Information Sciences: Signal Processing

Lecture 2: Filters

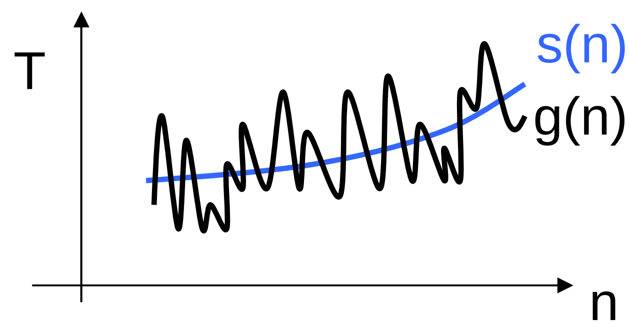
- In this lecture we will talk about a special group of systems called filters
- The intuitive idea about a filter is that it is a system that is able to remove some components from the input signal leaving some others unchanged. An example is the tone control that we saw in the previous lecture.
- Another example, more related to computer science, is an antispam filter. In this case, we want to remove the advertisement while keeping the desired messages

Let see one more example related to signal processing. This is the moving average

- Suppose that you have a certain quantity that is measured every day. For example, you want to study global warming and you measure the temperature at the same place every day. Let's call g(n) this quantity (n is the time index)
- g(n) is not exactly what we would like to have. In fact, g(n) is influenced by the short-term weather conditions. We can consider these as a source of measurement errors. If we call s(n) the "correct" value, we have that

$$g(n)=s(n)+e(n)$$

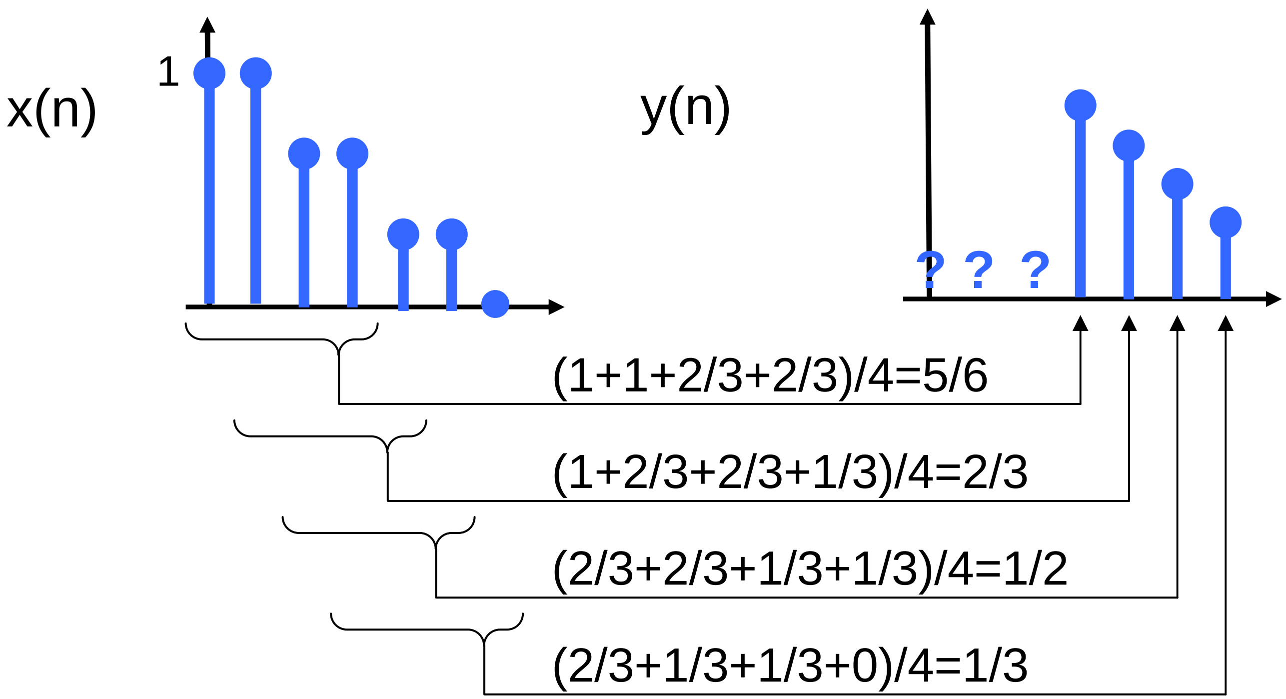
i.e. the measured value g(n) is the sum of the correct value s(n) and an error term e(n)



- Suppose that we observe g(n) for a certain time (i.e n=0,1,2,..) is there a way to reduce the error term e(n) and obtain s(n)?
- A common way to reduce the effect of errors from a sequence is to compute the average. The assumption is that the errors e(n) tend to have mean equal to zero.
- However, the mean gives a unique value for the whole sequence and we loose the dynamics of s(n) (if we want to study global warming, we are interested in how s(n) evolves over time)
- A method is to compute the moving average on L samples. This is defined by

y(n)= "average of the most recent L samples of g"

- Example (L=4):



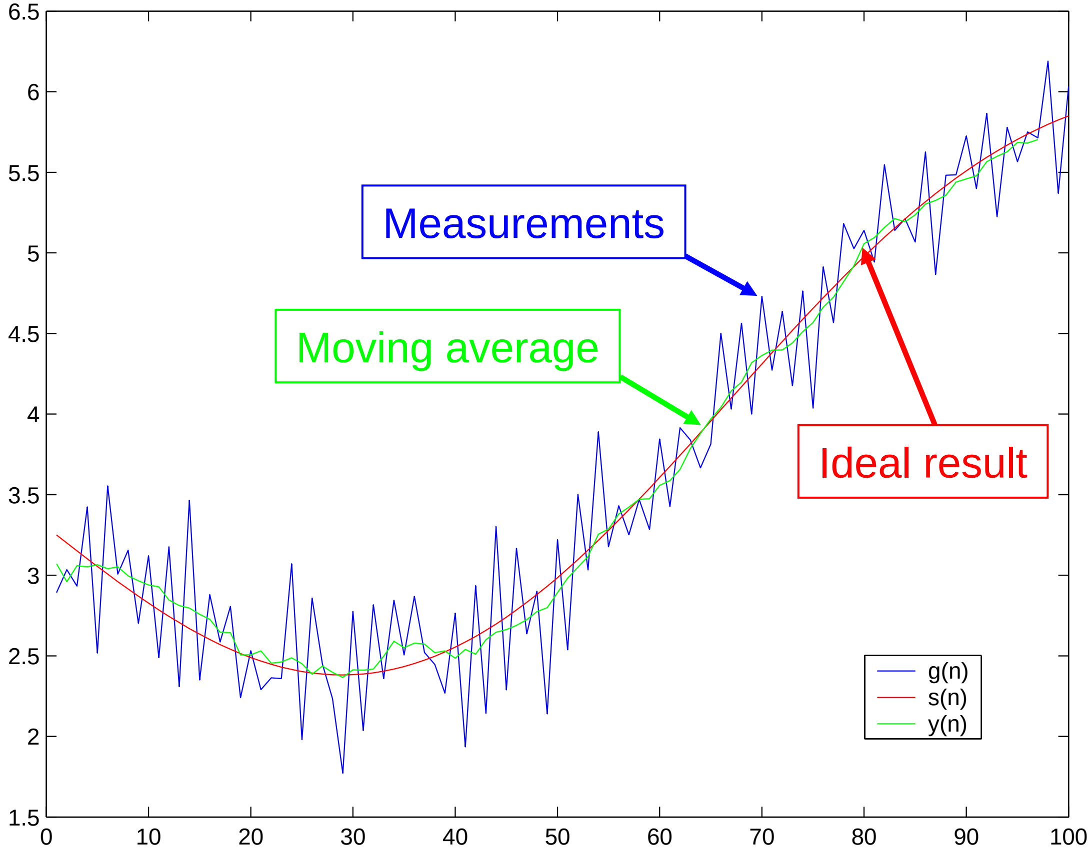
- At time n the most recent samples are g(n-L+1), …,g(n-1), g(n); therefore, the moving average is

