Rochester Institute of Technology
RIT Digital Institutional Repository

Theses

5-2021

Spatial vs. Graph-Based Formula Retrieval

Robin Avenoso
rxa8448@rit.edu

Follow this and additional works at: https://repository.rit.edu/theses

Recommended Citation
Avenoso, Robin, "Spatial vs. Graph-Based Formula Retrieval" (2021). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

Spatial vs. Graph-Based Formula Retrieval

by

Robin Avenoso

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science
in Computing Sciences

B. Thomas Golisano College of Computing and
Information Sciences
Rochester Institute of Technology

May 2021

MS IN COMPUTING AND INFORMATION SCIENCES
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

MS DEGREE THESIS

The MS degree thesis of Robin Avenoso
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
MS degree in Computing and Information Sciences

Richard Zanibbi, Chair

Zack Butler, Reader

Edith Hemaspaandra, Observer

Date

Spatial vs. Graph-Based Formula Retrieval

by

Robin Avenoso

Submitted to the
B. Thomas Golisano College of Computing and Information Sciences
Department of Computer Science
in partial fulfillment of the requirements for the
Master of Science Degree
at the Rochester Institute of Technology

Abstract

Recently math formula search engines have become a useful tool for novice users learning a new topic. While systems exist already with the ability to do formula retrieval, they rely on prefix matching and typed query entries. This can be an obstacle for novice users who are not proficient with languages used to express formulas such as LaTeX, or do not remember the left end of a formula, or wish to match formulas at multiple locations (e.g., using '$\int \quad dx$' as a query). We generalize a one dimensional spatial encoding for word spotting in handwritten document images, the Pyramidal Histogram of Characters or PHOC, to obtain the two-dimensional XY-PHOC providing robust spatial embeddings with modest storage requirements, and without requiring costly operations used to generate graphs. The spatial representation captures the relative position of symbols without needing to store explicit edges between symbols. Our spatial representation is able to match queries that are disjoint subgraphs within indexed formulas. Existing graph and tree-based formula retrieval models are not designed to handle disjoint graphs, and relationships may be added to a query that do not exist in the final formula, making it less similar for matching. XY-PHOC embeddings provide a simple spatial embedding providing competitive results in formula similarity search and autocompletion, and supports queries comprised of symbols in two dimensions, without the need to form a connected graph for search.

Acknowledgments

First I would like to give thanks to my advisor Dr.Richard Zanibbi for all the help in my academic and professional pursuits. From working with me on my independent study my second year of undergrad to welcoming me in to the lab to do my thesis. These years working together have been a learning experience and I have grown in many ways with your guidance. Also I would like to thank my thesis committee Dr.Zack Butler and Dr.Edith Hemaspaandra. Both of whom, for serving on my committee and for teaching some of my favorite classes I have taken during my time at Rochester Institute of Technology. I will never forget the great experience we had in the classroom. Thank you to all the members of the Document and Pattern Recognition Lab at Rochester Institute of Technology. The weekly meetings we have during the year have lead to great improvements in my work and source of relief during some of the toughest times. A special thanks to Behrooz Mansouri for your help in getting my system running on the data as well as offering insight into the impact on the work present in this thesis. A very special thank you to Yancarlos Diaz, my friend and partner in crime. This year we have spent days working together in zoom calls, making sure we get our work done, and made sure we stayed sane through it all. Thank you for being a great friend and someone I could always talk to. Additionally a big thank you to my friends who supported me through everything, hanging out, relaxing and even helping proofread this document. I would like to thank my parents and siblings for all your years of love and support. And finally thank you to my Girlfriend Katie, who helped me stay strong through the last stretch.

Contents

Contents **v**

List of Figures **vii**

List of Tables **x**

1 Introduction **1**
1.1 Contributions 4
1.2 Summary 4

2 Related Work **5**
2.1 Input Data 6
2.2 Output Representations 9
2.3 Similarity Measures 13
2.4 Inverted Index 15
2.5 Summary 16

3 XY-PHOC Retrieval Model **17**
3.1 Motivation 17
3.2 Embedding 19
3.2.1 Indexing 20
3.2.2 Retrieval 21
3.3 Summary 22

4 Experimental Design **25**
4.1 Baselines 25
4.1.1 Bag-of-Symbols 26

4.1.2 PHOC 26
4.1.3 Tangent 28
4.1.4 Tangent-V 28
4.1.5 Tangent-S 30
4.1.6 Tangent-CFTED 31
4.2 Datasets 32
4.3 Formula Search and Autocompletion Experiments 32
4.3.1 Formula Similarity Search 34
4.3.2 Formula Autocompletion 35
4.4 Summary 38

5 Results **39**
5.1 Formula Similarity Search 39
5.2 Formula Autocompletion 43
5.3 Summary 51

6 Conclusion **52**
6.1 Summary of Findings 52
6.2 Future Work 53

Bibliography **55**

Appendices **59**

A Autocompletion Benchmark Query Set **60**

B Rsaved Graphs **74**

C Mean Reciprocal Rank Graphs **91**

List of Figures

1.1 Symbol Layout Tree for the equation $2y^8 = \sqrt{x}$ where the symbols are represented in the nodes and the relationships are labeled on the edges between symbols 1
1.2 Comparison between 2 use cases for formula autocompletion using the Pythagorean Theorem. (a) Symbols in circles represent a symbols placed in space and the gray symbols represent autocomplete results. 3

2.1 Pipeline illustration of the flow of data through these systems . 5
2.2 PHOC representation with levels 1, 2, and 3 for the word "place." The right side represents the alphabet as a vector were a 1 indicates the presence of the character, the green lines indicate a split and the level number corresponds to the number of regions used to represent that level, where only characters present in that split are indicated with a 1. 7
2.3 Visualization of the extract of a PHOC from a given math formula at levels 1,2 and 3. Similar to Figure 2.2 the right side represents a binary vector indicating the presence of a symbol where the vector will include all valid math symbols recognized by the system. 9
2.4 Example of original image **(a)** and normalized projection profile **(b)** [19] 11
2.5 A LoS Graph [3] 12

3.1 PHOC Representation with the addition of horizontal splits to capture the relative vertical positions of symbols, these splits are represented by Levels $2'$ and $3'$ where the number of splits is the same as on 2 and 3 but the line is along the horizontal axis instead of vertical axis. 18
3.2 29-bit XY-PHOC embedding for a query formula. Where the first bit represents Level 1 and the second two bits represents Level 2 21
3.3 Computing the match score between a query formula and a candidate formula. The numerator represents the total number of common bits between the query and the candidate. The denominator represents the normalization constant for this candidate formula. 23

4.1 Example of inputting the Equation $\int_0^\infty \frac{\sin(x)}{x}\,dx$ with 3 symbols when the order entered is a) Left-to-Right, b) Right-to-left, c) Alternating from Left-Right from the Outside-in, d) Alternating Left-Right from the Middle-Out. 36

