Lecture 22

Quality Factor of Cavities, Mode Orthogonality

Cavity resonators are important for making narrow band filters. The bandwidth of a filter is related to the Q or the quality factor of the cavity. A concatenation of cavity resonators can be used to engineer different filter designs. Resonators can also be used to design various sensing systems, as well as measurement systems. We will study the concept of Q in this lecture.

Also, before we leave the lectures on waveguides and resonators, it will be prudent to discuss mode orthogonality. Since this concept is very similar to eigenvector orthogonality found in matrix or linear algebra, we will relate mode orthogonality in waveguides and cavities to eigenvector orthogonality.

22.1 The Quality Factor of a Cavity—General Concept

The quality factor of a cavity or its Q measures how ideal or lossless a cavity resonator is. An ideal lossless cavity resonator will sustain free oscillations forever, while most resonators sustain free oscillations for a finite time. This is because of losses coming from radiation, dissipation in the dielectric material filling the cavity, or resistive loss of the metallic part of the cavity.

22.1.1 Analogue with an LC Tank Circuit

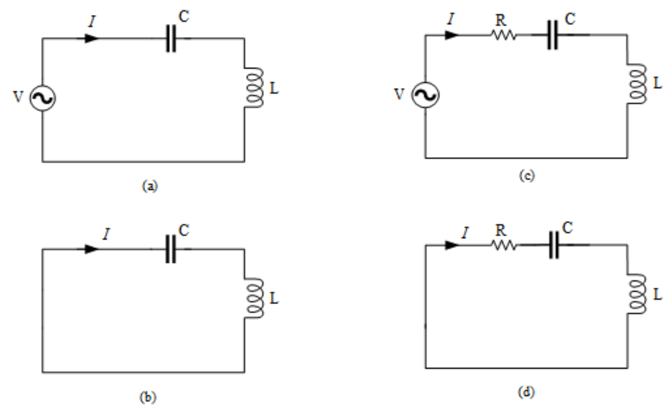

Figure 22.1: For the circuit on the left, it will resonate forever even if the source is turned off. But for the circuit on the right, the current in the circuit will decay with time due to dissipation in the resistor.

One of the simplest resonators imaginable is the LC tank circuit. By using it as an analogue, we can better understand the resonance of a cavity. When there is no loss in an LC tank circuit, it can oscillate forever. Moreover, if we turn off the source, a free oscillation solution exists.[1]

One can write the voltage-current relation in the circuit as

$$I(\omega) = \frac{V(\omega)}{j\omega L + 1/(j\omega C)} = V(\omega)Y(\omega) \tag{22.1.1}$$

where

$$Y(\omega) = \frac{1}{j\omega L + 1/(j\omega C)} \tag{22.1.2}$$

The above $Y(\omega)$ can be thought of as the transfer function of the linear system where the input is $V(\omega)$ and the output is $I(\omega)$. When the voltage is zero or turned off, a non-zero current exists or persists at the resonance frequency of the oscillator. The resonant frequency is when the denominator in the above equation is zero, so that I is finite despite $V = 0$.[2] This resonant frequency, obtained by setting the denominator of Y to zero, is given by $\omega_R = 1/\sqrt{LC}$.

When a small resistor is added in the circuit to give rise to loss, the voltage-current relation

[1] This is analogous to the homogeneous solution of an ordinary differential equation.
[2] We take advantage of the fact that zero divided by zero is undefined.

becomes

$$I(\omega) = \frac{V(\omega)}{j\omega L + R + 1/(j\omega C)} = V(\omega)Y(\omega) \tag{22.1.3}$$

$$Y(\omega) = \frac{1}{j\omega L + R + 1/(j\omega C)} \tag{22.1.4}$$

Now, the denominator of the above functions can never go to zero for real ω. But there exists complex ω that will make Y become infinite. These are the complex resonant frequencies of the circuit. The homogeneous solution (also called the natural solution, or free oscillation) can only exist at the complex resonant frequencies. With complex resonances, the voltage and the current are decaying sinusoids.