$$y(n) = \frac{1}{L}\sum_{i=0}^{L-1} g(n-i)$$

- How do we choose the parameter L?
  - If L is large, many terms are included in the average. We expect that the error term is reduced more. However, the signal s(n) is also averaged and we loose the short-term variations of s(n)
  - On the opposite, if L is small, the average is less effective in reducing the errors, but the variations of s(n) are preserved

The choice of L is a tradeoff between noise attenuation and ability to preserve the dynamics of s(n)

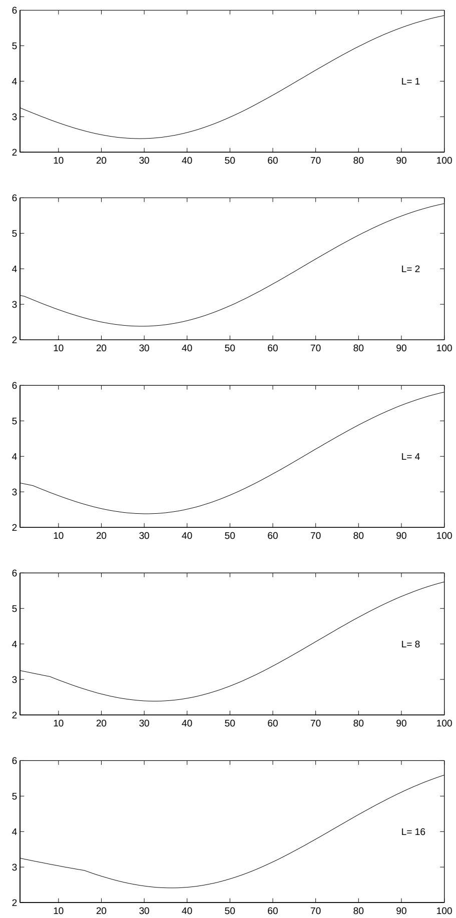
- Example: simulation (L=8)



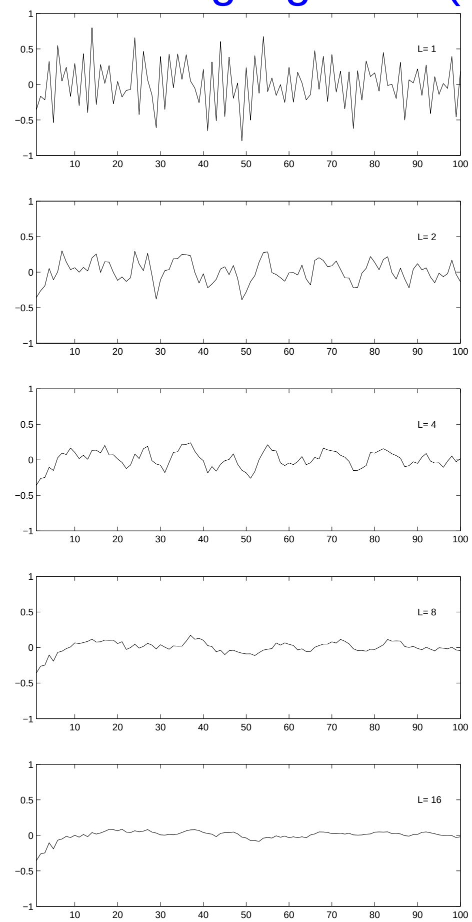
- What happens if L is changed?

Increasing L

Signal of interest - s(n)



Perturbing signal - e(n)



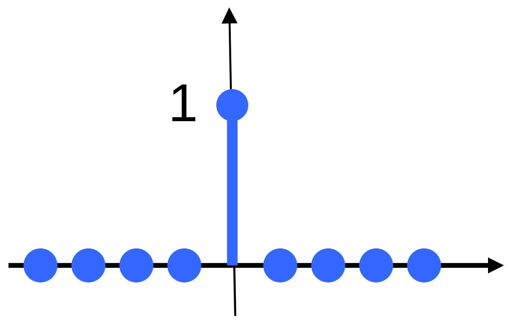
- We have seen 3 examples of filters (tone control, antispam, moving average), are there common properties? Yes!

- The 3 filters apply always the same scheme, i.e. if we send the same input at different instants, we obtain the same output. We call this property time-invariance. This means that the system does not adapt and does not “age” with time
- At a certain time instant, the output of the filters is determined only by the past samples of the input, i.e. the systems are not able to use the future of the signal. We call this property causality

- In the following, we want to study this properties for the special case of discrete time linear systems

