Digital Communications

Romain Couillet
ST-Ericsson, Supélec, FRANCE
romain.couillet@supelec.fr

Supélec

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Notion of information theory

C. Shannon, "A mathematical theory of communications," Bell System Technical Journal, vol. 27, no. 7, pp. 379-423, 1948.

- information theory deals with
  - a mathematical description of information
  - lossless information compression, e.g. ZIP files
  - lossy information compression, e.g. JPEG, MPEG etc.
  - transmitting information over lossy channels, i.e.
    - how to decode a signal corrupted with noise
    - how to optimally encode a signal bound to be corrupted by noise
- in one fundamental paper in 1948, Shannon provided
  - optimal information source encoding
  - optimal transmission rate over a lossy medium

This paper marked the birth of the field of *Information Theory*.

Notion of information theory

C. Shannon, “A mathematical theory of communications,” Bell System Technical Journal, vol. 27, no. 7, pp. 379-423, 1948.

- information theory deals with
  - a mathematical description of information
  - lossless information compression, e.g. ZIP files
  - lossy information compression, e.g. JPEG, MPEG etc.
  - transmitting information over lossy channels, i.e.
    - how to decode a signal corrupted with noise
    - how to optimally encode a signal bound to be corrupted by noise
- in one fundamental paper in 1948, Shannon provided
  - optimal information source encoding
  - optimal transmission rate over a lossy medium

This paper marked the birth of the field of *Information Theory*.

Notion of information theory

C. Shannon, "A mathematical theory of communications," Bell System Technical Journal, vol. 27, no. 7, pp. 379-423, 1948.

- information theory deals with
  - a mathematical description of information
  - lossless information compression, e.g. ZIP files
  - lossy information compression, e.g. JPEG, MPEG etc.
  - transmitting information over lossy channels, i.e.
    - how to decode a signal corrupted with noise
    - how to optimally encode a signal bound to be corrupted by noise
- in one fundamental paper in 1948, Shannon provided
  - optimal information source encoding
  - optimal transmission rate over a lossy medium

This paper marked the birth of the field of *Information Theory*.

Information theory: from theory to practice

- Shannon's results
  - Shannon does tell us how much information can be reliably transmitted over a medium, but *does not* tell us how to achieve such a rate!
  - Shannon assumed digital transmissions over the medium, which may not have any physical meaning.
- in this course, we shall
  - derive practical ways to transmit information over classical mediums at a given rate, e.g. copper wires, open air, fiber optics, etc.
    - information mapping onto sequences of bits
    - mapping of bits on *symbol constellations*
    - transmitting symbols on continuous waveforms
  - derive methods to decode a signal embedded in noise
    - filter the continuous waveforms (matched filtering)
    - infer transmitted symbols: signal decoding
    - map inferred symbols back onto bits
  - *we will only briefly mention how to encode a signal source so to protect it from noise*

Information theory: from theory to practice

- Shannon's results
  - Shannon does tell us how much information can be reliably transmitted over a medium, but *does not* tell us how to achieve such a rate!
  - Shannon assumed digital transmissions over the medium, which may not have any physical meaning.
- in this course, we shall
  - derive practical ways to transmit information over classical mediums at a given rate, e.g. copper wires, open air, fiber optics, etc.
    - information mapping onto sequences of bits
    - mapping of bits on *symbol constellations*
    - transmitting symbols on continuous waveforms
  - derive methods to decode a signal embedded in noise
    - filter the continuous waveforms (matched filtering)
    - infer transmitted symbols: signal decoding
    - map inferred symbols back onto bits
  - *we will only briefly mention how to encode a signal source so to protect it from noise*

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Reminders on finite-energy signals

- A signal $x(t)$ is said to be of finite energy if

$$\mathcal{E}_x \triangleq \int |x(t)|^2 dt < \infty$$

- We denote ϕ_x its autocorrelation function as

$$\phi_x(\tau) \triangleq \int x(t)x(t-\tau)^* dt$$

- We denote $S_x(f)$ the *energy spectral density* at frequency f,

$$S_x(f) \triangleq \mathcal{F}[\phi_x](f) = \int \phi_x(t) e^{-2\pi ift} dt$$

with $\mathcal{F}$ the Fourier transform operator. Note that

$$S_x(f) = |X(f)|^2$$

where X is the Fourier transform of x,

$$X(f) \triangleq \mathcal{F}[x](f) = \int x(t) e^{-2\pi ift} dt$$

Reminders on finite-power signals

- A signal $x(t)$ is said to be of finite power if

$$\mathcal{P}_x \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |x(t)|^2 dt < \infty$$

- We denote ϕ_x its autocorrelation function as

$$\phi_x(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)x(t-\tau)^* dt$$

- We denote $S_x(f)$ the *power spectral density* at frequency f,

$$S_x(f) \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} \phi_x(t)e^{-2\pi ift} dt$$

From Parseval's identity, for both finite-energy/power signals, we have

$$\phi_x(0) = \int S_x(f)df$$

Correlation

- We will often deal in this course with Linear Time Invariant (LTI) filters h correlating input signals x as,

$$y(\tau) = h(\tau) * x(\tau) \triangleq \int h(t)x(t-\tau)dt$$

- From the Fourier transform, we have

$$Y(f) = H(f)X(f)$$

hence,

$$|Y(f)|^2 = |H(f)|^2 |X(f)|^2$$

taking the inverse Fourier transform, this gives

$$\phi_f(\tau) = \phi_h(\tau) * \phi_x(\tau)$$

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of *M*-ary signals
4. Digital modulation
   - *M*-PAM modulation
   - *M*-QAM modulation
   - *M*-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Case of random signals

- In case of random processes, for a realization $x(t;\omega)$ of a continuous process, $\omega \in \Omega$ the universe attached to the random process, $\phi_x(\tau)$ is defined by
- In the mean sense, we define ϕ_x as

$$\phi_x(\tau) \triangleq \mathrm{E}_\omega \left[x(t;\omega)x^*(t-\tau;\omega)\right] = \int x(t;\omega)x^*(t-\tau;\omega)dP(\omega)$$

This definition only makes sense when ϕ_x is independent of t. Note that, if x is an ergodic process, $\phi_x(\tau) = \int x(t;\omega)x^*(t-\tau;\omega)dt$ for any $\omega \in \Omega$.

We also define the Power Spectral Density as

$$S_x(f) \triangleq \mathcal{F}[\phi_x](f) = \mathrm{E}\left[\lim_{T\to\infty} \frac{1}{T}\left|\int_{-T/2}^{T/2} x(t;\omega)e^{-2\pi i f t}dt\right|^2\right]$$

with $X(f) \triangleq \mathcal{F}[x](f)$.

- We call Wide Sense Stationary (WSS) such processes for which $\phi_x(\tau)$ does not depend on the time position t.
- In the case of WSS signals, we define the energy/power spectral power density

$$S_x(f) \triangleq \mathcal{F}[\phi_x](f) = \int \phi_x(\tau)e^{-2\pi i f \tau}d\tau$$

ST ERICSSON

Case of random signals

- In case of random processes, for a realization $x(t;\omega)$ of a continuous process, $\omega \in \Omega$ the universe attached to the random process, $\phi_x(\tau)$ is defined by
- In the mean sense, we define ϕ_x as

$$\phi_x(\tau) \triangleq \mathrm{E}_\omega \left[x(t;\omega)x^*(t-\tau;\omega)\right] = \int x(t;\omega)x^*(t-\tau;\omega)dP(\omega)$$

This definition only makes sense when ϕ_x is independent of t. Note that, if x is an ergodic process, $\phi_x(\tau) = \int x(t;\omega)x^*(t-\tau;\omega)dt$ for any $\omega \in \Omega$.

We also define the Power Spectral Density as

$$S_x(f) \triangleq \mathcal{F}[\phi_x](f) = \mathrm{E}\left[\lim_{T\to\infty} \frac{1}{T}\left|\int_{-T/2}^{T/2} x(t;\omega)e^{-2\pi i f t}dt\right|^2\right]$$

with $X(f) \triangleq \mathcal{F}[x](f)$.