By the same token, because of losses, the free oscillation in a cavity has electromagnetic field with time dependence as follows:

$$\mathbf{E} \propto e^{-\alpha t}\cos(\omega t + \phi_1), \quad \mathbf{H} \propto e^{-\alpha t}\cos(\omega t + \phi_2) \tag{22.1.5}$$

That is, they are decaying sinusoids. The total time-average stored energy, which is proportional to $\frac{1}{4}\epsilon\,|\mathbf{E}|^2 + \frac{1}{4}\mu\,|\mathbf{H}|^2$ is of the form

$$\langle W_T\rangle = \langle W_E\rangle + \langle W_H\rangle \doteq W_0 e^{-2\alpha t} \tag{22.1.6}$$

If there is no loss, $\langle W_T\rangle$ will remain constant. However, with loss, the average stored energy will decrease to $1/e$ of its original value at $t = \tau = \frac{1}{2\alpha}$. The Q of a cavity is defined as the number of free oscillations in radians (rather than cycles) that the field undergoes before the energy stored decreases to $1/e$ of its original value (see Figure 22.2). In a time interval $\tau = \frac{1}{2\alpha}$, the number of free oscillations in radians is $\omega\tau$ or $\frac{\omega}{2\alpha}$; hence, the Q is defined to be [33]

$$Q \doteq \frac{\omega}{2\alpha} \tag{22.1.7}$$

Q is an approximate concept, and makes sense only if the system has low loss.

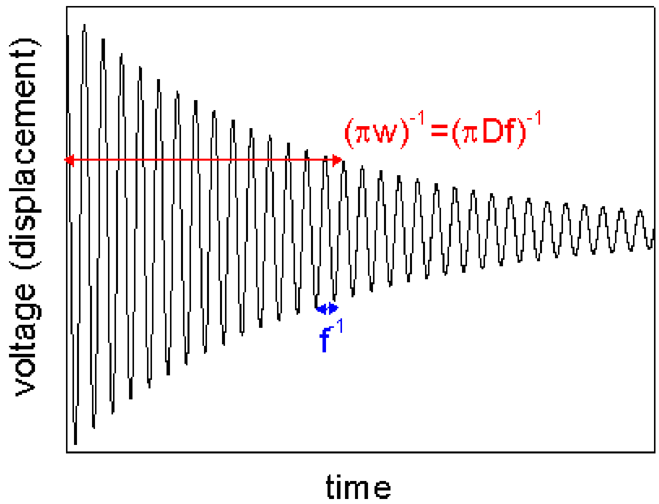

Figure 22.2: A typical time domain response of a high Q system (courtesy of Wikipedia).

Furthermore, by energy conservation, the decrease in stored energy per unit time must be equal to the total power dissipated in the losses of a cavity. In other words,

$$\langle P_D \rangle = -\frac{d\langle W_T \rangle}{dt} \tag{22.1.8}$$

By further assuming that W_T has to be of the form in (22.1.6), then

$$-\frac{d\langle W_T \rangle}{dt} \doteq 2\alpha W_0 e^{-2\alpha t} = 2\alpha \langle W_t \rangle \tag{22.1.9}$$

From the above, we can estimate the decay constant

$$\alpha \doteq \frac{\langle P_D \rangle}{2\langle W_T \rangle} \tag{22.1.10}$$

Hence, we can rewrite equation (22.1.7) for the Q as

$$Q \doteq \frac{\omega \langle W_T \rangle}{\langle P_D \rangle} \tag{22.1.11}$$

By further letting $\omega = 2\pi/T$, we lend further physical interpretation to express Q as

$$Q \doteq 2\pi \frac{\langle W_T \rangle}{\langle P_D \rangle T} = 2\pi \frac{\text{total energy stored}}{\text{Energy dissipated/cycle}} \tag{22.1.12}$$

In a cavity, the energy can dissipate in either the dielectric loss or the wall loss of the cavity due to the finiteness of the conductivity. It has to be re-emphasized the Q is a low-loss, asymptotic concept, and hence, the above formulas are only approximately true.

22.1.2 Relation to Bandwidth and Pole Location

The resonance of a system is related to the pole of the transfer function . For instance, in our previous example of the RLC tank circuit, the admittance Y can be thought of as a transfer function in linear system theory: The input is the voltage, while the output is the current. If we encounter the resonance of the system at a particular frequency, the transfer function becomes infinite. This infinite value can be modeled by a pole of the transfer function in the complex ω plane. In other words, in the vicinity of the pole in the frequency domain, the transfer function $Y(\omega)$ in (22.1.4) of the system can be approximated by a single pole which is

$$Y(\omega) \sim \frac{A}{\omega - \omega_R} = \frac{A}{\omega - \omega_0 - j\alpha} \tag{22.1.13}$$

