An Axiomatic Approach to Quantitative Differentiation

Alistair Barton

December 14, 2018

Abstract

This is a project for MATH 606: Topics in PDEs at UBC. In this course, we covered many modern methods for analyzing PDEs, one of which is the so-called 'quantitative differentiation' as popularized by J. Cheeger for analyzing characteristics of the geometry of solutions (e.g. singularity points).

Throughout this class (and in all the resources I could find), the technique was always dealt with in a specific context: a certain PDE, over a certain space, examining a specific characteristic. My project is intended to study the technique in a minimalist setting to see how it might be generalized or its conditions relaxed.

For this project, I wish to present an abstract survey of the technique of quantitative differentiation. The goal here is to develop a deeper understanding of the mechanics of the theory in a general, abstract formulation, and perhaps relax some common constraints.

In Section 1, I provide a rigourous formulation of the theory that is largely an abstraction of Cheeger's argument in his 'General Formulation' [1]. I follow this up by attempting to relaxing the conditions on the energy and set structure—making a conjectured connection with Minkowski dimension along the way—as well as discuss substitutions that can be made for the measure. Lastly, in Section 3, I present a quick survey of some quantitative differentiation results and how they fit in the context of the abstract theory, as well as a simple, original one based on lagrangian action.

MATH 606: Topics in PDEs
December 14, 2018 Alistair Barton

1 General Formulation

Quantitative differentiation is a way to measure how well a given function f (or collection of functions satisfying some property) is approximated by some class of ideal functions. It requires 5 structures—although (i) and (iii) will later be relaxed:

(i) A set of 'ideal' functions $\mathcal{L} \subseteq \mathcal{F} := \{f : X \to Y\}$. For now, we will assume that $0 \in \mathcal{L}$—this assumption will later be removed.

(ii) A *deviation* $\alpha(f, A)$ that takes as input the function and a set A, and generally quantifies how well approximated f is by a function in $\mathcal{L}$ on the set A. A scaled metric is an example.

(iii) An energy function $\mathcal{E} : \mathcal{A} \times \mathcal{F} \to \mathbb{R}_+$. By energy function, we require that $\mathcal{E}$ satisfy two properties: (1) $A_{x,r} \mapsto \mathcal{E}(A_{x,r}, f)$ be σ-additive, and (2) $f \mapsto \mathcal{E}(A_{x,r}, f)$ be strictly convex and lower semi-continuous for all $A_{x,r} \in \mathcal{A}$, to ensure minimizers are unique.
Note that this energy differs from Cheeger's energy in [1] by being scale invariant.

(iv) A σ-finite measure space (X, Σ, λ).

(v) A collection of sets $\mathcal{A} = \{A_{x,r}\} \subseteq \Sigma$ such that for all r $\mathcal{A}_r := \{A_{x,r}\}_x$ form a disjoint cover of X up to a λ-null set, and $\lambda(A_{x,r}) < +\infty$ for all x, r. We will also assume that it exhibits the hierarchical structure that for all x, r, and $s < r$ there exists a y such that $A_{x,r} \subseteq A_{y,s}$. Furthermore, we will only be dealing with x, r countable.
It is common to refer to x as 'locations' and r as 'scales'. We will initialize $r = 0$ with $A_0 := X$ and take higher r to mean finer scales.

There is some convenient notation to use to help us gain some intuition on the concepts. First, we define the '*total defect*' of f on a set $A_{x,r}$ at magnification s to be

$$V_s(A_{x,r}) := \sum_{A_{y,s+r} \subseteq A_{x,r}} \inf_{\ell \in \mathcal{L}} \{\mathcal{E}(A_{y,s+r}, f - \ell)\}. \tag{1.1}$$

(This is different from Cheeger's, which will be used in Application 3.3: I suspect this is more generally applicable.) Since the infimum exists and is unique, we may refer to the minimizing ℓ as $\ell_{y,s+r}$. Note that the interrelation between scale and magnification in this definition, along with the hierarchical structure of $\mathcal{A}$ required in (iv) yield the important relation

$$\sum_x V_s(A_{x,r}) = V_{s+r}(X). \tag{1.2}$$

To compare the magnitude of defects between magnifications, we define the '*relative defect*'

$$D_s(A_{x,r}) := V_0(A_{x,r}) - V_s(A_{x,r}). \tag{1.3}$$

If the relative defect is large, then many defects are apparent at 0 magnification that disappear by s magnification, this implies that f does not exhibit ideal structure on the set $A_{x,r}$. However it is possible that D_s might be small just because the scale r is large (making $A_{x,r}$ small will make $\mathcal{E}(A_{x,r})$ small due to σ-additivity). To counter this, we introduce the '*scaled* relative defect':

$$\hat{D}_s(A_{x,r}) := \begin{cases} \frac{D_s(A_{x,r})}{\lambda(A_{x,r})} & \lambda(A_{x,r}) > 0 \\ +\infty & \lambda(A_{x,r}) = 0 \end{cases} \tag{1.4}$$