- We call Wide Sense Stationary (WSS) such processes for which $\phi_x(\tau)$ does not depend on the time position t.
- In the case of WSS signals, we define the energy/power spectral power density

$$S_x(f) \triangleq \mathcal{F}[\phi_x](f) = \int \phi_x(\tau)e^{-2\pi i f \tau}d\tau$$

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Finite size bandwidth

- we are interested into transmitting signals in finite bandwidths
- these bandwidths correspond to frequencies that the medium can carry, e.g.
  - microwaves for medium-range mobile phone communications
  - centimeter wavelengths for large range radio transmissions
  - acoustic frequencies (300-3000 Hz) for human voice communications
- transmitted signals can then be modelled under the form

$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

where f_c is the central bandwidth frequency and where most energy of $a(t)$ is supposed to be concentrated around frequency 0.

Baseband complex equivalent

- instead of the real transmitted signal

$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

it is convenient to represent it as

$$s(t) = x_I(t)\cos(2\pi f_c t) - x_Q(t)\cos(2\pi f_c t)$$
$$\Re\left[x(t)e^{2\pi i f_c t}\right]$$

with

$$x(t) = x_I(t) - i \cdot x_Q(t)$$

and

$$x_I(t) = a(t)\cos(\theta(t))$$
$$x_Q(t) = a(t)\sin(\theta(t))$$

This is merely another representation. It is equivalent to work with the pair $(a(t), \theta(t))$ or the pair $(x_I(t), x_Q(t))$

- it is called *baseband complex representation* because $x_I(t)$ and $x_Q(t)$ are signals with frequency centered at $f = 0$.

Properties of baseband complex equivalent

- Denoting $S(f) = \mathcal{F}[s](f)$ and $X(f) = \mathcal{F}[x](f)$,
$$S(f) = \frac{1}{2}(X(f - f_c) + X^*(-f - f_c))$$
- if $s(t)$ is real, then
$$S(f) = S^*(-f)$$
- for LTI filters $h(t)$ with central frequency f_c, denoting
$$C(f - f_c) \triangleq H(f)1_{f>0}(f)$$
$$C^*(-f - f_c) \triangleq H^*(-f)1_{f<0}(f)$$
we have
$$H(f) = C(f - f_c) + C^*(-f - f_c)$$
$$h(t) = 2\Re[c(t)e^{2\pi i f_c t}]$$
- for $y(t) = h(t) * s(t)$, we finally have
$$Y(f) = \frac{1}{2}(X(f - f_c)C(f - f_c) + X^*(-f - f_c)C^*(-f - f_c))$$
- for a deterministic or WSS process $s(t)$, we also have
$$S_x(f) = \frac{1}{4}(S_s(f - f_c) + S_s(-f - f_c))$$
with $S_s(f) = \mathcal{F}[\phi_s](f)$ and $S_x(f) = \mathcal{F}[\phi_x](f)$.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Shannon's sampling theorem

- From now on, we consider complex baseband equivalent signals $x(t)$ of some passband signal $s(t)$.
- Shannon's sampling theorem states that

Theorem

If $X(f)$ has support $[-W/2, W/2]$, then $x(t)$ is entirely determined by the samples

$$x_i \triangleq x(i/W)$$

and can be reconstructed from

$$x(t) = \sum_{i=-\infty}^{\infty} x_i \operatorname{sinc}(Wt - i)$$

Proof of sampling theorem

- sampling in the time domain engenders spectrum duplicates in the frequency domain. The spectrum is given by

$$\mathcal{F}[x(t) \cdot \sum_i \delta(t - i/W)] = X(f) * W \sum_i \delta(f - iW) = W \sum_i X(f - iW)$$

- if $X(f)$ is wider than W, this engenders overlapping spectra.
- if $X(f)$ has support W, then

$$X(f) = \frac{1}{W} 1_{-W/2 \leq f \leq W/2}(f) \cdot W \sum_i X(f - iW)$$

and we have

$$x(t) = \mathcal{F}^{-1}\left[1_{-W/2 \leq f \leq W/2}(f) \cdot \mathcal{F}\left[x(t) \cdot \sum_i \delta(t - i/W)\right]\right]$$
$$= \text{sinc}(W(t - i/W))x(t) \cdot \sum_i \delta(t - i/W)$$

Relation with the discrete Fourier transform

- it will often be convenient to work on the discrete samples x_i themselves to derive spectrum content properties. For this, we introduce the *discrete Fourier transform* defined for the normalized frequency $-1/2 \leq f_0 \leq 1/2$, as

$$X_d(f_0) \triangleq \sum_i x_i e^{-2\pi j f_0 i}$$

this can be shown to be equal to

$$X_d(f_0) = W \sum_i X(f_0 W + iW) = W \sum_i X(f + iW)$$

with $f = f_0 W$.

- discrete Fourier transform folds the spectrum of $X(f)$ every W. It is therefore equivalent to work with Fourier transforms or discrete Fourier transforms.

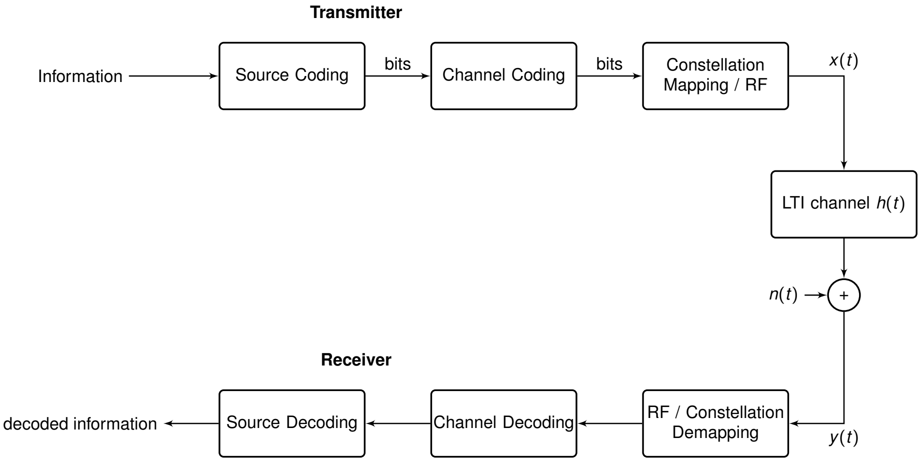
Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Signal source and channel



Introduction to communication theory

- The classical transmission chain is composed of
  - An information source, which is the abstract content we wish to transmit (e.g. an idea, a video/audio content, a definite data)
  - A source (en)coder, which aims at compressing information so to transmit it fast. Source encoders are of two types,
    - lossless encoders, that encode the information without loss, e.g. ZIP encoders. Those are required for data which require high fidelity at decoding, e.g. file storage on a hard drive.
    - lossy encoders, that accept data/quality loss to enforce compression above the lossless limit, e.g. JPEG encoders. Those are accepted when data content can be altered, e.g. JPEG/MPEG compression do not impact user experience much.
  - A channel encoder, which protects the bit stream against channel errors. This is performed by using redundancy codes, the most elementary of which being the repetition code, e.g.

$$0010110 \mapsto 000.000.111.000.111.111.000$$

    for a rate $1/3$ code. For well designed codes, a certain number of errors can be detected and a certain number of errors can be corrected, e.g. repetition codes can detect up to 2 consecutive decoding errors, and correct up to 1 decoding error.
  - A constellation mapper, which maps encoded bits onto symbols, e.g.
    - BPSK modulation: $100110 \mapsto 1, -1, -1, 1, 1, 0$
    - QPSK modulation: $10, 01, 00 \mapsto 1 - i, -1 + i, -1 - i$
  - A Radio Front-end (RF) interface, that turn the symbols $a_1, a_2, \ldots$ into modulated waveforms (through DAC)

$$x(t) = \sum_i a_i \phi(t - iT)$$

    these waveforms are sent as electromagnetic wave (or other) onto the medium.

