SIGNAL DENOISING AND ERROR BOUNDING USING ANALYTIC METHODS

NOAH STOCKWELL

ABSTRACT. It is shown that with normalized convolution kernels, one can find smooth derivatives that interpret any continuous system sampled at discrete points.

CONTENTS

1. INTRODUCTION

This paper provides a specific formulation of a method to find smooth trends in discrete data for both the Riemann integral and then the Lebesgue integral with bound consideration. The motivation is the existence of the engineering question: *what do real-world processes look like when they are sampled and how do we remove sampling error without destroying the signal?* This paper considers the result of applying a convolution to a set of 2-dimensional discrete points, and how accurate we can expect the convolution to be compared to the process that is being sampled.

2. Formulation for $S \in C_0(\mathbb{R})$ using the Riemann Integral

Theorem 2.1. *Given a set $T = (T_x, T_y)$ of points separated by an identical T_x distance λ, the generating function $S \in C_0(\mathbb{R})$ can be approximated by a smooth function.*

Proof. First, define the piecewise function made up of constant functions that passes through all points in T:

$$f(x) \mid (x, f(x)) \in T$$

With

$$f(t) \equiv T_{y_m} \qquad \forall\, t \in [T_{x_m}, T_{x_{m+1}})$$

Next, define the function with consideration width w:

(1)
$$\delta_n(x, w) := \begin{cases} \left(1 - \left(\frac{x}{w}\right)^n\right)^2 & x \in (-w, w) \\ 0 & x \notin (-w, w) \end{cases}$$

Normalization is trivial:

$$N(w) := \int_{-w}^{w} \delta_n(t, w)\, dt = C(w)$$

where C is a constant. Define and note that

(2)
$$\phi(x, w) := \frac{\delta_n(x, w)}{N(w)}$$

Is now a probability function: the total area under ϕ is now 1 across $\mathbb{R}$. Define the convolution:

$$f \approx (f \circledast \phi)(x, w) = \int_{-\infty}^{\infty} f(\tau) * \phi(x - \tau, w)\, d\tau$$

Which is exactly equivalent to

(3)
$$(f \circledast \phi)(x, w) = \begin{cases} \int_I f(\tau) * \phi(x - \tau, w)\, d\tau & x \in I \\ 0 & x \notin I \end{cases}$$

For $I = (\min(T_x) - w, \max(T_x) + w)$. Because f is constant and piecewise, we can redefine the integral part as such:

(4)
$$\int_I f(\tau) * \phi(x - \tau)\, d\tau = \sum_{a \in T_x} f(a) \int_a^{a+\lambda} \phi(x - \tau, w)\, d\tau$$

We do not care about extrapolation which means we can throw out the extra convolution width w and tails rewriting the convolution as:

$$f(x) \approx (f \circledast \phi)(x, w) = \sum_{a \in T_x} f(a) \int_a^{a+\lambda} \phi(x - \tau, w)\, d\tau$$

For $x \in [\min(T_x), \max(T_x)]$. By construction, ϕ is smooth and thus a linear combination of ϕ's is smooth. $\qquad \square$

Corollary 2.2. *Given a set $T = (T_x, T_y)$ of points separated by an identical T_x distance λ, the generating derivative can be approximated by a smooth function.*

Proof. From our last result,

$$(5) \qquad f(x) \approx \sum_{a \in T_x} f(a) \int_a^{a+\lambda} \phi(x - \tau, w) \, d\tau$$

$$(6) \qquad \implies \frac{d}{dx} f(x) \approx \sum_{a \in T_x} f(a) \, * \, \frac{d}{dx} \int_a^{a+\lambda} \phi(x - \tau, w) \, d\tau$$

$$(7) \qquad \implies f'(x) \approx \sum_{a \in T_x} f(a) * (\phi(x - a - \lambda, w) - \phi(x - a, w))$$

The approximation for f' can be convolved with an arbitrary probability kernel h, as well, to find a smoother interpretation of the derivative:

$$(8) \qquad f'(x) \approx \left(\frac{d}{dx} (f \circledast \phi)(x, w) \circledast h \right)(x)$$

□

It is clear that if we choose not only normalized functions ϕ_n and h to be approximations to the identity, we have that our representation will approach the function itself at every point [2].