where we have assumed that $\omega_R = \omega_0 + j\alpha$, the resonant frequency is complex. In principle, when $\omega = \omega_R$, the transfer function $Y(\omega)$ becomes infinite, but this does not happen in practice because ω_R is complex, and ω, the operating frequency is real. In other words, the pole is displaced slightly off the real axis to account for loss. Using a single pole approximation, it is quite clear that $|Y(\omega)|$ would peak at $\omega = \omega_0$. At $\omega = \omega_0 \pm \alpha$, the magnitude of $|Y(\omega)|$ will be $1/\sqrt{2}$ smaller, or that the power which is proportional to $|Y(\omega)|^2$ will be half as small. Therefore, the half-power points compared to the peak are at $\omega = \omega_0 \pm \alpha$. Thus, the full-width half maximum (FWHM) bandwidth is defined to be $\Delta\omega = 2\alpha$. And the Q can be written as in terms of the half-power bandwidth $\Delta\omega$ of the system, viz.,

$$Q = \omega/(2\alpha) \doteq \omega_0/\Delta\omega \tag{22.1.14}$$

The above implies that the narrower the bandwidth, the higher the Q of the system. Typical plots of transfer function versus frequency for a system with different Q's are shown in Figure 22.3.

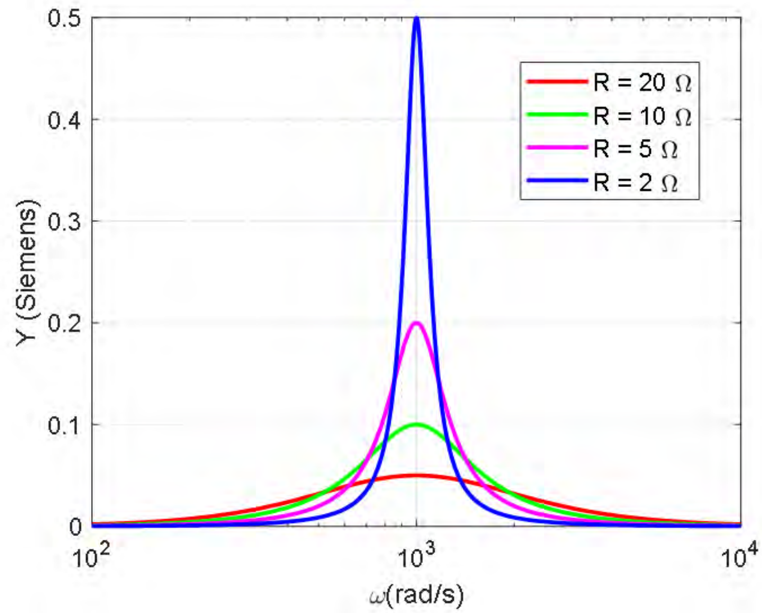

Figure 22.3: A typical system response versus frequency for different Q's using (22.1.4). The Q is altered by changing the resistor R in the circuit.

22.1.3 Wall Loss and Q for a Metallic Cavity—A Perturbation Concept

To estimate the Q of a cavity, we will need to calculate the loss inside the cavity as well as the energy stored according to (22.1.11). We can use perturbation concept to estimate the Q. First, we assume a lossless cavity so that the cavity wall is made from PEC. In this case, $\hat{n}\times\mathbf{E}=0$ and no power can be absorbed by the waveguide wall. Then we assume a small loss by assuming now that the cavity wall is made of imperfect conductors and hence, $\hat{n}\times\mathbf{E}\neq 0$ but small. We can assume that the magnetic field $\mathbf{H}$ remains unchanged before and after we have introduced loss.

If the cavity is filled with air, then the loss comes mainly from the metallic loss or copper-loss from the cavity wall. In this case, the power dissipated on the wall is given by [33]

$$\langle P_D\rangle = \frac{1}{2}\Re e \oint_S (\mathbf{E}\times\mathbf{H}^*)\cdot\hat{n}\,dS = \frac{1}{2}\Re e \oint_S (\hat{n}\times\mathbf{E})\cdot\mathbf{H}^*\,dS \tag{22.1.15}$$

where S is the surface of the cavity wall.[3] Here, $(\hat{n}\times\mathbf{E})$ is the tangential component of the electric field which would have been zero if the cavity wall is made of ideal PEC. Also,

[3]We have used the cyclic identity that $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})$ in the above (see Some Useful Mathematical Formulas).

$\hat{n}$ is taken to be the outward pointing normal at the surface S. The β (or k) vector in the transmitted medium is very large due to the high-conductivity of the wall. Due to the phase-matching condition, the transmitted wave β vector is almost normal to the interface. Therefore, we can approximate the transmitted wave as a wave propagating normal to the interface. In the metal, it decays predominantly in the direction of propagation which is normal to the surface as well.