ST ERICSSON

Introduction to communication theory (2)

- The transmitted waveform $x(t)$ is
  - convolved by an LTI channel $h(t)$
  - corrupted by additive noise $n(t)$
  - received on the receiver side as

$$y(t) = h(t) * x(t) + n(t)$$

- The classical reception chain is composed of
  - An RF interface/Constellation demapper, that turns the analog waveforms $y(t)$ back into a digital form. We shall see in this course that the received $y(t)$ can be filtered to obtain a digital *sufficient statistic* for the symbol decision. We then infer the transmitted symbol from this statistics,
    - either by taking each symbol at a time and perform *symbol-by-symbol* (hard) decision
    - either by keeping score of the *likelihood* of every possible transmitted symbol. This is called soft decision.
  - A channel decoder, that takes as an input the (hard/soft) decision symbols and decide on the *sequence* of symbols that must have been sent, e.g. if 000.111.000.010.111 is received from a repetition code of rate $1/3$, the channel decoder will decide that the transmitted symbol was 000.111.000.000.111, and therefore will decide that the sequence 01001 was transmitted.
  - A source decoder, that turns the bit content back into physically meaningful data, e.g. the output JPEG image, data to be stored, etc.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Shannon's theorems

- The first theorem of Shannon deals with source encoding. Shannon shows that a source $a_1, \ldots, a_N$ of N symbols with respective probability (for non-random source, probability is considered as frequency) of appearance $P_1, \ldots, P_N$ ($\sum_i P_i = 1$) can be represented in

$$-\sum_i P_i \log_2(P_i)$$

bits par channel use.

  - this can be understood as follows: if a_i has a strong probability of appearance, it must be encoded into as little bits as possible, so inversely proportional to P_i.
  - since the joint probability for an independent source is the product of the probabilities, the number of bits to transmit must be additive for probability products, hence the log function.
  - in the mean, we then have the desired result.

Shannon's theorems (2)

- The second theorem of Shannon is more important to the present course. This one deals with channel coding and states

Theorem

Consider a channel with additive complex white Gaussian noise process (AWGN) $n(t)$, i.e. $n(t)$ is a Gaussian process with $\phi_n(\tau) = N_0\delta(\tau)$. Note that, for this process, $S_n(f) = \mathcal{F}[\phi_n](f) = N_0$ is constant over f. Then the maximum rate R (i.e. number of bits per second) for which there exists a channel code which ensures as low decoding error as desired satisfies

$$R = W \log_2 \left(1 + \frac{RE_b}{N_0 W}\right)$$

with

- *W the channel bandwidth*
- *E_b the average energy used for a single transmitted bit*
- *N_0 is the noise spectral density*

- This theorem means that there exists a maximum rate for which
  - under this rate, there exists a code that can be decoded with almost no error
  - above this rate, no reliable communication is possible
- this provides bounds on the achievable transmission rates for practical systems

Shannon's theorems (3)

- Shannon's theorem is often written in the form

$$C = \log_2(1 + \mathrm{SNR})$$

where

- $C = R/W$ is called the channel capacity, in bits/s/Hz,
- SNR is called the signal-to-noise ratio, which is the total energy $E_b R$ consumed by the transmitter every second over the noise spectral density N_0 cumulated on the transmission bandwidth W.

However, Shannon

- does not tell us how to achieve such rates
- only considers communications of bits, and not of bits *though waveforms*.

The purpose of this course is to give an introduction of how we fill the gap.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. **Signal detection in AWGN**
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Hypothesis testing

- consider a set of events $\mathcal{X} = \{1, \dots, M\}$
- take $\mathcal{Y}$ to be the set of observations of the events in $\mathcal{X}$, characterized by the joint probability distribution on $\mathcal{X} \times \mathcal{Y}$
- we call H_m the *hypothesis* that m is the actual event
- the situation is as follows,
  - a transmitter (e.g. human voice, wireless phone) sends one of M messages, indexed by $\mathcal{X}$. Say the transmitter sends message x.
  - the message is received by the receiver through a communication channel as an output $y \in \mathcal{Y}$, characterized by the joint probability distribution on $\mathcal{X} \times \mathcal{Y}$

$$P(x, y)$$

  - the receiver has to decide on the best hypothesis among $H_1, \dots, H_M$ by minimizing the expectation of a cost function $r(x, y)$ on $\mathcal{X} \times \mathcal{Y}$
  - this is, the receiver must uniquely assign to every y an index $\hat{x} \in \mathcal{X}$ such that the expected risk function $\mathrm{E}[r(x, y)]$ is minimized.

Minimizing the average error probability

- it is classical to minimize the expectation of making a wrong decision, i.e. minimizing the probability of $\hat{x} \neq x$.
- we compute the error probability as

$$P(e) = P(\hat{X} \neq X)$$

- this extends into

$$\begin{aligned} P(e) &= 1 - P(\hat{X} = X) \\ &= 1 - \sum_{x=1}^{M} P(X = x) P(\hat{X} = x | X = x) \\ &= 1 - \sum_{x=1}^{M} \int_{\mathcal{D}_x} p(y|x) p(x) dy \end{aligned}$$

  with $\mathcal{D}_x$ the subset of $\mathcal{Y}$ on which the decision $\hat{x}$ is taken.
- in order to minimize $P(e)$, one must then find the *decision regions* $\mathcal{D}_x$ associated to every x.

Decision regions and MAP/ML detection

- reminder: we wish to minimize

$$P(e) = 1 - \sum_{x=1}^{M} \int_{\mathcal{D}_x} p(y|x)p(x)dy$$

  so we wish to maximize

$$\sum_{x=1}^{M} \int_{\mathcal{D}_x} p(y|x)p(x)dy$$

- if $p(y|x)p(x) = \max_{x'} p(y|x')p(x')$, then $\mathcal{D}_x$ must contain this y. Then the decision regions are then given by

$$\mathcal{D}_x = \left\{ y \in \mathcal{Y} : p(y|x)p(x) = \max_i p(y|i)p(i) \right\}$$

- the optimal hypothesis decision is called Maximum a Posteriori (MAP) and reads

$$\hat{x} = \arg\max_{x \in \mathcal{X}} p(y|x)p(x) = \arg\max_{x \in \mathcal{X}} p(x|y)$$

- it will often turn out that $p(x)$ is constant over $\mathcal{X}$ (uniform probability distribution of the transmitted symbols). In this case, the rule becomes the Maximum Likelihood (ML) estimate, given by

$$\hat{x} = \arg\max_{x \in \mathcal{X}} p(y|x)$$

Example: binary symmetric channel

- assume the following binary symmetric channel (BSC) with
  - $\mathcal{X} = \{0,1\}$
  - $\mathcal{Y} = \{0,1\}$

  and transition probabilities,

$$P(0|0) = 1 - p$$
$$P(1|0) = p$$
$$P(0|1) = p$$
$$P(1|1) = 1 - p$$

  where $p < 1/2$.

- under uniform probabilities, the optimal decision rule is

$$\hat{x} = y$$

Example: Gaussian noise

- we now consider a transmission with Gaussian noise
  - $\mathcal{X} = \{0, 1\}$
  - $\mathcal{Y} = \mathbb{R}$

  and transition probabilities,

$$P(y|0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu_0)^2}{2\sigma^2}}$$

$$P(y|1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu_1)^2}{2\sigma^2}}$$

  with $\mu_0 < \mu_1$.
- under uniform probabilities on $\mathcal{X}$, the optimal decision rule is to check that, for $x = 0$,

$$p(y|0) \geq p(y|1)$$

  which leads to

$$y < \frac{\mu_0 + \mu_1}{2}$$

Upper bound on detection errors

- it is often convenient to have detection bounds on the MAP/ML detectors to assess the quality of the detectors as a function of the channel parameters, e.g. as a function of the SNR.
- for this, we consider the *pairwise error probability* (PEP) of inferring x' while x was transmitted, as

$$P(x \to x')$$

which is the probability of the set $\{x \to x'\} \triangleq \{y \in \mathcal{Y} | p(y|x)p(x) < p(y|x')p(x')\}$.
- we then have the upper bound,

$$P(\hat{x} \neq x) = P(\cup_{x' \neq x}\{x \to x'\}) \leq \sum_{x' \neq x} P(x \to x')$$

- averaging over $\mathcal{X}$, this leads to

$$P(e) \leq \sum_{x=1}^{M} \sum_{x' \neq x} p(x)P(x \to x')$$

which becomes

$$P(e) \leq \frac{1}{M} \sum_{x=1}^{M} \sum_{x' \neq x} P(x \to x')$$

in case of uniform prior distribution on $\mathcal{X}$.