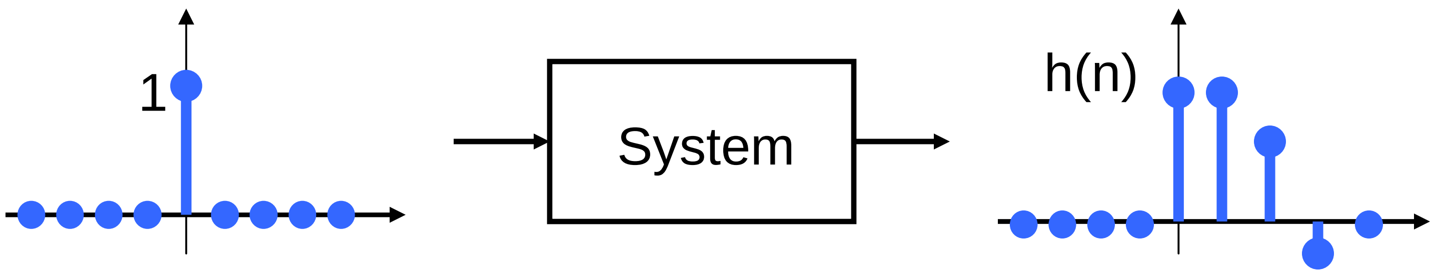
- We saw that a linear function has always the form f(x)=mx and the parameter m identify the linear function. We want to find a similar representation for linear system. What is the quantity that plays the role of m?
- Let's first define a special signal called Kronecker delta (or pulse):

$$\delta(n) = \begin{cases} 1 & \text{if} \quad n = 0 \\ 0 & \text{if} \quad n \neq 0 \end{cases} \quad \forall n \in \mathbb{Z}.$$

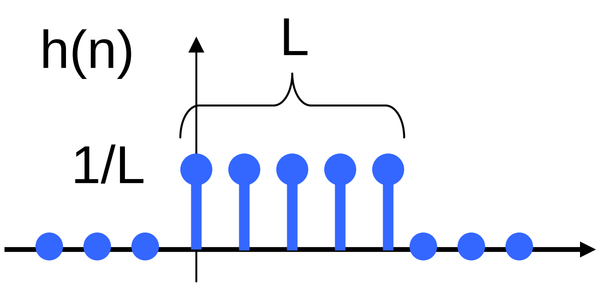


- What happens if we send the pulse to the input of a linear system, such as the tone control or the moving average? The systems produces an output signal that we call impulse response

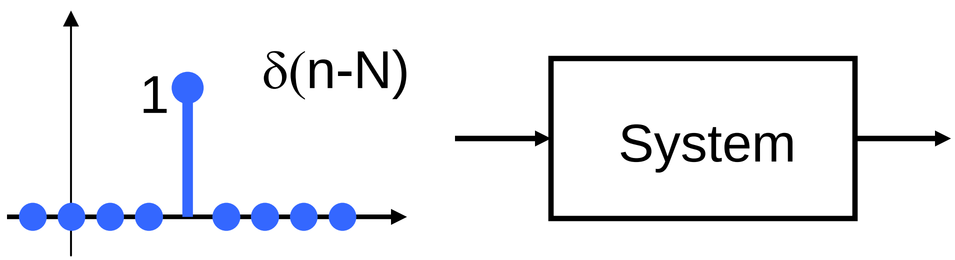
- Let's call h(n) the impulse response. Graphically, we have



- Example: impulse response of the moving average.
  We set $g(n)=\delta(n)$ and we compute the output $y(n)=h(n)$.
  Clearly, $h(n)=0$, for $n<0$ and $n>L-1$ (average of L zeros)
  while $h(n)=1/L$, for $n=0, \ldots, L-1$ (average of L-1 zeros and one 1), i.e.



- We want to study time invariance of a system. Let's suppose that the input signal is a delayed pulse δ(n-N), where N is a constant. Call $\bar{h}(n, N)$ the output signal. Graphically we have

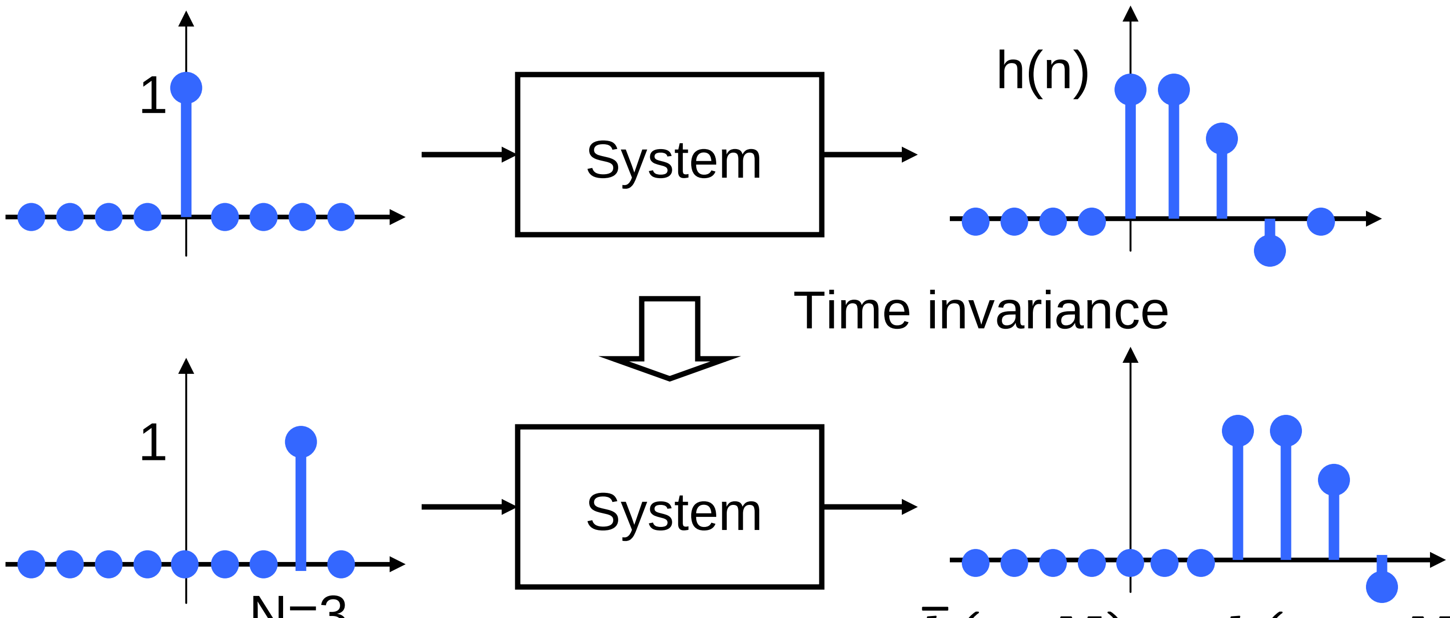


$\bar{h}(n, N)$

- We can expect that, if we delay the input signal, the output signal is also delayed. This is the property of time invariance. Since the output for δ(n) was h(n), we have that the system is time-invariant if

