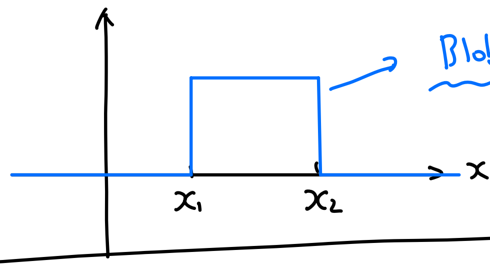
Scale-Invariant Feature Transform

- ① Detector
- ② Descriptor

① Detector Framework → Scale-Invariant Keypoints

idea: Blob Detection → LoG

$$\mathrm{Blob}_{x_1,x_2}(x) = \mathrm{Step}(x-x_2) - \mathrm{Step}(x-x_1)$$



Firstly, consider $U(x) = \mathrm{Step}(x-x_2)$ $\boxed{\text{Edge Detection}}$

Filter → LoG $= \dfrac{d^2}{dx^2} G_\sigma(x)$

$$\rightarrow \mathrm{Step}(x-x_2) * \frac{d}{dx^2} G_\sigma(x) = \frac{-(x-x_2)}{\sqrt{2\pi}\,\sigma^3} \exp\left(-\frac{1}{2\sigma^2}(x-x_2)^2\right)$$

a) Zero-Crossing : x_2 (edge detection)

b) Extrema : $x_2 \pm \sigma$

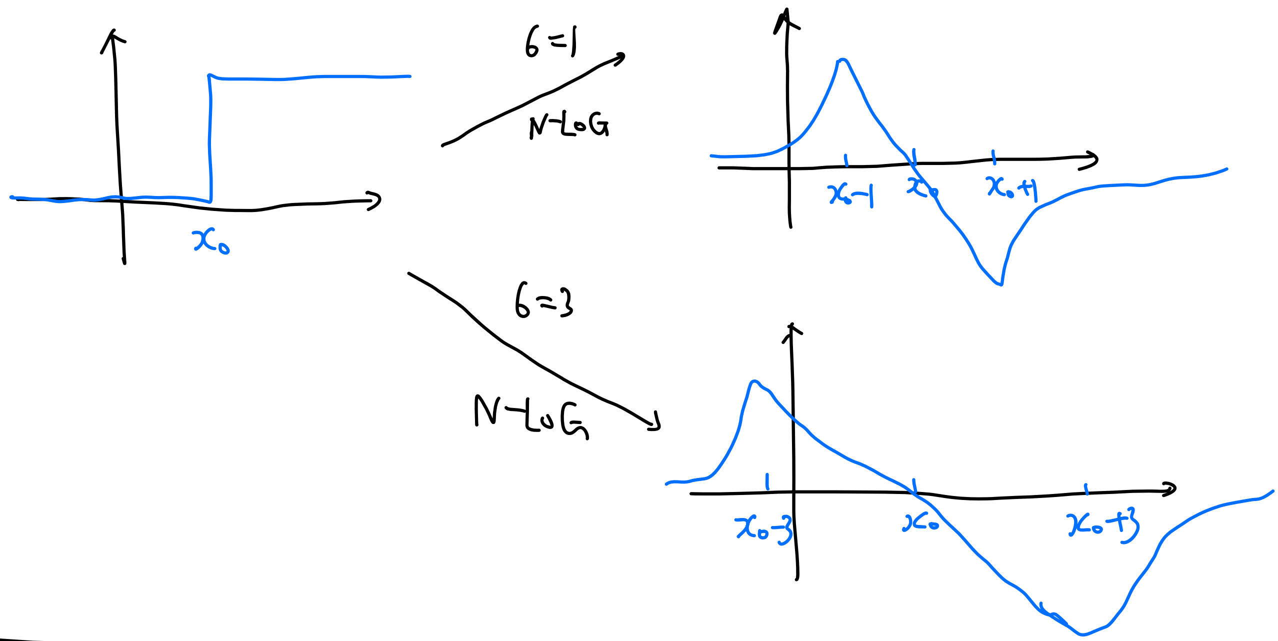
Corresponding response value : $\pm \dfrac{1}{\sqrt{2\pi}\,\sigma^2} \exp\left(-\dfrac{1}{2}\right)$

