KTH Computer Science
and Communication

2E1395 - Pattern Recognition
Solutions to Introduction to Pattern Recognition, Chapter 1: Matched Filters

Preface

This document[1] is a solution manual for selected exercises from "Introduction to Pattern Recognition" by Arne Leijon. The notation followed in the text book will be fully respected here. A short review of the issue discussed in the corresponding chapter of the text book is given here as a reference. For a complete proof of these results and for the problem text refer to the text book.

Problem definition and optimization

A two-category signal classification system (fig. 1) consists of three parts: a source, a channel, and a receiver (classifier).

- The source can generate one of two possible discrete-time signals, $s_0(n)$ or $s_1(n)$, according to a random selector. The signal $s_0(n)$ is selected with probability p_0 and $s_1(n)$ with probability $p_1 = 1 - p_0$.
- The channel introduces some addictive noise $W(n)$ whose stochastic characteristic is known. Note that this part of the system can be thought of as being part of the source without any change in the following discussion.
- the receiver consists of a linear filter and a threshold-based classifier.

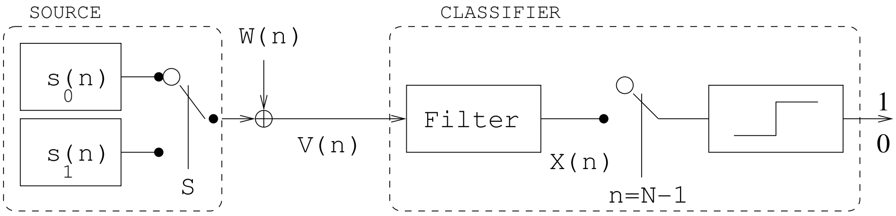

Figure 1. A two-category classification system

Giampiero Salvi (giampi@kth.se)

Given the source and the channel, we have to find the optimal classifier in terms of minimizing the probability of classification error. This problem has been solved under a number of assumptions[2]:

- The shape of $s_0(n)$ and $s_1(n)$ is known and $s_i(n) = a_i e(n)$, i.e. the signals are proportional to each other.
- The length of $s_i(n)$ is finite and equal to N, i.e. $s_i(n) = 0 \quad \forall n > N-1$ and $n < 0$.
- The noise introduced by the channel is a wide-sense stationary process with zero mean: $E[W(n)] = 0$.
- The noise introduced by the channel is white: $r_W(k) = E[W(n+k)W(n)] = 0; k \neq 0$ (but we will see that this assumption is not necessary).
- The time at which the source signal $s_i(n)$ is generated is known.
- The variance of the noise $\sigma_W^2 = r_W(0)$ and the probabilities of selecting one of the source signals p_0 and p_1 are known.

When the previous conditions are verified, the problem is solved employing a finite impulse response filter of order N (see fig. 2). The output of this filter $X(n)$ is a linear combination of N input samples $V(n)$, $V(n-1)$, ..., $V(n-N+1)$, with coefficients $b(0)$, $b(1)$, ..., $b(N-1)$. We note how, at the time step $n = N-1$, all and only the non zero samples of the source signal $s_i(n)$ contribute to the output. It is then convenient to consider the output of the classifier only at this particular time step (the maximum amount of useful information is used). This will also ensure that, in case $s_i(n)$ is followed or preceded by another signal, this will not affect results (see exercise 1.1b). The output $X(N-1)$ of the filter at this particular time step is a random variable whose probability distribution has to be computed if we want to characterize the performance of the system. Using vector notation:

$$\mathbf{e} \doteq \begin{pmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{pmatrix}, \quad \mathbf{W} \doteq \begin{pmatrix} W(0) \\ W(1) \\ \vdots \\ W(N-1) \end{pmatrix}, \quad \mathbf{b} \doteq \begin{pmatrix} b(N-1) \\ b(N-2) \\ \vdots \\ b(0) \end{pmatrix}$$

the expected value of $X(N-1)$, depending on the state of the source is:

$$\mu_0 = E[X(N-1)|s_0] = E[\mathbf{b}^t(a_0\mathbf{e} + \mathbf{W})] = a_0\mathbf{b}^t\mathbf{e}$$

$$\mu_1 = E[X(N-1)|s_1] = E[\mathbf{b}^t(a_1\mathbf{e} + \mathbf{W})] = a_1\mathbf{b}^t\mathbf{e}$$

The variance of $X(N-1)$ does not depend on the state of the source:

$$\sigma_X^2 = \begin{cases} \mathbf{b}^t\mathbf{C}\mathbf{b} & \text{gaussian noise} \\ \|\mathbf{b}\|^2\sigma_W^2 & \text{gaussian white-noise} \end{cases}$$

After optimization and normalization (normalized matched filter):

$$\mathbf{b} = \frac{\mathbf{e}}{\|\mathbf{e}\|}$$

[2] the assumptions mentioned here might seem too restrictive, but they can be satisfied by important classes of problem. One such example is the RADAR system, in which the echos received are a perturbation of the signal that had been transmitted and that is known by the receiver