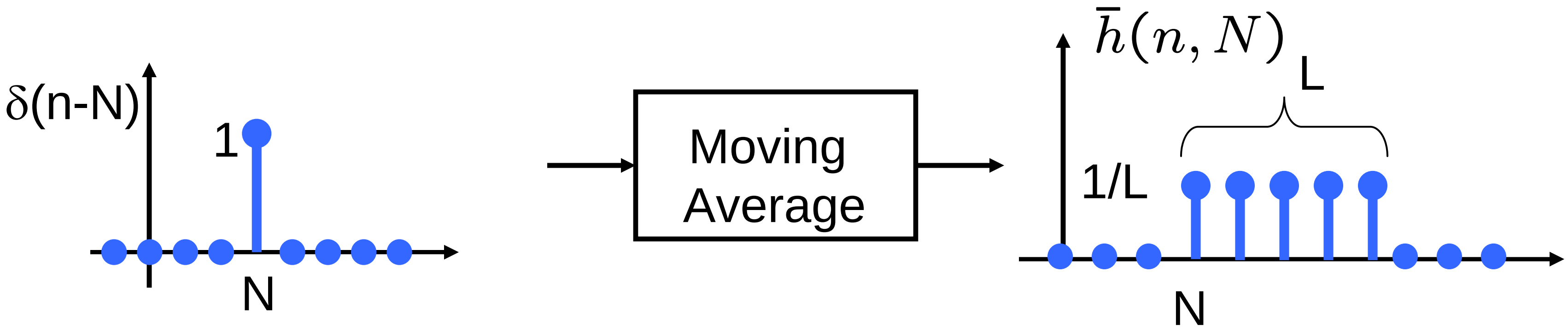
$$\bar{h}(n, N) = h(n - N)$$

- Example:



$$\bar{h}(n, N) = h(n - N)$$

- Example: is the moving average time-invariant?



Yes! $\bar{h}(n, N) = h(n - N)$

- Remark that we defined time invariance by considering the pulse signal only. Later we will see that this condition implies time invariance for any input signal

- A filter is a system that is
  - Linear
  - Time-invariant

- Since a filter is time-invariant, we are able to compute the output to a delayed pulse $\delta(n-N)$. This is obtained by delaying the impulse response $h(n)$.

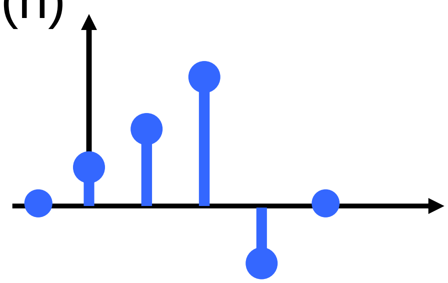
- We show that the impulse response is sufficient to compute the output signal corresponding to any input signal. i.e. the impulse response describes completely the system

- The operator that allows to compute the output signal is called convolution

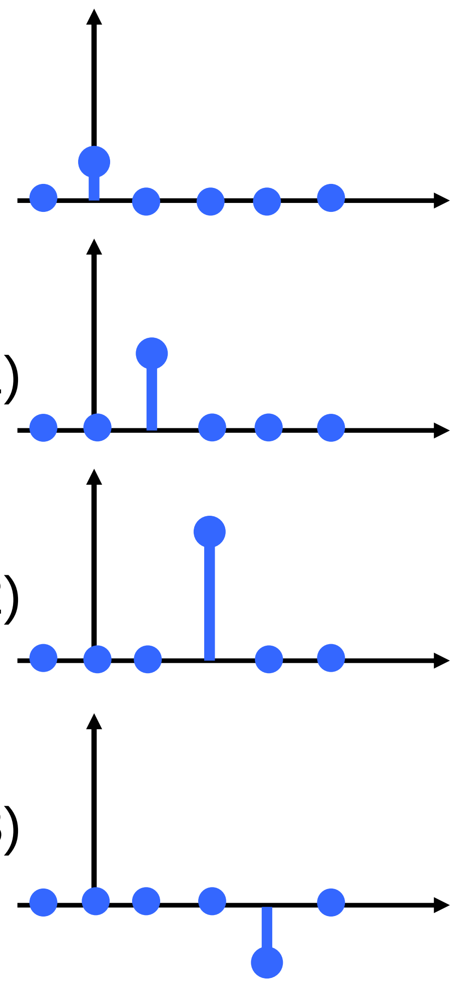
- Let's take an arbitrary input signal x(n):

We can always decompose x(n) as a sum of delayed pulse scaled by the sample amplitudes

x(n)



