Datasta Tietoon, Autumn 2011

SOLUTIONS TO EXERCISES 1

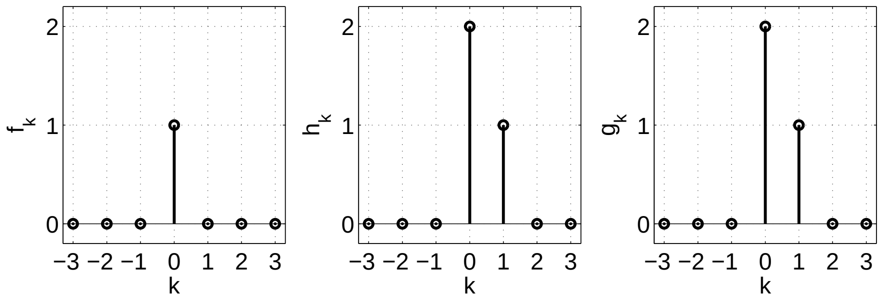
H1 / Problem 1.
Convolution sum is computed as

$$g_k = \sum_{m=-\infty}^{\infty} f_m s_{k-m} = \ldots + f_{-2}s_{k+2} + f_{-1}s_{k+1} + f_0 s_k + f_1 s_{k-1} + f_2 s_{+2} + \ldots$$

a) Now

$$f_0 = 1,\ f_m = 0 \text{ otherwise;} \quad (1)$$
$$s_0 = 2,\ s_1 = 1,\ s_n = 0 \text{ otherwise} \quad (2)$$

Thus $g_k = f_0 s_{k-0} = s_k$, which is $g_0 = 2$, $g_1 = 1$, and $g_k = 0$ elsewhere.



The other sequence f_k was an identity sequence (only one at $k = 0$, zero elsewhere), so it just copies the other sequence s_k into the output.

b) Now

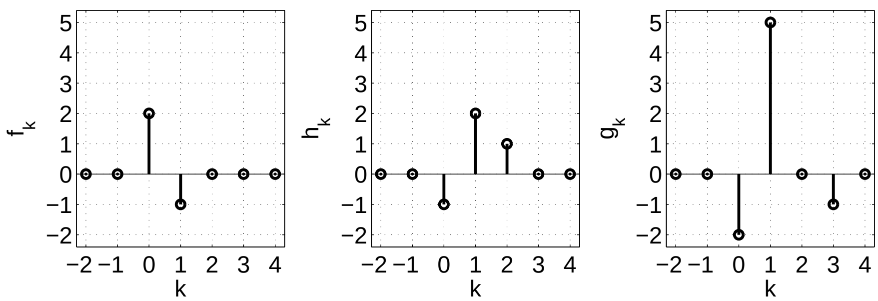
$$f_0 = 2,\ f_1 = -1,\ f_m = 0 \text{ otherwise;} \quad (3)$$
$$s_0 = -1,\ s_1 = 2,\ s_2 = 1,\ s_n = 0 \text{ otherwise.} \quad (4)$$

Thus

$$g_k = f_0 s_{k-0} + f_1 s_{k-1} = 2s_k - s_{k-1}$$

and we get

$$g_0 = 2s_0 - s_{-1} = -2 \quad (5)$$
$$g_1 = 2s_1 - s_0 = 4 + 1 = 5 \quad (6)$$
$$g_2 = 2s_2 - s_1 = 2 - 2 = 0 \quad (7)$$
$$g_3 = 2s_3 - s_2 = -1 \quad (8)$$
$$g_k = 0 \quad \text{otherwise} \quad (9)$$



Sequence $f_k = \{\underline{2}, -1\}$ was now a sum sequence of an identity filter multiplied by two ($f_0 = 2$) and a shifted identity filter multiplied by -1 ($f_1 = -1$). Therefore the output consisted of a sum of s_k multiplied by two and a shifted s_k multiplied by -1.

$$\begin{aligned} 2s_k - s_{k-1} &= 2 \cdot \{\underline{-1}, 2, 1\} - 1 \cdot \{\underline{0} - 1, 2, 1\} \\ &= \{\underline{-2+0}, 4+1, 2-2, 0-1\} = \{\underline{-2}, 5, 0, -1\} \end{aligned}$$

See more examples in the computer session **T1**.