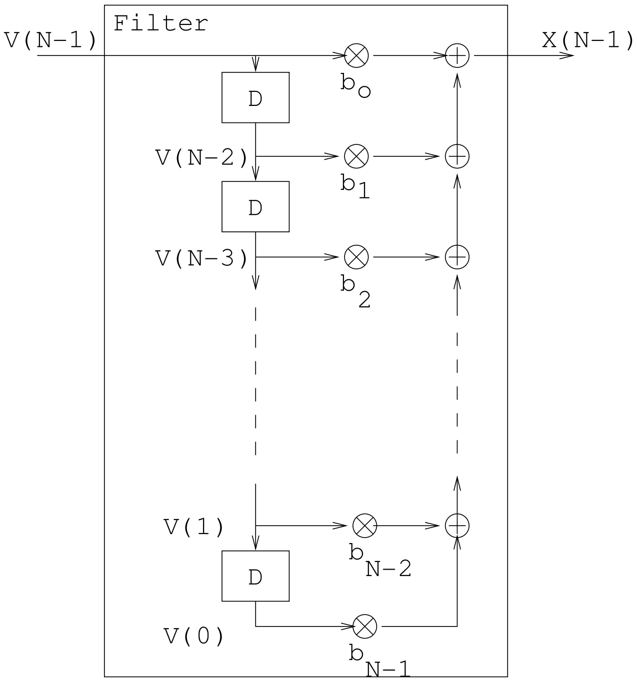

Figure 2. A finite impulse response linear filter

$$\begin{aligned}
\mu_0 &= a_0\|\mathbf{e}\| \\
\mu_1 &= a_1\|\mathbf{e}\| \\
\sigma_X^2 &= \begin{cases} \frac{1}{\|\mathbf{e}\|^2}\mathbf{e}^t\mathbf{C}\mathbf{e} & \text{gaussian noise} \\ \sigma_W^2 & \text{gaussian white-noise} \end{cases}
\end{aligned}$$

Additional information about the shape of the probability distribution for the noise $W(n)$ is needed if we want to compute the distribution $f_X(x)$. For example in the case of Gaussian white noise, $f_X(x)$ is also Gaussian:

$$f_{X|S}(x|s_i) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_i)^2}{2\sigma_W^2}}$$

Note that this probability distribution depends on the signal the source selects.

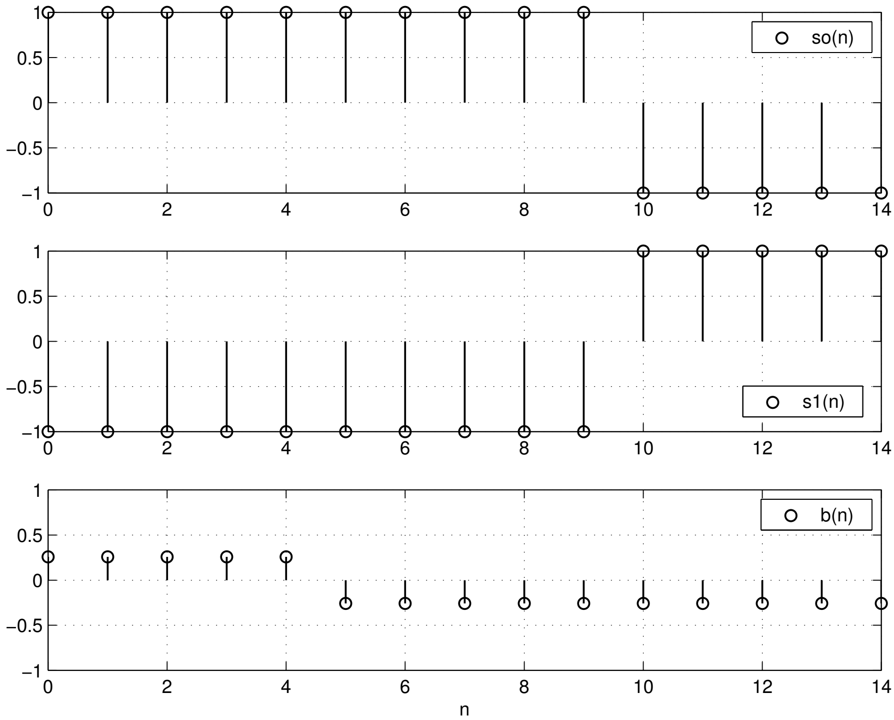

Figure 3. Input signals $s_0(n)$, $s_1(n)$ and filter coefficients $b(n)$.

Exercise 1.1

This is a standard problem of two-category classifier solved in the text book. The source signals are:

$$s_0(n) = \begin{cases} a & 0 \le n \le 9 \\ -a & 10 \le n \le 14 \\ 0 & \text{otherwise} \end{cases}$$

$$s_1(n) = -s_0(n)$$

a) Following the notation in the text book (chapter 1), we can write

$$\begin{aligned} s_0(n) &= a_0 e(n) = -e(n) \qquad (a_0 = -1) \\ s_1(n) &= a_1 e(n) = e(n) \qquad (a_1 = 1) \end{aligned}$$

where $e(n)$ is the sequence defined by[3]:

$$e(n) = \begin{cases} -a & 0 \le n \le 9 \\ a & 10 \le n \le 14 \\ 0 & \text{otherwise} \end{cases}$$

According to the book, the normalized matched filter coefficients are defined as:

$$b(n) = \frac{e(N-1-n)}{\|e(n)\|} = \begin{cases} 1/\sqrt{15} & 0 \le n \le 4 \\ -1/\sqrt{15} & 5 \le n \le 14 \\ 0 & \text{otherwise} \end{cases}$$

[3]Note that the choice for a_0 and a_1 is totally arbitrary: we might as well have set $a_0 = -a$, $a_1 = a$ and $e(n) = \pm 1$.

