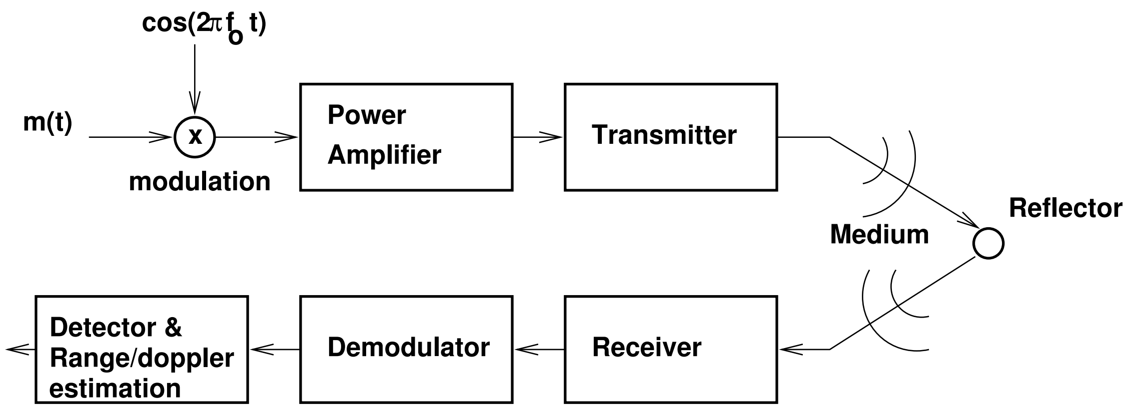
1 Modulation/demodulation and complex envelopes

A large number of signal processing operations in sonar, radar and seismic systems, especially active ones, involve modulated and/or narrow band signals. The general system operation is illustrated in Fig. 1. A low pass signal with bandwidth, W, is translated or frequency shifted, to be positioned about a carrier frequency, f_o. In most, but certainly not all cases the bandwidth of the modulating signal is very much less than the carrier frequency and this is referred to as a narrow band signal. For practical purposes, usually associated with the modulation and demodulation operations, the criterion $10W < f_o$ is often adopted. The modulating signal is then amplified and injected into the propagating medium by a radiating transducer. The channel, or propagating medium,, introduces its own filtering operation. Sometimes this is as simple as an attenuation, time delay and phase shift; alternatively, it can be linear filter which introduces time and Doppler spreading. The reflector, or target, can also introduce a linear filter of varying degrees of complexity.



Typical narrowband modulation system

Modulation is done for a number of reasons; however, the principal ones are: i) transducer and propagation medium properties; ii) spectral allocations and iii) angular resolution considerations. Transducers are are often resonant devices which only achieve high conversion efficiency around a narrow frequency band. In fact, it is often very difficult to make an efficient, broadband, so called "low Q" transducer. In addition, the propagation medium and reflector have filtering properties that are strongly dependent upon frequencies used and the cross section, or target strength, of reflectors have like properties. In radar many users want to use the electromagnetic propagation medium simultaneously. The only practical solution is keep the users orthogonal by allocating the spectrum by frequency. While this is not a common problem in sonar or seismic systems, there are certain popular bands where conflicts do arise. Finally, angular resolution scales with frequency so narrow beamwidths with fixed array apertures are achieved by high carrier frequencies. Generally, one increases the carrier as much as possible consistent with the attenuation of the propagation medium and the capabilities of the transducers.

The choice of carrier frequency is strongly influenced by the desired operating range, available physical dimensions on the platforms, power and beamwidth considerations. Low frequencies propagate well in the ocean; however, angular resolution is low; consequently, there is a tradeoff often made. Some of the ones typically used are noted below: Sonar operating frequencies

- $20 - 100\ Hz$ - long range propagation studies, passive Navy sonars;
- $200 - 400\ Hz$ - acoustic tomography, very low frequency sonars;
- $3.5\ kHz$ - deep water echo sounders, active Navy sonars;
- $8 - 15\ kHz$ - commercial fathometers, bathymetric mappers, acoustic navigation of submersibles, acoustic well logging;
- $10 - 70\ kHz$ - acoustic telemetry systems;
- $20 - 30\ kHz$ - fish finding sonars, under ice sonars;
- $> 100\ kHz$ - acoustic imaging systems;
- $200\ kHz$ - recreational depth finders.