⇒



$x(0)\delta(n)$ +

$x(1)\delta(n-1)$ +

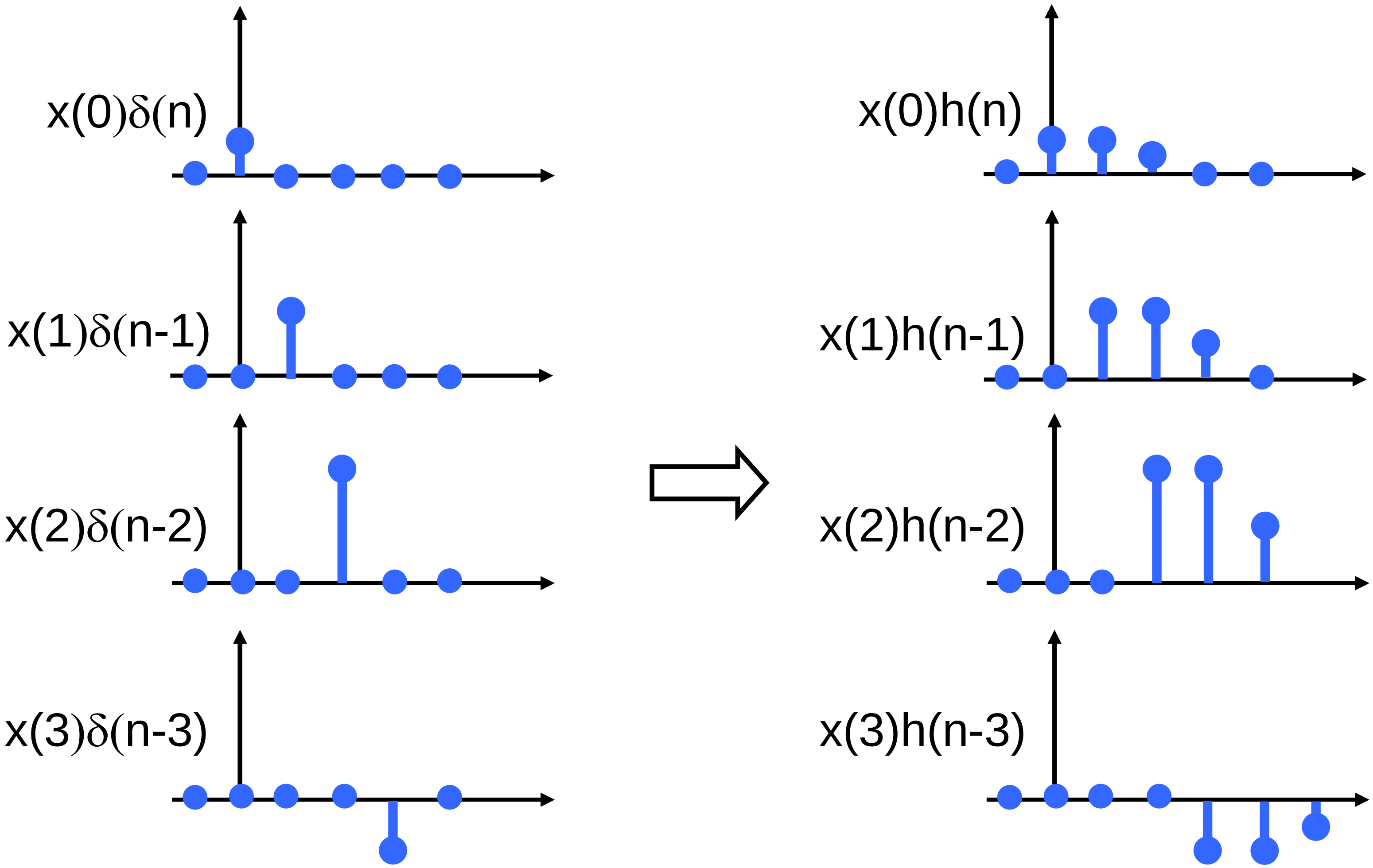
$x(2)\delta(n-2)$ +

$x(3)\delta(n-3)$ +

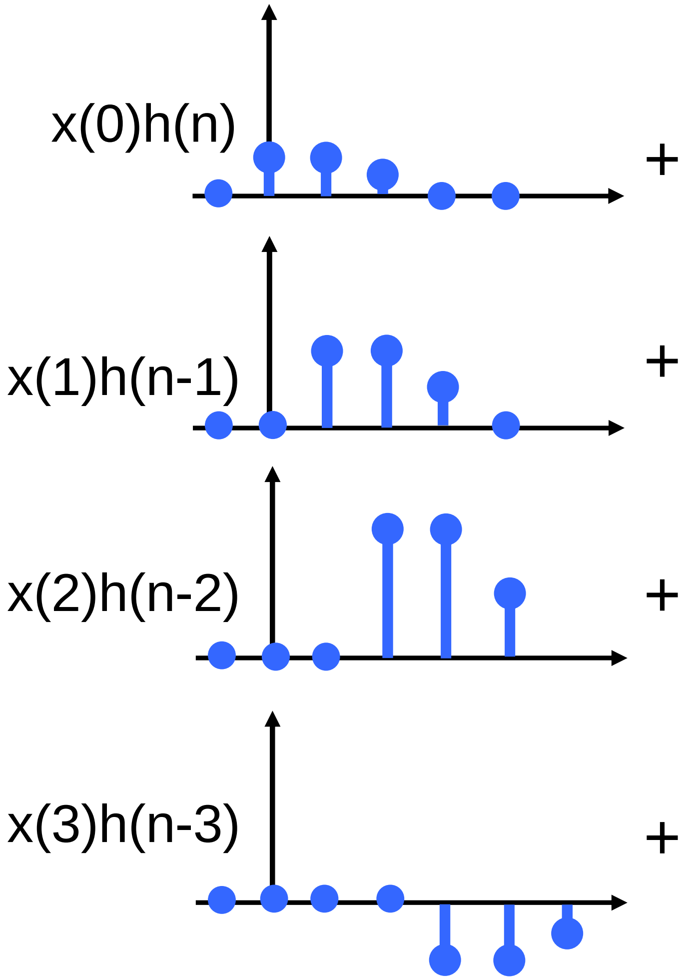
Formally,

$$x(n) = \sum_{m=-\infty}^{\infty} x(m)\delta(n-m)$$

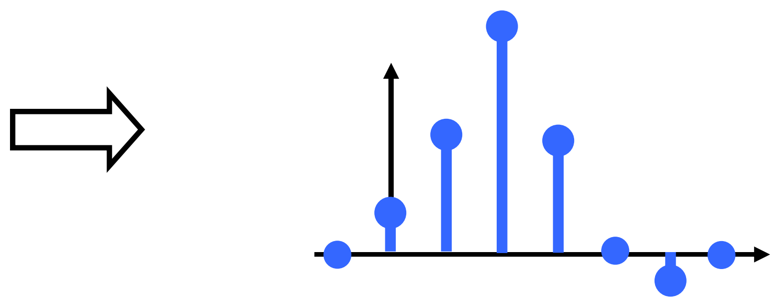
- We know how to compute the output corresponding to each delayed pulse: we delay and scale the impulse response h(n)



- Due to linearity, the output signal can be obtained by adding the output signals corresponding to the delayed and scaled pulses:



$$y(n) = \sum_{m=-\infty}^{\infty} x(m)h(n-m)$$



- Conclusion: the output signal y(n) can be computed from the input signal x(n) if we know the impulse response h(n)
- The procedure to compute the output signal is called the convolution product between x and h:

$$y(n) = \sum_{m=-\infty}^{\infty} x(m)h(n-m) = h * x(n)$$

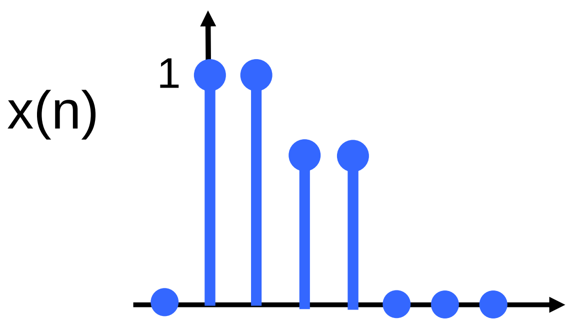
- The convolution product is a “good” product, i.e. it enjoys the properties of a product (e.g. for real numbers):

