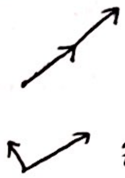
Convolutions

Let's begin by reviewing the dot product: $\langle \vec{f}, \vec{g} \rangle$

If $\vec{f} \in \mathbb{R}^n$ and $\vec{g} \in \mathbb{R}^n$ (e.g. $\vec{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix}$), then $\langle \vec{f}, \vec{g} \rangle = f_0 g_0 + f_1 g_1 + \dots f_n g_n$

The dot product is a scalar measure of how "close" the vectors $\vec{f}$ and $\vec{g}$ are.



Now, what if $\vec{f}$ and $\vec{g}$ were infinitley long vectors? They could represent continuous functions. We'll refer to these vectors as $\vec{f}(x)$ and $\vec{g}(x)$.

$$\vec{f}(x) = \begin{bmatrix} f(-\infty) \\ \vdots \\ f(0) \\ \vdots \\ f(\infty) \end{bmatrix} = \begin{bmatrix} \vdots \\ f(-2h) \\ f(-h) \\ f(0) \\ f(h) \\ f(2h) \\ \vdots \end{bmatrix}$$

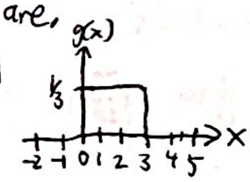
, where h is an infinitley small increment.

The dot product between $\vec{f}(x)$ and $\vec{g}(x)$ can be written as:

$$\langle \vec{f}(x), \vec{g}(x) \rangle = \vec{g}^T(x)\vec{f}(x) = \int_{-\infty}^{\infty} f(x)g(x)\,dx$$

, where $f(x), g(x)$ are simply the continuous functions that generate $\vec{f}(x), \vec{g}(x)$.

So now we have a way of measuring how "close" or "similar" two functions are. One good use of this is an averaging filter. If we set $g(x) = \begin{cases} 1/3 & \text{if } x \in [0,3] \\ 0 & \text{else} \end{cases}$



If we take $\langle \vec{f}(x), \vec{g}(x) \rangle$ or $\int_{-\infty}^{\infty} f(x)g(x)\,dx$, which are equal, we get the average of the values of $f(x)$ from $x=0$ to $x=3$. In other words, we are finding how "close" $f(x)$ is to the boxy shape of $g(x)$. However, what if we want to find the averages for other ranges of $f(x)$? We'd have to slide $g(x)$ around a lot. In fact, we can keep tabs on the dot product between $f(x)$ and $g(x)$ for all shifted versions of $g(x)$ by writing the following:

$$F(h) = \begin{bmatrix} \int_{-\infty}^{\infty} f(x)g(x)\,dx \\ \int_{-\infty}^{\infty} f(x)g(x-h)\,dx \\ \int_{-\infty}^{\infty} f(x)g(x-2h)\,dx \\ \vdots \end{bmatrix} = \begin{bmatrix} \langle \vec{f}(x), \vec{g}(x) \rangle \\ \langle \vec{f}(x), \vec{g}(x-h) \rangle \\ \langle \vec{f}(x), \vec{g}(x-2h) \rangle \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} g^T(x) \text{ ——} \\ g^T(x-h) \text{ ——} \\ g^T(x-2h) \text{ ——} \\ \vdots \text{ ——} \\ \vdots \text{ ——} \\ \vdots \end{bmatrix}} \begin{bmatrix} | \\ \vec{f}(x) \\ | \end{bmatrix}$$

→ This matrix is a catalog of all shifted versions of $g(x)$.

So instead of one dot product, we now have an infinite amount of dot products in the form of a matrix multiplication! The function $F(h)$ gives us a dot product for every shift "h". Notice how this matrix switches us from the "x" domain to the "h" domain.

This matrix multiplication is called a <u>correlation</u>.

(Side note: The "convolution" performed in Convolutional Neural Networks is really <u>correlation</u>, so it's a little misleading, but they're all just fancy words at the end of the day.)

So not only is this correlation going to give us alot of dot products and slides, but it is also capable of ~~changing~~ transforming the basis and domain of the original function.