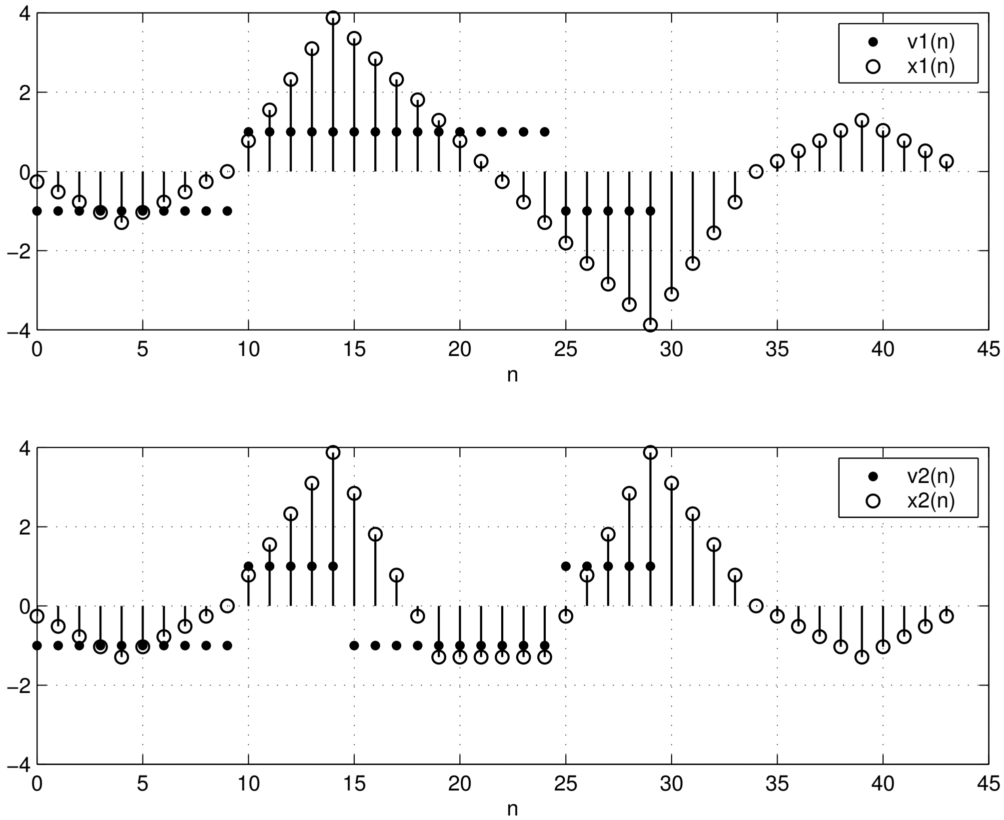

Figure 4. Input $V_i(n)$ and output $X(n)$ signals to the matched filter.

The plots of $s_0(n)$, $s_1(n)$ and $b(n)$ are in fig 3 for $a = 1$.

b)We know that the noise is not present in this case ($\sigma_W = 0$). The input to the filter will thus be:

$$V_1(n) = s_1(n) + s_0(n - N)$$
$$V_2(n) = s_1(n) + s_1(n - N)$$

The output is the convolution of the input with the impulse response of the filter (as for any linear system):

$$X_1(n) = V_1(n) \star b(n) = \sum_{k=-\infty}^{+\infty} V_1(k)b(n-k)$$
$$X_2(n) = V_2(n) \star b(n) = \sum_{k=-\infty}^{+\infty} V_2(k)b(n-k)$$

The solution to these sums is depicted in fig 4 for $a = 1$, where the black dots indicate the input $V_i(n)$ to the filter and the circles its output $X_i(n)$. It's worth noticing the values of the output $X(n)$ at the time steps $N - 1 = 14$ and $2N - 1 = 29$:

$$X_1(14) = 3.87 = \sqrt{15} = a_1\|\mathbf{e}\| = \mu_1$$
$$X_1(29) = -3.87 = -\sqrt{15} = a_0\|\mathbf{e}\| = \mu_0$$

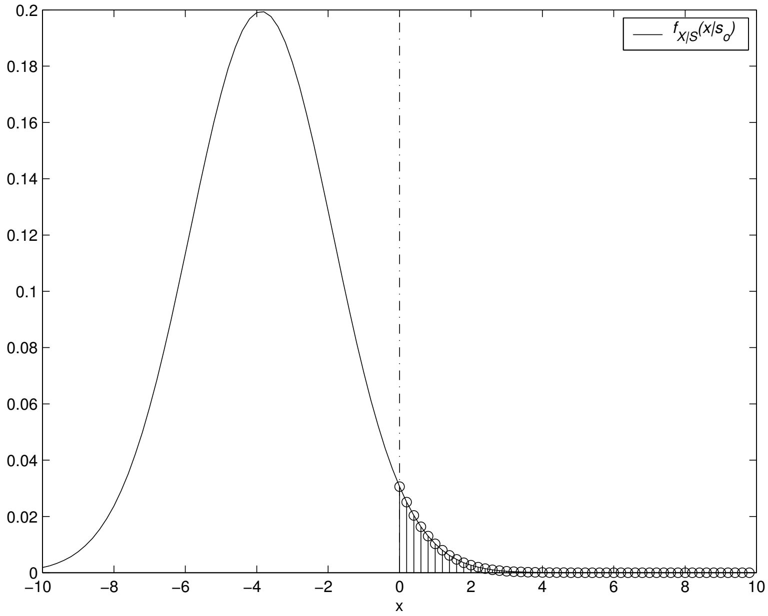

Figure 5.

$$X_2(14) = 3.87 = \sqrt{15} = a_1\|\mathbf{e}\| = \mu_1$$
$$X_2(29) = 3.87 = \sqrt{15} = a_1\|\mathbf{e}\| = \mu_1$$

showing that $X(mN-1)$ is not influenced by other samples than the ones in the signal we want to classify.