In electromagnetics the carrier frequencies can be as low as $100\ kHz$ for long range, over the horizon systems and global navigation nets to as high as $10 - 25\ gHz$ for weather radars. These are commonly designated by a letter according to their band of operation. Electromagnetic band designation (Ref: M. Skolnik, *Introduction to Radar Systems,* Mc-Graw Hill Book Co,)

- HF $3 - 30\ MHz$
- VHF $30 - 300\ MHz$
- UHF $300 - 1000\ Mhz$
- L band $1 - 2\ GHz$
- S band $2 - 4\ GHz$
- C band $4 - 8\ GHz$
- X band $8 - 12.5\ GHz$
- K band $12.5 - 40\ GHz$ [1]

There are radar systems in all these bands. Generally, one uses low frequencies for long range propagation; however, they have the disadvantage of requiring large antenna apertures to achieve directionality.

[1] The K band is subdivided into several sub bandands.

2 Narrowband signals

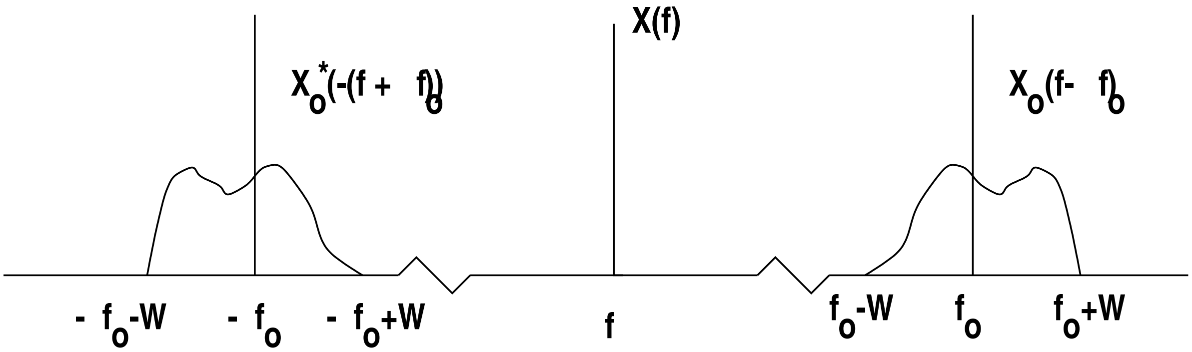
The Fourier transform for a narrowband signal is illustrated in Fig. 2. The symmetry constraint for real signal requires that[2]

$$X(f) = X^*(-f) \tag{1}$$

If we require $f_o > W$, this implies that we can represent $X(f)$ as

$$X(f) = X_o(f - f_o) + X_o^*(-(f + f_o)). \tag{2}$$

Note that $X_o(f)$ is bandlimited to $\pm W$. It does not have any special symmetry properties and generally leads to a complex signal $\tilde{x}_o(t)$. We soon demonstrate the $\tilde{x}_o(t)$ is the complex envelope of $x(t)$ for the carrier f_o. More importantly, all the important resolution properties such as estimating travel times and Doppler shifts are governed by $\tilde{x}_o(t)$.



Narrowband signals

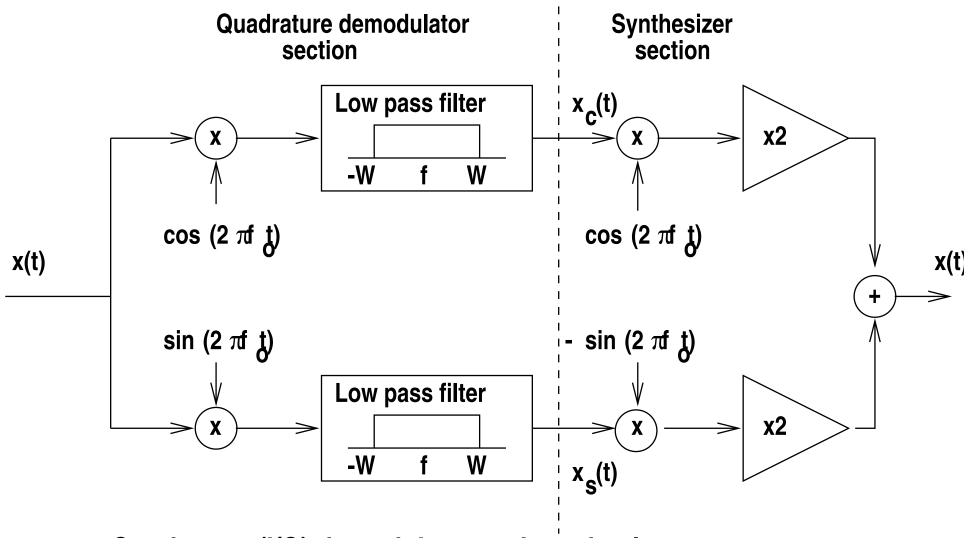
In most applications it is impractical to process the narrowband signal directly. This is especially true for digital implementations because sampling rates on the order of $2(f_o+W)$ are needed. While in sonars these rates can be now be obtained, it is quite difficult for high frequency radars; moreover, digital processing at such high sampling rates is very inefficient for purposes at hand. This leads to a "quadrature demodulation" [3] operation illustrated in Fig. 3. There are two important aspects of this operation: i) the narrowband, but high frequency, signal can be demodulated into two low pass signals: $x_c(t)$, often termed the "cosine," or "in phase (I)," component, and $x_s(t)$, the "sine," or "quadrature (Q), component and ii) the original narrowband signal can be recovered by the modulation operation illustrated in Fig. 3.