If you're already familiar with the Fourier Transform, and all this "flipping domain" business reminds you of that transform, then it should! We can flip between the time and frequency domains with the method used before.

$$F(w) = \int_{-\infty}^{\infty} f(t) e^{-iwt} dt = \begin{bmatrix} \int_{-\infty}^{\infty} f(t) e^{-i0t} dt \\ \int_{-\infty}^{\infty} f(t) e^{-i1t} dt \\ \int_{-\infty}^{\infty} f(t) e^{-i2t} dt \\ \vdots \end{bmatrix} = \begin{bmatrix} e^{-i0t} & \text{---} \\ e^{-i1t} & \text{---} \\ e^{-i2t} & \text{---} \\ \vdots & \end{bmatrix} \begin{bmatrix} | \\ \vec{f}(t) \\ | \end{bmatrix}$$

☆ The negative sign in the exponent is necessary to properly correlate with complex functions.

☆ The Fourier Transform is Unitary, which carries a long a bunch of other neat properties.

Let's play with correlation idea some more. Can the derivative be expressed as this shifted matrix multiplication?

If $\frac{d}{dt} f(t) = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}$, then perhaps

$$\frac{d}{dt} f(t) = \lim_{h \to 0} \frac{1}{h} \underbrace{\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & 0 & -1 & 1 & \\ \vdots & \vdots & \vdots & & \ddots & \ddots \end{bmatrix}}_{\text{Let's call this } D} \begin{bmatrix} f(t_0) \\ f(t_0+h) \\ f(t_0+2h) \\ f(t_0+3h) \\ \vdots \end{bmatrix} = \lim_{h \to 0} \frac{1}{h} D \vec{f}(t)$$

If we want the second, third, or n^{th} derivative, we can just take D^2, D^3, D^n.
And if you take successive multiplications of D, you will start to see the binomial theorem sequence!

$$D^2 = \begin{bmatrix} 1 & -2 & 1 & & \\ 0 & 1 & -2 & 1 & \\ & & & \cdots & \\ & & & \ddots & \end{bmatrix} \quad D^3 = \begin{bmatrix} 1 & -3 & 3 & -1 & \\ & 1 & -3 & 3 & -1 \\ & & & & \ddots \end{bmatrix} \quad D^n = \begin{bmatrix} n^{th} \text{ row of} \\ \text{Pascal's triangle} \\ \text{and all shifted versions} \end{bmatrix}$$

This should make sense if you are ~~so~~ familiar with how Pascal's triangle is related to multiplying polynomials, or if you can see how ~~the rows of Pascal's triangle are~~ taking the next derivative is the same process as generating successive rows of Pascal's triangle. Try working it out yourself.

e.g. $f''(t) = \lim_{h \to 0} \frac{f(t+2h) - 2f(t+h) + f(t)}{h^2}$

So we've talked about dot products, and correlations, but where is the <u>convolution</u>? In fact, don't the correlation and convolution look really similar?

$$F(h) = \int_{-\infty}^{\infty} f(x)\, g(x-h)\, dx \Longrightarrow \text{Correlation}$$

There are a couple differences you can spot just by looking at these integrals.

$$\Rightarrow \text{Convolution}$$

$$F(x) = \int_{-\infty}^{\infty} f(h)\, g(x-h)\, dh$$

notation for convolution $\hookrightarrow f(x) \circledast g(x)$

Correlation	Convolution
Function of <u>shift</u> "h"	Function of <u>original variable</u> "x"
Shifts $g(x)$	<u>Flips</u> $g(x)$ <u>before</u> shifting $(g(-(t-\tau)))$

But these small ~~[illegible]~~ differences don't give us the full picture. It's not obvious why you'd want to flip $g(x)$ before shifting, or why the domain stays the same ~~[illegible]~~ before and after convolution.

The reasons that the convolution has these quirks becomes a little ~~[illegible]~~ clearer in the context of time-dependent systems. For instance, if you <u>don't</u> flip $g(t)$ before the shift, you have an operator that acts on values in the future, which is absurd, especially since this operator describes the behavior of real physical systems.

I'll skip the details on LTI systems and concepts such as impulse response, and just get to the part where you really see the convolution.