H1 / Problem 2.
a) From Problem 1 b

$$f_0 = 2,\ f_1 = -1,\ f_m = 0 \text{ otherwise;} \tag{10}$$
$$s_0 = -1,\ s_1 = 2,\ s_2 = 1,\ s_n = 0 \text{ otherwise} \tag{11}$$

we get using the definition

$$F(\omega) = \sum_{m=-\infty}^{\infty} f_m e^{-i\omega m}$$

$$F(\omega) = f_0 \cdot e^{-i\omega 0} + f_1 e^{-i\omega 1} = 2 - e^{-i\omega} \tag{12}$$
$$S(\omega) = s_0 \cdot e^{-i\omega 0} + s_1 \cdot e^{-i\omega 1} + s_2 \cdot e^{-i\omega 2} = -1 + 2e^{-i\omega} + e^{-2i\omega} \tag{13}$$

Convulution of two sequences in time-domain corresponds multiplication of two transforms in transform/frequency-domain. The real argument ω gets normally values $-\pi \ldots \pi$ or $0 \ldots \pi$

$$G(\omega) = F(\omega)S(\omega) \tag{14}$$
$$= (2 - e^{-i\omega}) \cdot (-1 + 2e^{-i\omega} + e^{-2i\omega}) \tag{15}$$
$$= -2 + 5e^{-i\omega} - e^{-3i\omega} \tag{16}$$

We find out that the coefficients $\{-2, 5, 0, -1\}$ of the polynomial $G(\omega)$ are equal to the sequence g_k.

Remark. There are several integral transforms that are used in specific cases:

- *Fourier series*, where signal $f(t)$ is analog and periodic (Ω_0), gives discrete and aperiodic Fourier series coefficients F_n with multiples of the fundamental angular frequency Ω_0
- (Continuous-time) Fourier transform, where signal $f(t)$ is analog and aperiodic, gives continuous and aperiodic transform $F(\Omega)$
- Discrete-time Fourier transform, where signal f_k is discrete and aperiodic, gives continuous and periodic transform $F(\omega)$ as above
- Discrete Fourier transform (DFT), where signal f_k is discrete and periodic (length N), gives discrete and periodic transform F_n (length N)

H1 / Problem 3.

a) Substitute $F(\omega)$ into the integral:

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\sum_{m=-\infty}^{\infty} f_m e^{-i\omega m}]e^{i\omega n}d\omega = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} f_m \int_{-\pi}^{\pi} e^{i\omega(n-m)}d\omega$$

with $i = \sqrt{-1}$ the imaginary unit (sometimes also denoted j).

For the integral we get (note that $n, m \in \mathbb{Z}$)

$$\int_{-\pi}^{\pi} e^{i\omega(n-m)}d\omega = \begin{cases} 2\pi & \text{if } n = m, \\ /_{-\pi}^{\pi} \frac{1}{i(n-m)}e^{i\omega(n-m)} = \frac{1}{i(n-m)}\left(e^{i\pi(n-m)} - e^{-i\pi(n-m)}\right) & \text{if } n \neq m \end{cases}$$

We can easily see that $e^{i\pi(n-m)} = e^{-i\pi(n-m)}$ because $e^{i\pi} = e^{-i\pi} = -1$. Thus the integral is 2π if $n = m$ and zero otherwise. Substituting this into the full expression gives $I = f_n$ which was to be shown.