The quadrature components can be determined from $\tilde{x}_o(t)$ and have useful interpretations. Examining the cosine quadrature first, the signal at the input to the low pass filter in the quadrature demodulator is given by

$$x_c(t) = x(t)\cos(2\pi f_o t) = x(t)\left(\frac{e^{j2\pi f_o t} + e^{-j2\pi f_o t}}{2}\right). \tag{3}$$

[2] * indicates the complex conjugate.

[3] Quadrature demodulation is sometimes called "I and Q" filtering.



Quadrature (I/Q) demodulator and synthesizer

By using the modulation theorem the Fourier transform, $X_c(f)$, is given by

$$X_c(f) = \frac{1}{2}\left(X(f - f_o) + X(f + f_o)\right) = \tag{4}$$
$$\frac{1}{2}\left[X_o(f - 2f_o) + X_o^*(-f) + X_o(f) + X_o^*(-(f + 2f_o))\right].$$

The low pass filter rejects the components centered about $\pm 2f_o$, so we have

$$X_c(f) = \frac{1}{2}\left[X_o(f) + X_o^*(-f)\right] \; = \; Even[X_o(f)]. \tag{5}$$

Note that $X_c(f) = X*_c(-f)$, so it has real inverse transform as suggested by the quadrature demodulator structure. Note if also has the form of a complex generalization of the definition of the even part of a function, *i.e.* $Even[X_o(f)] = Even^*[X_o(-f)$. Similarly, the Fourier transform of the sine quadrature is given by

$$X_s(f) = \frac{j}{2}\left[X_o(f) - X_o^*(-f)\right] \; = \; jOdd[X_o(f)]. \tag{6}$$

Note that $X_s(f) = X_s(-f)$, so it is also has real inverse transform. The leading factor of j multiplies a term which has the form of the odd part of a complex function, *i.e.*, $Odd[X_o(f)] = -Odd^*[X_o(-f)]$ We note that

$$X_o(f) = X_c(f) \; - \; jX_s(f), \tag{7}$$

and this leads to defining $\tilde{x}_o(t)$ as the *complex envelope* in terms of the quadrature components, or [4]

$$\tilde{x}_o(t) \overset{\mathcal{F}}{\Longleftrightarrow} X_o(f) \tag{8}$$

[4]One needs to use the four quadrant arctan function $\arctan(y, x)$, for representing the phase process. One should note that there are different conventions regarding the sequence of the arguments (y, x).

with

$$\tilde{x}_o(t) = x_c(t) \ - \ jx_s(t) = \sqrt{x_c^2(t) + x_s^2(t)}e^{-j\arctan(x_s(t), x_c(t))} \tag{9}$$

Often it is useful to use a magnitude and phase representation for the complex envelope since it is related to amplitude and phase modulations.