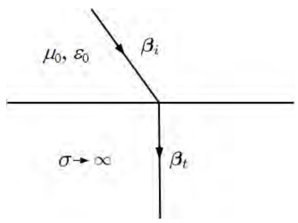

Figure 22.4: Due to high σ in the metal, and large $|\boldsymbol{\beta}_t|$, phase matching condition requires the transmitted wave vector to be almost normal to the surface. Then the field can be assumed to be a normally incident plane wave field in the transmitted or the metal region. Hence, the tangential electric field is almost zero, and the tangential magnetic field remains the same using this perturbative concept.

For such a wave, we can approximate $\hat{n} \times \mathbf{E} = \mathbf{H}_t Z_m$ where Z_m is the intrinsic impedance for the metallic conductor, as shown in Section 8.1, which is $Z_m = \sqrt{\frac{\mu}{\varepsilon_m}} \approx \sqrt{\frac{\mu}{-j\frac{\sigma}{\omega}}} = \sqrt{\frac{\omega\mu}{2\sigma}}(1+j)$,[4] where we have assumed that $\varepsilon_m \approx -\frac{j\sigma}{\omega}$, and $\mathbf{H}_t$ is the tangential magnetic field. From these equations, we can see that the tangential $\mathbf{E}$ field is small but tangential $\mathbf{H}$ field is not small.

This relation between $\mathbf{E}$ and $\mathbf{H}$ will ensure that power is flowing into the metallic surface. Hence,

$$\langle P_D \rangle = \frac{1}{2}\Re e \oint_S \sqrt{\frac{\omega\mu}{2\sigma}}(1+j)\,|\mathbf{H}_t|^2\, dS = \frac{1}{2}\sqrt{\frac{\omega\mu}{2\sigma}} \oint_S |\mathbf{H}_t|^2\, dS \tag{22.1.16}$$

By further assuming that the stored electric and magnetic energies of a cavity are equal to each other at resonance, the stored energy can be obtained by

$$\langle W_T \rangle = \frac{1}{2}\mu \int_V |\mathbf{H}|^2 dV == \frac{1}{2}\varepsilon \int_V |\mathbf{E}|^2 dV \tag{22.1.17}$$

Written explicitly, the Q becomes

$$Q = \sqrt{2\omega\mu\sigma}\frac{\int_V |\mathbf{H}|^2\, dV}{\oint_S |\mathbf{H}_t|^2\, dS} = \frac{2}{\delta}\frac{\int_V |\mathbf{H}|^2\, dV}{\oint_S |\mathbf{H}_t|^2\, dS} \tag{22.1.18}$$

[4]When an electromagnetic wave enters a conductive region with a large β, it can be shown that the wave is refracted to propagate normally to the surface as shown in Figure 22.4, and hence, this formula can be applied.

In the above, δ is the skin depth of the metallic wall. Hence, the more energy stored there is with respect to the power dissipated, the higher the Q of a resonating system. Also, the lower the metal loss, or the smaller the skin depth, the higher the Q would be.

Notice that in (22.1.18), the numerator is a volume integral and hence, is proportional to volume, while the denominator is a surface integral and is proportional to surface. Thus, the Q, a dimensionless quantity, is roughly proportional to

$$Q \sim \frac{V}{S\delta} \tag{22.1.19}$$

where V is the volume of the cavity, while S is its surface area. From the above, it is noted that a large cavity compared to its skin depth has a larger Q than an small cavity.

It is easy to make large cavities in optics as the wavelength is small. Also, dielectric resonator cavity can be made out of glass where the primary loss will be from the material. Quality factor $\approx 10^8 \sim 10^9$ is possible [144]. A dielectric resonator using total internal reflection to trap the wave as it bounces around is called a whispering gallery mode resonator. The glass can be made with very low loss. This together with the large size of the cavity compared to wavelength gives the resonator very high Q. The large cavity size increases the stored energy as well, which is good for increasing its Q.

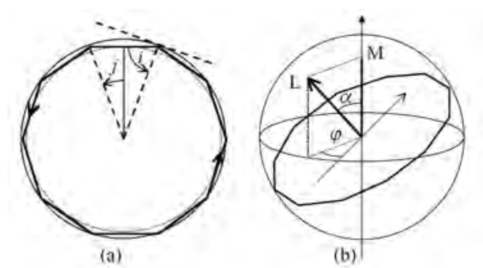

Figure 22.5: A dielectric resonator can be made by using total internal reflection of the bouncing wave within the resonator [144]. Such a mode is called a whispering gallery mode. It has high Q because the glass making the cavity has little loss, and that the cavity can be very large compared to wavelength, increasing the stored energy.