Lower bound on detection errors

- we can alternatively say that

$$P(\hat{x} \neq x) \geq \max_{x' \neq x} P(x \to x')$$

- averaging over $\mathcal{X}$, this leads to

$$P(e) \geq \sum_{x=1}^{M} p(x) \max_{x' \neq x} P(x \to x')$$

- under uniform probability again, we have the ML lower bound,

$$P(e) \geq \frac{1}{M} \sum_{x=1}^{M} \max_{x' \neq x} P(x \to x')$$

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Transmission model

- Transmission model
  - suppose that we wish to transmit one of M symbols through M different complex waveforms $x_1(t), \dots, x_M(t)$.
  - assume the transmitted sequence suffers from an additive white Gaussian noise (AWGN) $n(t)$
  - assume the convolution channel only induces a phase rotation, e.g. case of wired transmissions, line-of-sight wireless transmissions.
- The waveform arising at the receiver, for the choice of symbol m at the transmitter, reads

$$r(t) = x_m(t)e^{2\pi i\theta} + n(t)$$

- We shall consider *coherent detection*, i.e. all system parameters have been accurately estimated by the receiver (e.g. this can be performed using training sequences). As such, since the complex noise distribution is invariant by rotation, we can assume $\theta = 0$ without generality restriction.

$$r(t) = x_m(t) + n(t)$$

Sufficient statistics

- consider that $x_m(t)$ is given by
$$x_m(t) = a_m\phi(t)$$
for $a_1, \dots, a_M$ complex *symbols* of a finite alphabet and $\phi(t)$ a finite-time finite-energy signal, with $\int |\phi(t)|^2 dt = \mathcal{E}$.
- we want here for the detector to infer on m from the received waveform
$$r(t) = x_m(t) + n(t)$$
with $n(t)$ white Gaussian, with $\phi_n(0) = N_0$.
- we then wish to fall back on a Bayesian hypothesis testing. However the output $r(t)$ lies into a function space, which is not as convenient as a scalar space.
- in order to turn the problem into an hypothesis test, consider the any orthonormal basis $\{\xi_i(t)\}_{i=1,2,\dots}$ of the space of finite-time finite-energy signals (with inner product $< x(t)|y(t) >= \int x^*(t)y(t)dt$), such that
$$\xi_1(t) = \frac{1}{\sqrt{\mathcal{E}}}\phi(t)$$
- from orthogonality, we have that $r(t)$ can be represented in this basis, as a vector $\mathbf{r}$, such that
$$\mathbf{r} = \begin{pmatrix} a_m\sqrt{\mathcal{E}} + \nu_1 \\ \nu_2 \\ \nu_3 \\ \vdots \end{pmatrix}$$
where $\nu_i = \int \xi_i^*(t)n(t)dt$

ST ERICSSON

Sufficient statistics (2)

- we can then now work with r_1 alone!
- since $n(t)$ is Gaussian, by the orthonormal character of the basis $\{\xi_i(t)\}$, $\nu_1, \nu_2, \ldots$ are all Gaussian (by linearity of the inner product) each with zero mean and variance N_0.
- we then have

$$p(r_1|m) = p(y(t)|m)$$

  in which case r_1 is called a *sufficient statistics* for m, i.e. we do not loose any information on m by filtering $r(t)$ into the scalar r_1.
- *generally speaking, a sufficient statistics will be a function $f(r(t))$ of the input $r(t)$ such that the conditional probability on m can be written in the form*

$$p(f(r(t))|m) = p(r(t)|m)g(r(t))$$

  which does not alter the MAP/ML decisions.
- it is therefore equivalent to work here with the *scalar* model

$$y = a_m\sqrt{\mathcal{E}} + \nu$$

  where we denoted $\nu \triangleq \nu_1$ for readability.

ST ERICSSON

Hypothesis test

- we now need to study the hypothesis test with hypotheses $H_1, \dots, H_M$, defined as

$$H_m = \text{“}a_m \text{ was transmitted''}$$

under the transmission model

$$y = a_m\sqrt{\mathcal{E}} + \nu$$

with $\nu \sim \mathcal{CN}(0, N_0)$, which we know how to deal with.

- assume the simpler case $M = 2$, and $a_1 = +1$, $a_2 = -1$.
- the decision rule for a_1 is associated with the inequality,

$$p(1)\frac{1}{\pi N_0}e^{-\frac{1}{N_0}|y-\sqrt{\mathcal{E}}|^2} \geq p(2)\frac{1}{\pi N_0}e^{-\frac{1}{N_0}|y+\sqrt{\mathcal{E}}|^2}$$

which, after taking the log, is equivalent to

$$\Re(y) \geq \frac{N_0}{4\sqrt{\mathcal{E}}}\log\frac{p(2)}{p(1)}$$

- in case $p(1) = p(2) = 1/2$, we then decide a_1 in the case

$$\Re(y) \geq 0$$

Error analysis

- the error analysis is easy in the binary case. We have indeed

$$P(e) = p(1)p(1 \to 2) + p(2)p(2 \to 1)$$

where

$$p(1 \to 2) = P(\sqrt{\mathcal{E}} + \Re(\nu) < \tau)$$
$$p(2 \to 1) = P(-\sqrt{\mathcal{E}} + \Re(\nu) > \tau)$$

with $\tau = \frac{N_0}{4\sqrt{\mathcal{E}}} \log \frac{p(2)}{p(1)}$.

- because of the Gaussian distribution for $\Re(\nu)$ ($\mathcal{N}(0, N_0/2)$), this is

$$p(1 \to 2) = \int_{-\infty}^{\tau-\sqrt{\varepsilon}} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-t^2}{N_0}} dt = \int_{\sqrt{\varepsilon}-\tau}^{\infty} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-t^2}{N_0}} dt = Q\left(\frac{1}{\sqrt{N_0/2}}\left[\sqrt{\mathcal{E}} - \tau\right]\right)$$

with $Q(x)$ the Gaussian-queue function

$$Q(x) \triangleq \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- identically,

$$p(2 \to 1) = \int_{\tau+\sqrt{\varepsilon}}^{\infty} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-t^2}{N_0}} dt = Q\left(\frac{1}{\sqrt{N_0/2}}\left[\sqrt{\mathcal{E}} + \tau\right]\right)$$

Error analysis (2)