If we generalize the notions of the "Even" and "Odd" parts of a function to complex functions, we see that the transform of the cosine quadrature is the "Even" part and the sine the "Odd" part.[5] The quadratures, $x_c(t)$ and $x_s(t)$, are given by

$$\begin{array}{llllll} Even[X_o(f)] & = X_c(f) & \Longleftrightarrow & x_c(t) & = & Re[\tilde{x}_o(t)] \\ j\ Odd[X_o(f)] & = X_s(f) & \Longleftrightarrow & x_s(t) & = & -Im[\tilde{x}_o(t)] \end{array} \tag{10}$$

It is straightforward to synthesize the original narrowband signal by modulating the quadratures. The output of the synthesizer section is given by

$$\hat{x}_o(t) = 2\left[x_c(t)\cos(2\pi f_o t) + x_s(t)\sin(2\pi f_o t)\right] = 2Re\left[\tilde{x}_o(t)e^{j2\pi f_o t}\right]. \tag{11}$$

The equivalence to original signal can easily be found using the real part operation

$$Re[x] = \frac{x + x^*}{2};$$

the first term leads the positive frequency components while the conjugate, the negative frequency ones. Consequently, $\hat{x}_o(t) = x(t)$ and we no no longer distinguish the two. [6]

It is useful to interpret the narrowband signal in terms of the magnitude and phase representation of the complex envelope. Using the above we have

$$x(t) = 2A_{x_o}(t)\cos(2\pi f_o t + \theta_o(t)) \tag{12}$$

where

$$A_{x_o}(t) = \sqrt{x_c^2(t) + x_s^2(t)} \quad : \text{amplitude modulation}$$

$$\theta_o(t) = -\arctan(x_s(t), x_c(t)) \quad : \text{phase modulation}$$

The term $A_{x_o}(t)$ is the real envelope of the narrowband signal; what one sees on an oscilloscope is the oscillation between $\pm 2A_{x_o}(t)$.

Finally, it is useful to relate the energy in the narrowband signal to the complex envelope. This leads to

$$E = \int x^2(t)dt = 4\int A_{x_o}^2(t)\cos^2(2\pi f_o t + \theta_o(t))dt = 2\int A_{x_o}^2(t)\left(1 + \cos(4\pi f_o t + 2\theta_o(t))\right) \tag{13}$$

[5] One can use Hilbert transforms to relate the two quadratures when $X_o(f) = 0, f < 0$.

[6] One needs to exercise care with factors of 2, complex conjugates and minus signs in the sine, or imaginary quadrature, with narrowband representations. Signal processing programs such as MATLAB produce strange results if these are not implemented correctly.

Within the integrand there are components near $f = 0$ and $f = \pm 2f_o$; we argue that the high frequency ones integrate to zero because the sinusoidal modulation is rapid compared to the low frequency variation of the envelope of the integrand. We have then

$$E = 2\int [x_c^2(t) + x_s^2(t)]dt = 2\int |\tilde{x}_o(t)|^2 dt = 2E_o, \tag{14}$$

or the energy in the narrowband signal is twice that in the complex envelope.

Example 1: Narrowband short pulse

The simplest transient narrowband signal is a short pulse given by

$$x(t) = \left[\begin{array}{ll} \sqrt{\frac{2E}{T}}\cos(2\pi f_o t), & |t| \le \frac{T}{2} \\ 0 & |t| > \frac{T}{2} \end{array} \right. \tag{15}$$

This signal has a *sinc* spectrum centered at $\pm f_o$. The complex envelope is easily found by identifying the amplitude and phase modulation and compensating by a factor of two. This leads to

$$\tilde{x}_o(t) = \left[\begin{array}{ll} \sqrt{\frac{E}{2T}}, & |t| \le \frac{T}{2} \\ 0, & |t| > \frac{T}{2} \end{array} \right. \tag{16}$$

Only the cosine quadrature is present, so the complex envelope is real and $\theta_o(t) = 0$.

Example 2: Doppler and phase shifts

It is straightforward to generalize to the inclusion of Doppler and phase shifts. We have

$$x(t) = \left[\begin{array}{ll} \sqrt{\frac{2E}{T}}\cos(2\pi(f_o + f_d)t + \phi), & |t| \le \frac{T}{2} \\ 0 & |t| > \frac{T}{2} \end{array} \right. \tag{17}$$

which leads to the complex envelope

$$\tilde{x}_o(t) = \left[\begin{array}{ll} \sqrt{\frac{E}{2T}}e^{j(2\pi f_d t + \phi)}, & |t| \le \frac{T}{2} \\ 0. & |t| > \frac{T}{2}. \end{array} \right. \tag{18}$$