One impulse gives a $g(t)$. Multiple scaled impulses give multiple stacked versions of $g(t)$.

$$F(t) = f(t) \circledast g(t) = \begin{bmatrix} g(0) & g(-1) & \text{usually nothing} & \\ g(1) & g(0) & & \\ g(2) & g(1) & g(0) & \\ g(3) & g(2) & g(1) & g(0) \\ \vdots & & & \\ g(n) & & & \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \\ \vdots \end{bmatrix}$$

The affects of $f(0)$ haven't finished though! It cascades down $t \to \infty$.

Same with all the effects of $f(1)$!

A new $g(0)$ occurs for every new f thing done to the system.

$\hookrightarrow$ a new $g(n)$ is added at every timestep, because timestep n is the first time this value can appear.

Imagine $f(0)$ as 1, and $f(10) = 1$, everything else is 0.

$f(0)$ "does" $g(0 \ldots \text{something})$ to the system

☆ Think about neurons

stacking voltage on top of every spike

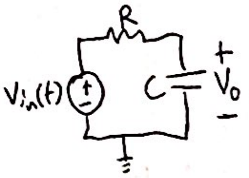
☆ Or slapping a chicken. You can't cook the chicken if you slap it too slowly.

Bonus: Convolution in an RC circuit

The capacitor is a great physical model for exploring convolution. Yes, you can mathematically demonstrate that the capacitor integrates the voltage being fed in, but you should also intuitively recognize that a capacitor is a device that STORES and ACCUMULATES charge, which is exactly what an integral does.

So even if you didn't know the transfer function of a capacitor, it's important to feel like something is being accumulated, which should give you a hint that maybe an integral/convolution is involved somewhere!!

In the next part, I will show that if you solve the differential equation for the circuit, you will indeed, end up with a convolution. We will only focus on the 0-state response, because that's where the convolution is.



$$V_{in}(t) = RC\frac{dV_0}{dt} + V_0(t), \quad V_0(0) = 0$$

Time Domain

① Re-arrange: $\frac{1}{RC}V_{in}(t) = \frac{dV_0}{dt} + \frac{1}{RC}V_0(t)$

② Use integrating factor to solve. It's $e^{t/RC}$, trust me.

$$\frac{e^{t/RC}}{RC}V_{in}(t) = e^{t/RC}\frac{dV_0}{dt} + \frac{e^{t/RC}}{RC}V_0(t) \rightarrow \text{this is product rule}$$

$$\frac{e^{t/RC}}{RC}V_{in}(t) = \frac{d}{dt}\left[e^{t/RC}V_0(t)\right]$$

Integrate

③ $$\frac{1}{RC}\int e^{t/RC}V_{in}(t)\,dt = e^{t/RC}V_0(t) + K_1$$

↳ This constant is 0 b/c $V_0(0)=0$ (work it out, hint: the integral goes to 0)

④ $$V_0(t) = \frac{1}{RC}e^{-t/RC}\int_0^t e^{\tau/RC}V_{in}(\tau)\,d\tau \text{ ☆}$$

→ This ~~convolution~~ integral is a function of t, so we do a harmless switching of variables.

⑤ $$\boxed{V_0(t) = \frac{1}{RC}\int_0^t e^{-(t-\tau)/RC}\,V_{in}(\tau)\,d\tau}$$

Laplace Domain

① $V_{in}(s) = RCs\,V_0(s) + V_0(s) - \cancel{V_0(0)}^{\,0}$

② $V_{in}(s) = (RCs+1)V_0(s)$

③ $V_0(s) = \frac{1}{RCs+1}V_{in}(s)$

④ Take inverse Laplace, using $\mathcal{L}\{f * g\} = F(s)G(s)$, and using a Laplace transform table because I didn't memorize them lol.

$$V_0(t) = \frac{e^{-t/RC}}{RC} * V_{in}(t)$$

⑤ $$\boxed{V_0(t) = \frac{1}{RC}\int_0^t e^{-(t-\tau)}\,V_{in}(\tau)\,d\tau} \text{ ☆}$$

They match!

☆ Although convolution is defined $\int_{-\infty}^{\infty} f(t)g(\tau - t)\,dt$, we can shrink these bounds because we only care about $[0, t]$.