5.1 `rsaved` Values By Target Formula Size Using a) Left-to-Right b) Right-to-Left c) Outside-in d) Middle-out Input Order for the Centroid Membership condition. The Quicker Convergence to 1 and lower number of outliers in the Outside-in condition shows how its scores across the board are better. 47
5.2 Mean Reciprocal Rank (MRR) of all 4 input orders by the number of symbols entered and the % of target formula symbols entered for the Centroid Membership condition. The Quicker Convergence to 1 in the Outside-in condition shows how its scores across the board are better. 48

B.1 Level 1 (BoS) 75
B.2 Level 2 76
B.3 2' 77
B.4 Level 3 78
B.5 Level 3' 79
B.6 Level 4 80
B.7 Level 4' 81

B.8 Level 5 82
B.9 5' 83
B.10 Levels 1-2 84
B.11 Levels 1-3 85
B.12 Levels 1-4 86
B.13 PHOC Embedding 87
B.14 XY-PHOC Embedding 88
B.15 Top-Left membership condition 89
B.16 Vertical Center membership condition 90

C.1 Level 1 (BoS) 91
C.2 Level 2. The Quicker Convergence to 1 in the Outside-in condition shows how its scores across the board are better 92
C.3 Level 2' 92
C.4 Level 3 92
C.5 Level 3' 93
C.6 Level 4 93
C.7 Level 4' 93
C.8 Level 5 94
C.9 Level 5' 94
C.10 Levels 1-2 94
C.11 Levels 1-3 95
C.12 Levels 1-4 95
C.13 PHOC Embedding 95
C.14 XY-PHOC Embedding 96
C.15 Top-Left membership condition 96
C.16 Vertical Center membership condition 96

List of Tables

4.1 Bag-of-Symbols Conjunctive Query in the Annotated ARQ-Math Collection for Equations 4.1 and 4.2 27
4.2 PHOC Conjunctive Query in the Annotated ARQMath Collection for Equations 4.1 and 4.2 29

5.1 Single Level of XY-PHOC using Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection. Prime ($'$) levels are split vertically. 40
5.2 Increase Depth of XY-PHOC on Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection. The Ranges Include Both the Horizontal and Vertical levels (e.g 1-2 includes 1, 2, and $2'$) 41
5.3 Different Membership conditions using Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection. 41
5.4 ARQMath Task 2 Formula Similarity Search Benchmark 43
5.5 Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With a single level where the number represents the number of sections in that level. Prime ($'$) levels are split vertically. 44
5.6 Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With Increasing Depth Represent As A Range Of Levels Included. The Ranges Include Both the Horizontal and Vertical levels (e.g 1-2 includes 1, 2, and $2'$) 45

5.7 Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With Different Membership Conditions. 45
5.8 Autocompletion Benchmark on ARQMath Annotated. 46
5.9 XY-PHOC Conjunctive Query in the Annotated ARQMath Collection, Where the Target is Returned at Rank 1 49
5.10 XY-PHOC Conjunctive Query in the Annotated ARQMath Collection, Where the Target is Returned at Rank 5 50
5.11 Comparing Conjunctive Queries on Annotated ARQMath and Full ARQMath collection of the first 19 queries of the Autocomplete Benchmark. 50
5.12 XY-PHOC Conjunctive Query on the Annotated ARQMath Collection and Full ARQMath Collection 51

Chapter 1

Introduction

Query autocompletion provides users with suggestions of possible search terms based on the query. It helps users to formulate their queries correctly when they have a need for information but no clear way to fully express it [20]. A user will prefer recognition over recall [11], as it takes less mental work to recognize the desired search query than to recall the entire query. Giving the user possible suggestions can speed up their workflow and reduce the amount of effort needed. Autocompletion systems are designed to reduce the user's effort to increase their satisfaction with the search engine [21].

While text query autocompletion is a well-researched area there is a lack of literature for math formulas [20]. This has caused an increasing gap as math

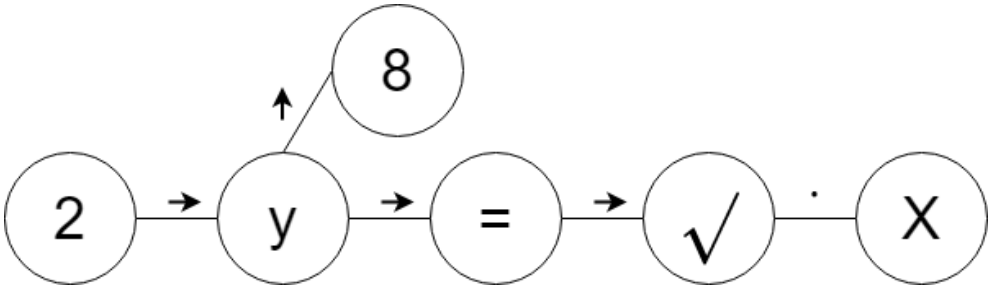


Figure 1.1: Symbol Layout Tree for the equation $2y^8 = \sqrt{x}$ where the symbols are represented in the nodes and the relationships are labeled on the edges between symbols

search engines are developed but lack the extensive autocompletion provided in text search engines. While there exist math search engines with autocompletion, like Symbolab* and WolframAlpha†, these systems are closed source and rely on prefix matching [20], which restricted to being input from the root, leftmost symbol, of a Symbol Layout Tree (SLT). A SLT is a commonly used representation for math formulas that captures the branching structure of the formula where the nodes are the symbols and the edges between the nodes represent the relationship between symbols (Next, Above, Pre-above, Below, Pre-Below, Over, Under, Within and Element).

An example of an SLT is given in Figure 1.1 where the arrows indicate the relationships of Next and Above, and $\cdot$ represents an element of. e.g in $\surd \cdot 8$ means that 8 is an element of $\surd$. Additional autocompletion methods have been proposed on LaTeX in [20]. A Tree-based system Tangent-CFT [16] mentions that it could be suitable for autocompletion. LaTeX based systems are limited in that there are many ways in LaTeX to generate the same expression and rendering these to a SLT requires that the symbols given do not form a disconnected subgraph like that seen in Equation.1.1, below, which shows an example of a rendered LaTeX formula that can not be made into a SLT.

$$^{2} +\ ^{2} =\ ^{2} \tag{1.1}$$

This equation can not be made into a useful SLT because it is a disconnected subgraph of a full equation. The missing symbols remove missing nodes and edges required to accurately represent the formula as a tree. Tree-based systems are not designed to account for disconnected graphs and will not be able to provide formulas of a similar representation.

We propose a new system that will use spatial information from a math formula query to perform similarity search and autocompletion. Techniques from word spotting will be applied as a measure of visual similarity between the query and the index. Word Spotting involves the exact matching of a query word in scanned documents, starting from the need to provide searchable archives of historical documents. The goal of this thesis is to study the behavior of spatial vs. graph-based autocompletion with the hope of finding a robust representation that can be used across modalities (e.g., handwritten

*https://www.symbolab.com

†https://www.wolframalpha.com/

(a) Visual query (b) WolframAlpha query

Figure 1.2: Comparison between 2 use cases for formula autocompletion using the Pythagorean Theorem. (a) Symbols in circles represent a symbols placed in space and the gray symbols represent autocomplete results.

and typeset).