22.1.4 Example: The Q of TM_{110} Mode

For the TM_{110} mode, as can be seen from the previous lecture, the only electric field is $\mathbf{E} = \hat{z}E_z$, with $\partial/\partial z = 0$. Then

$$E_z = E_0 \sin\left(\frac{\pi x}{a}\right) \sin\left(\frac{\pi y}{b}\right) \tag{22.1.20}$$

The magnetic field can be derived from the electric field using Maxwell's equation or Faraday's law, and

$$H_x = \frac{j\omega\epsilon}{\omega^2\mu\epsilon}\frac{\partial}{\partial y}E_Z = \frac{j\left(\frac{\pi}{b}\right)}{\omega\mu}E_0 \sin\left(\frac{\pi x}{a}\cos\right)\left(\frac{\pi y}{b}\right) \tag{22.1.21}$$

$$H_y = \frac{-j\omega\epsilon}{\omega^2\mu\epsilon}\frac{\partial}{\partial x}E_Z = -\frac{j\left(\frac{\pi}{a}\right)}{\omega\mu}E_0 \cos\left(\frac{\pi x}{a}\right)\sin\left(\frac{\pi y}{b}\right) \tag{22.1.22}$$

Therefore[5]

$$\begin{aligned}\int_V |\mathbf{H}|^2\, dV &= \int_{-d}^{0}\int_0^b\int_0^a dx\,dy\,dz\left[|H_x|^2 + |H_y|^2\right]\\ &= \frac{|E_0|^2}{\omega^2\mu^2}\int_{-d}^{0}\int_0^b\int_0^a dx\,dy\,dz\\ &\quad \left[\left(\frac{\pi}{b}\right)^2 \sin^2\left(\frac{\pi x}{a}\right)\cos^2\left(\frac{\pi y}{b}\right) + \left(\frac{\pi}{a}\right)^2 \cos^2\left(\frac{\pi x}{a}\right)\sin^2\left(\frac{\pi y}{b}\right)\right]\\ &= \frac{|E_0|^2}{\omega^2\mu^2}\frac{\pi^2}{4}\left[\frac{a}{b}+\frac{b}{a}\right]d \end{aligned} \tag{22.1.23}$$

A cavity has six faces, finding the tangential exponent at each face and integrate

$$\begin{aligned}\oint_S |\mathbf{H}_t|\, dS &= 2\int_0^b\int_0^a dx\,dy\left[|H_x|^2 + |H_y|^2\right]\\ &\quad + 2\int_{-d}^{0}\int_0^a dx\,dz\,|H_x(y=0)|^2 + 2\int_{-d}^{0}\int_0^b dy\,dz\,|H_y(x=0)|^2\\ &= \frac{2|E_0|^2}{\omega^2\mu^2}\frac{\pi^2 ab}{4}\left[\frac{1}{a^2}+\frac{1}{b^2}\right] + \frac{2\left(\frac{\pi}{b}\right)^2}{\omega^2\mu^2}|E_0|^2\frac{ad}{2} + \frac{2\left(\frac{\pi}{a}\right)^2}{\omega^2\mu^2}|E_0|^2\frac{bd}{2}\\ &= \frac{\pi^2|E_0|^2}{\omega^2\mu^2}\left[\frac{b}{2a}+\frac{a}{2b}+\frac{ad}{b^2}+\frac{bd}{a^2}\right] \end{aligned} \tag{22.1.24}$$

Hence the Q is

$$Q = \frac{2}{\delta}\frac{\left(\frac{ad}{b}+\frac{bd}{a}\right)}{\left(\frac{b}{2a}+\frac{a}{2b}+\frac{ad}{b^2}+\frac{bd}{a^2}\right)} \tag{22.1.25}$$

The result shows that the larger the cavity, the higher the Q. This is because the Q, as mentioned before, is the ratio of the energy stored in a volume to the energy dissipated over the surface of the cavity.

[5]Since the electric field is simpler than the magnetic field, it is easier to find the energy stored using the electric field. Like a LC tank circuit, the magnetic field energy stored and the electric field energy stored are equal to each other.

22.2 Mode Orthogonality and Matrix Eigenvalue Problem

It turns out that the modes of a waveguide or a resonator are orthogonal to each other. This is intimately related to the orthogonality of eigenvectors of a matrix operator.[6] Thus, it is best to understand this by the homomorphism between the electromagnetic mode problem and the matrix eigenvalue problem. Because of this similarity, electromagnetic modes are also called eigenmodes. Thus it is prudent that we revisit the matrix eigenvalue problem (EVP) here.