b)

$$h_n = \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{i\omega n}d\omega = \frac{1}{2\pi} /_{-\omega_0}^{\omega_0} \frac{1}{in}e^{i\omega n} \tag{17}$$

$$= \frac{1}{2\pi i n}(e^{i\omega_0 n} - e^{-i\omega_0 n}) \tag{18}$$

$$= \frac{1}{2\pi i n}[\cos(\omega_0 n) + i\sin(\omega_0 n) - \cos(\omega_0 n) + i\sin(\omega_0 n)] \tag{19}$$

$$= \frac{1}{\pi n}\sin(\omega_0 n). \tag{20}$$

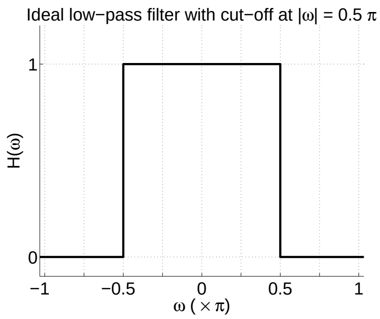
Using the cut-off frequency $\omega = \pi/2$ we get

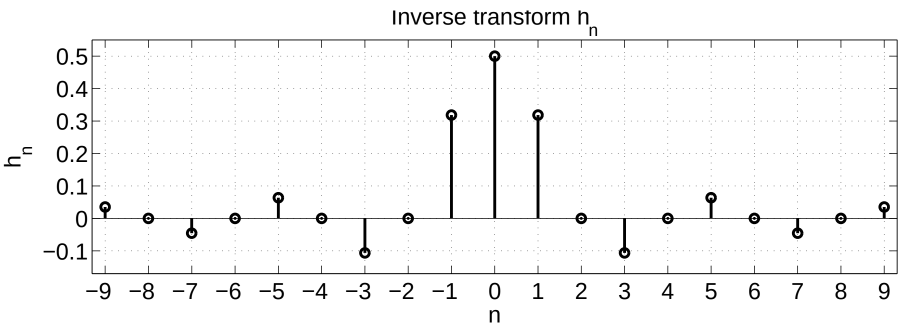
$$h_n = \frac{1}{\pi n}\sin(\frac{\pi n}{2})$$

which is sometimes written as $h_n = (1/2)\text{sinc}(n/2)$, where sinc function is $\text{sinc}(\omega n) = \sin(\pi\omega n)/(\pi\omega n)$. Some values: $h_0 = 0.5$, $h_1 = 1/\pi$, $h_2 = 0$.

Note that at $n = 0$ we end up to $0/0$. It can be solved, e.g., either Taylor series $(1/x)\sin(x/2) = (1/2)(2/x)\sin(x/2) = (1/2) - (x^2/48) + \ldots$, or l'Hospital's rule by derivating both sides. Thus at zero the value is 0.5. In addition, $\text{sinc}(0) = 1$.

Note also that the sequence h_n is infinitely long.





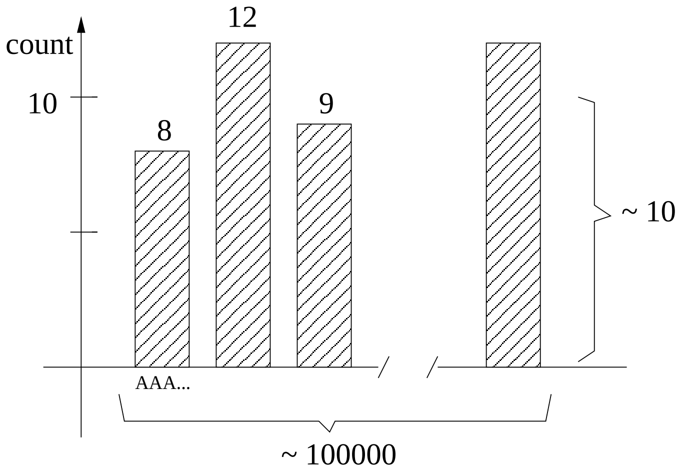
H1 / Problem 4.