Thesis Statement A relative spatial embedding for math formulas can provide competitive results for similarity search and autocomplete tasks, while using less embedded relational information than graph-based systems.

This proposed system for math formula retrieval is designed to capture use cases missed by other system. The system is useful in a math search engines for nonexperts, who may not be well versed in LaTeX syntax. This system can directly use symbols in space rather, allowing a user to drag and drop symbols onto a canvas or select symbols from a formula to perform search with.

In Figure 1.2 there is an input query using symbols on a canvas of $^2+\ ^2=\ ^2$ which could be a valid search for a novice user learning about the Pythagorean Theorem. In this case, they remember that there are several symbols raised to the second power but can not remember the variables used. This same query provides no autocompletion results from WolframAlpha since it will not match the prefix. That query is also typed out requiring that the user knows that the ^ symbol is used for superscript. This can be a hurdle for new users and require more effort than simply writing the query by hand.

There are possible situations when a valid Symbol Layout Tree can not be generated when the symbols are sparsely present on the canvas used for generating queries, seen in Equation 1.1. In these situations a user should still be able to get a valid autocompletion result with this system as it will be matching on the visual similarity, symbols, and their positions in space, to find matches.

1.1 Contributions

This work makes the following contributions:

1. The XY-PHOC encoding and associated retrieval model, which supports math formula search engine using any input order for symbols.
2. Competitive results on the ARQMath formula search task, using a simple spatial retrieval model.
3. A new metric *rsaved* to measure autocompletion results.
4. An expansion on word spotting techniques that provides an embedding capable of representing two dimensional symbol layouts.

1.2 Summary

Chapter 2 summarizes the previous work done in the fields of Word Spotting, Math Spotting, and Math Formula retrieval. The work has been broken down based on the input data, the output representation, the scoring function and discussion on inverted indices.

Chapter 3 details the proposed XY-PHOC retrieval model. The work is broken down based on our motivations, the a discussion on the embedding, techniques used to index the collection and the way retrieval is performed on the index.

Chapter 4 presents the experiments used to evaluate the system. Several baseline systems are detailed, along with the ARQMath data that will be tested on. Experiments are run for both formula similarity search and autocompletion tasks to better understand which PHOC configurations work best for each task. These experiments consider different amounts of data in embeddings as well as different ways to embed symbols in a formula.

Chapter 5 illustrates and discusses the results of running our experiments. The results are broken up based on the similarity search results followed by the autocomplete results. We identify the best model for each task and elaborate on the strengths and weakness of the model for each task.

Chapter 6 summarizes the contributions and outcomes of this work and provides recommendation on future work.

Chapter 2

Related Work

The ideas behind spatial retrieval presented in this paper come from the field of exact matching of math formulas, called math spotting. Math spotting is an adaptation of word spotting, the exact matching of words in documents and scenes [19]. Drawing from inspiration of how handwritten and typeset word spotting has been done, we can build on these systems in order to handle the task of similarity search and autocompletion of math queries. The main differences between these techniques are the input data, the output representation, and ranking metrics. In this chapter, we will look at some of the approaches taken by previous systems based on their inputs, outputs, and rankings. This organization of the different types of systems is inspired by the survey on word spotting by Giotis et al. [10]. The way the data moves through these systems is illustrated in Figure 2.1. In order for any retrieval system to function at scale there is a need for information retrieval optimizations, such as an inverted index, which is used in many of the systems we looked at.

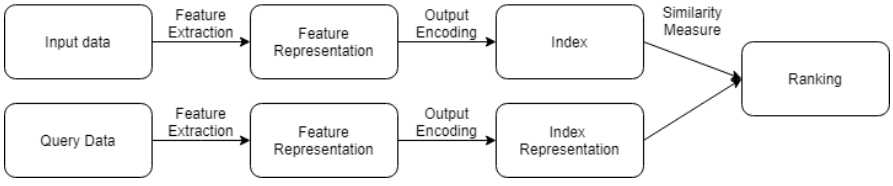


Figure 2.1: Pipeline illustration of the flow of data through these systems

2.1 Input Data

Several different features have been used in approaches to word spotting. Some previously common features include scale-invariant feature transform, the extraction of local key points in the image, and histogram of gradients, which are used to describe an image as a count of gradient orientations at localized portions of the image, that extract information about local areas in the image [25]. Other features directly use pixel values and positions. These systems do not rely on a recognition system and operate directly on the image data. Another set of features used in previous work are character primitives. These systems rely on an initial character recognition system and operate on the recognized characters directly. These approaches are limited by the accuracy of their recognizer but can work with limited symbol classes combining visually similar classes to improve recall [5]. The features used will impact the visual information that is present in the output embedding.

Scale-Invariant Feature Transform, SIFT features are similar to another set of features, Histogram of gradients, HoG. HoG features were utilized in [25], where they compared against SIFT features, this was interesting as the local features captured by HoG were better for determining word matches then SIFT features which were more popular at the time. Both of these methods rely on capturing visual features about the image from the pixel values but this is not the only way pixel values are used.

Pixel values are used directly as features in several approaches as well. Column wise pixel projects, histograms representing the intensity values in the image column, are used to create word profiles in [19]. This requires word-level segmentation as the pixel projections are done across the entire word. Column wise pixel projections are used on math data in [27] where scanned document images were segmented with X-Y cutting as described in [18] and than extracted the profiles of the formula. X-Y cutting is the process of cutting along gaps in the pixel projects. Cuts are alternated between vertical and horizontal directions made at the largest projection gap. In the XY-trees that are created by the X-Y cuts each symbol is only represented once in the leaves of the tree, giving no redundancy in the representation which can help improve robustness.

Pixel values are directly used in Convolutional Neural Network (CNN) approaches, as well, where the input is the entire word image. SIFT features

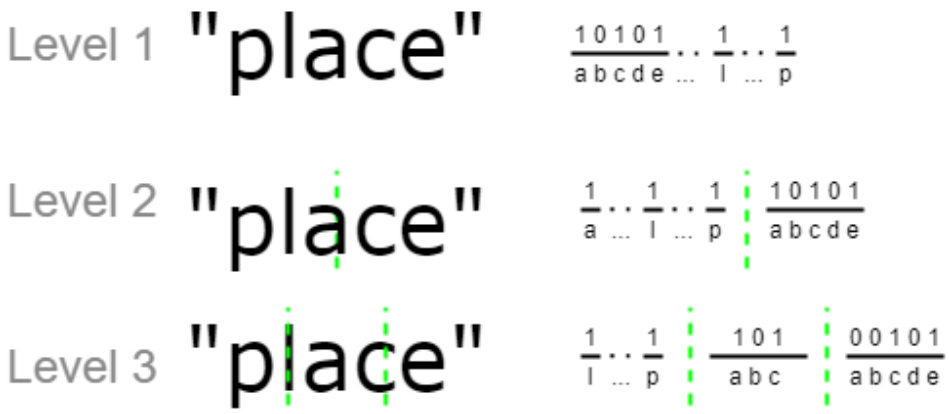


Figure 2.2: PHOC representation with levels 1, 2, and 3 for the word "place." The right side represents the alphabet as a vector were a 1 indicates the presence of the character, the green lines indicate a split and the level number corresponds to the number of regions used to represent that level, where only characters present in that split are indicated with a 1.

have been compared to the new CNN techniques in [28] where it was shown how while historically SIFT features were useful, state-of-the-art instance retrieval, which refers to the task of returning images of the same class of the query, systems are using CNNs. An example of instance retrieval is using a picture of a house as a query and similar house images are return.