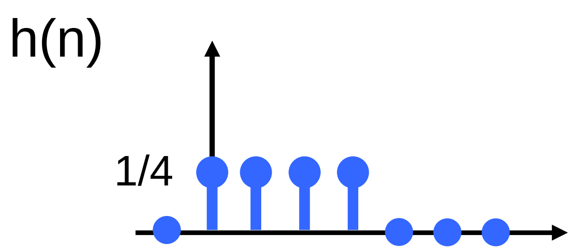
$$h * x = x * h$$ Commutativity

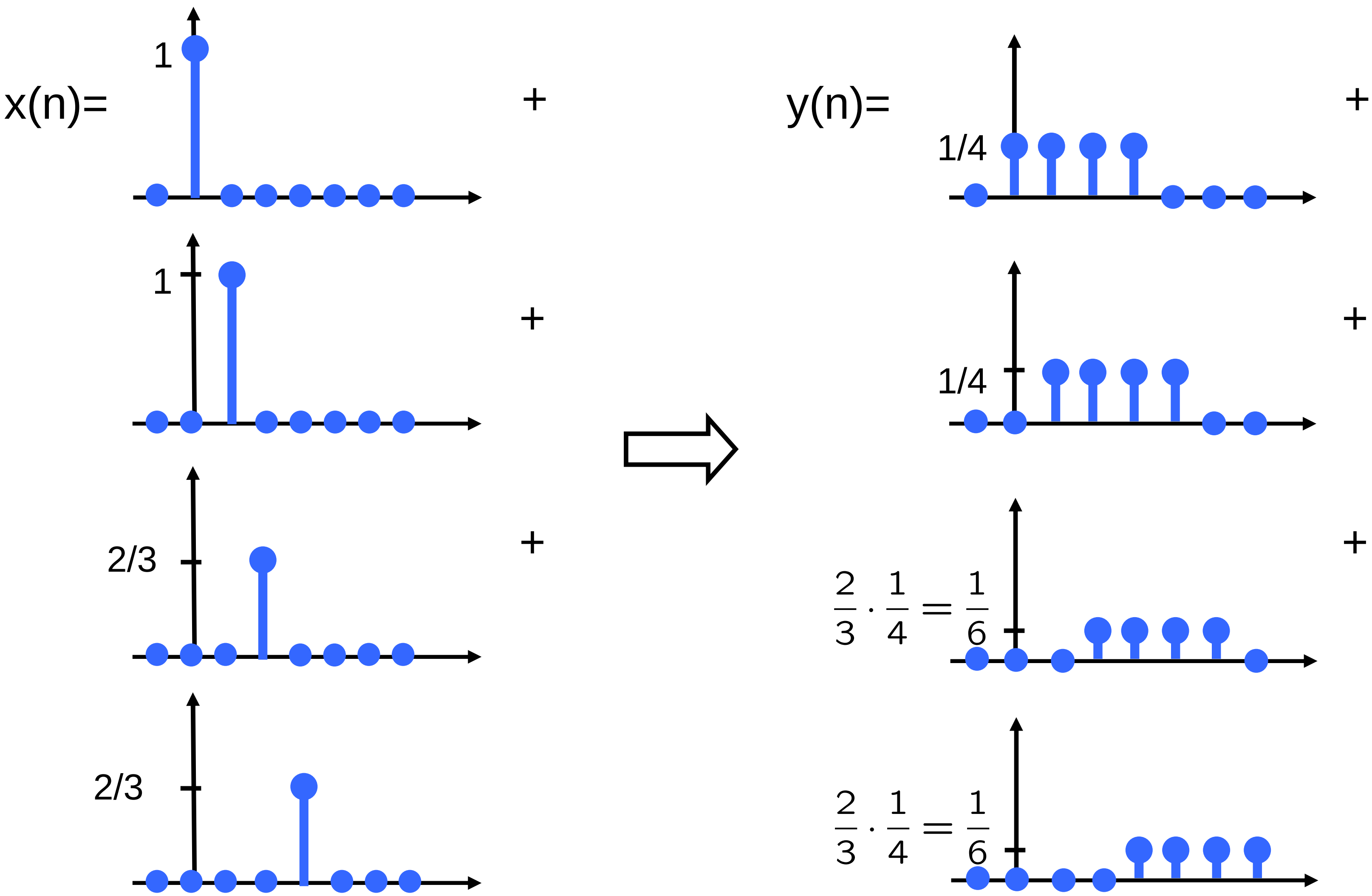
$$h * (x_1 + x_2) = h * x_1 + h * x_2$$ Distributivity

$$h_1 * (h_2 * x) = (h_1 * h_2) * x$$ Associativity

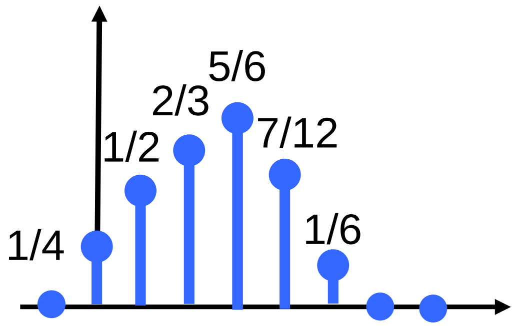
- Computation of the convolution: the simplest way is to decompose the input signal.
- Example: the moving average (L=4)






x(n)=
1
+
1
+
2/3
+
2/3
y(n)=
1/4
+
1/4
+
$\frac{2}{3}\cdot\frac{1}{4}=\frac{1}{6}$
+
$\frac{2}{3}\cdot\frac{1}{4}=\frac{1}{6}$

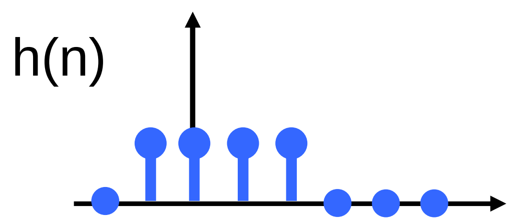
- Result:



- The convolution gives the same result of the direct computation of the moving average. Try!

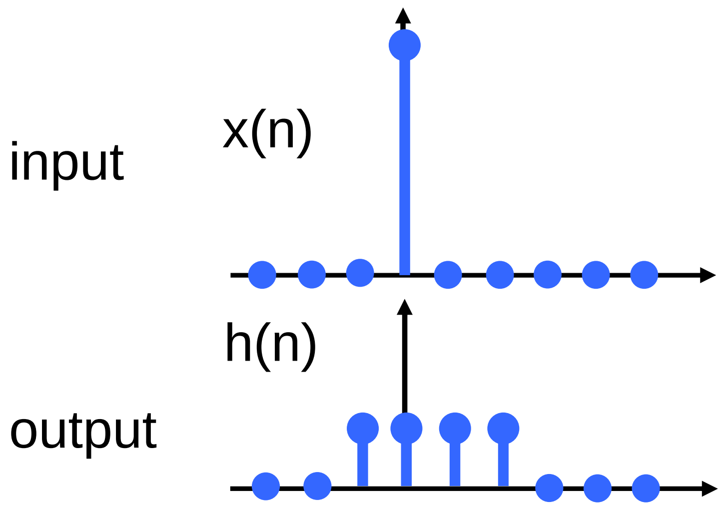
- If we know the impulse response of a linear time-invariant system, we can compute the output signal for any input signal by computing the convolution
- The impulse response describes completely the behavior of the system! i.e. we can measure a system (e.g. microphone, loudspeaker) by measuring the impulse response
- Let's consider the opposite. Suppose we are given an arbitrary function h(n), can we build a linear time-invariant system that has h(n) as impulse response?

e.g. can we take h(n)



and build a system with impulse response h(n)?