3. MOTIVATION FOR $S \in C_0(\mathbb{R})$ USING THE LEBESGUE INTEGRAL

Of the many properties of the Lebesgue integral that we do not have with the Riemann integral is the ability to ignore points on a set of measure zero. It is clear that a process of engineering interest is happening continuously in the real world but can be only sampled at finitely many points. Furthermore, it is possible to have significant noise or sensor malfunctions at some points that drive S incredibly far away from the generating signal.
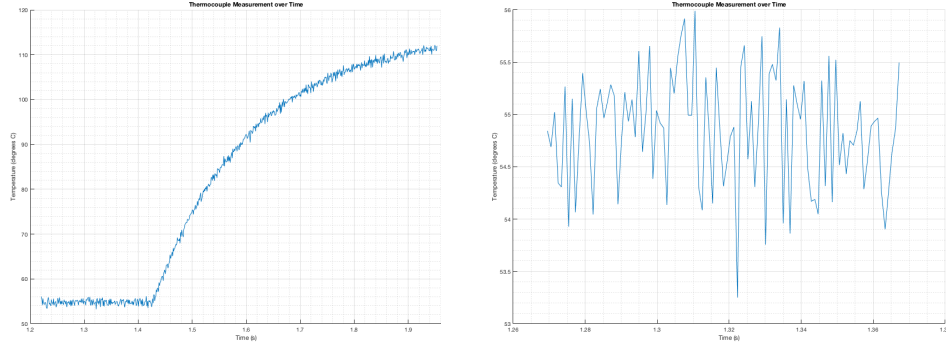
(A) The signal over multiple samples   (B) The signal over few samples

FIGURE 1. A sample thermocouple response at different sampling levels

Above in Figure 1(A), it is clear to us as humans the general shape of the generating signal. However, the true data points being collected look more like Figure 2(B) and it is often a trade-off between maintaining a close relationship to the original generating signal's value and maintaining a close relationship to the original generating signal's behavior. For example, as seen in section (2), the convolution with δ_n will either produce a function that has very few quick changes in its derivative, or a function that approximates the original function closely by value. It is in the Engineer's interest to ask the question: *how much data can I reasonably destroy without jeopardizing the conclusion?*

With this in mind, we will construct a method and identify how 'bad' it can be: how far can it be away from the generating signal?

4. FORMULATION FOR $S \in C_0(\mathbb{R})$ USING THE LEBESGUE INTEGRAL

4.1. **Construction of problem.** Let S be our observed signal. We can visualize S as $S(t) = f(t) + \eta(t) + \rho(t)$ where f is the exact representation of the real process, η is white noise upon the signal, and ρ is extreme instantaneous variation induced by the signal. To replicate the real world, let ρ vanish everywhere except on a set of Lebesgue measure 0. The motivation for this representation is that it is possible to have extreme variation when recording a signal that has no physical meaning

and does not repeat. As white noise on a signal is zero in the time domain, for all E of sufficient measure, we would like $\int_E \eta = 0$. We only need to consider S supported on a set of finite measure as the time values sampled are discrete and thus extrapolation beyond local areas is not desired. As such, we only require S to belong to L^1_{loc}.

4.2. **Requirements.** Without loss of generality, we can split η into η^+ and η^- such that:

$$\int S = \int f + \int \eta$$

$$\int S = \int f + \int \eta^+ - \int \eta^-$$

Where η^+ and η^- are both strictly nonnegative functions on the signal interval I and since we know η should be zero in the time domain [1], we want to create a process such that:

$$\int_I \eta^+ = \int_I \eta^-$$

Since the $\eta^\pm$ are zero when the other is positive, we can let

$$I = E^+ \bigcup E^-$$

with E^+ and E^- disjoint and thus we have:

$$\int_I \eta^i = \int_I \chi_{E^i} \cdot \eta^i = \int_{E^i} \eta^i$$

Since $E^i \subset I$. We can begin constructing the function by defining:

$$E_a^+ = \{x \,:\, S(x) - f(x) > a\}$$
$$E_a^- = \{x \,:\, f(x) - S(x) > a\}$$

Clearly, $E_a^+ \nearrow E^+$ and $E_a^- \nearrow E^-$ as $a \to 0$. For almost everywhere compactly supported functions, we have

$$m(E_a^\pm) \le m(E^\pm) \quad \forall\, a > 0$$

Similarly, for finitely supported functions, we have

$$|E_a^\pm| \le |E^\pm| \quad \forall\, a > 0$$

With these, we want to optimize f such that:

(1) The quantity of noise is largely time independent:

$$\lim_{a \to 0} \left| m(E_a^+) - m(E_a^-) \right| < \epsilon_a$$

(2) The amount of noise is largely time independent:

$$\lim_{a \to 0} \left| \int_{E_a^+} \eta - \int_{E_a^-} \eta \right| < \epsilon_a$$

We can rewrite (2) as:

$$\lim_{a \to 0} \left| \int_{E_a^+} S(x) - f(x)\,dx - \int_{E_a^-} f(x) - S(x)\,dx \right| < \epsilon_a$$

With these in hand, we now want to see how we can bound this difference. It is clear that we desire some type of average over value and time so that we can balance both requirement (1) and (2). This paper makes the choice to consider a convolution ϕ that is normalized and can be considered part of a sequence approaching an approximation to the identity. In fact, we will chose the exact ϕ_n that was selected earlier:

$$\delta_n(x, w) := \begin{cases} \left(1 - \left(\frac{x}{w}\right)^n\right)^2 & x \in (-w, w) \\ 0 & x \notin (-w, w) \end{cases}$$

$$N(w) := \int_{-w}^{w} \delta_n(t, w)\,dt = C(w)$$

$$\phi(x, w) := \frac{\delta_n(x, w)}{N(w)}$$

The question is how far away from f is $S \circledast \phi$? Two bounding options are considered.

4.3. **Error Bounding using a Naive Approach.** Because we require our convolution kernels to be normalized, we have:

$$(S \circledast \phi_n) = \int_E S(\tau) * \phi(x - \tau, w)\,d\tau$$
$$\leq \sup_{x \in E} S(x)$$

However, this bound is really not helpful. A smarter approach to bounding should be considered. One possible method would be to restrict ourselves to $A \subset E$:

$$(S \circledast \phi_n) = \int_A S(\tau) * \phi(x - \tau, w)\,d\tau$$
$$\leq \sup_{x \in A_w} S(x)$$

Where A_w is a w-dilation of A. This is a little more helpful but again A has to be large enough to absorb η.

4.4. **Error Bounding using the Maximal Function.** The Hardy-Littlewood maximal function S^* is defined to be:

$$S^*(t) := \sup_{B_r(t) \ \forall t \in \mathbb{R}} \fint_B S(t) \, dt$$

It is possible to put a bound on the maximal function, especially because we only consider a single-dimensional independent variable on E with $m(E) < \infty$ [2]:

$$m(x \in E \ : \ S^*(x) > \alpha) \le \frac{3}{\alpha} \, \|S\|_{L^1(E)}$$

It is apparent that we can consider the maximal function as the upper bound for the convolution if we require that the convolution kernel has most density around the origin and is only a small value away from vanishing outside of a small ball centered at the origin. In fact, with our choice of ϕ_n have:

$$(S \circledast \phi_n)(t) \le (1 + c) \, S^*(t)$$

With $c > 0$ small. In fact, because our convolution vanishes outside of a ball of radius w around the origin and our convolution kernel is less than the box kernel K of width $\le w$ outside of a smaller ball, we have:

$$(S \circledast \phi_n) = \int_E S(\tau) * \phi(x - \tau, w) \, d\tau$$

$$\le (1 + c) \sup_{\tilde{w} \in (0,w]} \int_E S(\tau) * K(x - \tau, \tilde{w}) \, d\tau$$

$$= (1 + c) \sup_{B_r(t) \ \forall t \in (0,w]} \fint_B S(t) \, dt$$

$$= (1 + c) \, S_w^*(x)$$

Where $S_w^*(x)$ is the maximal function taken over balls of size up to w around x and $c > 0$ small. This provides us a much nicer bound on the convolution.