- if $p(1) = p(2)$, this is finally

$$P(e) = Q\left(\sqrt{\frac{2\mathcal{E}}{N_0}}\right)$$

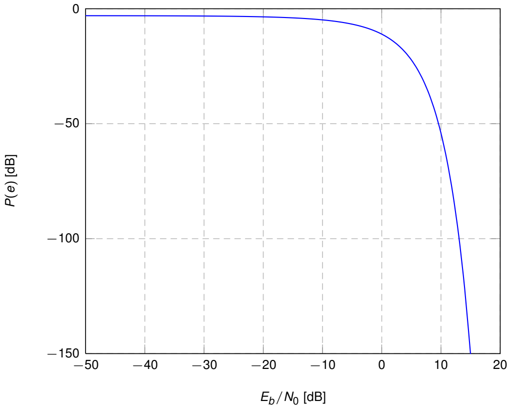


Figure: Probability of decoding error as a function of E_b/N_0

Matched filter

- to obtain the filtered sample y from the source $r(t)$, we performed the filtering operation

$$y = \int \xi_1(t)^* r(t) dt$$

where we had chosen $\xi_1(t) \stackrel{\Delta}{=} \phi(t)/\sqrt{\mathcal{E}}$. This filter is called a matched-filter, in the sense that it is matched to the modulation waveform $\phi(t)$.
- this can be written in the form of an LTI filter $h(t)$ at time instant T, as

$$y = \int_0^T \frac{\phi^*(t)}{\sqrt{\mathcal{E}}} r(t) dt \quad (1)$$

$$= \int_0^T h(T-t) r(t) dt \quad (2)$$

hence the matched-filter $h(t) = \phi^*(T-t)/\sqrt{\mathcal{E}}$
- for general purpose h filter, defining the signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = \frac{|\int_0^T h(T-t)x(t)dt|^2}{\mathrm{E}|\int_0^T h(T-t)n(t)dt|^2}$$

it can be shown that the SNR maximizing filter $h(t)$ is the matched-filter. In which case,

$$\mathrm{SNR} = \frac{\mathcal{E}}{N_0}$$

(this is a direct application of the Cauchy-Schwartz inequality)

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

System model

- we now treat the general model, already introduced,

$$r(t) = x_m(t) + n(t)$$

  for the M functions $(x_1(t), \dots, x_M(t))$.
- call N the dimension of the space $\mathrm{Span}(x_1(t), \dots, x_M(t))$ in the space of finite-time (T) finite-energy functions.
- in the same way as previously, consider $\{\xi_i(t)\}_{i\geq 1}$ an orthonormal basis for the inner product $< x(t), y(t) >= \int x^*(t)y(t)dt$, such that $\xi_1(t), \dots, \xi_N(t)$ is the orthonormal basis of $\mathrm{Span}(x_1(t), \dots, x_M(t))$.
- denote

$$y_1 = \int \xi_1^*(t)x_m(t)dt + \int \xi_1^*(t)n(t)dt \triangleq x_{m,1} + \nu_1$$

$$\vdots$$

$$y_N = \int \xi_N^*(t)x_m(t)dt + \int \xi_N^*(t)n(t)dt \triangleq x_{m,N} + \nu_N$$

$$y_{N+1} = \int \xi_{N+1}^*(t)x_m(t)dt + \int \xi_{N+1}^*(t)n(t)dt \triangleq \nu_{N+1}$$

$$\vdots$$

M-ary detection

- under white Gaussian assumption of the noise, $r_1, \ldots, r_N$ is a sufficient statistics for m.
- we can then work from now on with the vectorial model

$$\mathbf{y} = \mathbf{x}_m + \nu$$

with $\mathbf{y} = (y_1, \ldots, y_N)^\mathsf{T}$ and $\nu = (\nu_1, \ldots, \nu_N)^\mathsf{T}$.

- the vector ν is a (N)-multivariate complex white Gaussian process with power N_0,

$$p(\nu) = \frac{1}{(\pi N_0)^N} e^{-\frac{1}{N_0}\nu^\mathsf{H}\nu}$$

- in this case, the MAP decision $\hat{m}$ is given by

$$\begin{aligned}\hat{m} &= \arg\max_{1\leq m\leq M} p(m)p(\mathbf{y}|m)\\ &= \arg\max_{1\leq m\leq M} p(m)\frac{1}{(\pi N_0)^N} e^{-\frac{1}{N_0}\|\mathbf{y}-\mathbf{x}_m\|^2}\end{aligned}$$

- taking the log, this is

$$\hat{m} = \arg\max_{1\leq m\leq M} \log p(m) - \frac{1}{N_0}\|\mathbf{y} - \mathbf{x}_m\|^2$$

- in the case where $p(m) = 1/M$ for all m,

$$\hat{m} = \arg\min_{1\leq m\leq M} \|\mathbf{y} - \mathbf{x}_m\|^2$$

M-ary detection (2)

- since $\|\mathbf{y}\|^2$ does not intervene in the minimization, this can be rewritten

$$\hat{m} = \arg \max_{1 \leq m \leq M} 2\Re(\mathbf{y}^{\mathrm{H}}\mathbf{x}_m) - \|\mathbf{x}_m\|^2$$

  where we remind

$$\mathbf{x}_m^{\mathrm{H}}\mathbf{y} = \int x_m(t)^* r(t) dt$$

  and

$$\mathbf{x}_m^{\mathrm{H}}\mathbf{x}_m = \int |x_m(t)|^2 dt$$

- therefore the detection method here sums up to
  1. computing all correlations $r_m \triangleq \int x_m(t)^* r(t) dt$, for $1 \leq m \leq M$
  2. compute the values $2\Re(r_m) - \|\mathbf{x}_m\|^2$
  3. $\hat{m}$ is that m for which $2\Re(r_m) - \|\mathbf{x}_m\|^2$ is maximal

Error analysis

- as previously, we define $P(e)$ as

$$P(e) = \sum_{m=1}^{M} p(m) P(\hat{m} \neq m)$$
$$= \sum_{m=1}^{M} \sum_{m' \neq m} p(m) P(m \to m')$$

- in the case when $p(m) = 1/M$,

$$P(m \to m') = P(\|\mathbf{y} - \mathbf{x}_m\|^2 \geq \|\mathbf{y} - \mathbf{x}_{m'}\|^2 | m)$$
$$= P(2\Re([\mathbf{x}_{m'} - \mathbf{x}_m]^{\mathrm{H}} \nu) \geq \|\mathbf{x}_{m'} - \mathbf{x}_m\|^2)$$

- now, ν is Gaussian and then, by linearity, $2\Re([\mathbf{x}_{m'} - \mathbf{x}_m]^{\mathrm{H}} \nu)$ is also Gaussian, with zero mean and covariance matrix $2N_0\|\mathbf{x}_{m'} - \mathbf{x}_m\|^2$.
- this finally gives

$$P(m \to m') = Q\left(\sqrt{\frac{\|\mathbf{x}_{m'} - \mathbf{x}_m\|^2}{2N_0}}\right)$$

- summing up and averaging over m, we finally have the upper bound

$$P(e) \leq \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m} Q\left(\sqrt{\frac{\|\mathbf{x}_{m'} - \mathbf{x}_m\|^2}{2N_0}}\right)$$

Error analysis (2)

- observe that the error depends here on the distances between the constellation points $\mathbf{x}_m$ which are known in advance. These distances are function of the energy every waveform $x_m(t)$ carries.
- denoting E_s the mean energy per symbol, i.e.

$$E_s = \sum_{m=1}^{M} p(m) \|\mathbf{x}_m\|^2$$

which in the uniform case is

$$E_s = \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{x}_m\|^2$$

- since there are M symbols, according to Shannon's first theorem, these symbols can carry (without distortion) a maximum of

$$-\sum_{m=1}^{M} p(m) \log_2(p(m)) \text{ bits.}$$

- in the uniform case, this is simply

$$\log_2(M) \text{ bits.}$$

in which case the mean energy E_b transported by every bit is

$$E_b = E_s / \log_2(M)$$

Error analysis (3)

- from this, we can rewrite the upper bound on $P(e)$ as a function of E_b/N_0, as

$$P(e) \leq \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m} Q\left(\sqrt{\frac{\log_2(M) d^2(m,m')}{2} \frac{E_b}{N_0}}\right)$$

with $d(m,m')$ the energy-normalized distance between symbols $\mathbf{x}_m$ and $\mathbf{x}_{m'}$ in $\mathbb{C}^N$

$$d(m,m') = \frac{1}{\sqrt{E_s}} \|\mathbf{x}_{m'} - \mathbf{x}_m\|$$

- we similarly have the lower bound

$$P(e) \geq \frac{1}{M} \sum_{m=1}^{M} Q\left(\sqrt{\frac{\log_2(M) \min_{m' \neq m} d^2(m,m')}{2} \frac{E_b}{N_0}}\right)$$