Alternatively, we have for $|t| \le \frac{T}{2}$

$$\begin{aligned} x_c(t) &= \sqrt{\frac{E}{2T}}\cos(2\pi f_d t + \phi); \\ x_s(t) &= -\sqrt{\frac{E}{2T}}\sin(2\pi f_d t + \phi); \\ A_{x_o}(t) &= \sqrt{\frac{E}{2T}}; \\ \theta_o(t) &= (2\pi f_d t + \phi). \end{aligned} \tag{19}$$

Note that the transforms $X_c(f)$ and xsf contain *sinc* terms centered at $\pm f_d$. Both the signals in Examples 1 and 2 have the same amplitude modulation whereas they differ in the phase modulation. Note that for $f_d \neq 0$ both quadratures are present; this necessary to represent just a single tone about $\pm f_o$ instead of the situation for real signals when they appear as pairs.

Example 3: Phase modulated M sequences

A signal often used for acoustic tomography is a phase modulated "M Sequence." $m(t)$ is a periodic sequence which toggles between ± 1 at times $n\Delta T$. The period is given by $T = N\Delta T$ with $N = 2^k - 1$; typically, $k = 7, 8, 9$ or 10 in tomographic experiments. The details of the sequence generation are not relevant here; however, the important property for $m(t)$ is its autocorrelation function. One can demonstrate that it is periodic with period T and for the first period, $|\tau| < T/2$,

$$R_m(\tau) = \frac{1}{T}\int_{t-T/2}^{t+T/2} m(t')(m(t'-\tau)dt' = \left[\begin{array}{rl} 1 - (1+\frac{1}{N})\frac{|\tau|}{\Delta T} & |\tau| < \Delta T \\ -\frac{1}{N} & |\tau| \geq \Delta T \end{array}\right. \tag{20}$$

The narrowband signal is given by

$$x(t) = \sqrt{2P}\cos(2\pi f_o t + \Theta_{max} m(t) + \Delta\Theta) \tag{21}$$

where Θ_o is the phase deviation and $\Delta\Theta$ is chosen such that the transitions of $m(t)$ occur at a point of continuity for the overall phase process. Θ_{max} determines the allocation of power between the carrier and the modulation. In addition, ΔT is chosen to be $\Delta T = l/f_o$, where l is the "number of cycles per digit." Consequently, the bandwidth of the signal is approximately $W \approx f_o/l$. For this signal the complex envelope is given by

$$\begin{array}{rcl} A_{x_o}(t) & = & \sqrt{2P} \\ \theta_o(t) & = & \Theta_{max} m(t) + \Delta\Theta \end{array} \tag{22}$$

The quadratures $x_c(t)$ and $x_s(t)$ can easily be determined by finding the real and imaginary parts from the amplitude and phase representation.

Example 4: Linear frequency modulation

Frequency modulated signals are a very important narrowband class of waveforms for sonar, radars and seismics. In addition, many animals such as whales, dolphins and bats use frequency modulated signals. They are used to resolve travel times when one has transducers which are peak power limited, yet have wide bandwidths. In this example we consider a linear frequency modulation (LFM), the simplest of all FM signals. [7] We have

$$x(t) = \left[\begin{array}{r} \sqrt{\frac{2E}{T}}\cos(2\pi(f_o t + \frac{\mu t^2}{2})), \quad |t| \leq \frac{T}{2} \\ 0, \quad |t| > \frac{T}{2} \end{array}\right. \tag{23}$$

The phase modulation is a quadratic function of time; the factor μ is usually termed the "chirp rate" and has the dimensions of Hz/sec. One defines the instantaneous frequency

[7] There are a large class of nonlinear frequency modulations in literature. We discuss some of them later in the context of sonars and radars.

of a signal in terms of the derivative of the total phase, or

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}.$$

For the LFM signal $f(t) = f_o + \mu t$, so the frequency is swept from $f_o - \mu T/2$ to $f_o + \mu T/2$ for a total bandwidth of $2W = \mu T$. The complex envelope is given by

$$\tilde{x}_o(t) = \begin{bmatrix} \sqrt{\frac{E}{2T}} e^{j2\pi\frac{\mu t^2}{2}}, & |t| \le \frac{T}{2} \\ 0, & |t| > \frac{T}{2} \end{bmatrix} \tag{24}$$

One can find $x_c(t)$ and $x_s(t)$ simply by extracting the real and imaginary parts of $\tilde{x}_o(t)$. The amplitude, $A_{x_o}(t) = \sqrt{\frac{E}{2T}}$, while the phase $\theta_o(t) = 2\pi\frac{\mu t^2}{2}$. The transforms, $X_c(f)$ and $X_s(f)$ are more difficult to obtain and can be related to Fresnel integrals for this case of constant amplitude modulation; alternatively, approximate expressions can be obtained using the method of stationary phase