It was claimed earlier that S could be represented as the sum of three functions and ρ vanishes everywhere except a set of Lebesgue measure 0. There are multiple methods to remove outliers for data: apply one or multiple and now we have:

$$\int_E S = \int_E f + \eta$$

as the Lebesgue integral ignores values on a set of measure zero. We can now assume that we can simply sum the convolution as we did in the beginning:

$$S(x) \approx (S \circledast \phi)(x, w) \le \epsilon_0(x) + \sum_{a_n \in T_x} S(a_n) \int_{a_n}^{a_{n+1}} \phi(x - \tau, w) \, d\tau$$

$$= \epsilon_0(x) + \sum_{a_n \in T_x} \left(f(a_n) + \eta(a_n) \right) \int_{a_n}^{a_{n+1}} \phi(x - \tau, w) \, d\tau$$

If we set

$$b_{a_n(x),w} = \int_{a_n}^{a_{n+1}} \phi(x - \tau, w) \, d\tau$$

we have

$$(S \circledast \phi)(x, w) \le \epsilon_0(x) + \sum_{a_n \in T_x} \left(f\left(a_n\right) + \eta\left(a_n\right) \right) * b_{a_n,w}(x)$$

$$= \epsilon_0(x) + \left(\sum_{a_n \in T_x} f\left(a_n\right) * b_{a_n,w}(x) \right) + \left(\sum_{a_n \in T_x} \eta\left(a_n\right) * b_{a_n,w}(x) \right)$$

Since η sums to zero in the time domain, we get:

$$\sum_{a_n \in T_x} \eta\left(a_n\right) * b_{a_n,w}(x) \equiv 0$$

Leaving us with only

$$(S \circledast \phi)(x, w) \le \epsilon_0(x) + \sum_{a_n \in T_x} f\left(a_n\right) * b_{a_n,w}(x)$$

Which leads us back to the Riemann conclusion. At this point, the question is can we recover $f(t)$ exactly? Clearly we want equality, so the process would have to start with identifying ϵ_0:

$$\epsilon_0(x) = \min\left(\left| (1 + c) * S_w^*(\tilde{x}) - S(x) \right|, \left| \sup_{\tilde{x} \in B_w(x)} S(\tilde{x}) - S(x) \right| \right)$$

It should be clear at least with this choice of bounding, $f(t)$ cannot be recovered exactly. However, for certain classes of functions, we can approximate f very closely. A further question of importance is given a certain type of function, is it possible to significantly improve the bound, i.e. given a first-order ODE, is it possible to be within only an epsilon range of f?

5. CONCLUSION

We do not have the ability to recreate f with this formulation, however we may be able to get close for certain functions, especially simple functions. This process was implemented in MATLAB(™) on an experiment involving temperature measurements of a thermocouple being moved from one environment to another [3]:
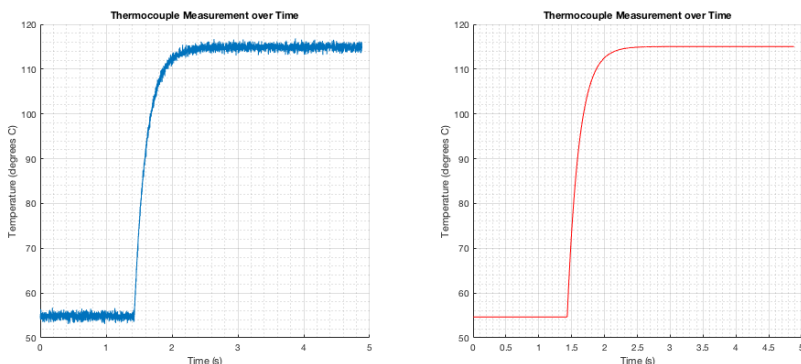

FIGURE 2. Thermocouple response: Noisy and Convolved

Figure 2 demonstrates the potential of a well-tuned convolution in denoising signals for Engineering analysis. It is clear that a correct choice of parameters allows for a close interpretation of the signal. For the mathematical question of how far away is the convolution, more work should be done to better identify the constant c and if superior bounding methods exist. Furthermore, it may be of interest to study only certain classes of functions, as we are modeling real-world processes and we should expect a process to be near a simple PDE, for example.

A potential method to attempt to analyze S in the future could be trying to interpret $f + \eta^+$ and $f - \eta^-$ because these functions would appear to be a sawtooth-wave style upon a smooth function. It could then be possible to use the property that the sawtooth frequency and amplitude should be equal in the time domain to reduce the error in the approximation to f.

REFERENCES

aoki_simultaneous_1994

[1] Naoki Saito. Simultaneous Noise Suppression and Signal Compression using a Library of Orthonormal Bases and the Minimum Description Length Criterion, 1994.

stein_real_2006

[2] Elias Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Number 3 in Princeton Lectures in Analysis. Princeton University Press, 2006.

engr_132

[3] ENGR 132 Team. Engineering 132 Thermocouple Data, 2018.