The chief challenge in quantitative differentiation is to conclude that if this quantity is sufficiently small at sufficiently high scales, then $\alpha(f, A_{x,r})$ will be small (so-called coercivity).

It is also necessary to introduce a measure $\mathcal{C}_\lambda$ on collections of $A_{x,r}$ (that is, a measure on sets $B \subseteq \mathcal{A}$). If the set of scales r is countable, then a natural choice is

$$\mathcal{C}_\lambda(B) = \sum_{A \in B} \lambda(A) \tag{1.5}$$

Theorem 1 (Quantitative Differentiation via an Energy)**.** *If we can find an energy $\mathcal{E}$ satisfying the following properties:*

1. ***Boundedness:*** $\mathcal{E}(X, f) < +\infty$.
2. ***Monotonicity:*** $V_s(A) \searrow 0$ *as* $s \to \infty$ *for all* $A \in \mathcal{A}$.
3. ***Coercivity:*** *There exist* $S(\epsilon), \eta(\epsilon)$ *such that if* $s > S$ *and* $\hat{D}_s(A, f) < \eta$ *then* $\alpha(A, f) < \epsilon$.

then for all $\epsilon > 0$

$$\mathcal{C}_\lambda(\{A; \alpha(f, A_{x,r}) > \epsilon\}) < +\infty \tag{1.6}$$

Remark. The boundedness of $\mathcal{E}$ ensures that $V_s(A)$ is bounded for all s, A. The σ-additivity of $\mathcal{E}$ ensures that V_s is monotonic decreasing in this setting, however it does not guarantee convergence to 0. The monotonicity of V_s ensures that $\hat{D} \geq 0$.

Proof. We shall work our way backwards. As a consequence of coercivity, we can directly write inequality 1.6 as an inequality on the measure of $\hat{D}_r$, and then use Markov's inequality (M) to frame this in terms of an integral over all sets:

$$\begin{aligned}\mathcal{C}_\lambda(\{A; \alpha(f, A_{x,r}) > \epsilon\}) \leq& \mathcal{C}_\lambda(\{A; \hat{D}_s(A_{x,r}) > \eta\}) \\ &\overset{\text{(M)}}{\leq} \eta^{-1} \int \hat{D}_s(A) d\mathcal{C}_\lambda(A).\end{aligned} \tag{1.7}$$

Unpacking this integral, we get

$$\begin{aligned}\int \hat{D}_s(A) d\mathcal{C}_\lambda(A) =& \sum_{x,r} \hat{D}_s(A_{x,r}) \mathcal{C}_\lambda(\{A_{x,r}\}) = \sum_{x,r} D_s(A_{x,r}) \\ =& \sum_{x,r} V_0(A_{x,r}) - V_s(A_{x,r})\end{aligned} \tag{1.8}$$

Here we use the convention that $\infty \cdot 0 = 0$. Using eq. 1.2, this comes to

$$\begin{aligned}\int \hat{D}_s(A) d\mathcal{C}_\lambda(A) =& \sum_r V_r(X) - V_{s+r}(X) \\ =& \sum_{r=0}^{s} V_r(X) \leq sV_0(X).\end{aligned} \tag{1.9}$$

Monotonicity is used here to give convergence of the telescoping series, and boundedness ensures the finiteness: since $0 \in \mathcal{L}$, $V_0(X) \leq \mathcal{E}(X, f) < \infty$, completing the proof. □

The obvious corollary to quantitative differentiation is that f maybe be 'approximated' (in the sense of α) by a $\ell \in \mathcal{L}$ λ-a.e. (a form of Rademacher's theorem). We define this formally:

Definition 1 (α-Differentiability)**.** *We say f is α-differentiable at a point $x \in X$ if there exists a sequence of $A_n \in \mathcal{A}$ with positive measure decreasing to $\{x\} \cup N$ where $\lambda(N) = 0$, such that $\alpha(f, A_n) \to 0$.*

