-dimensional operators (i.e. operators on functions), one can construct pathological counterexamples where this is not the case, but these are rarely relevant to physical problems—the counterexamples typically exhibit solutions that oscillate infinitely fast as they approach some point. So, physicists typically assume that all Hermitian problems have a basis of eigenfunctions. A more rigorous approach can be found in any book on functional analysis, e.g. Gohberg *et al.* [7], but rigorous treatment of non-constant coefficients c is often quite limited.

[3] In an infinite domain, functional analysts would call some of these solutions "generalized" eigenfunctions. e.g. if $c = 1$ on an infinite domain, the solutions are planewaves $u = e^{i\mathbf{k}\cdot\mathbf{x}}$ for real vectors $\mathbf{k}$ (disallowing solutions that grow exponentially at infinity), but these are not normalizable: $\langle u, u\rangle$ diverges, and u lives in a "rigged" Hilbert space rather than a Hilbert space. I will gloss over this and similar distinctions.

[4] We use xz and not xy for easy generalization to 3d: we always let z denote the propagation direction (the direction with continuous translational symmetry) while x and y are the transverse directions.

[5] The consequences of symmetry are a deep consequence of group representation theory [1, 8]: whenever a group of symmetry operations commutes with $\hat{A}$ (the problem is "symmetrical"), it follows that the eigenfunctions are partners of "irreducible representations" of the symmetry group. For continuous translational symmetry in z, this leads to e^{ikz} dependence.

eigensolutions of the form

$$u(x,z,t) = u_k(x)e^{i(kz-\omega t)},$$

for real "propagation constants" k (sometimes denoted β), which correspond to waves that are *propagating* in the z direction with phase velocity ω/k and group velocity $d\omega/dk$, where the eigenfrequencies $\omega(k)$ are the "dispersion relation" and satisfy a "reduced" eigenproblem:

$$\hat{A}_k u_k = \omega(k)^2 u_k$$

with

$$\hat{A}_k = e^{-ikz}\hat{A}e^{ikz} = c^2\left(-\frac{\partial^2}{\partial x^2} + k^2\right).$$

That is, for each k we have a 1d eigenproblem for the "mode profile" $u_k(x)$. Note that $\hat{A}_k = \hat{A}_k^* \geq 0$ is also self-adjoint positive semidefinite, like $\hat{A}$. Hence, for *each* value of k the eigenfunctions u_k are orthogonal: given $\hat{A}_k u_{k,1} = \omega_1(k)^2 u_{k,1}$ and $\hat{A}_k u_{k,2} = \omega_2(k)^2 u_{k,2}$ with $\omega_1 \neq \omega_2$, it follows that $\langle u_{k,1}, u_{k,2}\rangle = 0$, where here we have the 1d inner product $\langle u(x), v(x)\rangle = \int \frac{\bar{u}v}{c^2}dx$.

Waveguide modes (also called "guided" or "bound" modes) are solutions for which u_k is localized; at the very least so that $\int |u_k|^2 dx < \infty$, and typically $u_k(x)$ will decay at least exponentially fast with $|x|$. To obtain such solutions, we need some mechanism to trap waves. Obviously, guided modes will not occur if c = constant in an infinite domain $\mathbb{R}^2$, since in that case u_k will simply be a sinusoid [yielding planewave solutions $u = e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$]. The simplest way to construct guided modes is to have "hard walls" in the x direction, such as a finite domain $x \in [0, L]$ with Dirichlet boundary conditions $u_k(0) = u_k(L) = 0$, in which case c = constant gives eigenfunctions $u_k(x) = \sin(n\pi x/L)$ for $n = 1, 2, \ldots$ and $\omega_n(k) = c\sqrt{(n\pi/L)^2 + k^2}$. Physically, this corresponds to waves propagating within a "hollow tube" (in 2d). Instead, in the next section we will trap waves in an infinite domain by a "slow" region: a region of smaller c surrounded by a region of larger c.