Looser bounds and explicit approximation of $P(e)$

- note that the upper-bound can be further upper-bounded by replacing all distances $d(m, m')$ by the minimum distance

$$d_{\min} \triangleq \min_{\substack{1 \leq m \leq M \\ 1 \leq m' \leq M \\ m \neq m'}} d(m, m')$$

  in which case, we have

$$P(e) \leq (M-1) Q\left(\sqrt{\frac{\log_2(M) d_{\min}^2}{2} \frac{E_b}{N_0}}\right)$$

- as for the lower bound, it can be further lower bounded only considering the terms m in the sum that have minimum distance $d(m, m')$ over m' that is exactly $d_{\min}$. Denote then

$$M_{\min} \triangleq \# \left\{ 1 \leq m \leq M, \min_{\substack{1 \leq m' \leq M \\ m' \neq m}} d(m, m') = d_{\min} \right\}$$

  in which case,

$$P(e) \geq \frac{M_{\min}}{M} Q\left(\sqrt{\frac{\log_2(M) d_{\min}^2}{2} \frac{E_b}{N_0}}\right)$$

Looser bounds and explicit approximation of $P(e)$

- this finally leads to an approximation of the probability of error by considering, instead of $(M-1)$ or $M_{\min}/M$, the mean number of points that are at distance $d_{\min}$ of each m, defined as

$$K \triangleq \frac{1}{M}\sum_{m=1}^{M} \#\left\{1 \leq m' \leq M, m' \neq m, d(m,m') = d_{\min}\right\}$$

then

$$P(e) \simeq KQ\left(\sqrt{\frac{\log_2(M)d_{\min}^2}{2}\frac{E_b}{N_0}}\right)$$

- using now the classical bounds on the Q function,

$$\frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} < Q(x) < \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

we finally have for large ratios E_b/N_0

$$P(e) = O\left(e^{-\frac{1}{4}\log_2 M d_{\min}^2 \frac{E_b}{N_0}}\right)$$

which is compliant with both lower and upper bounds.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. **Digital modulation**
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Introduction

- for practical applications, in *coherent detection* scheme, the signal set $x_1(t), \dots, x_M(t)$ is simply given by
$$x_m(t) = a_m\phi(t)$$
in which case the dimension space $\mathrm{Span}(x_1(t), \dots, x_M(t))$ is generated by $\xi_1(t) = \phi(t)/\sqrt{E_s}$ and a single matched-filter can be applied.
- the complex points
$$x_m \triangleq \int \xi_1(t)^* x_m(t)dt = a_m\sqrt{E_s}$$
are called the constellation symbols. They constitute the constellation
$$\mathcal{C} = \{x_1, \dots, x_M\}$$
- to ensure consistency with the SNR definition, we assume $\sum_{m=1}^M p(a_m)|a_m|^2 = 1$.
- in the following, we introduce classical constellations vectors $(x_1, \dots, x_M)$.
- intuitively, for fixed E_s, the larger M, the larger $P(e)$, but the larger M, the larger the number of transmitted *bits per channel use*. So there is a trade-off between large constellations (and then large transmission rates) and secured communications (low decoding error rate).

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. **Digital modulation**
   - **M-PAM modulation**
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

4-PAM modulation

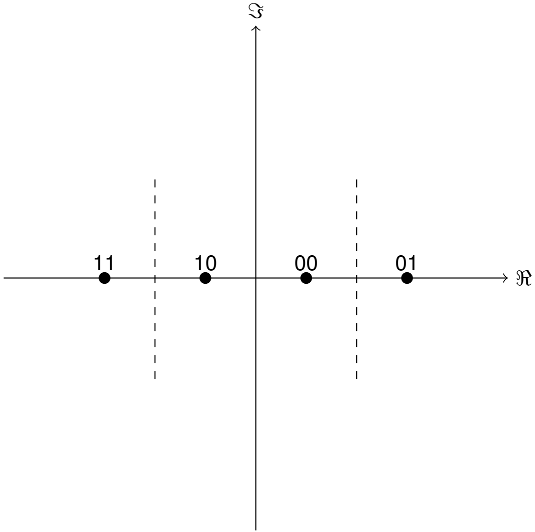


Figure: 4-PAM modulation constellation

M-PAM modulation

- the alphabet of an M-PAM (Pulse Amplitude Modulation) is described as

$$\mathcal{C} = \left\{ 2d\left(m - \frac{M-1}{2}\right), m \in \{0, \dots, M-1\} \right\}$$

with $2d$ the distance between neighboring constellation points, i.e.

$$d_{\min} = 2d$$

- by definition of E_s, the mean energy per symbol,

$$E_s = \log_2(M) E_b = \frac{1}{M} \sum_{m=0}^{M-1} d^2 (2m - M + 1)^2 = d^2 \frac{M^2 - 1}{3}$$

hence the upper-bound

$$P(e) \leq 2Q\left(\sqrt{\frac{6\log_2(M)}{M^2-1}\frac{E_b}{N_0}}\right)$$

where the coefficient in front is 2 instead of $M-1$, since all points further than $d_{\min}$ are completely inside the decision region of points at $d_{\min}$.

- an exact value of $P(e)$ can in fact be found,

$$P(e) = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6\log_2(M)}{M^2-1}\frac{E_b}{N_0}}\right)$$

(this is obtained by considering errors on inner points and outer points independently)

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

M-QAM modulation

- when performing PAM modulation, all the complex imaginary dimension is left unused, and decisions are made based on $\Re(y)$.
- this is a waste of the available space, that can be exploited further using the so-called QAM (Quadrature Amplitude Modulation).
- the constellation set is here

$$\mathcal{C} = \left\{ 2d\left(m - \frac{M-1}{2} + i(m' - \frac{\sqrt{M}-1}{2}) \right), m, m' \in \{0, \ldots, \sqrt{M}-1\} \right\}$$

- obviously, E_s is now

$$E_s = 2d^2 \frac{M-1}{3}$$

and $d_{\min} = 2d$.

- the calculus of $P(e)$ is somewhat different here. We notice that the real and imaginary parts of the incoming signal are totally independent for white Gaussian noise *(we in fact need to add the assumption that the noise is circularly symmetric, i.e.* $\mathrm{E}[\Re(\nu)\Im(\nu)] = 0$*)*. Therefore

$$P(e) = 1 - (1 - P_{\mathrm{PAM}}(e))^2$$

where $1 - P_{\mathrm{PAM}}(e)$ is the correct detection probability of a $\sqrt{M}$-PAM.

- note importantly that this is not $P(e) = P_{\mathrm{PAM}}(e)^2$! Indeed, to be erroneous, one needs to be erroneous *either* in the in-phase or in the quadrature component, while to be correct one needs to be *both* correct in the two components, so $1 - P(e) = (1 - P_{\mathrm{PAM}}(e))(1 - P_{\mathrm{PAM}}(e))$ by independence (for independent A, B, $P(AB) = P(A)P(B)$).