An example of a system that uses a CNN in this way is PhocNet [22]. Pyramidal Histogram of Characters (PHOC) is a binary vector indicating the presence of a character at N horizontal regions. Starting at level 1 with one region, level 2 split in half and a further split at each level. For each split, the binary vectors indicating the presence of a symbol are concatenated together. These systems have been used for exact matching, which inspired it for us in autocompletion, which normally requires identification of the exact match-upon while also requiring more symbols to appear in the match than in the query. An example of this representation for a word is given in Figure 2.2. A similar system to PhocNet was proposed in [17] where a Pyramid of bidirectional character sequences (PBCS) are generated from a CNN. The differences in these representations will be discussed further in Output Representations

Section 2.2. Both of these previous CNN models required pre-segmenting the words, this is not the case in the R-PHOC network proposed in [9] where a ROI, Region of Interest, pooling layer. With the inclusion of the ROI pooling layer, the segmentation is done as a part of the CNN.

Binary pixel values can be used as connected components. Connected components represent all of the pixels that are touching above a certain threshold, in a binary image these would be the black pixels. Connected components are utilized in the graph-based keyword spotting system presented in [1]. Then each connected component represent a character in the image. The graphs used in [1] work similarly to the strokes of online handwritten where a series of x-y points are connected together in space to represent characters, where several algorithms were tried to generate key points and an order in which to connect the points.

Character primitives can also be used as features. In these systems, an external recognizer is used to get recognition results that are then used as features. In [7] an HMM-based recognizer is trained per symbol class on a combination of geometrical features and global features. This Character primitive models are chained together in a sequence that is used in another HMM designed to do word matching. This technique is capable of partial matching within a text-line, this is different than the tree structure of math formulas, which would require the HMM to be able to traverse a sparse subgraph representation of the tree. Another system using character primitives is presented in [2] where a CNN is used to locate words in a scene and character recognition is used to generate a PHOC embedding.

Two systems that use character primitive and are designed for math formulas are Tangent-V [5] and Tangent-S [4]. The Tangent-V system is designed to either work with born digital PDFs in which it can use the exact symbols labels and bounding boxes or uses a pre-trained classifier for labeling connected components which are input to the system. Tangent-S system uses a symbolic or LaTeX input.

In born digital PDFs, symbols labels and bounding boxes are used by Tangent-V [5]. Here an external system SymbolScrapper [13] which takes a born-digital PDF and extracts the exact bounding boxes for the symbols. The symbol classes are known as they are contained in the pdf data and the bounding boxes are generated from information about font and font size. Using the known labels and bounding boxes for born digital documents allows you to full

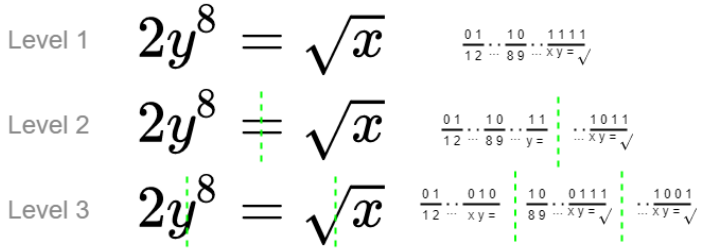


Figure 2.3: Visualization of the extract of a PHOC from a given math formula at levels 1,2 and 3. Similar to Figure 2.2 the right side represents a binary vector indicating the presence of a symbol where the vector will include all valid math symbols recognized by the system.

test the output representation without the variance of a recognizer. In the next Section 2.2 we will explain more about the different output representations.

2.2 Output Representations

Different systems return different output representations of the data used for retrieval. These output representations can be classified into three different groups. There are fixed-length vector representations, where any length input will map to a fixed-length output. This is useful as all the outputs can be directly compared for ranking. Another representation is the variable-length vector which can also be seen as a sequence. This representation has a different length based on the size of the input. In variable-length vectors two vectors of different lengths can still be representing the same word. Graphs are an output representation that is able to represent the relationship between symbols. Graphs are useful in math-related tasks since math formulas are tree structures, where there are several different relationships between symbols as seen in the SLT Figure 1.1.

Systems using SIFT return a bag-of-visual words. This is a vector with a set number of key points and key point descriptors (e.g., orientation) that make up the image. The fixed-length vector represents the count of the gra-

dients at different orientations. Bag-of-Symbols (BoS) is a also a fixed length embedding which captures which symbols are present. PHOCs, and other similar embedding methods such as Spatial Pyramid of Characters (SPOC), an expansion of PHOC that counts the number of symbols instead of using a binary vector, and Discrete Cosine Transform of Words (DCToW), where the symbols are represented as a one-hot vector over the alphabet and then packed sequentially into a matrix, then discrete cosine transform is applied per row and only the highest three values per row are concatenated into a vector to form the DCToW descriptor, which is a minimal representation of the sequence of the word. Both SPOC and DCToW are fixed-length vectors [23]. The vectors represent the symbols and their relative spatial relationship to one another. Embedding spatial information can be important for reducing confusion in the model, this is expanded on in later examples in Section 4.1.2.

There is an important difference between words and math formulas and that is the layout structure they represent. A word can be thought of as a sequence that moves from left to right. Math, on the other hand, is a tree-like structure so additional information may need to be packed into the output to correctly match math formulas with these types of representation. An example of a PHOC on a math formula is given in Figure 2.3 where at the first level there is no split and is represented by a binary vector across all the valid math symbols, in this figure a small list of symbols is used for illustrative purposes. Level 2 has a single split along the middle of the formula splitting it into the left and right half, the encoding at this level has two binary vectors one for each area separated by the split. Any symbols that are on the split points appear in both regions. This can continue for any number of spits in this example we went to level 3 and after this process, all of the vectors are concatenated together into one vector. Similar to PHOC, PBCS [17] represents the word in a hierarchical structure but with some differences. One difference is that it encodes the characters from in the forwards and backward direction, this can add additional structure to the representation, which could be helpful with the commutative nature of the formulas. Unlike the PHOC and SPOC representations that have a vector the length of the alphabet, the PBCS representation uses a predetermined fixed-length output vector of the longest word the system can process. Each index in the vector represents the class of symbol that appears at this position in the sequence, and filler characters are used if the word is shorter than the max length. This

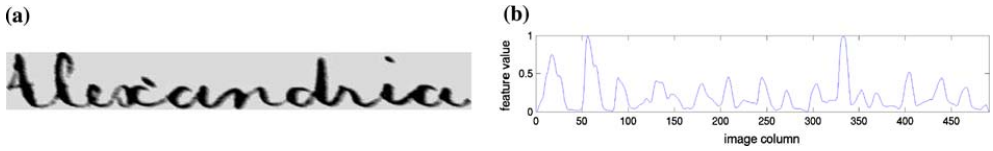


Figure 2.4: Example of original image **(a)** and normalized projection profile **(b)** [19]

does reduce the size of the embedding but also sets a max length on the number of symbols which would need to be the length of the largest formula in the index.