c) We now consider the case of noisy channel. $W(n)$ has Gaussian distribution and any two samples $W(n)$ and $W(m)$, $n \neq m$ are uncorrelated (white noise). Since $X(n)$ is a linear combination of samples from $W(n)$, its distribution is also Gaussian, with mean $\mu_i = a_i\|\mathbf{e}\|$ and variance $\sigma_X = \sigma_W$ (see text book). The probability of the joint event that $s_0(n)$ was generated and the classifier decides "1" ($D = 1$), can be written as:

$$P_{DS}(1 \cap s_0) = P_S(s_0)P_{D|S}(1|s_0) = P_S(s_0)P_{X|S}(X(N-1) > 0\,|s_0)$$

because the classifier decides 1 if the output of the filter exceeds the threshold 0. This can be written, according to the properties of the probability distributions (see also fig. 5), as:

$$\begin{aligned} P_{DS}(1 \cap s_0) &= P_S(s_0)\int_0^\infty f_{X|S}(x|s_0)dx \\ &= P_S(s_0)\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_W} e^{-\frac{(x-\mu_0)^2}{2\sigma_W^2}}dx \end{aligned} \tag{1}$$

As well known, this integral is not solvable analitically, and can be solved by defining (and tabling) a function called *error function*. The different definitions of the solving function are reported in table 1. In any case the procedure is to find the proper substitution that simplifies the integral we want to solve into one of those forms, and then to look up the values on a table or compute them with a calculator.

name	definition	solution method
error function	$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$	matlab: `erf`
complementary error function	$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$	matlab: `erfc`
distribution function	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$	BETA, pag. 405

Table 1.

In this manual we will always use the definition from BETA[4]. In particular, operating the substitution $t = \frac{x-\mu_0}{\sigma_W}$ in eq. 1 we obtain:

$$P_{DS}(1 \cap s_0) \quad = \quad P_S(s_0) \left[1 - \Phi \left(-\frac{\mu_0}{\sigma_W} \right) \right]$$

Substituting $P_S(s_0) = 1/2$, $\sigma_W = 2$ and $a = 1$ ($\mu_0 = -\sqrt{15}$), we obtain the numerical value: $P_{DS}(1 \cap s_0) \approx 0.0132$.

Exercise 1.2

In this exercise we see that even if we don't know exactly the shape of the source signals $s_i(n)$, but provided that we know their energy, we can still deduct some information about the system performance. We know that $s_1(n) = -s_0(n)$, that $s_0(n) = s_1(n) = 0 \quad \forall n < 0$ and $n > N-1$, and that the energy is fixes:

$$E = \sum_{n=0}^{N-1} s_0(n) = \sum_{n=0}^{N-1} s_1(n)$$

We also know that the additive zero-mean white Gaussian noise, disturbing the signal, has standard deviation σ_W.

a) From the normalized matched-filter optimization we know that

$$\mu_0 = a_0 \|e\| = -\sqrt{E}$$

$$\mu_1 = a_1 \|e\| = \sqrt{E}$$

NOTE that $E = a_i^2 \|\mathbf{e}\|^2$

b) As in exercise 1.1c, the probability of the combined event that s_1 was actually generated and the classifier decides "0" is given by:

$$P_{DS}(0 \cap s_1) = P_S(s_1)\, P_{D|S}(0|s_1)$$

where $P_S(s_1)$ is the probability that the source generates s_1, and $P_{D|S}(0|s_1)$ is the probability that the classifier decides "0" when the input is s_1. The source generates $s_0(n)$ and $s_1(n)$ with the same probability $P_S(s_1) = P_S(s_0) = \frac{1}{2}$. Moreover $P_{D|S}(0|s_1)$ is the probability that the output variable X assumes values lower than the threshold, when the input is s_1. We know that, under the last condition, and for a normalized matched filter, the variable X has a Gaussian distribution

[4]Beta, mathematics handbook, Studentlitteratur

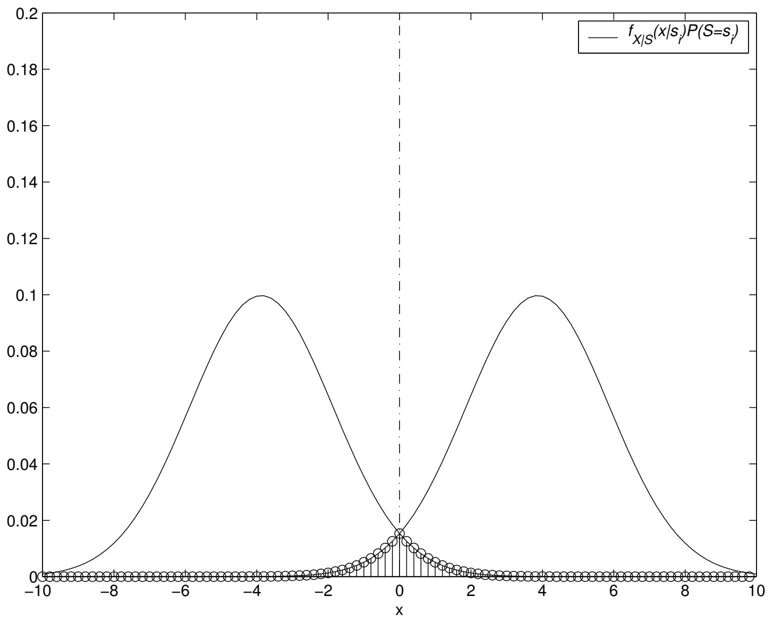

Figure 6.