22.2.1 Matrix Eigenvalue Problem (EVP)

It is known in matrix theory that if a matrix is hermitian, then its eigenvalues are all real. Furthermore, their eigenvectors with distinct eigenvalues are orthogonal to each other [79]. Assume that an eigenvalue and an eigenvector exists for the hermitian matrix $\overline{\mathbf{A}}$. Then

$$\overline{\mathbf{A}} \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{22.2.1}$$

Dot multiplying the above from the left by $\mathbf{v}_i^\dagger$ where $\dagger$ indicates conjugate transpose, then the above becomes

$$\mathbf{v}_i^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_i^\dagger \cdot \mathbf{v}_i \tag{22.2.2}$$

Since $\overline{\mathbf{A}}$ is hermitian, or $\overline{\mathbf{A}}^\dagger = \overline{\mathbf{A}}$, then the quantity $\mathbf{v}_i^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{v}_i$ is purely real. Moreover, the quantity $\mathbf{v}_i^\dagger \cdot \mathbf{v}_i$ is positive real.[7] So in order for the above to be satisfied, λ_i has to be real.

To prove orthogonality of eigenvectors, now, assume that $\overline{\mathbf{A}}$ has two eigenvectors with distinct eigenvalues such that

$$\overline{\mathbf{A}} \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{22.2.3}$$

$$\overline{\mathbf{A}} \cdot \mathbf{v}_j = \lambda_j \mathbf{v}_j \tag{22.2.4}$$

Left dot multiply the first equation with $\mathbf{v}_j^\dagger$ and do the same to the second equation with $\mathbf{v}_i^\dagger$, one gets

$$\mathbf{v}_j^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_j^\dagger \cdot \mathbf{v}_i \tag{22.2.5}$$

$$\mathbf{v}_i^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{v}_j = \lambda_j \mathbf{v}_i^\dagger \cdot \mathbf{v}_j \tag{22.2.6}$$

Taking the conjugate transpose of (22.2.5) in the above, and since $\overline{\mathbf{A}}$ is hermitian, their left-hand sides (22.2.5) and (22.2.6) become the same. Subtracting the two equations, we arrive at

$$0 = (\lambda_i - \lambda_j) \mathbf{v}_j^\dagger \cdot \mathbf{v}_i \tag{22.2.7}$$

[6] This mathematical homomorphism is not discussed in any other electromagnetic textbooks.
[7] Convince yourself that this is the case if you are dubious.

For distinct eigenvalues, $\lambda_i \neq \lambda_j$, the only way for the above to be satisfied is that

$$\mathbf{v}_j^\dagger \cdot \mathbf{v}_i = C_i \delta_{ij} \tag{22.2.8}$$

Hence, eigenvectors of a hermitian matrix with distinct eigenvalues are orthogonal to each other. The eigenvalues are also real.

22.2.2 Homomorphism with the Waveguide Mode Problem

We shall next show that the problem for finding the waveguide modes or eigenmodes is analogous to the matrix eigenvalue problem as well. The governing equation for a waveguide mode is a BVP involving the reduced wave equation previously derived in (18.1.14), or

$$\nabla_s^2 \psi_i(\mathbf{r}_s) + \beta_{is}^2 \psi_i(\mathbf{r}_s) = 0, \Rightarrow -\nabla_s^2 \psi_i(\mathbf{r}_s) = \beta_{is}^2 \psi_i(\mathbf{r}_s) \tag{22.2.9}$$

with the pertinent homogeneous Dirichlet or Neumann boundary condition, depending on if TE or TM modes are considered. In the above, the differential operator ∇_s^2 is analogous to the matrix operator $\overline{\mathbf{A}}$, the eigenfunction $\psi_i(\mathbf{r}_s)$ is analogous to the eigenvector $\mathbf{v}_i$, and β_{is}^2 is analogous to the eigenvalue λ_i.

In the above, we can think of ∇_s^2 as a linear operator that maps a function to another function, viz.,

$$\nabla_s^2 \psi_i(\mathbf{r}_s) = f_i(\mathbf{r}_s) \tag{22.2.10}$$

Hence ∇_s^2 is analogous to a matrix operator $\overline{\mathbf{A}}$, that maps a vector to another vector, viz.,

$$\overline{\mathbf{A}} \cdot \mathbf{x} = \mathbf{f} \tag{22.2.11}$$

We shall elaborate this further next.