4 Slow-wave guiding

Consider a 2d domain $\mathbb{R}^2$ with

$$c(x) = \begin{cases} c_1 & |x| < h/2 \\ c_0 & |x| \geq h/2 \end{cases},$$

where $c_1 < c_0$. That is, there is a width-h strip of "slower" material running along the z axis. The dispersion relation $\omega(k)$ of the eigensolutions of $\hat{A}_k$ for this c are shown in Fig. 1(left), and can be divided into two categories:

- The **light cone** (so-called in optics [1]): a continuous spectrum $\omega \geq c_0|k|$, consisting of all of the solutions that can exist in the homogeneous c_0

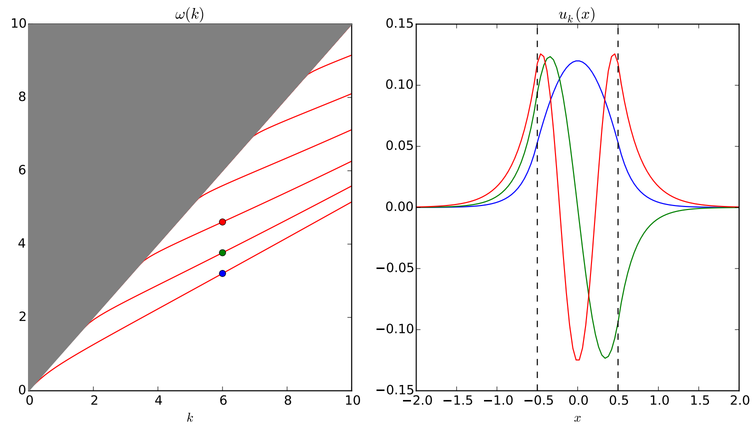

Figure 1: *Left:* Dispersion relation $\omega(k)$ of all of the eigenvalues ω^2 of $\hat{A}_k$ for a width-h ($h = 1$) strip of "slow" material $c_1 = 0.5$ surrounded by "fast" material $c_0 = 1$. The shaded region $\omega \geq c_0|k|$ is the "light cone," the continuous spectrum of all solutions that can propagate in the homogeneous c_0 region infinitely far from the strip, whereas the red curves are guided modes localized in the strip (and exponentially decaying in c_0). *Right:* Plot of three of the guided modes $u_k(x)$ at $k = 6$ (corresponding to the dots in the dispersion relation at left).

medium infinitely far from the strip. Intuitively, if you are very far from the strip, the solutions are just those of the homogeneous c_0 medium: planewaves $u_k e^{ikz} \approx e^{i(k_\perp x + kz)}$ for any real $k_\perp$, with a continuum of frequencies $\omega = c_0\sqrt{k^2 + k_\perp^2} \geq c_0|k|$.

- The **guided modes**: a discrete set of "bands" $\omega_n(k)$ lying below the light cone ($\leq c_0|k|$), which are *localized* in the vicinity of the strip. In the c_0 region, frequencies below the light cone correspond to purely *imaginary* $k_\perp = \pm i\sqrt{k^2 - (\omega/c_0)^2}$, and hence these solutions (shown at right in Fig. 1) are *exponentially decaying* in c_0.

Intuitively, decreasing c leads to a decrease of the eigenvalues of $-c^2\nabla^2$, so the $c_1 < c_0$ region "pulls down" solutions from the light cone, and once a solution falls below the light cone it is necessarily guided. This can be proved mathematically in a couple of ways (aside from numerical demonstrations). First, this particular 2d problem happens to be simple enough to be solvable analytically: essentially, you write u_k as a sine or cosine in the c_1 region and as decaying exponentials in the c_0 regions, and then match the solutions at $|x| = h/2$ to enforce continuity of u_k and u_k'. After tedious algebra, this leads to a transcendental equation that you can analyze to find $k_\perp$ and ω. However, that approach, besides being somewhat messy, is quite limited to this specific $c(x)$ in two dimensions. Instead, in the next section we will derive the existence of guided modes *without* solving the PDE, but in a *much more general* setting (that can even be extended to three dimensions).

4.1 Variational proof of slow-wave guiding

We want to show that the fundamental (lowest-ω) mode $\omega_1(k)$ is $< c_0|k|$, and hence is guided (exponentially decaying in c_0). To do this, we don't actually need to *compute* ω_1, we only need to find an *upper bound* for ω_1 and show that this bound is $< c_0|k|$. The key to finding an upper bound is the **min–max theorem**, also known as the **variational theorem**: for any self-adjoint operator $\hat{A}_k$, it is easy to show [1, 2] that the smallest eigenvalue ω_1^2 minimizes the Rayleigh quotient $R\{u\} = \langle u, \hat{A}_k u\rangle / \langle u, u\rangle$ over all $u(x)$, and in particular that $\omega_1^2 \leq R\{u\}$ for *any* $u(x)$. Hence, we just need to find *some* $u(x)$ such that $R\{u\} < c_0^2 k^2$ and it will follow that $\omega_1 < c_0|k|$: that a guided mode exists.