- The answer is no. Of course, many things are possible on the blackboard, but in reality the variable "n" represents time. Let's see what happens when we measure the impulse response…



n=-3: The input was zero since "always" and so was the output. Nothing special happened so far…

n=-2: One more input sample equal to zero. The output continue to be zero…

n=-1: Something strange happens! The input continues to be zero, but the output is not. The system is doing something with no apparent reason!

n=0: Now we see why the system was active! The input sample is 1. How did the system know one sample ahead that we were going to send 1 at the input?
n=1,2,…: Nothing special on this part of the impulse response…

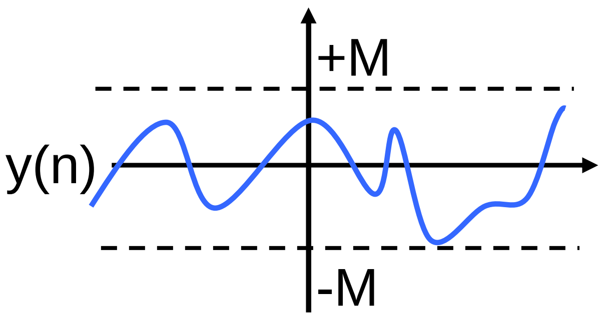
- During the part n<0 of the impulse function, we can suppose that the system reaches an equilibrium state and the output is zero.
- To have a non zero output for n<0, the system should be able to predict the arrival of the pulse! We can't build a system like that.
- We say that a system is causal if

$$h(n) = 0, \quad \forall n < 0$$

- In order to build a system this has to be at least causal (necessary condition). But this is not sufficient…

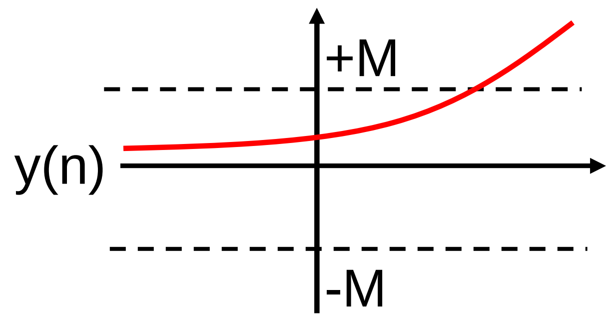
- Can the output of a system be unbounded?

Bounded function



We can choose M (arbitrarily!), such that |y(n)|<M

Unbounded function



The graph of y impinges on the forbidden region, whatever M we choose

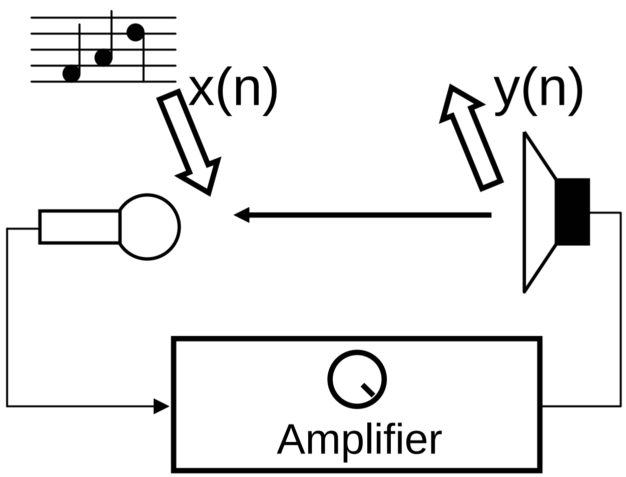
- The output of the system can be unbounded, if the input is also unbounded. For example, the moving average of an unbounded signal is also unbounded
- However, if the input of a system is bounded, we want to impose that the output is also bounded. In such a case, we say that the system is stable

- Formally, a system is stable (Bounded Input, Bounded Output) if

$$\forall x(n),\ |x(n)| < N\ \forall n \in \mathbb{Z} \quad \Longrightarrow \quad |y(n)| < M\ \forall n \in \mathbb{Z}$$

For an appropriate choice of the positive constants N and M (you are free to choose them!)

- Example: an unstable system



If the system has a loop gain larger than 1, any non zero input signal will trigger an unlimited output signal (theoretically…) This is called the Larsen effect

- If we consider linear time-invariant systems, the impulse response is a complete description of the system and stability depends on the impulse response
- A necessary condition for stability is that the impulse response is bounded. This is because the impulse response is the output for a bounded input (i.e. the pulse).
- A necessary and sufficient condition is that

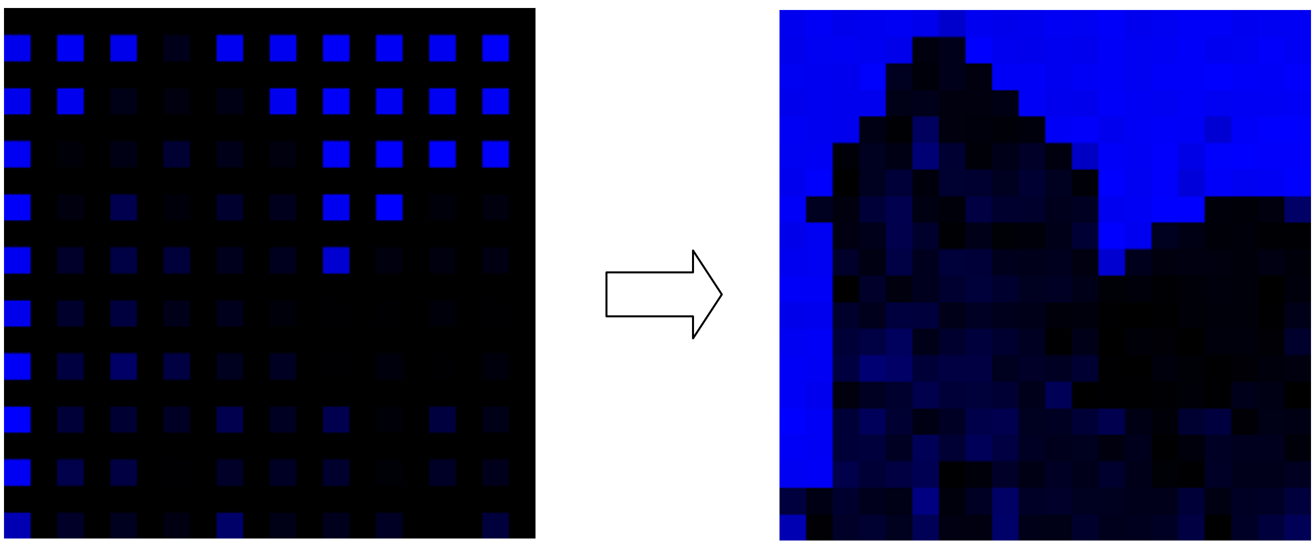
$$\sum_{n=-\infty}^{\infty} |h(n)|$$ Converges to a finite value