with mean μ_1 (see point a) and variance $\sigma_X = \sigma_W$. We can hence compute $P_{D|S}(0|s_1)$ integrating the p.d.f. of X between $-\infty$ and 0 (the threshold value):

$$
\begin{aligned}
P_{DS}(0 \cap s_1) &= P_S(s_1) \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\,\sigma_W} e^{-\frac{(x-\mu_1)^2}{2\sigma_W^2}}\, dx \\
&= P_S(s_1)\Phi\left(-\frac{\mu_1}{\sigma_W}\right) \\
&= P_S(s_1)\left[1-\Phi\left(\frac{\mu_1}{\sigma_W}\right)\right] \\
&= P_S(s_1)\left[1-\Phi\left(\frac{\sqrt{E}}{\sigma_W}\right)\right]
\end{aligned}
$$

c) There is an error either because $s_0(n)$ was selected AND the classifier outputs "1" or because $s_1(n)$ was selected AND the classifier outputs "0". Being the two events disjoint, the total error probability can be written as

$$P_E = P_S(s_0)\,P_{D|S}(1|s_0) + P_S(s_1)\,P_{D|S}(0|s_1)$$

Since the problem is fully symmetric (see fig 6)

$$P_{D|S}(1|s_0) = P_{D|S}(0|s_1)$$

and

$$P_E = 2P_S(s_1)\,P_{D|S}(0|s_1)$$

that was computed in point b (beside the factor 2).

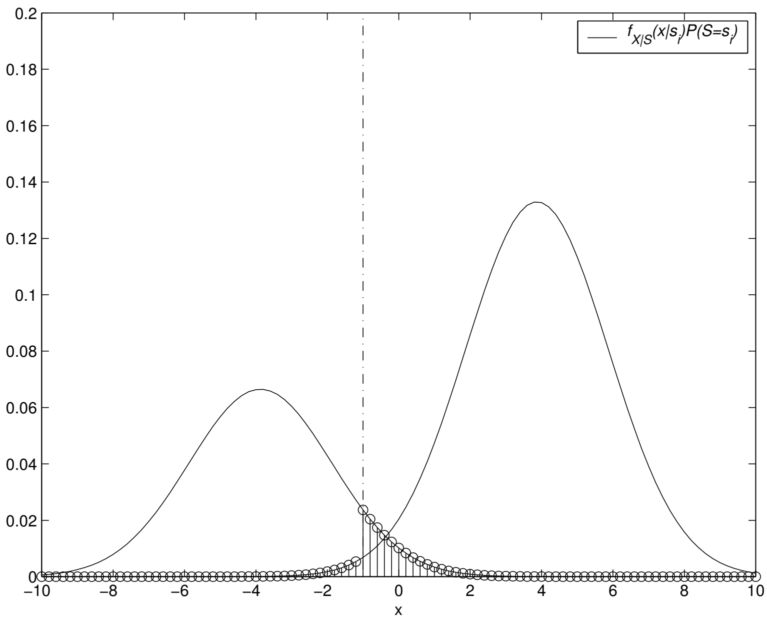

Figure 7.

Exercise 1.3

In this exercise the symmetry of exercise 1.2 is removed: the source generates the signal $s_0(n)$ with probability p_0 and the signal $s_1(n)$ with probability $p_1 = 1 - p_0 \neq p_0$. As a consequence the threshold value that leads to the minimum decision error is not 0 as in the previous case. Let's call the threshold value x_t. We have to express the decision error probability as a function of x_t. Remember that $f_{X|S}(x|s_i)$ does not depend on the value of $p_i = P_S(s_i)$ because it is the distribution of X given that the signal transmitted is known. We know that the total probability of error can be written as:

$$
\begin{aligned}
P_E &= P_S(s_0)\, P_{D|S}(1|s_0) + P_S(s_1)\, P_{D|S}(0|s_1) \\
&= p_0\, P_{X|S}(x > x_t|s_0) + (1 - p_0)\, P_{X|S}(x < x_t|s_1) \\
&= p_0 \int_{x_t}^{+\infty} f_{X|S}(x|s_0)dx + (1 - p_0) \int_{-\infty}^{x_t} f_{X|S}(x|s_1)dx \\
&= \int_{x_t}^{+\infty} \frac{p_0}{\sigma_W\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma_w^2}} dx + \int_{-\infty}^{x_t} \frac{1-p_0}{\sigma_W\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_w^2}} dx
\end{aligned}
$$

The two integrand functions are depicted in fig. 7, where the sum of the two areas corresponds to the value of P_E. Note how, in this case, the symmetry of fig. 6 is lost. Note also that, after integration, P_E will be a function of the variable x_t and that $P_E(x_t)$ can be optimized (i.e. minimized) imposing

$$\frac{dP_E(x_t)}{dx_t} = 0$$

To compute this derivative we refer to the general property of continuous functions:

$$\text{if} \quad F(\alpha) = \int_{-\infty}^{\alpha} f(x)dx \quad \Rightarrow \quad \frac{dF(\alpha)}{d\alpha} = f(\alpha)$$

then

$$\frac{dP_E(x_t)}{dx_t} = -\frac{p_0}{\sigma_W\sqrt{2\pi}}e^{-\frac{(x_t-\mu_0)^2}{2\sigma_W^2}} + \frac{1-p_0}{\sigma_W\sqrt{2\pi}}e^{-\frac{(x_t-\mu_1)^2}{2\sigma_W^2}}$$

where the minus sign in the first term comes from inverting the limits in the integral. Simplifying:

$$\begin{aligned}
p_0 e^{-\frac{(x_t-\mu_0)^2}{2\sigma_W^2}} &= (1-p_0)e^{-\frac{(x_t-\mu_1)^2}{2\sigma_W^2}} \\
\frac{p_0}{1-p_0}e^{\frac{(x_t-\mu_1)^2-(x_t-\mu_0)^2}{2\sigma_W^2}} &= 1 \\
(x_t-\mu_1)^2-(x_t-\mu_0)^2 &= 2\sigma_W^2 \ln\frac{1-p_0}{p_0} \\
2x_t\mu_0 - 2x_t\mu_1 &= 2\sigma_W^2 \ln\frac{1-p_0}{p_0}
\end{aligned}$$

and finally, since $\mu_1 = -\mu_0 = \sqrt{E}$,

$$x_t = \frac{\sigma^2}{2\sqrt{E}}\ln\frac{p_0}{1-p_0}$$

Comments:

- the optimal value for x_t is the point in which the two curves in fig 7 intersect. This result could be predicted looking at the figure and noticing that for any other choice of x_t the area that represents P_E increases
- when $p_0 = p_1 = 1/2$ the formula gives $x_t = 0$ as the symmetry of the problem suggests

Exercise 1.4

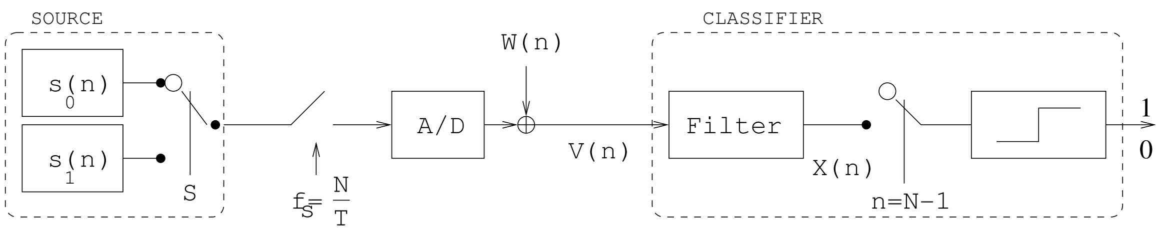

Figure 8.

This exercise is designed to point out how the performance of the system varies when we change the number of samples that influence the decision. We have two continuous signals

$$s_0(t) = \begin{cases} a & 0 \le t < T \\ 0 & \text{otherwise} \end{cases}$$

$$s_1(t) = -s_0(t)$$

but we take only N samples out of them (A/D converter with sampling rate $f_s = N/T$, see fig. 8). Note that, after the A/D converter, the problem is exactly the same as in the previous examples. We can then write:

$$\mu_0 = -\sqrt{E} = -a\sqrt{N}$$

$$\mu_1 = \sqrt{E} = a\sqrt{N}$$

$$\sigma_X = \sigma_W$$

and for the probability of error:

$$\begin{aligned}
P_E &= P_S(s_0)\int_0^{+\infty} f_{X|S}(x|s_0)dx + P_S(s_1)\int_{-\infty}^{0} f_{X|S}(x|s_1)dx \\
&= \frac{1}{2}\int_0^{+\infty} \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma_X^2}}dx + \frac{1}{2}\int_{-\infty}^{0} \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_X^2}}dx \\
&= \frac{1}{2}\left[1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu_0}{\sigma_X}} e^{-\frac{t^2}{2}}dt + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu_1}{\sigma_X}} e^{-\frac{t^2}{2}}dt\right] \\
&= \frac{1}{2}\left[1 - \Phi\left(-\frac{\mu_0}{\sigma_X}\right) + \Phi\left(-\frac{\mu_1}{\sigma_X}\right)\right] \\
&= \frac{1}{2}\left[1 - \Phi\left(\frac{a\sqrt{N}}{\sigma_W}\right) + \Phi\left(-\frac{a\sqrt{N}}{\sigma_W}\right)\right] \\
&= 1 - \Phi\left(\frac{a\sqrt{N}}{\sigma_W}\right)
\end{aligned}$$

substituting the values ($a = 1$, $\sigma_W = 10$ and $N = 400$), we obtain $P_E \simeq 0.0228$.

b) If we consider P_E as a function of N, we can easily prove that $P_E(N)$ goes to zero for N approaching infinity, see also fig. 9. This would suggest that increasing the sampling frequency

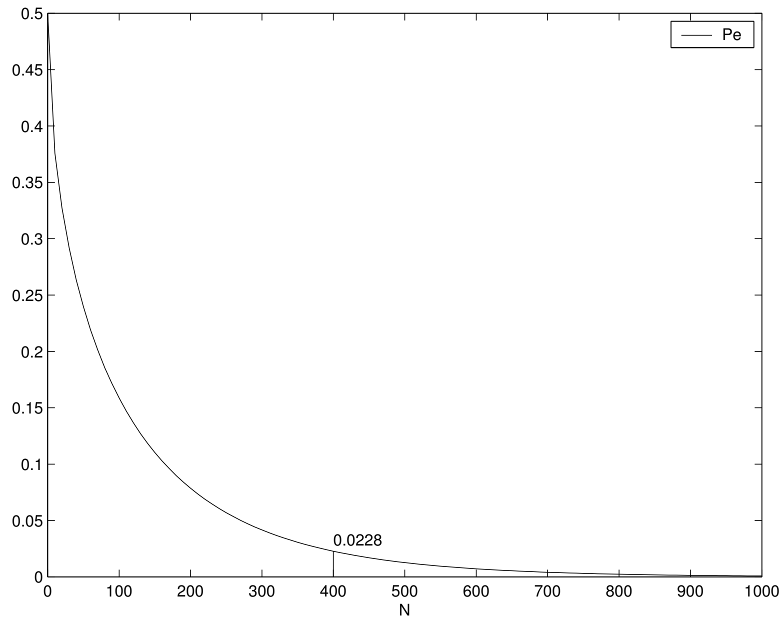

Figure 9. Probability of error in function of the number of samples N

is a good method for improving the system performance. In reality there are some limitations to this:

- there are physical and technological problems in creating a sampler and A/D converter at high frequencies: we have to expect that the higher the frequency is, the higher the noise introduced by the A/D converter (σ_W) would be. At a certain point the contribution of $\sqrt{N}$ in the formula would be lower than the contribution of σ_W and P_E would start rising again.