Now the number of bins is at most 100000, because the average number of substrings in a bin must be at least 10. The number of different substrings of length n is 4^n. We get

$$4^n \leq 100000$$

giving $n \leq 8$.

An example of a histogram of a data sample given below. It is assumed that letters are drawn independently from uniform distribution, i.e., the total amount of each letter is the same.



Another example on building a histogram with the sequence 'AAGTACCGTGACGGAT'.

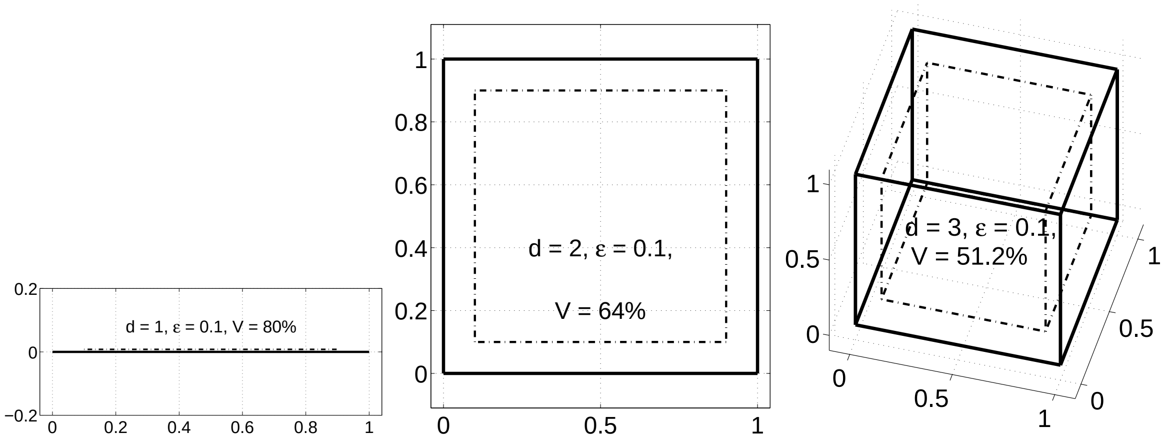
If $n = 1$, all possible substrings are 'A', 'C', 'G', and 'T', shortly A, C, G, T. The number of substrings is $4^1 = 4$. The count for each substring: $|A| = 5$, $|C| = 3$, $|G| = 5$, and $|T| = 3$.

If $n = 2$, all possible substrings are 'AA', 'AC', 'AG', 'AT', 'CA', 'CC', 'CG', 'CT', 'GA', 'GC', 'GG', 'GT', 'TA', 'TC', 'TG', 'TT', that is, $4^2 = 16$ substrings. The count for each substring: $|AA| = 1$, $|AC| = 2$, $|AG| = 1$, $|AT| = 0$, etc.

H1 / Problem 5.

The volume of the unit hypercube is 1 and the volume of the set of inner points is $V_d = (1 - 2\epsilon)^d$. For any ϵ, this tends to 0 as $n \to \infty$.

Below an illustration of hypercubes in dimensions $d = 1,\ 2,\ 3$ with $\epsilon = 0.1$. We can see that the volume of inner points decreases when the dimension increases.



H1 / Problem 6.

Now the small hypercubes are similar, hence all have the same volume which must be $\frac{1}{n}$ times the volume of the large unit hypercube. (This is only possible for certain values of (n, d); for $d = 2$, n must be 4, 9, 16, ...; for $d = 3$, n must be 8, 27, 64 ... etc.)

Also, we assume here a special distance which is not Euclidean distance but $D(\mathbf{x}_1, \mathbf{x}_2) = \max_i |x_{i1} - x_{i2}|$, that is, the largest distance along the coordinate axes.

Then it is easy to see that the distance of the centres of the small hypercubes is equal to the length of their side s. Because the volume is $s^d = \frac{1}{n}$, we have $s = \frac{1}{n}^{\frac{1}{d}}$.

The case of $d = 2, n = 4$ is shown below.