Variable-length vectors are another output that is returned by systems that perform feature extraction across the entire variable-sized word images. HoG features like the ones used in [25] are variable-length as the histogram representations are collected across the entire image, counting the features at each position. The output of the system in [19] is a word profile from the image. This output is directly proportional to the size of the input image. There are two word profiles used in this system, an upper and lower projection which represent the tops and bottoms of each character moving across the image. Figure 2.4 shows what the normalized word profile for the original image would be. This is the same figure presented in [19]. These profiles would not work directly in math in cases where there are symbols that overlap other symbols. In cases where there are fractions the upper profile and lower profiles would be describing two different sub-formulas. The use of an additional pixel projection along the rows would give additional information for vertical layout [27] but these projections would have large gaps still in case of autocompletion where the goal is to match exactly the characters present with more characters in the formula.

Graph-based outputs are also used to represent the output in previous systems. The connected components used in [1] are extracted into several graph representations. The graphs represent nodes and edges that could be used to draw the handwritten text. These graphs are similar to the data capture when recording online handwriting, capturing the x-y points, and the sequence in which to connect them. Graphs are a strong representation for math formulas because you can capture the relationship between symbols and not just their relative position to each other captured in spatial representations.

A line-of-sight graph is the output representation used for formula search of the Tangent-V system [5]. The line of sight graph is used directly as well as symbol pair tuples that are made based on the symbols that can "see", a line can be drawn from the convex hull of one symbol to another symbols with out crossing the convex hull of an additional symbol. An example of the Line-of-Sight graph is presented in Figure 2.5 where the symbols are isolated and line of sight lines connect the center of symbols that can see each other. A more restricted graph representation is the SLT. This output structure represents the tree structure that makes up a math formula with relationships between symbols represented by edge labels (Next, Above, Pre-above, Below, Pre-Below, Over, Under, Within, and Element). SLT outputs are used by [4], one problem with this representation is that a valid auto completion query may have a sparse subgraph representation which is not easily handled in a tree, an example of this is shown in Equation 1.1. An example of a SLT can be seen in Figure 1.1 where the relationships are represented as the labels on the edges. Operator Trees (OPT) that represent the order of operation for a math formula are also used to represent the formulas with their syntactic structure. Like SLTs, OPTs are used as a part of the Tangent-S [4] and Tangent-CFTED [15] systems. Where tuples are used to embed the graphs into vectors, which are then indexed.

HMM-based systems return an output that represents a probability of similarity. The system described in [7] utilizes a sequence of HMMs trained for symbol recognition. Based on the results of these HMMs together a probability of the desired word is returned. This could be useful when trying to do partial matching between a formula query and autocompletion results. An HMM can

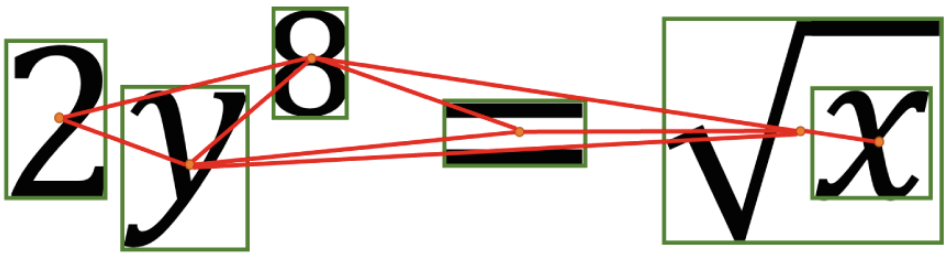


Figure 2.5: A LoS Graph [3]

be designed to traverse a SLT of the formula modeling the likelihood that the symbols given are similar to the index by traversing the tree and giving the likelihood of the symbols and relationships of the query are represented by the formula SLT. Building on the same input features used by [7] another system was purposed using a BLSTM in [8], where the likelihood of a given word being in the sequence is given as the output.

2.3 Similarity Measures

Based on the different output representations different similarity measures must be used. Fixed-length vectors can be directly compared to each other. Being fixed-length they have the same dimensionality so standard distance measures can be applied. Common distance measures that have been used for BoS and other fixed-length vectors include Cosine similarity and Euclidean distance. The PhocNet system presented in [22] uses Bray-Curtis dissimilarity Equation 2.1 as the distance measure for the fixed-length PHOCs. Given $\mathbf{a}$ and $\mathbf{b}$ represent a fixed-length vector embedding and a_i and b_i represent the value at the ith index of the embedding.

$$BC(\mathbf{a},\mathbf{b}) = \frac{\sum_i |\mathbf{a}_i - \mathbf{b}_i|}{\sum_i |\mathbf{a}_i + \mathbf{b}_i|} \tag{2.1}$$

The Bray-Curtis dissimilarity represents the proportion of the number of mismatched bits to the total number of on bits in either vector. In a follow up paper [24] the cosine distance, Equation2.2, was used as it worked well for comparison of large vectors.

$$d_{cos}(\mathbf{a},\mathbf{b}) = 1 - \frac{\mathbf{a}^T\mathbf{b}}{||\mathbf{a}|| \cdot ||\mathbf{b}||} \tag{2.2}$$

The cosine distance formula represent the angle between the two vectors $\mathbf{a}$ and $\mathbf{b}$. This is done by taking the inner product of the vectors normalized to a length of 1. Subtracting this value from 1 gets the angular distance rather then angular similarity.