$\Longrightarrow$ Introduce "Normalized LoG" $\boxed{\sigma^2 \cdot \dfrac{d^2}{dx^2} G_\sigma(x)}$

↓

$\boxed{\text{Scale invariant}}$

E.g.



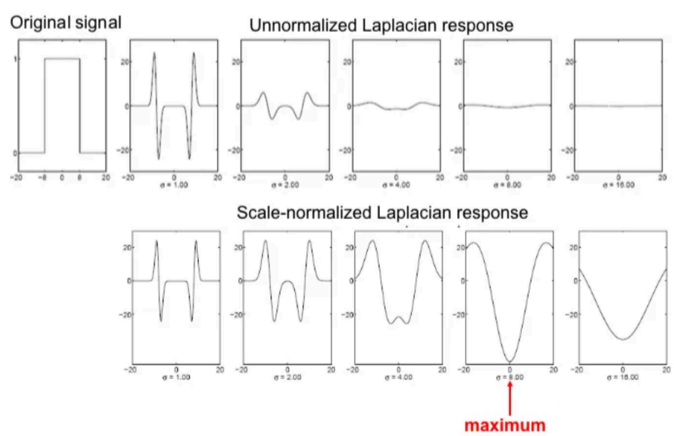
Then, consider $\text{Blob}_{x_1,x_2}(x)$ and Normalized LoG (N-LoG)

when $\sigma = \dfrac{x_2-x_1}{2}$, $\text{Blob}_{x_1,x_2}(x) * \text{N-LoG}(x)$

reaches its minima at $x = x_1+\sigma$

$= x_2-\sigma$

[E.g.]



Note: this is a extrema in scale-space (x, y, σ), not just in function domain (x, y)

SIFT Approximate $\sigma^2 (\nabla^2 G_\sigma (x,y))$ $\longrightarrow$ $\boxed{\text{Scale-Invariant LoG}}$

through $[G_{k\sigma} (x,y) - G_\sigma (x,y)]$ $\longrightarrow$ $\boxed{\text{DoG}}$

Pf. observation : $G_\sigma (x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2} (x^2+y^2))$

then $\boxed{\sigma \nabla^2 G_\sigma (x,y) = \frac{\partial}{\partial \sigma} G_\sigma}$ (from direct calculation)

$$\Rightarrow \frac{\partial}{\partial \sigma} G_\sigma(x,y) = \lim_{k \to 1^+} \frac{G_{k\sigma} (x,y) - G_\sigma (x,y)}{k\sigma - \sigma}$$

$$\approx \frac{G_{k\sigma} (x,y) - G_\sigma (x,y)}{(k-1) \sigma}$$

$$\Rightarrow G_{k\sigma} (x,y) - G_\sigma (x,y) \approx (k-1) \sigma^2 \nabla^2 G_\sigma (x,y)$$

our interest is : extrema of $\sigma^2 \nabla^2 G_\sigma (x,y) := \boxed{F(x,y,\sigma) * I(x,y)}$

in (x,y,σ) space

can be approximately solved by : $(G_{k\sigma}(x,y) - G_\sigma (x,y)) * I (x,y)$

$\downarrow$

$\boxed{\text{extrema in } (x,y,\sigma) \text{ space}}$

$\longrightarrow$ This is what we do in SIFT , detector part

② Descriptor

a) orientation alignment

b) 128 (4×4×8 feature representation

based on gradient histgram (8 direction)

corresponding to " Dominant Direction "

↑

$\boxed{\text{orientation alignment}}$