
ĐỀ BÀI

1. Phép lấy chập (convolution) mà các em đã học trong xử lý tín hiệu số được thực hiện theo công thức sau đây:

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

- a) Hãy giải thích ý nghĩa của $y(n)$, $x(n)$ và $h(n)$. Giải thích chỉ bằng một câu về bản chất của phép lấy chập này.

Giải thích:

$y(n)$ là kết quả của phép chập

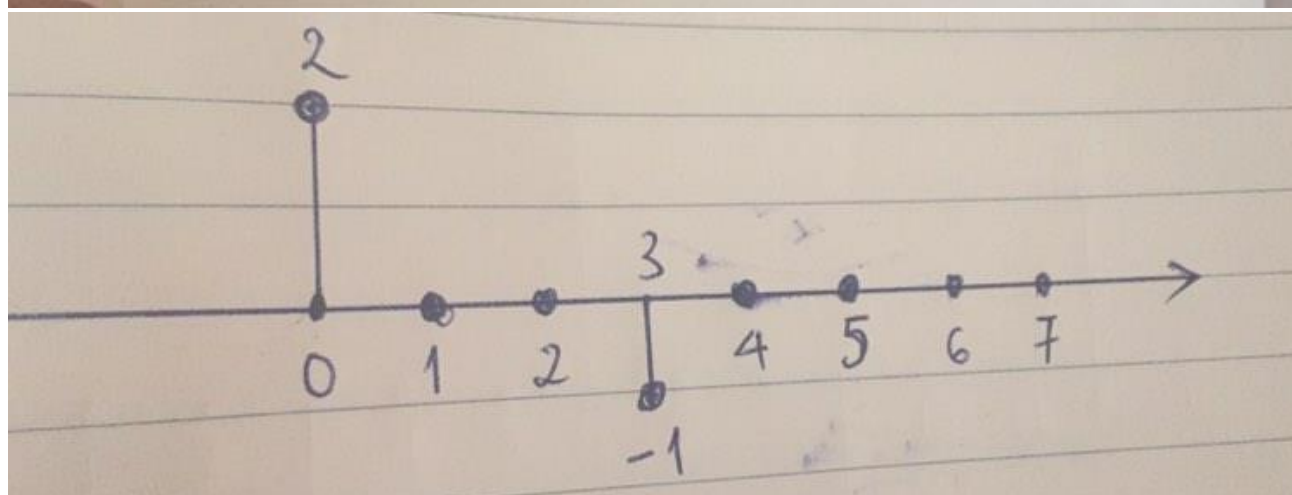
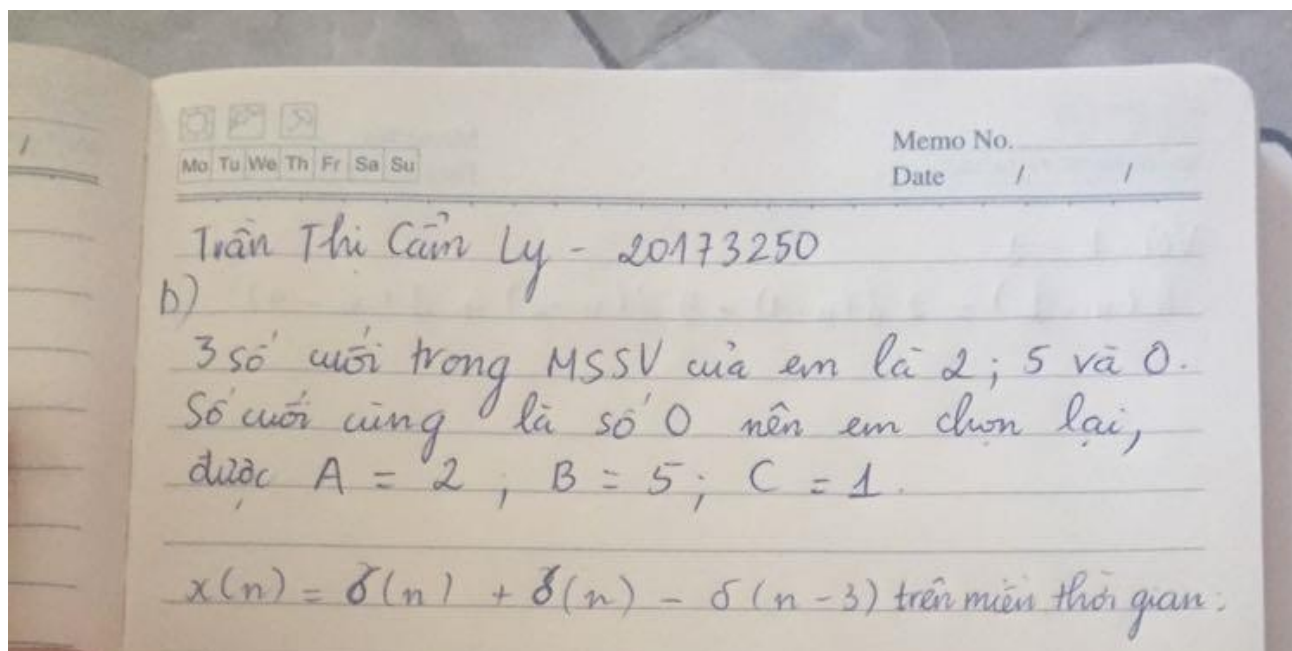
$x(n)$ là đầu vào của phép chập

