M668 Fourier Analysis - Final Project Write-up

Stefan Cline

Spring 2023

**Note that this write up is informal and was used to help keep track of the theory behind the code.*

1 Back Propagation for the Neural Net

A **mother wavelet** must satisfy

$$\int \psi(t)\ dt = 0 \tag{1}$$

$$||\psi(t)||_{L^2} = \left[\int |\psi(t)|^2\ dt\right]^{1/2} = 1 \tag{2}$$

To satisfy the first condition, we ensure each function that makes up ψ is odd. Then we know the function will have zero area (considering *p.v.*). This then allows us to use one of many possible general forms. In this case we'll use (letting $\varepsilon = 0.1$):

$$\hat{\psi}(t) = \exp\left(\frac{-t^2}{2}\right)(\varepsilon t + A\sin(Bt)) \tag{3}$$

This obviously satisfies the property: $\forall t \in \mathbb{R},\ \hat{\psi}(t) \neq 0$ as it is impossible to have $\varepsilon t = A\sin(Bt)$ if $\varepsilon \neq 0$. This ensures that as we learn a better wavelet, our wavelet doesn't simply become zero.

For the second condition, we simply note that this will be true if we divide $\hat{\psi}$ by its L^2 norm. Effectively (and noting we're assuming real functions)

$$\psi(t) = \frac{\hat{\psi}(t)}{||\hat{\psi}(t)||_{L^2}}$$

$$= \frac{\exp\left(\frac{-t^2}{2}\right)(\varepsilon t + A\sin(Bt))}{\left(\int_{\mathbb{R}} \left|\exp\left(\frac{-t^2}{2}\right)(\varepsilon t + A\sin(Bt))\right|^2\ dt\right)^{1/2}}$$

$$= \frac{\exp\left(\frac{-t^2}{2}\right)(\varepsilon t + A\sin(Bt))}{\left(\int_{\mathbb{R}} \exp(-t^2)(\varepsilon t + A\sin(Bt))^2\ dt\right)^{1/2}}$$

The above is our **mother wavelet**. Just to be clear, we have some function $\hat{\psi}(t)$ divided by a definite integral, i.e., some quantity. As such, we effectively have (where $L_2 \equiv$ const.) our mother wavelet in the form

$$\psi = \frac{1}{L_2}\hat{\psi}(t) \tag{4}$$

Now, we want to be able to squish and stretch φ so we have a scale in the form of an exponential, j:

$$\varphi_j(t) = \frac{1}{\sqrt{a^j}}\psi\left(\frac{t}{a^j}\right) \tag{5}$$

$$= \frac{1}{\sqrt{a^j}} \frac{\exp\left(-\frac{1}{2}\frac{t^2}{a^{2j}}\right)\left(\frac{\varepsilon t}{a^j} + A\sin\left(\frac{Bt}{a^j}\right)\right)}{\left(\int_{\mathbb{R}} \exp\left(-\frac{t^2}{a^{2j}}\right)\left(\frac{\varepsilon t}{a^j} + A\sin\left(\frac{Bt}{a^j}\right)\right)^2\ dt\right)^{1/2}} \tag{6}$$

In this setup, we'll simply pick that $a = 2$ and $j \in \mathbb{Z}^+$. As such, scaling j creates a broader wavelet, and shrinking it brings it back to its highest frequency base value of $0 :\to 2^0 = 1$. It is worth pausing for a moment that we are writing these formulas out explicitly due to the fact that later we will be taking derivatives with respect to the parameters A and B and thus seeing them written out fully is valuable.

The wavelet transform (WT) then can be seen as

$$W_{j,t} = \varphi_j(t) * f(t) = \int_{\mathbb{R}} \varphi_j(t-\tau) f(\tau)\ d\tau \tag{7}$$

$$= \int_{\mathbb{R}} \left[\frac{1}{\sqrt{a^j}} \frac{\exp\left(-\frac{1}{2}\frac{(t-\tau)^2}{a^{2j}}\right)\left(\frac{\varepsilon(t-\tau)}{a^j} + A\sin\left(\frac{B(t-\tau)}{a^j}\right)\right)}{\left(\int_{\mathbb{R}} \exp\left(-\frac{(t-\tau)^2}{a^{2j}}\right)\left(\frac{\varepsilon(t-\tau)}{a^j} + A\sin\left(\frac{B(t-\tau)}{a^j}\right)\right)^2\ dt\right)^{1/2}} \right] f(\tau)\ d\tau \tag{8}$$

Where f is some arbitrary signal.