A common technique for similarity measure used in word spotting is Dynamic Time Warping, DTW. DTW is able to compare variable-length vectors. The variable-length vectors represent a sequence of local features describing the word image. DTW is used as it can squeeze sequences into a common

dimensionality for comparison. This is the technique that is used in [19]. When determining the DTW-distance, dist($\mathbf{X}$, $\mathbf{Y}$) between two time series $\mathbf{X} = (x_1, ..., x_m)$ and $\mathbf{Y} = y_1, ..., y_N$, a matrix $D\epsilon\mathbb{R}^{MxN}$ is built, where each entry $D(i, j)(1 \leq i \leq M, 1 \leq j, \leq N)$ is the cost of aligning the sub-sequences $\mathbf{X}_{1:i}$ and $\mathbf{Y}_{1:j}$ [19]

Each entry $D(i, j)$ is calculated from some $D(i', j')$ plus an additional cost d, which is usually some distance between the samples x_i and y_i. For instance the implementation of the algorithm in [19] uses

$$D(i,j) = min \begin{Bmatrix} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{Bmatrix} + d(x_i, y_i). \tag{2.3}$$

The algorithm determines a warping path composed of index pairs $((i_1, j_1)$, $(i_2, j_2), ..., (i_k, j_k))$, which aligns corresponding samples in the input sequences $\mathbf{X}$ and $\mathbf{Y}$. Once the necessary values of D have been calculated, the warping path can be determined by backtracking along the minimum cost path starting from (M,N), where M is the length of $\mathbf{X}$ and N is the length of $\mathbf{Y}$. We want the accumulated cost along the warping path, which is stored in D(M,N). This matching cost is normalized by the length of the warping path as cost for shorter paths would be lower for shorter sequences, which yields

$$dist(\mathbf{X}, \mathbf{Y}) = D(M, N)/K \tag{2.4}$$

Where K is the path length. This distance represents the warping score normalized by the length of the warping path.

Graph-based output representations have several common similarity measures. One common approach for graph representation is using an edit distance, a way of measuring the amount of change needed to take one graph and get another graph. Hausdorff edit distance is utilized in [1] where rules are assigned to several possible actions performed to get from one graph to the other. The possible actions include deletion, insertion, and substitution for edges and nodes. This works as a similarity measure in exact matching since two representations of the same formula should only require minimal editing to represent the same graph, but this approach would heavily penalize autocomplete results as many nodes would need to be added to complete the graph.

Tangent-V [5] uses a two-stage scoring. The first stage is based on the probability of classification results and angles for all symbol pairs that are matched to symbol pairs in the index. All symbols with a relationship were paired in tuples that were sorted lexicographically for this comparison. The second stage function is based on matching the node of the subgraphs through unification to find the largest matching. The first score would be useful in partial matching as longs as symbols in the query that see each other would see each other in the final match. Matching based on the largest line-of-sight graph might be less useful if the symbols in the query are on the leaf nodes in the final tree representation because having relationships in the query that don't exist in the final formula would make the line-of-sight graph more disjoint. This scoring process is expensive, having to unify the query graph with each candidate formula. Tangent-CFTED [15] uses an embedding of the tuple data of the graph using a linear regression model to combine the embedding into a single embedding for scoring. Followed by a Tree-edit distance measure for re-ranking the top candidates, this scoring process is elaborated on in Section 4.1.

For the HMM-based outputs that represent a probability of matching, Viterbi decoding probability is used. This ranks the likelihood that the word appears in a given text line. This would need to be modified based on changes to the HMM for math formulas to represent the likelihood that the symbols are a sub-tree of a larger formula tree. A text line represents a sequence of characters that are all at the same baseline level. When it comes to math formulas there is a baseline for the formula along with a branching structure of subscripts, superscripts, and above and below relationships. An improvement for HMM-based systems was proposed in [6] where a formal grammar for the language is used in the Markov Logic Network to improve the precision in word spotting. Grammars are often limited since they must be manually created which can negatively affect the results by not capturing all possible edge cases or being too strict to generalize to all cases.

2.4 Inverted Index

In any information retrieval (IR) system it is import that the system can not only function effectively, but also efficiently. There are several well known techniques used in IR to help improve the efficiency of the search. A common

approach is to begin with an inverted index mapping of the data as seen by the Tangent Systems where symbol pairs/tuples are used as the keys into the index where the postings consist of the formulaIDs that have that symbol pair. Using an inverted index can help to reduce the search time as you can quickly collect all the candidates for scoring with out need to traverse the entire indexed collection.

2.5 Summary

The previous systems looked at following the same data pipeline shown in Figure 2.1. The differences come down to the representation of the features, the index representation generated from those features, and how those outputs are ranked. Fixed-length vectors are easy to work with because they can be directly compared with distance metrics, without the required additional computation of DTW. For our approach, we will be looking at using a fixed-length vector representation and graph-based approaches. A fixed-length vector would reduce the computations required during ranking. If a strong representation can be quickly computed for a fixed-length approach it will reduce the needed steps of creating the graph. Graph representations are useful to look at further because of the tree nature of math formulas. Variable-length vectors while previously popular in the word spotting field have become less popular with improvements to CNN encoding as seen with the shift from earlier publications focused on DTW methods and recent papers focused on CNNs. We will be focusing on an improved output representation that is capable of capturing the spatial structure of math formulas, without the need to capture the graph relationship between symbols.

Chapter 3

XY-PHOC Retrieval Model

In this chapter we will describe XY-PHOC, our generalization of the PHOC embedding to capture both horizontal and vertical symbol positions. First we will discuss our motivation for designing this embedding, followed by the specification of the embedding, how we will construct an efficient index, and how we efficiently search on this index. In designing this embedding we hope to capture a new set of possible inputs for autocompletion as well as create a simple embedding capable of competitive results on similarity search.

3.1 Motivation

When using autocompletion for word based queries the left portion of the query is given and it is expected to complete the right part of the query. This works well for the sequential nature of text where typing out the query as a sentence you start from the beginning and can use suggestions to complete the end of the query. While writing a math formula you tend to write the symbols on the left before the symbols on the right, this is not always the case when there are fractions or exponentiation that are added after the fact. There are three possible areas that we would like to address when it comes to autocompletion of math formulas, symbols missing from the right (the leaf nodes of the SLT, which include subscripted subexpressions), left (root of the SLT), or the middle (inner nodes of the SLT).

Having missing symbols on the right, is the same condition which text query autocompletion systems already work for. We expect that this would

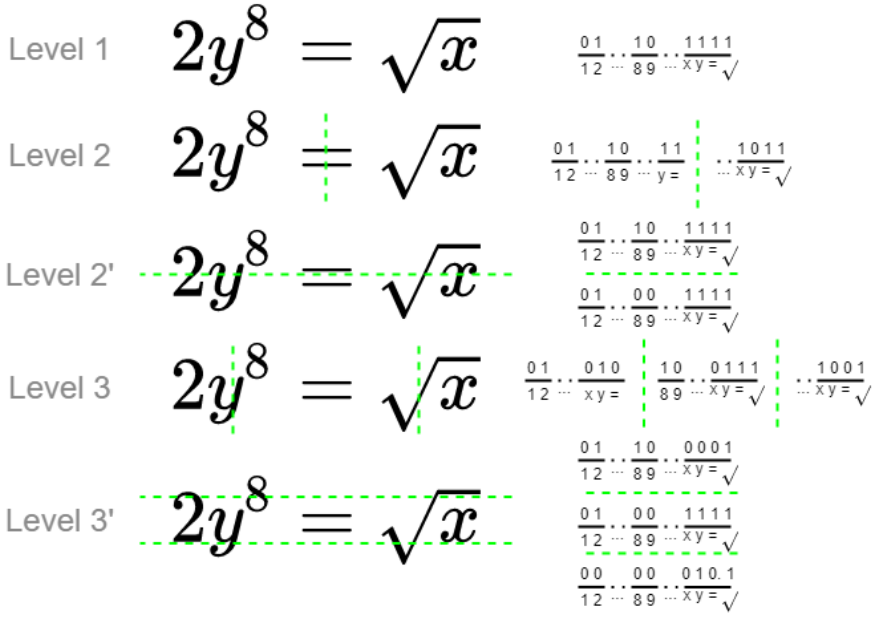


Figure 3.1: PHOC Representation with the addition of horizontal splits to capture the relative vertical positions of symbols, these splits are represented by Levels $2'$ and $3'$ where the number of splits is the same as on 2 and 3 but the line is along the horizontal axis instead of vertical axis.