Note that this is dependent on $\mathcal{A}$. It is necessary to use properties of α to extend this to broader classes of sets (for example, open intervals containing x in [1]). The version of Rademacher is thus the following:

Corollary 1 (A.E. α-Differentiability)**.** *Assuming Theorem 1 is satisfied, then f is λ-a.e. α-differentiable.*

Proof. Suppose not. There then exists a set B of positive measure $b := \lambda(B) > 0$ such that for all $\epsilon > 0, x \in B$ there exists infinite $B_{x,n} \in \mathcal{A}$ containing x (indexed by n, which may not be r) such that $\alpha(f, B_{x,n}) > \epsilon$. Since this holds for all $x \in B$, we have $\bigcup_{x\in B} B_{x,n} \supseteq B$, hence $\lambda(\bigcup_{x\in B} B_{x,n}) \geq b$ for all n. This ensures $\mathcal{C}_\lambda(\{B_{x,n}; x \in B, n \in \mathbb{N}\}) = \infty$, contradicting the conclusion of Theorem 1. □

2 Relaxations

There are several relaxations one can make to theorem 1.

2.1 Energy The energy plays an illustrative, but non-essential role. It is sufficient to have a 'total defect' function V_s that monotonically converges to 0, satisfying eq. 1.2, and such that $V_0(X)$ is bounded (to replace the boundedness condition on $\mathcal{E}$). This is the method used in Application 3.2 below (in [2] as we saw in class).

This method also has the benefit of removing the need of the function space $\mathcal{F}$ to be a vector space (and indeed the existence of any $\mathcal{L}$ at all), allowing us to speak of quantitative differentiation of functions to manifolds for example.

2.2 Sets Another relaxation that would be helpful is to allow multiple $A_{x,r}$ within the same scale r to overlap. The challenge here is that this invalidates eq. 1.2 so long as there exists a $A \in \mathcal{A}$ such that $A \subseteq A_{x,r} \cap A_{y,r}$ for some x, y, r (as this set will be double counted on the lefthand sum of eq. 1.2). This may be sidestepped by using relative complements to construct a suitable $\mathcal{A}$ (we can assume this $\mathcal{A}$ is then countable if our space is Lindelhöf). We may then use the original, overlapping sets to demonstrate the monotonicity property (**2**), so long as $A \mapsto V_s(A)$ is monotone increasing (that is $A \subseteq B$ implies $V_s(A) \leq V_s(B)$ for all s).

My main motivation to look into overlapping sets was because I wanted to look at tubular neighbourhoods of α-differentiable/indifferentiable points as we did in class. The above paragraph shows that this is impossible within this framework, however the following conjecture shows that there still might be a way to evaluate the Minkowski dimension of the non-α-differentiable points.

Conjecture 1 (Minkowski Dimension)**.** *Suppose $A \mapsto \alpha(f, A)$ is monotone increasing. Then, if $X = \mathbb{R}^n$, with λ Lebesgue and $A_{x,r}$ boxes of side length $1/r$, then the Minkowski dimension of the non-differentiable set is then bounded from above by*

$$dim_M(\{x; f \text{ not } \alpha\text{-differentiable}\}) \leq n + \lim_{r\to\infty} \frac{\log\left(V_r(X)\right)}{\log(r)} \tag{2.1}$$

Sketch of Proof. This is inspired by the definition of Minkowski dimension as

$$\dim_{\mathrm{M}}(B) = \lim_{r\to\infty} \frac{\log\left(N_B(\frac{1}{r})\right)}{\log(r)} \tag{2.2}$$

where $N_B(\frac{1}{r})$ is the number of boxes of side length $\frac{1}{r}$ needed to cover the set B. For fixed ϵ, we can find the sets $A_{x,r} \in \mathcal{A}_r$ with $\alpha(f, A_{x,r}) > \epsilon$ to have measure of at most $\eta(\epsilon)^{-1}V_r(X)$ by the methods used in Theorem 1 and we can multiply by r^n to get an upper bound $N_B(1/r)$. Plugging this into eq. 2.2 and taking $r \to \infty$ gives us an upper bound on the Minkowski dimension of points that are not (ϵ, α)-differentiable (in a quantitative sense). Taking $\epsilon \to 0$ yields all points that are not α-differentiable, completing the proof (ϵ's influence vanishes when we take $r \to \infty$ so this step does not change the computation).