After observing some images of 'high performing' wavelet transforms, we notice the surface of those transforms seems to mostly be zero. Therefore, if we can get the surface to be zero almost everywhere, we've succeeded at making a 'good' wavelet. As such, after the WT has been performed, we'll aim for a learning goal that seeks to minimize the value of the field, i.e., tell it to try to set every point to zero. Thinking ahead, we also don't want our values for A or B to become zero. Hence, we'll penalize an extremely steep partial parabolic term for each variable. As such, our loss function, C, will be a discrete field $C_{j,t}$ such that:

$$C_{j,t} = \left[\sum_{j=0}^{J} \sum_{t=t_0}^{T} W_{j,t} - \mathcal{W}_{j,t} \right]^2 + \mathcal{Q}_A + \mathcal{Q}_B \tag{9}$$

$$C_{j,t} = \left(\sum_{j=0}^{J} \sum_{t=t_0}^{T} W_{j,t} \right)^2 + \mathcal{Q}_A + \mathcal{Q}_B \tag{10}$$

where $\mathcal{W}_{j,t} = 0$ for all pairs of j and t (i.e., it's a zero matrix the size of W meaning we're doing a least squares optimization).

The discouraging function for A and B is defined along with its derivative, D as

$$\mathcal{Q}_\Gamma = \begin{cases} \Gamma \in [-l,l]: & -(l^{-1}\sqrt{\beta}\cdot\Gamma)^2 + \beta \\ \Gamma \notin [-l,l]: & 0 \end{cases}, \qquad D\mathcal{Q}_\Gamma = \begin{cases} \Gamma \in [-l,l]: & -2l^{-2}\beta\cdot\Gamma \\ \Gamma \notin [-l,l]: & 0 \end{cases} \tag{11}$$

Above, l is the lowest value desired for A and B, and β is the error you want to provide to the algorithm by choosing $A = 0$ or $B = 0$.

Here, because the parameters we'll be adjusting are A and B, we'll want to calculate the partial derivatives of A and B and find their downhill direction, i.e., the negative gradient for each parameter. We'll label the parameters arbitrarily as Γ, then solve for them more specifically once we've sufficiently adjusted the equation. This is how we will minimize our cost function, C, and thereby ideally minimize our loss.

$$\frac{\partial C}{\partial \Gamma} = \frac{\partial}{\partial \Gamma}[W^2 + \mathcal{Q}_A + \mathcal{Q}_B] \tag{12}$$

$$= \frac{\partial}{\partial \Gamma} \underbrace{\left(\int_{\Omega} \varphi_j(t-\tau) f(\tau)\ d\tau \right)^2}_{W_{j,t}^2} + D\mathcal{Q}_\Gamma \tag{13}$$

$$= 2W_{j,t} \cdot \frac{\partial}{\partial \Gamma} W_{j,t} + D\mathcal{Q}_\Gamma \tag{14}$$

Let's be mad real[1]. That above line is disgusting so let's head on over to the 'Essential Tool for Mathematics', Maple™.

