

The Structure of Mathematical Expressions

An ARXIV Case Study

Deyan Ginev and Bruce R. Miller

National Institute of Standards and Technology

March 21, 2012



Contents

Contents	2
1 Introduction	3
1.1 Motivation	3
1.2 Related Resources	3
1.3 Experimental Setup	4
2 A Study of Mathematical Syntax	5
2.1 Basics	5
2.2 Discrete math	5
2.3 Continuous math	6
2.4 Other fields	11
3 Discussion	13
4 Conclusion	15

Chapter 1

Introduction

In this study, we survey the notational diversity of present-day mathematical expressions, in order to uncover their linguistic phenomena. A practical motivation for this study is to provide a foundation for determining the boundary between syntactic and semantic phenomena in said expressions, from the perspective of language modeling. The ultimate goal of this project is to construct a grammar of mathematical expressions, which captures all relevant syntactic properties established in this study, and allows for the semantic analysis necessary to model and observe the semantic relationships.

1.1 Motivation

We want to enable machine-reading of formulas, in order to provide a variety of user-assistance services, such as semantic search, text-to-speech synthesis, semantic interactions (definition lookup), as well as computer algebra support (“evaluate subexpressions on demand”) and ultimately computer verification (“does that proof step really hold?”).¹

EdN:1

1.2 Related Resources

Notation census, beginnings of study are in Deyan’s thesis, Naproche and FMathL have examples, but no real systematic study.²

EdN:2

¹EdNOTE: expand

²EdNOTE: expand

Train1	Differential Geometry http://arxmliv.kwarc.info/files/9609/dg-ga.9609012
Train2	Quantum Physics http://arxmliv.kwarc.info/files/0910/0910.5733/
Train3	High Energy Physics - Theory http://arxmliv.kwarc.info/files/9407/hep-th.9407125/
Train4	Commutative Algebra http://arxmliv.kwarc.info/files/0809/0809.4873/
Train5	Statistics Theory http://arxmliv.kwarc.info/files/0905/0905.1486/
Train6	General Relativity and Quantum Cosmology http://arxmliv.kwarc.info/files/0807/0807.2507/
Train7	Cosmology and Extragalactic Astrophysics http://arxmliv.kwarc.info/files/0908/0908.2548
Train8	Exactly Solvable and Integrable Systems http://arxmliv.kwarc.info/files/0905/0905.2033
Train9	Geometric Topology http://arxmliv.kwarc.info/files/0809/0809.4477
Train10	Algebraic Geometry http://arxmliv.kwarc.info/files/0704/0704.0537

Table 1.1: Sandbox of Ten Random ARXIV Papers from Diverse Scientific Subfields

1.3 Experimental Setup

The primary corpus on which we base this investigation is the Cornell pre-print archive “ARXIV”³, consisting of over 700,000 articles in 37 scientific subfields.

arXiv Sandbox

4

As a secondary resource, we we will also consult entry-level literature on high-school mathematics, in order to exhibit basic phenomena, as well as to demonstrate phenomena apriori known to the authors.⁵

³EDNOTE: cite here

⁴EDNOTE: Say that, on the ARXIV front, we first start with the train sandbox from Deyan’s thesis

⁵EDNOTE: Wikipedia? PEMDAS?

Chapter 2

A Study of Mathematical Syntax

2.1 Basics

Foundations

6 7 8

EdN:6

EdN:7

EdN:8

High School

9 10

EdN:9

EdN:10

2.2 Discrete math

Set Theoretic Notations

11 12

EdN:11

EdN:12

Logical Operators

13

EdN:13

⁶EdNOTE: arithmetic, grouping fences and equality

⁷EdNOTE: basic relations and orderings

⁸EdNOTE: arithmetic and algebraic sequences?

⁹EdNOTE: geometry here, otherwise a separate geometry subsection

¹⁰EdNOTE: trigonometry, complex and rational numbers

¹¹EdNOTE: elementhood, inclusions, set constructors, overloaded arith ops

¹²EdNOTE: also maps : domains \rightarrow codomains, xRy notations

¹³EdNOTE: classic logic, HOL, type theories

Combinatorics

EdN:14 14 15
EdN:15

Number Theory

EdN:16 16 17 18 19
EdN:17
EdN:18
EdN:19

Graph Theory

EdN:20 20 21 22
EdN:21
EdN:22

Algebra

EdN:23 23 24 25 26
EdN:24
EdN:25
EdN:26

Functions Theory

EdN:27 27

2.3 Continuous math

Calculus

EdN:28 28

¹⁴EdNOTE: Infinite sums

¹⁵EdNOTE: binomials, combinations, permutations,

¹⁶EdNOTE: modulo modifiers

¹⁷EdNOTE: tuples

¹⁸EdNOTE: divisibility notations $a \mid b$ and b/a

¹⁹EdNOTE: DLMF sneaky notations

²⁰EdNOTE: edge and vertex notations

²¹EdNOTE: incidence and adjacency notations

²²EdNOTE: Wiki is very nice: http://en.wikipedia.org/wiki/Glossary_of_graph_theory

²³EdNOTE: vectors

²⁴EdNOTE: maps and complements

²⁵EdNOTE: groups

²⁶EdNOTE: lattices

²⁷EdNOTE: talk about associativity of application and composition, “;” and “o” as notation variants, discuss complex examples

²⁸EdNOTE: differentials, integrals, limits, remember brownian motion integral notations!

Probability

29 30

EdN:29

EdN:30

Interval Notation and Arithmetic

31

EdN:31

Topology

32

EdN:32

Differential Geometry

Some intro text?