2.3 Measure We can see that the measure $\mathcal{C}_\lambda$ cancels out itself when we take the expectation of $\hat{D}$ with respect to it. Thus, so long as the energy doesn't depend on λ, there are many alternative measures μ we can use that would leave the result unchanged. The sole constraint is that our new measure. This means not only that $\mu(A) = 0$ exactly when $\lambda(A) = 0$ for $A \in \mathcal{A}$, but also that $\mu(A) \neq +\infty$ for any $A \in \mathcal{A}$.

This may prove advantageous if coercivity is easier to prove for $\hat{D}$ for some choice of μ than others.

3 Applications

3.1 Cheeger's Lipschitz case In the general formulation presented by Cheeger [1], quantitative differentiation is applied to the class of 1-lipschitz functions on the unit interval $\mathcal{F} = \mathrm{Lip}([0,1])$, with energy $\mathcal{E}(I, f) = \int_I (f')^2\, dx$, λ lebesgue and $\mathcal{A}$ the dyadic intervals. The model structure $\mathcal{L}$ is the class of linear functions, and α is the scaled L^∞-norm.

3.2 Quantitative Differentiation at a point The other example of this framework that we saw in class was to use the boundedness of Almgren's frequency to control the quantitative symmetry of u at a point x [2, Remark 2.10]. In this case the measure takes the form of a dirac mass δ_x, with $A_{x,r} = B_{\gamma^r}(x)$. V_r takes the form of Almgren's frequency ([2, Theorem 2.4] showing monotonicity; this is an 'energy-less' differentiation) and quantitative symmetry plays the role of α (with [2, Theorem 2.8] showing coercivity) and 0-symmetric functions forming $\mathcal{L}$.

3.3 Action Differentiation In the framework introduced here, scalability isn't an issue. To demonstrate this, we can make a simple generalization of Cheeger's Lipschitz quantitative differentiation, applied to C^1 functions $\gamma : [0,1] \to \mathbb{R}^d$ is to use an action energy of the form

$$\mathcal{E}(A_{x,r}, \gamma) = \int_{A_{x,r}} L(t, \gamma(t), \dot{\gamma}(t))\, dt \tag{3.1}$$

where $L(t, x, v) = |v|^2 + M(t, x)$ for some positive $C^1([0,1] \times \mathbb{R}^d)$ function $M(t, x)$. We modify the total defect to be

$$V_r(A_{x,s}) := \sum_{A_{y,s+r} \subseteq A_{x,r}} \mathcal{E}(A_{x,r}, \gamma) - \mathcal{E}(A_{x,r}, \ell_{x,r}) \tag{3.2}$$

where $\ell_{x,r}$ is the action minimizing path sharing end points with γ on $\partial A_{x,r}$. Monotonicity is clear. α is as in the Lipschitz case (scaled L^∞-norm); to get coercivity we can largely apply the same technique as in [3], except with the condition that we go to scales sufficiently small that $M(t, \gamma(t)) \approx \mathit{cnst}$ and $M(t, \ell_{x,r}(t)) \approx \mathit{cnst}$.

The conclusion drawn from this work is that any γ of finite action can be approximated on small scales by piecewise action-minimizing curves.

3.4 Further Ideas I Ran Out Of Time For One of the things I wanted to look at was higher (or lower) order quantitative differentiation. An example might be

$$\alpha(f, A) = \inf_{\ell \in \mathcal{L}} \frac{\|f - \ell\|}{|A|^n}$$

where $n \geq 2$ to measure the distance to higher order polynomials $\mathcal{L}$. Unfortunately, coercivity proved difficult to show (I think Cheeger's framework with the $\ell_{i,n}$ is more conducive to showing coercivity in these cases).

Once I made the Minkowski Dimension conjecture I also wanted to come up with an example for it. The example would have to be 0-th order to match the monotonicity condition, so an example might be if α was some quantitative measure of f being locally bound by a polynomial of a given degree. Unfortunately I had the idea for the conjecture too close to the deadline to dedicate sufficient thought to this.

References

[1] Jeff Cheeger. Quantitative differentiation: A general formulation. *Communications on Pure and Applied Mathematics*, 65(12):1641–1670. doi: 10.1002/cpa.21424. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21424`.

[2] J. Cheeger, A. Naber, and D. Valtorta. Critical sets of elliptic equations. *ArXiv e-prints*, July 2012.

[3] Jeff Cheeger. Quantitative differentiation. Geometric and Algebraic Structures in Mathematics, 2011. URL `https://www.math.stonybrook.edu/Videos/dennisfest/`.