▶ 11.10 Solutions

Solution to exercise 11.1 (p.181). Introduce a Lagrange multiplier λ for the power constraint and another, μ, for the constraint of normalization of $P(x)$.

$$F = I(X;Y) - \lambda \textstyle\int \mathrm{d}x\, P(x) x^2 - \mu \textstyle\int \mathrm{d}x\, P(x) \tag{11.36}$$

$$= \int \mathrm{d}x\, P(x) \left[\int \mathrm{d}y\, P(y \,|\, x) \ln \frac{P(y \,|\, x)}{P(y)} - \lambda x^2 - \mu \right]. \tag{11.37}$$

Make the functional derivative with respect to $P(x^*)$.

$$\frac{\delta F}{\delta P(x^*)} = \int \mathrm{d}y\, P(y \,|\, x^*) \ln \frac{P(y \,|\, x^*)}{P(y)} - \lambda x^{*2} - \mu - \int \mathrm{d}x\, P(x) \int \mathrm{d}y\, P(y \,|\, x) \frac{1}{P(y)} \frac{\delta P(y)}{\delta P(x^*)}. \tag{11.38}$$

The final factor $\delta P(y)/\delta P(x^*)$ is found, using $P(y) = \int \mathrm{d}x\, P(x) P(y \,|\, x)$, to be $P(y \,|\, x^*)$, and the whole of the last term collapses in a puff of smoke to 1, which can be absorbed into the μ term.

Substitute $P(y \,|\, x) = \exp(-(y-x)^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$ and set the derivative to zero:

$$\int \mathrm{d}y\, P(y \,|\, x) \ln \frac{P(y \,|\, x)}{P(y)} - \lambda x^2 - \mu' = 0 \tag{11.39}$$

$$\Rightarrow \int \mathrm{d}y\, \frac{\exp(-(y-x)^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}} \ln\left[P(y)\sigma\right] = -\lambda x^2 - \mu' - \frac{1}{2}. \tag{11.40}$$

This condition must be satisfied by $\ln[P(y)\sigma]$ for all x.

Writing a Taylor expansion of $\ln[P(y)\sigma] = a + by + cy^2 + \cdots$, only a quadratic function $\ln[P(y)\sigma] = a + cy^2$ would satisfy the constraint (11.40). (Any higher order terms y^p, $p > 2$, would produce terms in x^p that are not present on the right-hand side.) Therefore $P(y)$ is Gaussian. We can obtain this optimal output distribution by using a Gaussian input distribution $P(x)$.

Solution to exercise 11.2 (p.181). Given a Gaussian input distribution of variance v, the output distribution is $\mathrm{Normal}(0, v + \sigma^2)$, since x and the noise are independent random variables, and variances add for independent random variables. The mutual information is:

$$I(X;Y) = \int \mathrm{d}x\, \mathrm{d}y\, P(x) P(y \,|\, x) \log P(y \,|\, x) - \int \mathrm{d}y\, P(y) \log P(y) \tag{11.41}$$

$$= \frac{1}{2} \log \frac{1}{\sigma^2} - \frac{1}{2} \log \frac{1}{v + \sigma^2} \tag{11.42}$$

$$= \frac{1}{2} \log \left(1 + \frac{v}{\sigma^2}\right). \tag{11.43}$$

Solution to exercise 11.4 (p.186). The capacity of the channel is one minus the information content of the noise that it adds. That information content is, per chunk, the entropy of the selection of whether the chunk is bursty, $H_2(b)$, plus, with probability b, the entropy of the flipped bits, N, which adds up to $H_2(b) + Nb$ per chunk (roughly; accurate if N is large). So, per bit, the capacity is, for $N = 100$,

$$C = 1 - \left(\frac{1}{N} H_2(b) + b\right) = 1 - 0.207 = 0.793. \tag{11.44}$$

In contrast, interleaving, which treats bursts of errors as independent, causes the channel to be treated as a binary symmetric channel with $f = 0.2 \times 0.5 = 0.1$, whose capacity is about 0.53.

Interleaving throws away the useful information about the correlatedness of the errors. Theoretically, we should be able to communicate about $(0.79/0.53) \simeq 1.6$ times faster using a code and decoder that explicitly treat bursts as bursts.

Solution to exercise 11.5 (p.188).

(a) Putting together the results of exercises 11.1 and 11.2, we deduce that a Gaussian channel with real input x, and signal to noise ratio v/σ^2 has capacity

$$C = \frac{1}{2} \log \left(1 + \frac{v}{\sigma^2}\right). \tag{11.45}$$

(b) If the input is constrained to be binary, $x \in \{\pm\sqrt{v}\}$, the capacity is achieved by using these two inputs with equal probability. The capacity is reduced to a somewhat messy integral,

$$C'' = \int_{-\infty}^{\infty} \mathrm{d}y \, N(y;0) \log N(y;0) - \int_{-\infty}^{\infty} \mathrm{d}y \, P(y) \log P(y), \tag{11.46}$$

where $N(y;x) \equiv (1/\sqrt{2\pi}) \exp[(y-x)^2/2]$, $x \equiv \sqrt{v}/\sigma$, and $P(y) \equiv [N(y;x) + N(y;-x)]/2$. This capacity is smaller than the unconstrained capacity (11.45), but for small signal-to-noise ratio, the two capacities are close in value.

(c) If the output is thresholded, then the Gaussian channel is turned into a binary symmetric channel whose transition probability is given by the error function Φ defined on page 156. The capacity is

$$C'' = 1 - H_2(f), \text{ where } f = \Phi(\sqrt{v}/\sigma). \tag{11.47}$$

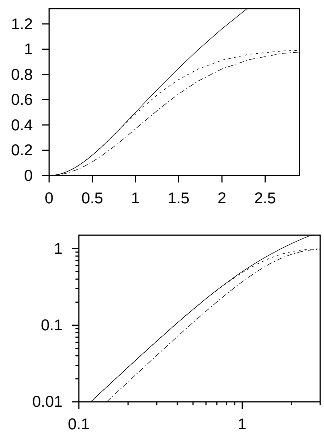

Figure 11.9. Capacities (from top to bottom in each graph) C, C', and C'', versus the signal-to-noise ratio $(\sqrt{v}/\sigma)$. The lower graph is a log–log plot.

Solution to exercise 11.9 (p.188). There are several RAID systems. One of the easiest to understand consists of 7 disk drives which store data at rate 4/7 using a (7, 4) Hamming code: each successive four bits are encoded with the code and the seven codeword bits are written one to each disk. Two or perhaps three disk drives can go down and the others can recover the data. The effective channel model here is a binary erasure channel, because it is assumed that we can tell when a disk is dead.

It is not possible to recover the data for *some* choices of the three dead disk drives; can you see why?

▷ Exercise 11.10.[2, p.190] Give an example of three disk drives that, if lost, lead to failure of the above RAID system, and three that can be lost without failure.

Solution to exercise 11.10 (p.190). The (7, 4) Hamming code has codewords of weight 3. If any set of three disk drives corresponding to one of those codewords is lost, then the other four disks can recover only 3 bits of information about the four source bits; a fourth bit is lost. [cf. exercise 13.13 (p.220) with $q = 2$: there are no binary MDS codes. This deficit is discussed further in section 13.11.]

Any other set of three disk drives can be lost without problems because the corresponding four by four submatrix of the generator matrix is invertible. A better code would be a digital fountain – see Chapter 50.