The Structure of Mathematical Expressions

An arXiv Case Study

Deyan Ginev and Bruce R. Miller

National Institute of Standards and Techonology

March 21, 2012

Contents

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

[1] EdNote: expand
[2] EdNote: 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 Cornel pre-print
EdN:3 archive "arXiv"[3], consisting of over 700,000 articles in 37 scientific subfields.

arXiv Sandbox

EdN:4 [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
EdN:5 phenomena apriori known to the authors.[5]

[3] EdNote: cite here

[4] EdNote: Say that, on the arXiv front, we first start with the train sandbox from Deyan's thesis

[5] EdNote: 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

[6] EdNote: arithmetic, grouping fences and equality
[7] EdNote: basic relations and orderings
[8] EdNote: arithmetic and algebraic sequences?
[9] EdNote: geometry here, otherwise a separate geometry subsection
[10] EdNote: trigonometry, complex and rational numbers
[11] EdNote: elementhood, inclusions, set constructors, overloaded arith ops
[12] EdNote: also maps : domains -¿ codomains, xRy notations
[13] EdNote: 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]

[14] EdNote: Infinite sums
[15] EdNote: binomials, combinations, permutations,
[16] EdNote: modulo modifiers
[17] EdNote: tuples
[18] EdNote: divisibility notations $a \mid b$ and b/a
[19] EdNote: DLMF sneaky notations
[20] EdNote: edge and vertex notations
[21] EdNote: incidence and adjacency notations
[22] EdNote: Wiki is very nice: http://en.wikipedia.org/wiki/Glossary_of_graph_theory
[23] EdNote: vectors
[24] EdNote: maps and complements
[25] EdNote: groups
[26] EdNote: lattices
[27] EdNote: talk about associativity of application and composition, ";" and "$\circ$" as notation variants, discuss complex examples
[28] EdNote: 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

[29] EdNote: Bayes formula with multiple denotations of P
[30] EdNote: Various conditional and joint probability notations
[31] EdNote: introduce interval notations, then move to interval arithmetic
[32] EdNote: manifold constructors and notations
[33] EdNote: Complex named enttity: "$U(1)$ Chern-Simons gauge theory."

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

Table 2.1: Differential Geometry Notations, Part 1

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

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.	$q_1, \ldots, q_g$	coordinate functions	enumerative sequence
	Discussion: [Train1]		
33.	$X_{q_1}, \ldots, X_{q_g}$	Hamiltonian vector fields	enumerative sequence
	Discussion: [Train1]		
34.	$q_1 \circ \Pi_{\mathcal{P}}, \ldots, q_g \circ \Pi_{\mathcal{P}}$	functions	enumerative sequence
	Discussion: sequence elements are applicative objects, [Train1]		
35.	$\gamma_1(\Lambda), \ldots, \gamma_g(\Lambda)$	basis for a homology group	enumerative sequence
	Discussion: [Train1]		
36.	$j_i(y) = \int\limits_{\gamma_i(\Lambda)} \theta,$ where $y = \Pi_{\mathcal{P}}(\Lambda)\,,$	definitional assignment	relation
	Discussion: integral has no binder, nat. lang. modifier, punctuation, [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

[34] [35]

EdN:34
EdN:35

2.4 Other fields

Quantum Physics

[36] [37] ⋮

EdN:36
EdN:37

[34] EdNote: Scripts give you new names or new objects
[35] EdNote: Prime scripts can be used for both naming and operating
[36] EdNote: Bra-ket notation
[37] EdNote: computer science, biology, chemistry...

Chapter 3

Discussion

Chapter 4

Conclusion