In summary, in most applications using narrow band signals it is more convenient theoretically and more efficient practically both to conduct an analysis and to implement a signal processing operations using the complex envelope. The carrier terms are cumbersome to include and usually do not directly impact the results; more importantly, sampling the complex envelope reduces the sampling rate requirements by approximately the ratio $f_o/2W$ which is typically quite large. There are two notes of caution: i) one must be careful of conjugation operations, which can be a nuisance, as well as the factors of two; ii) there is an implicit reference for the quadratures by the choice of the time origin and the use of cosine terms; simply, shifting the time origin by 1/4 of a carrier period leads to an exchange of the quadrature components and possibly their sign.

3 Narrowband signals and LTI systems

Many signal processing applications involve narrowband signals propagating through linear time invariant systems. In many applications we can approximate the output using the complex envelope in combination with the concepts of phase and group delay of the system. We consider a narrowband input to a real, linear, time invariant system with impulse response $h(t)$ and a transfer function $H(f)$. The output of the system can be found in terms of the inverse transform

$$y(t) = \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft}df \tag{25}$$

We use the complex envelope for representing $X(f)$ and this leads to

$$y(t) = \int_{f_o-W}^{f_o+W} H(f)X_o(f-f_o)e^{j2\pi ft}df + \int_{-f_o-W}^{-f_o+W} H(f)X_o^*(-(f+f_o))e^{j2\pi ft}df \tag{26}$$

Using the property that $H(f) = H^*(-f)$ it is direct to demonstrate that the second term is the complex conjugate of the first.

We now make some approximations to $H(f)$ which lead to useful interpretations for the output when they are valid. We first assume that the magnitude the $H(f)$ is approximately a constant, $|H(f_o)|$ across the band $f_o \pm W$. We next perform a Taylor series expansion for the phase up to the quadratic term. This implies that the phase process must be well represented by this quadratic expansion across the frequency band of interest. Combining the two approximations leads to

$$H(f) = |H(f_o)| e^{j(arg(H(f_o)) + \frac{d\, arg(H(f))}{df}|_{f_o}(f-f_o) + \frac{1}{2}\frac{d^2\, arg(H(f))}{df^2}|_{f_o}(f-f_o)^2)} \tag{27}$$

We substitute this into the expression above, shift integration variables by $f = f_o + \nu$ and we obtain

$$y(t) = 2Re[\tilde{y}_o(t) e^{j2\pi f_o t}] \tag{28}$$

with

$$\tilde{y}_o(t) = |H(f_o)| e^{-j2\pi f_o T_p(f_o)} \int_{-W}^{W} X_o(\nu) e^{-j2\pi(\nu T_g(f_o) + \frac{1}{2}\nu^2 \beta(f_o)))} e^{j2\pi\nu t} d\nu \tag{29}$$

where we have made the following definitions of phase delay, group delay and dispersion terms:

$$\begin{aligned} T_p(f_o) &= -\tfrac{1}{2\pi f_o} arg(H(f_o)) && \text{phase delay} \\ T_g(f_o) &= -\tfrac{1}{2\pi} \tfrac{d\, arg(H(f))}{df}|_{f_o} && \text{group delay} \\ \beta(f_o) &= -\tfrac{1}{2\pi} \tfrac{d^2\, arg(H(f))}{df^2}|_{f_o} && \text{dispersion} \end{aligned} \tag{30}$$

Note that the phase and group delay have the units of seconds while the dispersion in in seconds per Hertz. With these approximations we observe that the phase delay leads to a phase shift at the carrier frequency and the group delays a delay operator. If we ignore the dispersion term, $\beta(f_o)$, for the moment, we find that the complex envelope of the output is given by

$$\tilde{y}_o(t) = H(f_o)\tilde{x}_o(t - T_g(f_o)) = |H(f_o)| e^{-j2\pi f_o T_p(f_o)} \tilde{x}_o(t - T_g(f_o)) \tag{31}$$

or the input is simply amplified by $|H(f_o)|$, phase shifted by $T_p(f_o)$ and delayed by $T_g(f_o)$.