First we show the equation for W in Maple with a large note that it still needs to be housed inside of $\int_{\mathbb{R}} W\ d\tau$. This was done as the outside integral doesn't effect the partial derivatives, but is computationally excessive for Maple to perform:

[1] This is a reference to the episode, *The Mad Real World*, of the famous Chappelle's Show™.

$$W := \frac{e^{-\frac{(t-\tau)^2}{2 a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right) f(\tau)}{\sqrt{a^j} \sqrt{\int_{-\infty}^{\infty} e^{-\frac{(t-\tau)^2}{a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right)^2 d\tau}}$$

For A, this is DW_A (again still needing to be inside an integral):

> $EQA := diff(W, A)$

$$EQA := \frac{e^{-\frac{(t-\tau)^2}{2 a^{2j}}} \sin\left(\frac{B (t-\tau)}{a^j} \right) f(\tau)}{\sqrt{a^j} \sqrt{\int_{-\infty}^{\infty} e^{-\frac{(t-\tau)^2}{a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right)^2 d\tau}} - \left(e^{-\frac{(t-\tau)^2}{2 a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} \right.\right.$$

$$\left.\left. + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right) f(\tau) \left(\int_{-\infty}^{\infty} 2\, e^{-\frac{(t-\tau)^2}{a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right) \sin\left(\frac{B (t-\tau)}{a^j} \right) d\tau \right) \right) \Bigg/$$

$$\left(2 \sqrt{a^j} \left(\int_{-\infty}^{\infty} e^{-\frac{(t-\tau)^2}{a^{2j}}} \left(\frac{\varepsilon (t-\tau)}{a^j} + A \sin\left(\frac{B (t-\tau)}{a^j} \right) \right)^2 d\tau \right)^{3/2} \right)$$

For B, this is DW_B (again still needing to be inside an integral):

> $EQB := diff(W, B)$

$$EQB := \frac{e^{-\frac{(t-\tau)^2}{2 a^{2j}}} A\,(t-\tau)\cos\left(\frac{B\,(t-\tau)}{a^j}\right) f(\tau)}{\left(a^j\right)^{3/2}\sqrt{\int_{-\infty}^{\infty} e^{-\frac{(t-\tau)^2}{a^{2j}}}\left(\frac{\varepsilon\,(t-\tau)}{a^j}+A\sin\left(\frac{B\,(t-\tau)}{a^j}\right)\right)^2 d\tau}} - \left(e^{-\frac{(t-\tau)^2}{2 a^{2j}}}\left(\frac{\varepsilon\,(t-\tau)}{a^j}+A\sin\left(\frac{B\,(t-\tau)}{a^j}\right)\right) f(\tau)\left(\int_{-\infty}^{\infty}\frac{2\,e^{-\frac{(t-\tau)^2}{a^{2j}}}\left(\frac{\varepsilon\,(t-\tau)}{a^j}+A\sin\left(\frac{B\,(t-\tau)}{a^j}\right)\right) A\,(t-\tau)\cos\left(\frac{B\,(t-\tau)}{a^j}\right)}{a^j}\,d\tau\right)\right)\Bigg/\left(2\sqrt{a^j}\left(\int_{-\infty}^{\infty} e^{-\frac{(t-\tau)^2}{a^{2j}}}\left(\frac{\varepsilon\,(t-\tau)}{a^j}+A\sin\left(\frac{B\,(t-\tau)}{a^j}\right)\right)^2 d\tau\right)^{3/2}\right)$$

It's finally time to put it all together. Because we are sliding the functions in the convolutions by τ, this means that each discrete point in the field W will come from sliding across $\tau \in \Omega$ where $\Omega = [t_0, T]$ is our respective discrete time interval. Therefore, our gradients will look like (with $j \in ([0, J] \subset \mathbb{Z}^+)$ and a constant Δj between each value)

$$\nabla C_\Gamma = \begin{bmatrix} Q_{\Gamma,0} \\ \vdots \\ Q_{\Gamma,J} \end{bmatrix}_{j,t} \tag{15}$$

This means the gradient for each parameter will be a matrix sized j, t for discretized $t \in \Omega$, and as such

$$\nabla C_\Gamma = \sum_{j=0}^{J}\sum_{t=t_0}^{T}\left[\frac{\partial C}{\partial \Gamma}\right]_{j,t} \tag{16}$$

Finally, if we have that

$$\nabla C_\Gamma > 0 \Rightarrow \text{Subtract the Learning Rate} \to \Gamma - L_r \tag{17}$$
$$\nabla C_\Gamma < 0 \Rightarrow \text{Add the Learning Rate} \to \Gamma + L_r \tag{18}$$