You will see the proof in a couple of years…

In this lecture we have seen that:

- Systems transform signal into other signals
- Linear time-invariant systems:
  - The input-output relation that does not depend on time, i.e. a delay of the input signal correspond to an equal delay of the output signal
  - They are completely described by the impulse response. This is the output signal corresponding to pulse signal applied at the input.
  - The convolution product between the input signal and the impulse response allows to compute the output signal
- Realizable systems have to be causal, i.e. the output cannot anticipate the input
- Useuful systems have to be stable, i.e. the output signal does not “blow” when the input signal is bounded

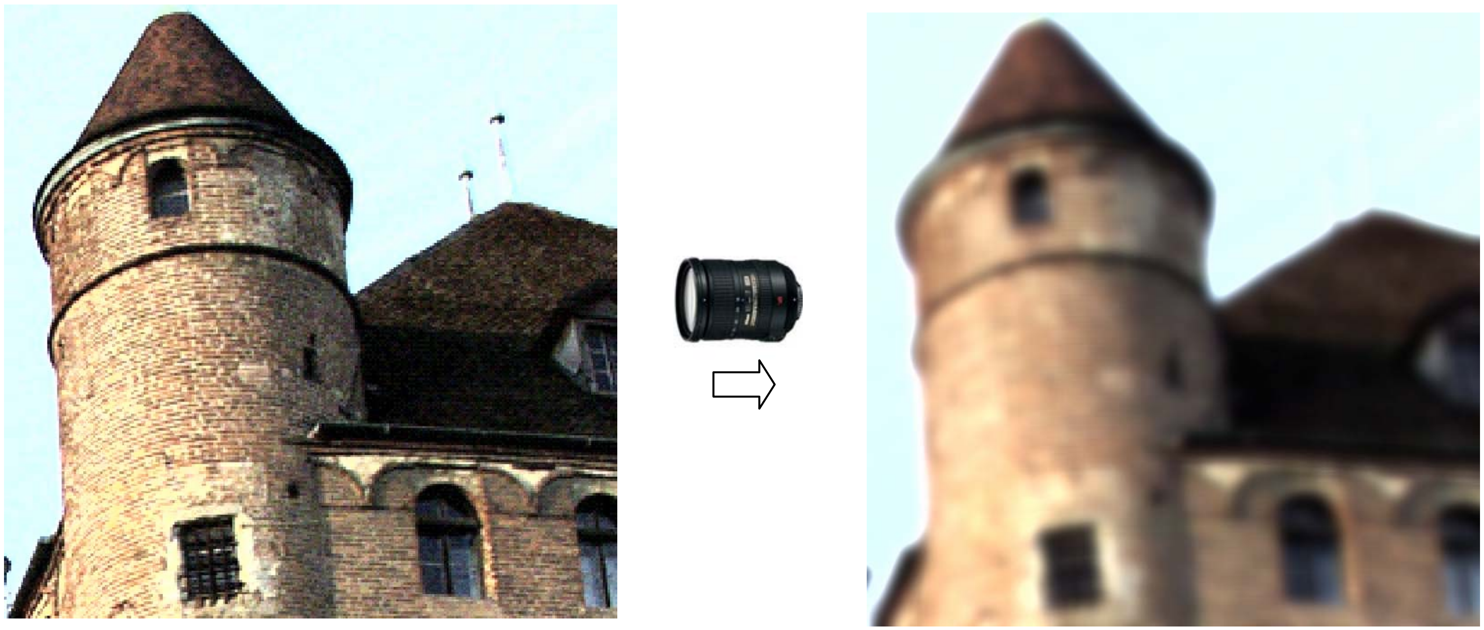
- We saw that an image can be considered as a signal, where the time coordinate is replaced by the position
- For these signals a system is simply an algorithm that takes an image and generates an image as a result
- We already saw the problem of reconstruction of color images when a single sensor is used. The interpolation of each color component is an example of system



- Another system present in a camera is the optics. This actually works on the continuous image

- The optics is designed taking into account the resolution of the sensor (large lenses for high resolution)
- Is this a linear system?

- Let’s check the two conditions for linearity:



- If the input image is scaled, the output image is scaled by the same factor (and this is valid for any input image)

- Is the output of the sum equal to the sum of the outputs?

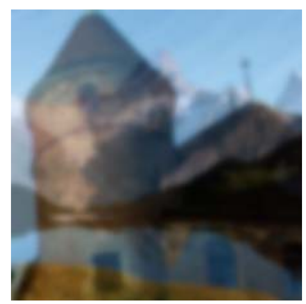
Input image 1

Output to image 1

Input image 2

Output to image 2

Sum of image 1
and image 2



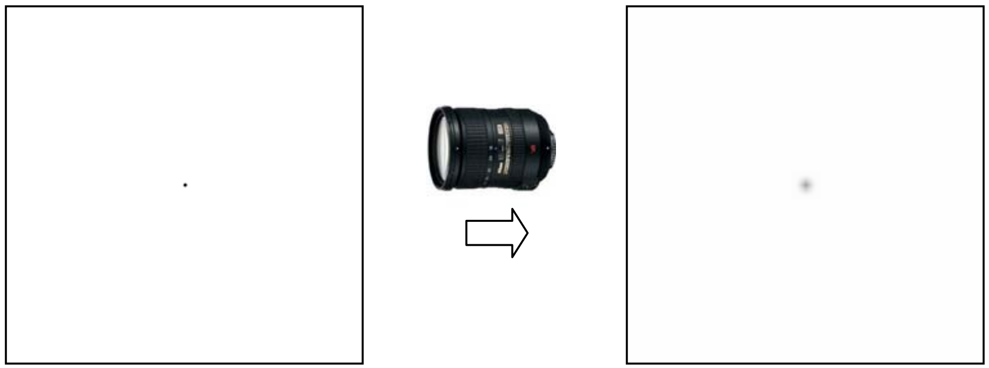
Output to the sum.
Is it equal to the sum
of the outputs? Yes!

- This is valid for any pair of input images. The optics is a linear system!
- Is it time-invariant? Here “time” is replaced by “space”… we should verify that a translation of the input image gives a translation of the output image

- Yes! (Again this is valid for any input image)

- We conclude that the optics is a filter. We can compute the impulse response

Pulse function



Impulse response

- The impulse response represents completely the optics (with a certain approximation…) and we can reproduce the same effect in image processing software

- What about other image processing effects? Are they linear, space-invariant?

Image Warping

Red eyes removal