- the formula for P_E was found under the assumption that the noise sample are uncorrelated. This is always true for an ideal white noise, but in reality the white noise is only an approximation, and when the samples start to be very near to one another (high sampling frequencies), the correlation can be grater that zero. This means that σ_X is not equal to σ_W any longer (σ_X is the one affecting the value on P_E). Since σ_X is always grater than σ_W (for a proof look at the derivation of σ_X in the text book), we expect that P_E will be negatively affected by this problem.

Finally note that (fig. 9), for $N = 0$ the system takes a blind decision that, in the case of $p_0 = p_1 = 1/2$, leads to a probability of error equal to one half (random chance).

Exercise 1.5

In this exercise the problem pointed out in ex. 1.4b is discussed thoroughly. In this case the noise is inserted before sampling (continuous Gaussian noise), and the assumption of ideal white noise is removed. The signal is the same as in ex. 1.4:

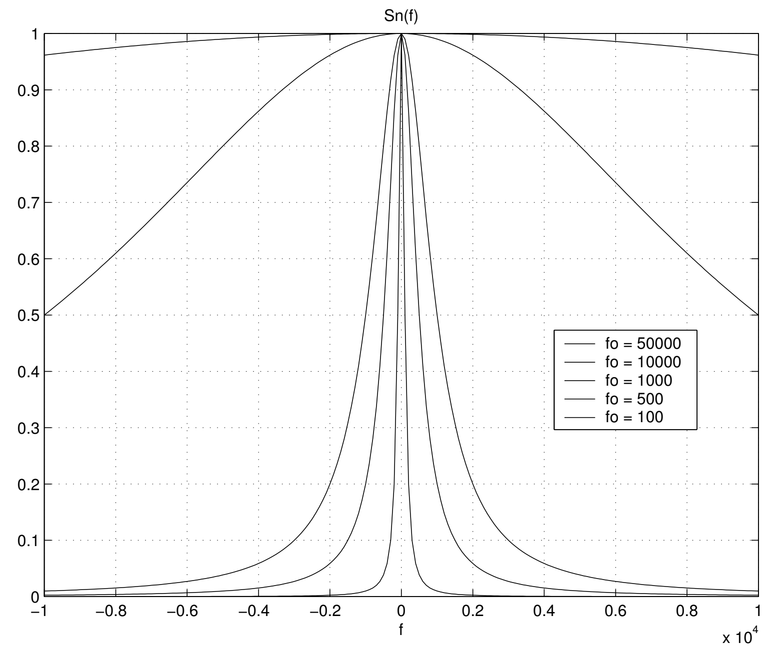

Figure 10. Power spectrum of the noise $N(t)$ for different values of f_0

$$s_0(t) = \begin{cases} a & 0 \le t < T \\ 0 & \text{otherwise} \end{cases}$$

$$s_1(t) = -s_0(t)$$

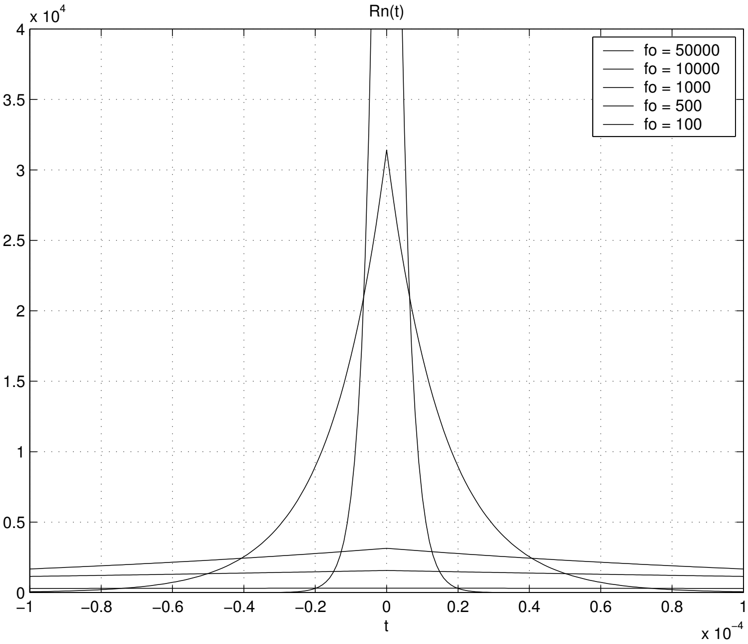

Figure 11. Autocorrelation function of the noise $N(t)$ for different values of f_0

but the noise has power density spectrum:

$$S_N(f) = \frac{N_0/2}{1 + (f/f_0)^2}$$

Looking at fig. 10 it can be seen that $S_N(f)$ is not flat (constant) in frequency. Its deviation from being "white" is controlled by the parameter f_0: the larger f_0 the larger the frequency interval in which $S_N(f)$ can considered to be almost white.