Discussion on Functional Space

To think of a function $\psi(\mathbf{r})$ as a vector, where $\mathbf{r}$ is a position vector in 2D or 3D space, one has to think in the discrete world.[8] If one needs to display the function $\psi(\mathbf{r})$, on a computer, one will evaluate the function $\psi(\mathbf{r})$ at discrete N locations $\mathbf{r}_l$, where $l = 1, 2, 3, \ldots, N$. For every $\mathbf{r}_l$ or every l, there is a scalar number $\psi(\mathbf{r}_l)$. These scalar numbers can be stored in a column vector in a computer indexed by l. The larger N is, the better is the discrete approximation of $\psi(\mathbf{r})$. In theory, N needs to be infinite to describe this function exactly.[9]

From the above discussion, a function is analogous to a vector and a functional space is analogous to a vector space in linear or matrix algebra. However, a functional space is infinite dimensional whereas a matrix vector $\mathbf{v}$ is usually in a finite dimensional space. But in order to compute on a computer with finite resource, such functions are approximated with finite length vectors. Or infinite dimensional vector spaces are replaced with finite dimensional ones to make the problem computable. Such an infinite dimensional functional space is also

[8] Some of these concepts are discussed in [36, 145].

[9] In mathematical parlance, the index for $\psi(\mathbf{r})$ is uncountably infinite or nondenumerable.

called a Hilbert space. Hilbert space are functional spaces where all the functions are square integrable, implying that they have finite energy. This is usually true for physical systems.

It is also necessary to define the inner product between two vectors in a functional space just as inner product between two vectors in an matrix vector space. The inner product (or dot product) between two vectors in matrix vector space is

$$\mathbf{v}_i^t \cdot \mathbf{v}_j = \sum_{l=1}^{N} v_{i,l} v_{j,l} \tag{22.2.12}$$

The analogous inner product between two vectors in function space is[10]

$$\langle \psi_i, \psi_j \rangle = \int_S d\mathbf{r}_s \psi_i(\mathbf{r}_s)\psi_j(\mathbf{r}_s), \text{ 2D problems}, \qquad \langle \psi_i, \psi_j \rangle = \int_V d\mathbf{r} \psi_i(\mathbf{r})\psi_j(\mathbf{r}), \text{ 3D problems} \tag{22.2.13}$$

where S here can denote the cross-sectional area of the waveguide, and V can denote the volume of a cavity, over which the integration is performed. The left-hand side is the shorthand notation for inner product in functional space or the infinite dimensional functional Hilbert space.

Another requirement for a vector in a functional Hilbert space is that it contains finite energy. In 2D, this can be expressed as

$$\mathscr{E}_f = \int_S d\mathbf{r}_s |\psi_i(\mathbf{r}_s)|^2 \tag{22.2.14}$$

is finite. The above is analogous to that for a matrix vector $\mathbf{v}$ as

$$\mathscr{E}_m = \sum_{l=1}^{N} |v_l|^2 \tag{22.2.15}$$

The square root of the above is often used to denote the "length", the "metric", or the "norm" of the vector. Finite energy also implies that the vectors are of finite length.

22.2.3 Proof of Orthogonality of Waveguide Modes[11]

Because of the aforementioned discussion, we see the similarity between a functional Hilbert space, and the matrix vector space. In order to use the result of the matrix EVP, one key step is to prove that the operator ∇_s^2 is hermitian. In matrix algebra, if the elements of a matrix is explicitly available, a matrix $\overline{\mathbf{A}}$ is hermitian if

$$A_{ij} = A_{ji}^* \tag{22.2.16}$$

[10]In many math books, the conjugation of the first function ψ_i is implied, but here, we follow the electromagnetic convention that the conjugation of ψ_i is not implied unless explicitly stated.

[11]This may be skipped on first reading.

But if an operator, such as a Laplacian operator, is not described by a matrix, and hence, a matrix element is not explicitly available, a different way to define hermiticity is required: we will use inner products to define it. A matrix operator is hermitian if

$$\mathbf{x}_i^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{x}_j = \left(\mathbf{x}_j^\dagger \cdot \overline{\mathbf{A}}^\dagger \cdot \mathbf{x}_i\right)^\dagger = \left(\mathbf{x}_j^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{x}_i\right)^\dagger = \left(\mathbf{x}_j^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{x}_i\right)^* \tag{22.2.17}$$

The first equality follows from standard matrix algebra,[12] the second equality follows if $\overline{\mathbf{A}} = \overline{\mathbf{A}}^\dagger$, or that $\overline{\mathbf{A}}$ is hermitian. The last equality follows because the quantity in the parenthesis is a scalar, and hence, its conjugate transpose is just its conjugate. Therefore, if a matrix $\overline{\mathbf{A}}$ is hermitian, then

$$\mathbf{x}_i^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{x}_j = \left(\mathbf{x}_j^\dagger \cdot \overline{\mathbf{A}} \cdot \mathbf{x}_i\right)^* \tag{22.2.18}$$

The above inner product can be used to define the hermiticity of the matrix operator $\overline{\mathbf{A}}$. It can be easily extended to define the hermiticity of an operator in an infinite dimensional Hilbert space where the matrix elements are not explicitly available, but an inner product is defined.