be the most useful autocompletion for users as it is similar to how other search systems work. When writing a very complex math formula it is possible that only portions of the formula are remembered when trying to search for it so in this case parts of the left and right may be entered and symbols will be missing from the middle. Even writing $^2+^2 =^2$ is possible for representing the Pythagorean Theorem. In these cases the system should be able to provide a valid autocompletion result.

One way to achieve this functionality is in using a canvas for query entries that lay out the symbols in roughly their correct spatial locations in the final query. Using the relative position on the canvas will help to establish where the portion of the equation should be represented in the final formula. The PHOC

representation utilizes relative positioning already with its cuts in space per-level. This notation can be used with partial formulas to match by utilizing the entire canvas as the dimensions of the spatial cuts and not normalizing the symbols to the size of the entire canvas space. This will be able to capture where the given symbols should be relatively in the formulas matches. An is shown in Fig.3.1. In our work we are using 5 levels in both the horizontal and vertical directions (only 3 are shown in Figure3.1). This embedding also be useful for similarity search as the redundant representation is capable of capturing the variance of formulas with similar symbols and structures.

In similarity search, current systems rely on graph-structured data, such as in Tangent-S and Tangent-CFTED. These systems require special indexing strategies in order to build inverted indexes on non traditional keys. The scoring of these systems also requires several operations or a trained embedding model to process tuples to be scored.

The XY-PHOC representation we propose will require less storage for the index as all keys are symbols with postings consisting of a bitstring and a formula, explained further in Section 3.2.1. Storing standard types in the index will allow for standard information retrieval tools and techniques to be utilized. Additionally the scoring of XY-PHOC embeddings is performed by an optimized function which is rank equivalent to the cosine similarity. A simple vector based scoring mechanism provides efficient rankings in less time than the scores on the more complicated graph data. Next we will describe in detail the embedding proposed that will be able to accomplish the cases we aim to cover.

3.2 Embedding

When encoding the binary XY-PHOC vectors, as shown in Figure 3.1, if any pixels from the symbol are included in that region the bit is set to 1, the same for the vertical regions. For each horizontal level n greater than 1, there will be a level n' which represents the vertical splits at that level. In [23] several variations on the PHOC embedding were tested, as explained in Section 2.2 and the conclusion they reached was more information than just the relative positions and symbols present would need to be included to improve results any further.

The criteria for being in a given region can be defined in several ways. In

our experiments we look at the effect of having region membership defined as any pixel in the region, the original definition, as well as having the centroid of the symbol, the top left of the symbol and the horizontal span of the symbol projected on the vertical center.

Improving the uniqueness of a formula in the embedding space will help to increase the reciprocal rank, which is defined as $\frac{1}{r}$ where r is the rank of the target, when doing autocompletion and exact matching. When ranking formulas you want the representation to capture enough information so the formulas will be close in embedding space to the query and the wrong queries are further away. The use of distance measure for scoring similarity allows for the distance in embedding space to be used in the ranking. The more details that make a formula unique are embedded in the space the better the reciprocal rank should be with a possible loss in similarity effectiveness.

In order to both improve the efficiency and efficacy of the results, an index and retrieval system is designed for XY-PHOC which is described bellow.

3.2.1 Indexing

The original PHOC representation is a (very) sparse vector representing the symbols present in each region. For math the symbol vocabulary is much larger than the Latin alphabet, making the vector even longer. To use XY-PHOC efficiently, we use an inverted index over symbols with each posting as a pair $(id,\ v)$ containing a formula identifier and a bit vector representing only the XY-PHOC regions where that symbol appears. We use five levels, making the symbol-specific XY-PHOC vector 29 bits long, which we store in a 32-bit integer. An example of this embedding is given in Figure 3.2.

An additional index maps formula ids to their original file, the normalization constant for formula's XY-PHOC vector as described in Equation 3.2, and the number of symbols in the formula.

To index the large collections of data, several technologies are utilized together. In order to handle distribution of the work to index, Apache Spark* is used to read in the data and perform a map-reduction to produce a text output index file. The formulas are rendered using the javascript library,

*https://spark.apache.org/

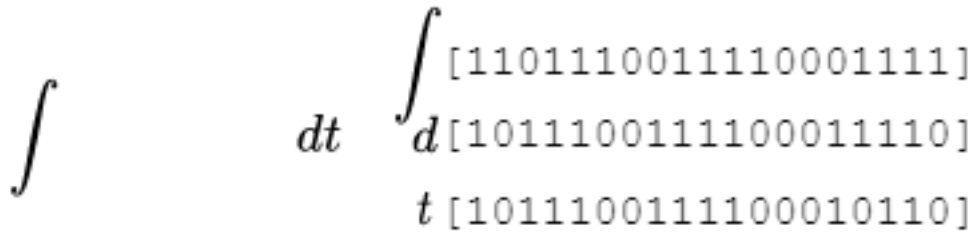


Figure 3.2: 29-bit XY-PHOC embedding for a query formula. Where the first bit represents Level 1 and the second two bits represents Level 2

MathJax[†]. The index is loaded into a Redis[‡] database to allow for efficient at scale look up into the inverted index.

An important advantage of the reduced representation of XY-PHOC is that standard information retrieval techniques and tools can be used to generate an efficient and robust system. Other math formula retrieval systems that work on graph representations need to use custom solutions in order to index the paths that make up the graph, where the XY-PHOC embedding easily fits into standard tools for text-based search engines.

3.2.2 Retrieval

We have designed retrieval models for both general formula retrieval, and formula autocompletion. We use both conjunctive and disjunctive retrieval over query formula symbols (i.e., requiring all (conjunctive) or at least one (disjunctive) symbol to be located in returned hits). When doing similarity retrieval there is no requirement for all the symbols in the query to appear in the relevant responses. Due to this, it is possible that when doing conjunctive queries that no formulas will be returned. When using disjunctive queries, a large net is cast on all the formulas with any common symbols and the scoring function will rank the formulas and any formula with few common symbols will have lower scores.