16-QAM modulation

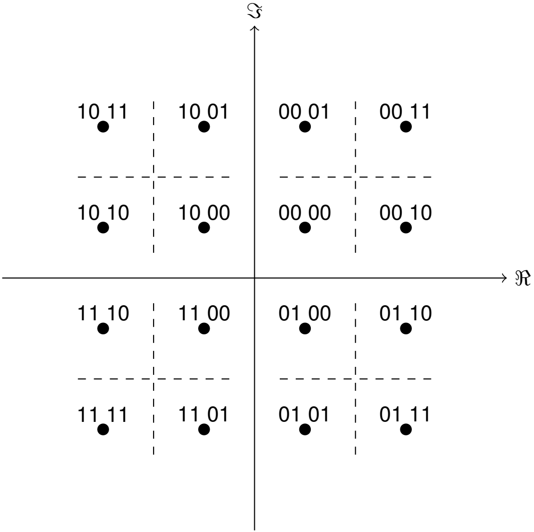


Figure: 16-QAM modulation constellation

Outline

1 Introduction and Reminders
- Course introduction
- Autocorrelation and power spectrum density
- Random Processes
- Baseband complex equivalent
- Signal sampling

2 Introduction to Information Theory
- Communication chain
- Spectral efficiency and Shannon Theorem

3 Signal detection in AWGN
- Bayesian hypothesis testing and ML detection
- Matched-filtering and detection of binary signals
- Coherent detection of M-ary signals

4 Digital modulation
- M-PAM modulation
- M-QAM modulation
- M-PSK modulation
- A word on non-coherent detection

5 Sequence detection in AWGN
- Symbol-by-symbol detector
- Nyquist condition
- ML detection and the Viterbi algorithm

8-PSK modulation

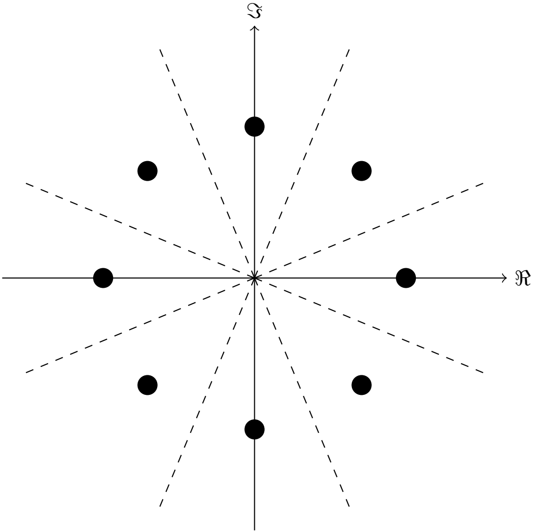


Figure: 8-PSK modulation constellation

M-PSK modulation

- *M*-PSK (Phase-Shift Modulation) symbols are encoded in the phase and have constant magnitude $\sqrt{E_s}$, (which is of interest for power amplifiers)
- the constellation is here

$$\mathcal{C} = \left\{ de^{2\pi i \frac{m}{M}}, 0 \leq m \leq M-1 \right\}$$

- $d_{\min}$ is given by trigonometric rules,

$$d_{\min} = 2d \sin(\pi/M)$$

  for $d = \sqrt{E_s}$.
- an upper-bound on the decoding error reads

$$P(e) \leq 2Q\left(\sqrt{2\log_2(M)\sin^2(\pi/M)\frac{E_b}{N_0}} \right)$$

  where the term 2 arises from the fact that decision regions for points further than $d_{\min}$ are included into the decision regions for points at $d_{\min}$.

Outline

1 Introduction and Reminders
- Course introduction
- Autocorrelation and power spectrum density
- Random Processes
- Baseband complex equivalent
- Signal sampling

2 Introduction to Information Theory
- Communication chain
- Spectral efficiency and Shannon Theorem

3 Signal detection in AWGN
- Bayesian hypothesis testing and ML detection
- Matched-filtering and detection of binary signals
- Coherent detection of M-ary signals

4 Digital modulation
- M-PAM modulation
- M-QAM modulation
- M-PSK modulation
- A word on non-coherent detection

5 Sequence detection in AWGN
- Symbol-by-symbol detector
- Nyquist condition
- ML detection and the Viterbi algorithm

Non-coherent detection

- remember the general AWGN communication model

$$y(t) = x_m(t)e^{2\pi i\theta} + n(t)$$

which, after matched-filtering, leads to the vectorial model

$$\mathbf{y} = \mathbf{x}_m e^{2\pi i\theta} + \nu$$

- it is possible to perform decoding even when θ is unknown.
- in this case, we have the MAP decision

$$\hat{m} = \arg\max_m p(m|\mathbf{y}) = \arg\max_m p(\mathbf{y}|m)p(m) = \arg\max_m p(m)\int_0^1 p(\mathbf{y}|m,\theta)p(\theta)d\theta$$

- in the case where $p(m)$ and $p(\theta)$ are uniform, one obtains

$$p(\mathbf{y}|m) = \frac{1}{2\pi(\pi N_0)^N} e^{-\frac{\|\mathbf{y}\|^2+\|\mathbf{x}_m\|^2}{N_0}} I_0\left(\frac{2}{N_0}|\mathbf{y}^{\mathrm{H}}\mathbf{x}_m|\right)$$

with I_0 the modified Bessel function.
- we finally have the MAP/ML decision rule

$$\hat{m} = \arg\max_m |\mathbf{y}^{\mathrm{H}}\mathbf{x}_m|^2$$

- note in particular that M-PSK modulation leads to no possible decision here.
- other constellations than those used here need be defined.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. **Sequence detection in AWGN**
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Introduction

- up to now, we have only dealt with the transmission of a unique symbol in time, taken among a constellation of M points.
- for practical communications, one needs to transmit a continuous waveform of symbols
- a first idea is to transmit successive finite-time finite-energy waveforms
- however, this comes with a very low transmission rate
- we study here the transmission of a continuous waveform of multiple symbols in time and the decoding procedure for such sequences.

Transmission model

- we consider here the transmission waveform

$$x(t;\mathbf{a}) = \sum_{n=0}^{N-1} a_n \phi(t - nT_s)$$

  where T_s is the symbol period, $\mathbf{a} = (a_0, \ldots, a_{N-1})$ belongs to a particular constellation (PAM, QAM etc.), and ϕ is a finite-energy waveform, i.e.

$$\int |\phi(t)|^2 dt = E_s < \infty$$

- through the AWGN channel, assuming coherent detection (hence the phase angle θ can be discarded), this becomes

$$y(t) = \sum_{n=0}^{N-1} a_n \phi(t - nT_s) + n(t)$$

Outline

1 Introduction and Reminders
- Course introduction
- Autocorrelation and power spectrum density
- Random Processes
- Baseband complex equivalent
- Signal sampling

2 Introduction to Information Theory
- Communication chain
- Spectral efficiency and Shannon Theorem

3 Signal detection in AWGN
- Bayesian hypothesis testing and ML detection
- Matched-filtering and detection of binary signals
- Coherent detection of M-ary signals

4 Digital modulation
- M-PAM modulation
- M-QAM modulation
- M-PSK modulation
- A word on non-coherent detection

5 Sequence detection in AWGN
- Symbol-by-symbol detector
- Nyquist condition
- ML detection and the Viterbi algorithm

Inter-Symbol Interference

- using the matched-filter at time instant n, the filtered waveform becomes,

$$y_n = \sum_{l=-L}^{L} g_l a_{n-l} + v_n$$

  where
  - $v_n = 1/\sqrt{E_s} \int \phi(t - nT_s)^* n(t)dt$
  - $y_n = 1/\sqrt{E_s} \int \phi(t - nT_s)^* y(t)dt$
  - L is such that, for all $t > 0$, $|\phi(t + LT_s)|^2 \simeq 0$, $|\phi(-t - LT_s)|^2 \simeq 0$
  - $g_l = 1/\sqrt{E_s} \int \phi(t - nT_s)^* \phi(t - lT_s)dt$ (in particular, $g_0 = E_s$)
- y_n is better rewritten

$$y_n = g_0 a_n + \sum_{l \neq 0} g_l a_{n-l} + v_n$$

  where a_n is the effective symbol due at instant n, and the term $\sum_{l \neq 0} g_l a_{n-l}$ is an interfering term, referred to as Inter-Symbol Interference (ISI).