Multiplying both sides of $R\{u\} < c_0^2 k^2$ by $\langle u, u\rangle$ and integrating the numerator $\langle u, \hat{A}_k u\rangle$ once by parts, we obtain

$$\int_{-\infty}^{\infty} \left(|u'|^2 + k^2|u|^2\right) dx < c_0^2 k^2 \int_{-\infty}^{\infty} \frac{|u|^2}{c^2} dx.$$

Now, if we can find *some* function $u(x)$ for which this inequality is true, we are done. It turns out that there are many choices of $u(x)$ that will work, but one of the easiest is $u(x) = e^{-|x|/L}$, where L is a free parameter that we can choose to enforce the inequality. Before we do this, let us be more precise about $c(x)$. We could use this technique to analyze the specific piecewise-constant $c(x)$ from the previous section, but it turns out that we can prove localization for a much wider range of $c(x)$ functions. In particular, we write

$$\frac{1}{c(x)^2} = \frac{1}{c_0^2} + \Delta(x),$$

so that $\Delta(x) > 0 \Leftrightarrow c(x) < c_0$. We will allow $\Delta(x)$ to be *any* function as long as there is a localized "slow" ($c < c_0$) region *on average*, as defined by the following two conditions:

$$\int_{-\infty}^{\infty} \Delta(x)\, dx > 0 \qquad \text{(slow on average)},$$

$$\int_{-\infty}^{\infty} |\Delta(x)|\, dx < \infty \qquad (c \neq c_0 \text{is localized}).$$

We will now show that these two conditions are sufficient to imply that $\omega_1(k) < c_0|k|$ for all $k \neq 0$.

Plugging this $1/c^2$ into our desired inequality on u, we immediately see that the $k^2|u|^2$ terms cancel on both sides. Computing the integral $\int |u'|^2 dx = \frac{2}{L^2}\int_0^{\infty} e^{-2x/L} = \frac{1}{L}$, we are left with:

$$\frac{1}{L} < c_0^2 k^2 \int_{-\infty}^{\infty} \Delta(x)\, e^{-2|x|/L} dx.$$

For any $k \neq 0$, this is true for a sufficiently large choice of L. To show this, it suffices to show that the inequality holds in the limit $L \to \infty$. The limit of the left-hand side is 0, while the limit of the right hand side is

$$\lim_{L\to\infty} c_0^2 k^2 \int_{-\infty}^{\infty} \Delta(x)\, e^{-2|x|/L} dx = c_0^2 k^2 \int_{-\infty}^{\infty} \Delta(x) \lim_{L\to\infty} e^{-2|x|/L} dx = c_0^2 k^2 \int_{-\infty}^{\infty} \Delta(x)\, dx.$$

Interchanging limits and integration is not allowed in general, but it is okay under various conditions on the integrand that one can look up in analysis textbooks. In this case, it is okay because of something called "Lebesgue's dominated convergence theorem" [9].[6] Thus, in the $L \to \infty$ limit, our inequality becomes

$$0 < c_0^2 k^2 \int_{-\infty}^{\infty} \Delta(x)\, dx,$$

which is true for $k \neq 0$ due to our assumption $\int \Delta > 0$. If the inequality holds for $L \to \infty$, then it must also hold for some sufficiently large (but finite) L. Hence, there exists a $u(x)$ such that $\omega_1(k)^2 \leq R\{u\} < c_0^2 k^2$ and hence we have at least one guided mode with a frequency ω_1 below the light cone. Q.E.D.

5 Generalizations

There are many wave-localization problems that can be subjected to a similar analysis, such as waveguiding in Maxwell's equations [5,6] or localized states of a potential well ($\int V < 0$, $\int |V| < \infty$) in the Schrödinger operator $-\nabla^2 + V(\mathbf{x})$ of quantum mechanics [2]. For the Schrödinger operator $-\nabla^2 + V(\mathbf{x})$, one similarly wants to show $\int |\nabla u|^2 < -\int |u|^2 V$ for some trial function u in order to prove that the eigenvalue is negative (hence localized if $V \to 0$ for large $|x|$), so by comparison we see that $-\Delta$ in the wave equation above is playing the role of a "potential well."