Hence, using the inner product definition in (22.2.13) for infinite dimensional functional Hilbert space, a functional operator ∇_s^2 is hermitian if

$$\langle \psi_i^*, \nabla_s^2 \psi_j \rangle = (\langle \psi_j^*, \nabla_s^2 \psi_i \rangle)^* \tag{22.2.19}$$

We can rewrite the left-hand side of the above more explicitly as

$$\langle \psi_i^*, \nabla_s^2 \psi_j \rangle = \int_S d\mathbf{r}_s \psi_i^*(\mathbf{r}_s) \nabla_s^2 \psi_j(\mathbf{r}_s) \tag{22.2.20}$$

and then the right-hand side of (22.2.19) can be rewritten more explicitly as

$$(\langle \psi_j^*, \nabla_s^2 \psi_i \rangle)^* = \int_S d\mathbf{r}_s \psi_j(\mathbf{r}_s) \nabla_s^2 \psi_i^*(\mathbf{r}_s) \tag{22.2.21}$$

To prove the above equality in (22.2.19), one uses the identity that

$$\nabla_s \cdot [\psi_i^*(\mathbf{r}_s) \nabla_s \psi_j(\mathbf{r}_s)] = \psi_i^*(\mathbf{r}_s) \nabla_s^2 \psi_j(\mathbf{r}_s) + \nabla_s \psi_i^*(\mathbf{r}_s) \cdot \nabla_s \psi_j(\mathbf{r}_s) \tag{22.2.22}$$

Integrating the above over the cross sectional area S, and invoking Gauss divergence theorem in 2D, one gets that

$$\int_C dl \hat{n} \cdot (\psi_i^*(\mathbf{r}_s) \nabla_s \psi_j(\mathbf{r}_s)) = \int_S d\mathbf{r}_s \left(\psi_i^*(\mathbf{r}_s) \nabla_s^2 \psi_j(\mathbf{r}_s)\right) + \int_S d\mathbf{r}_s \left(\nabla \psi_i^*(\mathbf{r}_s) \cdot \nabla_s \psi_j(\mathbf{r}_s)\right) \tag{22.2.23}$$

[12] $(\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} \cdot \overline{\mathbf{C}})^\dagger = \overline{\mathbf{C}}^\dagger \cdot \overline{\mathbf{B}}^\dagger \cdot \overline{\mathbf{A}}^\dagger$ [79].

where C the the contour bounding S or the waveguide wall. By applying the boundary condition that $\psi_i(\mathbf{r}_s) = 0$ or that $\hat{n} \cdot \nabla_s \psi_j(\mathbf{r}_s) = 0$, or a mixture thereof, then the left-hand side of the above is zero. This will be the case be it TE or TM modes.

$$0 = \int_S d\mathbf{r}_s \left(\psi_i^*(\mathbf{r}_s)\nabla_s^2\psi_j(\mathbf{r}_s)\right) + \int_S d\mathbf{r}_s \left(\nabla\psi_i^*(\mathbf{r}_s) \cdot \nabla_s\psi_j(\mathbf{r}_s)\right) \tag{22.2.24}$$

Applying the same treatment to (22.2.21), we get

$$0 = \int_S d\mathbf{r}_s \left(\psi_j(\mathbf{r}_s)\nabla_s^2\psi_i^*(\mathbf{r}_s)\right) + \int_S d\mathbf{r}_s \left(\nabla\psi_i^*(\mathbf{r}_s) \cdot \nabla_s\psi_j(\mathbf{r}_s)\right) \tag{22.2.25}$$

The above two equations indicate that

$$\langle \psi_i^*, \nabla_s^2 \psi_j \rangle = (\langle \psi_j^*, \nabla_s^2 \psi_i \rangle)^* \tag{22.2.26}$$

proving that the operator ∇_s^2 is hermitian. One can then use the above property to prove the orthogonality of the eigenmodes when they have distinct eigenvalues, the same way we have proved the orthogonality of eigenvectors. The above proof can be extended to the case of a resonant cavity. The orthogonality of resonant cavity modes is also analogous to the orthogonality of eigenvectors of a hermitian operator. It is to be noted that hermitian transpose of a matrix is similar to that of conjugate transpose. In math parlance, when an operator is equal to its hermitian transpose, it is called hermitian. But it is also called self-adjoint in the math literature.