The dispersion term incorporates the change in group delay across the frequency band $f_o \pm W$. In particular, we have

$$T_g(f_o + \nu) = T_g(f_o) + \beta(f_o)\nu;$$

so if the dispersion is positive, *i.e.* the phase function is convex down, the group delay increases with frequency and higher frequencies are delayed more. First we consider the conditions under which we can ignore it. We assume that its contribution is negligible if it introduces less than a phase shift of $\pi/4$, or

$$2\pi \frac{1}{2} \beta(f_o) W^2 \leq \frac{\pi}{4}, \tag{32}$$

or

$$W \leq \frac{1}{2\sqrt{\beta(f_o)}}.$$

More generally, one can inverse transform the dispersion operator which one could then use to convolve with the gain, phase shifted and delayed complex envelope for the output. We have

$$e^{-j2\pi\beta(f_o)\frac{\nu^2}{2}} \Longleftrightarrow \frac{1}{|\beta(f_o)|} e^{-j\pi\frac{t^2}{\beta(f_o)}} e^{j\ sgn(\beta(f_o))\frac{\pi}{4}}$$

The function has a quadratic dependence in time and is generally difficult to convolve, so one usually works with the simpler frequency domain formulation.

Example 5: Resonant system

Consider operating within the resonance region of a high Q system as illustrated in Figure 4. The transfer function for such a system is given by

$$H(f) = \frac{f_r^2}{f_r^2 - f^2 + j\frac{f*f_r}{Q}} \tag{33}$$

where f_r is the resonance frequency and $Q >> 1$. The resonance bandwidth is approximately f_r/Q, for the narrowband signal to be within the resonance band, one must have $|f_o \pm W - f_r| < f_r/2Q$. For a high Q system near resonance the transfer function is given approximately by

$$H(f) \approx \frac{-j}{2} \frac{f_r}{\frac{f_r}{2Q} + j(f - f_r)} \tag{34}$$

The gain at f_o is then given by

$$|H(f_o)| \approx \frac{1}{2} \frac{f_r}{\sqrt{(\frac{f_r}{2Q})^2 + (f_o - f_r)^2}}. \tag{35}$$

The phase delay is

$$T_p(f_o) \approx \frac{1}{2\pi f_o}[\frac{\pi}{2} + \arctan(2Q(\frac{f_o}{f_r} - 1))] \tag{36}$$

and the group delay is

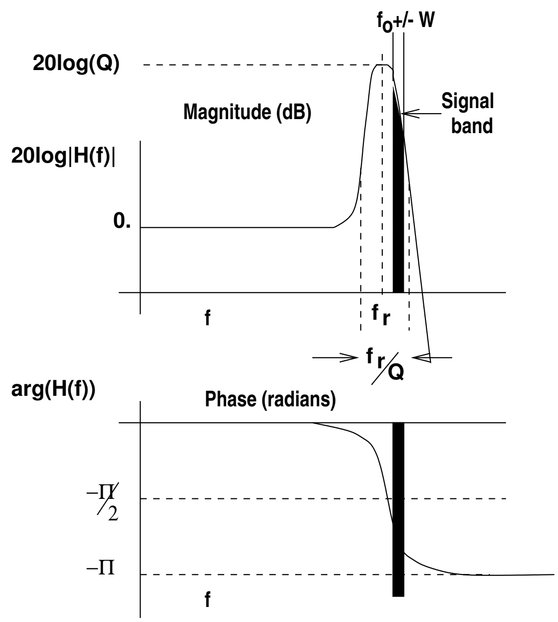
$$T_g(f_o) \approx \frac{1}{2\pi} \frac{2Q}{\sqrt{f_r^2 + (2Q)^2(f_o - f_r)^2}} \tag{37}$$

One may note that at resonance

$$T_g(f_r) \approx \frac{Q}{\pi f_r} = \frac{QT_r}{\pi}$$

where $T_r = 1/f_r$ is the period of carrier, *i.e* the group delay is approximately Q/π times the period of carrier. Similarly, it is easy to use the expressions above that at the edge of the resonance band, the group delay is 1/2 that at resonance. This implies that there may be a significant group delay in terms of the number of cycles for a system operating near resonance.

The linear system, $H(f)$ may represent many systems in sonar, radar and seismics. In sonar, especially at low frequencies and/or shallow water, dispersion effects can be very important. In all regimes transducers operating near resonance can have strong dispersive effects. One must determine if it is significant for the bandwidths used.



Magnitude and phase for a resonant system