a) As we know from signal theory, the autocorrelation function of a wide-sense stationary process is the inverse Fourier transform of the power spectrum:

$$R_N(\tau) = F^{-1}\left\{S_N(f)\right\} = \int_{-\infty}^{\infty} S_N(f) e^{j2\pi f \tau} df = \frac{\pi f_0 N_0}{2} e^{-2\pi f_0 |\tau|}$$

This function is depicted in fig. 11 for different values of f_0. We can see that if f_0 increases, $R_N(\tau)$ tends to a Dirac impulse (as for the white noise).

b) before sampling $R_N(\tau) = E[N(t+\tau)N(t)]$. After sampling

$$\begin{aligned} R_W(k) &= E[W(n+k)W(n)] = E[N(nT_s + kT_s)N(nT_s)] && (2) \\ &= R_N(kT_s) = \frac{\pi f_0 N_0}{2} e^{-2\pi f_0 |kT_s|} && (3) \\ &= \sigma_W^2 e^{-2\pi f_0/f_s |k|} && (4) \end{aligned}$$

where the variance ($R_W(0)$) for each sample is $\sigma_W^2 = \frac{\pi f_0 N_0}{2}$. If $f_s \ll f_0$ all the terms for $k \neq 0$ are almost zero if compared to the maximum (at $k = 0$). The autocorrelation function is hence a good approximation of the Dirac function (white noise).

c) If we consider M samples (where $M = Tf_s$) of the input signal, their joint distribution is gaussian, with covariance matrix

$$C = \left\{ c_{ij} = \sigma_W^2 e^{-2\pi \frac{f_0 T}{M}|i-j|} \right\}$$

We write again the formula we obtained in exercise 1.4c without the substitution of σ_X with σ_W:

$$P_E = 1 - \Phi\left(\frac{a\sqrt{M}}{\sigma_X}\right)$$

To find σ_X we note that X is a sum of gaussian variables with covariance matrix C. Using matrix notation:

$$X(M-1) = \mathbf{b}^T\mathbf{V}$$

The variace is then:

$$\sigma_X^2 = \mathbf{b}^T\mathbf{C}\mathbf{b}$$

Since we are using a normalized matched filter: $b(i) = 1/\sqrt{M}$. and the previous formula becomes:

$$\begin{aligned} \sigma_X^2 &= \frac{1}{M}\sum_i^M\sum_j^M c(i,j) \\ &= \frac{1}{M}\sigma_W^2\sum_i^M\sum_j^M e^{-2\pi\frac{f_0 T}{M}|i-j|} \end{aligned} \tag{5}$$

From this expression we can already say that for M that tends to infinity the exponential terms in the sums tend to 1 and $\sigma_X \propto \sqrt{M}$. This term will compete with the one in the total error function leading to some asymptotic value that depends on the parameter f_0. Expression 5 can be further simplified if we note that the elements on every diagonal of the covariance matrix have constant value, then:

$$\sigma_X^2 = \frac{1}{M}\sigma_W^2\left[M + 2\sum_{j=1}^{M-1}(M-j)e^{-2\pi\frac{f_0 T}{M}j}\right] \tag{6}$$

Note how the first term in parentesis corresponds to the white-noise case, and leads to $\sigma_X^2 = \sigma_W^2$ while the second term is a correction that takes into account the non-zero correlation. Solving the sum in 6 we can write P_E as a function of M. This function is decreasing in M, but for M that goes to infinity tends to a value $\neq 0$ and that depends on the parameter f_0. This means that in the beginning increasing the sampling frequency (number of samples) we add useful information to the classification problem. When the correlation between each sample starts to increase (due to the high sampling frequency), the new samples don't add any new information to the classifier. The total probability of error versus number of samples is plotted in fugure 12 for $\sigma_W = 10$, $a = 1$, $T = 10ms$ and for different values of f_0. Compare also with the similar plot in fig 9: you can see that for $f_0 = 50$kHz the noise is a good approximation of white-noise, for this problem.

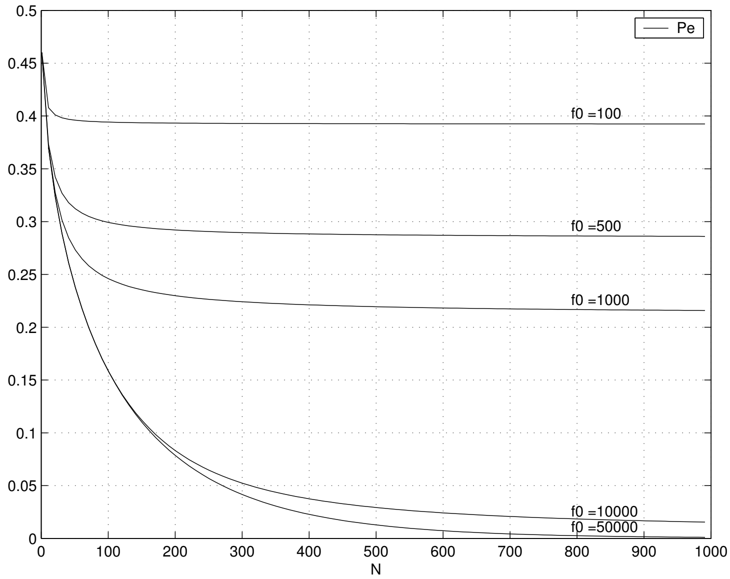

Figure 12. Total probability of error as a function of the number of samples (sampling frequency) for different values of the parameter f_0