Symbol-by-Symbol detector

- since the ISI interferes the term a_n to be decoded, it might as a first idea be considered as part of the ambient noise. For $a_n \in \{-1, 1\}$, we consider the estimate $\hat{a}_n = \mathrm{sign}(y_n)$.
- as a first approximation, considering that the ISI term is Gaussian (which is obviously incorrect), we have the detection error approximation

$$P(e) \simeq Q\left(\sqrt{\frac{|g_0|^2}{\frac{N_0}{2}g_0 + \frac{1}{2}(\sum_{l\neq 0}|g_l|^2)}}\right)$$

  in case of a binary antipodal transmission (i.e. $a_m \in \{-1, 1\}$).
- this approximation is however only valid for low SNR (i.e. small N_0). Indeed, asymptotically large SNR incur a non-zero error-rate, which is inconsistent (as long as $\sum_{l\neq 0}|g_l|^2 < |g_0|^2$)
- in case of binary antipodal constellation, we have exactly, with $I = \sum_{l\neq 0} a_l g_l$,

$$P(e|I) = \frac{1}{2}\left[Q\left(\frac{|g_0| - I}{\sqrt{|g_0|N_0/2}}\right) + Q\left(\frac{|g_0| + I}{\sqrt{|g_0|N_0/2}}\right)\right]$$

  hence, after integration over I,

$$P(e) = \frac{1}{2^{2L}}\sum_{\mathbf{a}\in\{-1,+1\}^{2L}} Q\left(\frac{|g_0| + \sum_{l\neq 0} a_l|g_l|}{\sqrt{|g_0|N_0/2}}\right)$$

Symbol-by-Symbol detector (2)

- considering the worst case scenario, i.e. for the configuration of $a_1, \dots, a_N$ that maximizes the error rate, we have

$$P(e) \leq Q\left(\sqrt{\frac{2|g_0|(1-D_p)^2}{N_0}}\right)$$

with D_p the *peak-distortion*,

$$D_p = \frac{\sum_{l \neq 0} |g_l|}{|g_0|}$$

- the parameter D_p characterizes the worst-case scenario of the eye opening in the eye-diagram.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

ST ERICSSON

ISI-free conditions

- the waveform $\phi(t)$ is said to fulfill Nyquist conditions, if $D_p = 0$, i.e. there is no ISI.
- in that case, the eye-diagram is fully opened, and symbol-by-symbol detection can be performed as in the single-symbol case.
- to ensure ISI-free conditions, we seek a waveform that is zero in all nT_s positions but 0 and that would be finite-time finite-energy.
- a first idea is to use the sinc function, but it is infinite-time and does not decay fast enough to be truncated without impairing the signal.
- another idea is to generate a finite-time waveform of length T_s. However this would increase the required bandwidth N-fold!
- a classically used waveform with such properties is the root-raised-cosine function, with Fourier transform

$$\mathcal{F}[\phi](f) = \begin{cases} \sqrt{T_s} & 0 \leq |f| \leq (1-\alpha)/(2T_s) \\ \sqrt{\frac{T_s}{2}}\sqrt{1-\sin(\pi T_s/\alpha(f-1/(2T_s)))} & (1-\alpha)/(2T_s) \leq |f| \leq (1+\alpha)/(2T_s) \end{cases}$$

  for $0 < \alpha < 1$ the roll-off (or excess-bandwidth) factor.
- the root-raised-cosine has bandwidth $W = (1+\alpha)/T_s$, where the choice of α is a trade-off between
  - setting α low to increase the spectral efficiency
  - setting α high to decrease the sensitivity to sampling period mismatch: a small error in the sampling instants should not raise large decoding errors, and α large ensures a larger opening of the eye along the horizontal axis.

Outline

1. Introduction and Reminders
   - Course introduction
   - Autocorrelation and power spectrum density
   - Random Processes
   - Baseband complex equivalent
   - Signal sampling
2. Introduction to Information Theory
   - Communication chain
   - Spectral efficiency and Shannon Theorem
3. Signal detection in AWGN
   - Bayesian hypothesis testing and ML detection
   - Matched-filtering and detection of binary signals
   - Coherent detection of M-ary signals
4. Digital modulation
   - M-PAM modulation
   - M-QAM modulation
   - M-PSK modulation
   - A word on non-coherent detection
5. Sequence detection in AWGN
   - Symbol-by-symbol detector
   - Nyquist condition
   - ML detection and the Viterbi algorithm

Introduction

- in the case of ISI-limited symbol sequences, the symbol-by-symbol detector (say, for a_n) is clearly suboptimal since it voluntarily discards the information about $a_1, \ldots, a_{n-1}, a_{n+1}, \ldots, a_N$.
- optimally, the MAP/ML detector would process the whole sequence $\mathbf{a} = (a_1, \ldots, a_N)$ to obtain

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{b}} p(\mathbf{b}|\mathbf{y})$$

  which leads to finding the best $\mathbf{b}$ that fits the particular behaviour observed, not only at time n but at all times.
- however, the problem here is largely more involved since we need to find the one solution among M^N hypothesis. For N large, an exhaustive search is out of the question.
- fortunately, we have the so-called Viterbi algorithm, which allows to significantly reduce the computational complexity.

The Viterbi Algorithm

- consider again the transmission model

$$y(t) = \sum_{n=0}^{N-1} a_n \phi(t - nT_s) + \nu(t)$$

- the ML decision is based on the metric

$$m(\mathbf{y}; \mathbf{b}) \triangleq 2\Re\left(\int y(t) \sum_{n=0}^{N-1} b_n^* \phi^*(t - nT_s)\right) - \int \left| \sum_{n=0}^{N-1} b_n \phi(t - nT_s) \right|^2 dt$$

which can be rewritten

$$m(\mathbf{y}; \mathbf{b}) = 2\Re\left(\sum_{n=0}^{N-1} b_n^* \left[y_n - \frac{1}{2} g_0 b_n - \sum_{l=1}^{L} b_{n-l} g_l \right]\right)$$

where we assumed again that g_l is zero for $l > L$.

The Viterbi Algorithm (2)

- for a fixed $1 \leq k \leq N$, denote

$$m_k(\mathbf{y}; b_0, \ldots, b_k) = 2\Re\left(\sum_{n=0}^{k} b_n^* \left[y_n - \frac{1}{2} g_0 b_n - \sum_{l=1}^{L} b_{n-l} g_l\right]\right)$$

- observe that this can be rewritten

$$m_k(\mathbf{y}; b_0, \ldots, b_k) = m_{k-1}(\mathbf{y}; b_0, \ldots, b_{k-1}) + 2\Re\left(b_k^* \left[y_k - \frac{1}{2} g_0 b_k - \sum_{l=1}^{L} b_{k-l} g_l\right]\right)$$

the second term being only a function of the L terms $b_{k-L}, \ldots, b_{k-1}$ and the additional term b_k.

The Viterbi Algorithm (3)

- now consider a state machine in N stages, with k-th stage a group of all possibles states of $(b_{k-1}, \ldots, b_{k-L})$ (there exist M^L of those for $k \geq L$, and M^k otherwise). This can be looked as a shift register with content $(b_{k-1}, \ldots, b_{k-L})$ at time k.
- to end up in the particular state $(b_k, \ldots, b_{k-L+1})$ at step k, one could have taken a maximum of M paths from $(b_{k-1}, \ldots, b_{k-L})$ at step $k-1$, each path corresponding to a particular value of b_{k-L}.
- obviously, the path with maximum value of $m_k(\mathbf{y}; b_0, \ldots, b_k)$ will dominate those with shorter values in the final $m(\mathbf{y}; \mathbf{b})$. Those can be already discarded. The remaining path is called the survivor.
- starting from stage 1, one can then compute all metrics on the paths leading to each of the M^2 states (b_1, b_0) for every possible b_0, which is a obviously unique link for every possible (b_1, b_0).
- one then moves to stage 2, 3 etc. until it reaches stage $L+1$.
- at stage $L+1$, the new states $(b_{L+1}, \ldots, b_2)$ could have arisen from M different values of b_1. Hence, M path metrics must be computed for every of the M^L vectors $(b_{L+1}, \ldots, b_2)$ arising from the M vectors $(b_L, \ldots, b_2, b_1)$ (one for each b_1), and one keeps only the surviving ones.
- subsequent paths behave the same, until we reach the final stage N.
- at the end of the trellis, i.e. at stage N, M^L surviving paths are subsisting. A trace-back is then applied, that consists in comparing those M^L path metrics. The path with largest metrics is the ML solution.