33

EdN:33

²⁹EDNOTE: Bayes formula with multiple denotations of P

³⁰EDNOTE: Various conditional and joint probability notations

³¹EDNOTE: introduce interval notations, then move to interval arithmetic

³²EDNOTE: manifold constructors and notations

³³EDNOTE: Complex named entity: "*U*(1) Chern-Simons gauge theory."

	Expression	Meaning	Syntax
1.	$(\mathcal{V}/\mathcal{Z}, k\omega)$ Discussion: [Train1]	symplectic torus	circumfix constructor
2.	\mathcal{Z} Discussion: [Train1]	self-dual lattice	atom abbreviation
3.	(\mathcal{V}, ω) Discussion: [Train1]	symplectic vector space	circumfix constructor
4.	$Lag(\mathcal{V})$ Discussion: [Train1]	Lagrangian Grassmannian	circumfix constructor
5.	$Lag_4(\mathcal{V})$ Discussion: [Train1]	4-fold covering space	applicative constructor
6.	\mathcal{M}_Σ Discussion: [Train1]	moduli space	scripted constructor
7.	Σ Discussion: [Train1]	Riemann surface	atom variable
8.	$H^1(\Sigma; \mathbb{R})$ Discussion: [Train1]	chomology space	applicative constructor
9.	$H^1(\Sigma; \mathbb{R})/H^1(\Sigma; \mathbb{Z})$ Discussion: [Train1]	torus	applicative constructor
10.	(M, ω) Discussion: [Train1]	symplectic manifold	circumfix constructor
11.	$f \in \mathcal{C}^\infty(M)$ Discussion: [Train1]	smooth function	modified atom
12.	X_f Discussion: [Train1]	field	scripted constructor
13.	\lrcorner Discussion: Formed via <code>\mathop</code> in <code>T_EX</code> , [Train1]	interior product	complex infix operator
14.	$[\omega] \in H^2(M; \mathbb{R})$ Discussion: [Train1]	cohomology class	modified complex object
15.	(\cdot, \cdot) Discussion: [Train1]	hermitian metric	notation pattern

Table 2.1: Differential Geometry Notations, Part 1

	Expression	Meaning	Syntax
16.	$-2\pi i \omega$ Discussion: [Train1]	curvature of ∇	arithmetic expression
17.	(\mathcal{L}, ∇) Discussion: [Train1]	prequantum line bundle	circumfix constructor
18.	$U \subset M$ Discussion: [Train1]	open subset	modified atom
19.	$\mathcal{L} _U$ Discussion: postfix restriction via “ $ _U$ ”, [Train1]	restricted line bundle	modified atom
20.	$s \in \Gamma(U; \mathcal{L})$ Discussion: [Train1]	nonzero section	modified atom
21.	$\nabla s = -2\pi i \theta s$ Discussion: [Train1]	equation	relation
22.	$\omega _U = d\theta$ Discussion: [Train1]	equation	relation
23.	$T_x M$ Discussion: invisible infix bundle-forming operator, [Train1]	bundle	applicative constructor
24.	$\omega _{\mathcal{P}_x} \equiv 0$ Discussion: [Train1]	equivalence	relation
25.	$\dim \mathcal{P}_x = \frac{1}{2} \dim T_x M$ Discussion: \dim has lower precedence than invisible bundle-formation, [Train1]	equality	relation
26.	$[X, Y] \in \mathcal{X}_{\mathcal{P}}(M)$ Discussion: used as verb phrase in sentence, [Train1]	commutator is in set	relation
27.	$\nabla^{\mathcal{P}}$ Discussion: big op?, [Train1]	covariant differentiation	scripted prefix op
28.	$\nabla^{\mathcal{P}} : \mathcal{X}_{\mathcal{P}}(M) \times \mathcal{X}_{\mathcal{P}}(M) \longrightarrow \mathcal{X}_{\mathcal{P}}(M)$ $(X, Y) \longmapsto \nabla_X^{\mathcal{P}} Y,$ Discussion: alignment splits type statement, trailing comma [Train1]	domain specification	typing modifier
29.	$(\nabla_X^{\mathcal{P}} Y) \lrcorner \omega = X \lrcorner d(Y \lrcorner \omega).$ Discussion: trailing dot, [Train1]	definitional assignment	relation
30.	$\Pi_{\mathcal{P}} : M \rightarrow M/\mathcal{P}$ Discussion: [Train1]	canonical projection map	typed modifier

Table 2.2: Differential Geometry Notations, Part 2

	Expression	Meaning	Syntax
31.	T^g	g -dimensional torus	complex object
	Discussion: script means dimensionality[Train1]		
32.			
	Discussion: [Train1]		
33.			
	Discussion: [Train1]		
34.			
	Discussion: [Train1]		
35.			
	Discussion: [Train1]		
36.			
	Discussion: [Train1]		
37.			
	Discussion: [Train1]		
38.			
	Discussion: [Train1]		
39.			
	Discussion: [Train1]		
40.			
	Discussion: [Train1]		
41.			
	Discussion: [Train1]		
42.			
	Discussion: [Train1]		
43.			
	Discussion: [Train1]		
44.			
	Discussion: [Train1]		
45.			
	Discussion: [Train1]		

Table 2.3: Differential Geometry Notations, Part 3

2.4 Other fields

Quantum Physics

³⁴ ³⁵ :

EdN:34
EdN:35

³⁴EDNOTE: Bra-ket notation

³⁵EDNOTE: computer science, biology, chemistry...

Chapter 3

Discussion

Chapter 4

Conclusion