$h(n-k)$ là phép dịch phải k mẫu trên $h(n)$

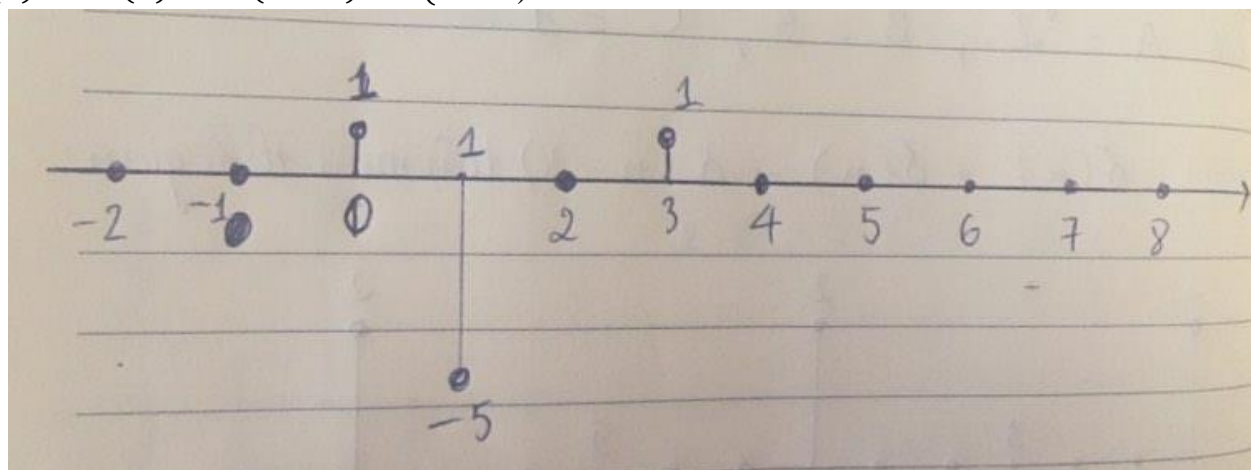
Bản chất của phép lấy chập: là phép dịch chuyển $h(n)$, mỗi lần dịch đi k mẫu, sau đó nhân với mẫu tương ứng trong input $x(n)$, sau đó cộng tất cả các giá trị $x(k)h(n-k)$ thì ta thu được output.

- b) Giả thiết $x(n) = \delta(n) + \delta(n) - \delta(n-3)$, $h(n) = A\delta(n) - B\delta(n-1) + C\delta(n-3)$ trong đó A, B và C là 3 số cuối tương ứng với mã sinh viên, nếu có số nào bằng 0 trong 3 số cuối của mã sinh viên thì cộng hoặc trừ một số nguyên nào đó (1, 2, 3...) sao cho trong A, B, C không có trùng số.

Hãy vẽ $x(n)$, $h(n)$ và tính $y(n)$ trong miền thời gian theo từng giá trị của k , với mỗi giá trị của k vẽ $h(n-k)$ tương ứng và nêu nhận xét so sánh về $h(n-k)$ với các giá trị k khác nhau trong trường hợp này.



$$h(n) = 2\delta(n) - 5\delta(n-1) + \delta(n-3):$$



- c) Giải thích chỉ bằng một câu về bản chất của phép lấy chập (convolution) được dùng trong xử lý ảnh. So sánh phép lấy chập nói ở mục a) với phép lấy chập được dùng trong xử lý ảnh (không lấy ví dụ bằng số để so sánh). Chỉ dùng một câu để giải thích tại sao phép lấy chập dùng trong xử lý ảnh cũng được dùng trong nhận dạng tiếng nói.

Giải thích: Bản chất của phép lấy chập (convolution) được dùng trong xử lý ảnh: là phép dịch một ma trận kernel lần lượt từ trái qua phải từ trên xuống dưới theo từng pixel trong ảnh, sau đó thực hiện phép nhân ma trận kernel với ma trận biểu diễn ảnh đầu vào rồi cuối cùng cộng lại thì ta sẽ thu được ma trận output.

So sánh: phép lấy chập nói ở mục a) tương tự với phép lấy chập được dùng trong xử lý ảnh:

Đây là công thức của phép lấy chập dùng trong xử lý ảnh:

$$g(x, y) = \omega * f(x, y) :$$

Trong đó: omega là ma trận kernel, $f(x, y)$ là ma trận ảnh input và $g(x, y)$ là ma trận ảnh output. Ta thấy omega thực hiện phép tích chập với $f(x, y)$ để cho ra $g(x, y)$ cũng giống như công thức trên: $y(n) = x(n) * h(n)$

Giải thích tại sao phép lấy chập dùng trong xử lý ảnh cũng được dùng trong nhận dạng tiếng nói: Phép lấy chập dùng trong xử lý ảnh, tùy theo giá trị của Kernel sẽ có tác dụng lọc nhiễu, làm nét ảnh hoặc xác định các đường biên trong ảnh, v.v...; trong nhận dạng tiếng nói, ta cũng cần thực hiện những công việc tương tự như: giảm tạp âm, VAD (voice activity detection)..., vì vậy cũng cần dùng đến phép lấy chập giống như trong xử lý ảnh.

2. Giải thích lý do dùng hàm tự tương quan để xác định tần số cơ bản của tín hiệu tiếng nói (Không dùng công thức toán để giải thích)

Giải thích:

Tự tương quan của một dãy luôn luôn có giá trị cực đại tại $n = 0$, bởi vì một dãy bao giờ cũng giống chính nó. Tìm F_0 chính là đi tìm cực đại nên ta có thể dùng hàm tự tương quan để xác định tần số cơ bản của tín hiệu tiếng nói.