The fact that the proof in the previous section was so easy (at the level of an undergraduate homework problem [2]), and would have worked for almost any reasonable choice of $u(x)$ where we have a parameter L to control the localization length, is essentially the consequence of a scaling relationship. Slightly rewritten, we wanted to show

$$\frac{\int |\nabla u|^2}{\int |u|^2} < c_0^2 k^2 \frac{\int |u|^2 \Delta}{\int |u|^2}.$$

Suppose that $\Delta \neq 0$ only in some region of diameter h, and $u = \psi(\mathbf{x}/L)$ for some square-integrable function ψ (so that L controls the lengthscale as above). Then the left-hand side scales as $1/L^2$, while the right-hand side scales as h/L (the fraction of $|u|^2$ in the $\Delta \neq 0$ region). Therefore, it is obvious that the inequality must hold for a sufficiently large L. In general, the consequence is that

[6] In particular, you can interchange limits and integration for $\int f(x)$ when $|f(x)| \leq g(x)$ for some $g(x)$ such that $\int g(x) < \infty$. In this case $|\Delta(x)\, e^{-2|x|/L}| \leq |\Delta(x)|$ and $\int |\Delta| < \infty$ by our assumption of a localized slow region.

localization is easy in one dimension—even for an arbitrarily weak "potential well" (small Δ and/or small k^2), one always has a localized eigenfunction, a familiar theorem in quantum mechanics.

In higher dimensions, however, the same scaling comparison tells us that matters are different. In d dimensions, the left-hand side still scales as $1/L^2$, but the right-hand side scales as $(h/L)^d$. This means that for $d > 2$ it is *not* advantageous to have a large L, so weak localization does not occur. For dimensions > 2, you must have a sufficiently strong "potential well" in order to localize a solution, again a well-known fact in quantum mechanics but true in other areas as well. For $d = 2$, both sides of the inequality scale in the same way with L, which tells us that two dimensions is a borderline case that requires a more delicate analysis.

In fact, it turns out that an arbitrarily weak potential well *does* produce a localized eigenfunction in 2d wave equations, but the proof is quite tricky. The first proof, in 1976, involved extremely sophisticated functional analysis [10]. Much simpler, variational-style proofs were discovered in the 1980s [3–5], but involved more complicated trial functions like $u(\mathbf{x}) \sim e^{-(|\mathbf{x}|+1)^\alpha}$ in the $\alpha \to 0$ limit, which are very tricky to integrate (in practice, one just bounds the integrals). The analogous problem for waveguides is a 3d waveguide in which the solutions are localized in two transverse dimensions (e.g. xy), and again it turns out that there is indeed a guided mode for every $k \neq 0$, but the proofs are again complicated, especially in the vectorial Maxwell wave equation and for complicated media [5, 6]. Another interesting case is to prove localization within a "gap" in the continuous spectrum of the surrounding medium (rather than for ω_1 below the infimum of the spectrum), which can happen when c_0 is periodic rather than constant, and this has been done for the Schrödinger operator with a related variational proof [4].

References

[1] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light*. Princeton University Press, second ed., February 2008.

[2] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*. Oxford: Butterworth-Heinemann, 3rd ed., 1977.

[3] K. Yang and M. de Llano, "Simple variational proof that any two-dimensional potential well supports at least one bound state," *Am. J. Phys.*, vol. 57, no. 1, pp. 85–86, 1989.

[4] A. Parzygnat, K. K. Y. Lee, Y. Avniel, and S. G. Johnson, "Sufficient conditions for two-dimensional localization by arbitrarily weak defects in periodic potentials with band gaps," *Physical Review B*, vol. 81, p. 155324, April 2010.

[5] A. Bamberger and A. S. Bonnet, "Mathematical analysis of the guided modes of an optical fiber," *SIAM J. Math. Anal.*, vol. 21, pp. 1487–1510, 1990.

[6] K. K. Y. Lee, Y. Avniel, and S. G. Johnson, "Rigorous sufficient conditions for index-guided modes in microstructured dielectric waveguides," *Optics Express*, vol. 16, pp. 9261–9275, June 2008.

[7] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Basic Classes of Linear Operators*. Basel: Birkhäuser, 2000.

[8] T. Inui, Y. Tanabe, and Y. Onodera, *Group Theory and Its Applications in Physics*. Heidelberg: Springer, 1996.

[9] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 3rd ed., 1987.

[10] B. Simon, "The bound state of weakly coupled Schrödinger operators in one and two dimensions," *Ann. Phys.*, vol. 97, no. 2, pp. 279–288, 1976.