[†]https://www.mathjax.org/
[‡]https://redis.io/

For autocompletion we use a conjunctive query with the additional constraint that returned formulas must have at least as many symbols as the query. These constraints come from how traditional autocompletion works. It can be assumed when a user inputs a symbol that they expect that it will appear in the autocompletion for that particular formula, as autocompletion is a type of exact matching retrieval. It also does not make sense to score any formulas with less symbols than that contained in the query, as the goal with an autocompletion system is to reduce the effort for the user, as modeled and talked about in [12]. If the user gave a query with the formulas x^2+ it would not make sense to return to them x^2, since this is not completing a possible formula and is just returning a formula that is similar to the query formula.

The work done in [23] demonstrated that cosine similarity worked well for ranking with PHOCs, and so we use it in our work. For query vector $\mathbf{a}$ and candidate formula vector $\mathbf{b}$ the cosine similarity is:

$$cos = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{||\boldsymbol{a}||\ ||\boldsymbol{b}||} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}} \tag{3.1}$$

A faster rank-equivalent similarity metric *bcos* is defined as:

$$cos(\mathbf{a}, \mathbf{b}) \overset{rank}{=} bcos(\mathbf{a}, \mathbf{b}) = |\mathbf{a} \wedge \mathbf{b}|_1 \frac{1}{\sqrt{|\mathbf{b}|_1}} \tag{3.2}$$

The dot product of two binary vectors is the Hamming weight (shared bits) in the logical AND of the vectors, which is equivalent to the L_1 norm ($||_1$). The normalization factor for query $\mathbf{a}$ is constant across candidate formulas, and so can be removed. To accelerate computation, the normalization factor for $\mathbf{b}$ is pre-computed and stored for lookup at retrieval time. In Figure 3.3 we show an illustration of how the scoring between the query and a candidate formula is computed. Each symbol has a corresponding bitstring for the formula which represents its relative position.

3.3 Summary

The XY-PHOC model is a relative spatial embedding which is able to capture both horizontal and vertical position. An index can be efficiently generated from the embedding by an inverted index of symbols with postings representing

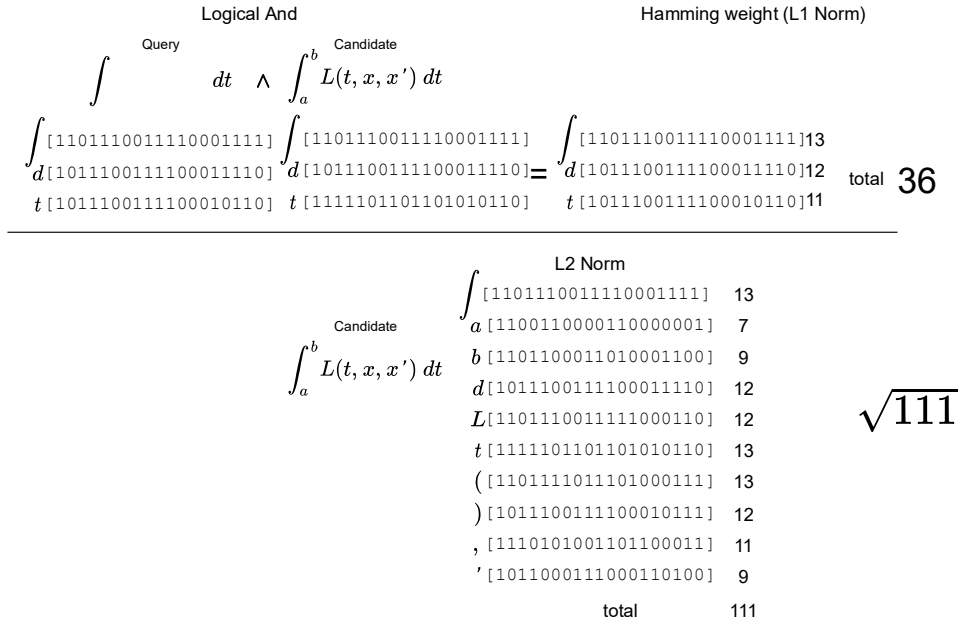


Figure 3.3: Computing the match score between a query formula and a candidate formula. The numerator represents the total number of common bits between the query and the candidate. The denominator represents the normalization constant for this candidate formula.

the bitstring of the symbol for the associated formula. With this representation in the inverted index efficient conjunctive and disjunctive queries can be used. In order to efficiently score candidate formulas a scoring function is proposed which is rank equivalent to cosine similarity with reduced computation.

Chapter 4

Experimental Design

In order to evaluate the efficiency and efficacy of the XY-PHOC encoding experimentation will be performed for the recognition task and the autocompletion task. In this chapter we will discuss the baseline systems that will be compared against, the datasets that will be indexed, the experimental conditions tested and metrics used. To the best of our knowledge there is little work done in the area of math formula autocompletion, no work done looking at using visual matching techniques for autocompletion. To help with evaluation of the autocompletion a new metric is proposed based on the work on esaved presented in [12].

4.1 Baselines

Comparing against baselines helps to better understand the impact of the work done for this model, for improving both efficiency and efficacy. A naive baseline, Bag-of-symbols (BoS), is useful in making sure that the system makes an improvement on the bare minimum. A comparison against the PHOC representation will help to determine if packing vertical spatial information into the embedding provides an improvement. Tangent-V [5], Tangent-S [4], and Tangent-CFTED [15] provide a comparison against graph based systems, Tangent-V using visual information and Tangent-S and Tangent-CFTED using features from the SLT and OPT.

4.1.1 Bag-of-Symbols

Bag-of-Symbols (BoS) is a naive embedding for formulas which indicate what symbols are present the formula. This embedding has no information of the spatial relationship between symbols. A similar binary vector used to represent the PHOC embedding, elaborated on in 4.1.2 can be used for BoS. The first level of the PHOC embedding is equivalent to the BoS for the formula. The LaTeX string representations for formulas can be extracted and converted into this embedding with minimum effort. A limitation of this representation is its lack of spatial information, meaning two formulas with the same number of unique symbols but completely different layouts will have the same embedding. An example of two formulas with the same symbols with different structures is show in equations 4.1 and 4.2

$$f(x) = \frac{x}{1+x} \tag{4.1}$$

$$f(x) = x + \frac{1}{x} \tag{4.2}$$

To help illustrate how these formulas would retrieve the same formulas from search when using BoS, the Conjunctive query, requiring all symbols present in the query in the candidates before scoring, results are presented in Table 4.1 from the annotated ARQMath collection described in Section 4.2. Both queries, returned the exact same results with the same scores, because with BoS the embedding both formulas are is (,), +, fraction(-), 1, =, f, x . There is little structural similarity between the queries and the retrieved formulas. Further more, all of these top-5 results also have the exact same embedding. To improve the retrieval, so that these formulas that have the same symbols but different structures will not be matched, spatial information must be added.

4.1.2 PHOC

Encoding spatial information about the symbols will be very important to make improvements on matching the correct formula. Originally, PHOC was designed to capture relative spacing between the symbols in words, by cutting the word spatially into different regions where multiple regions will contain overlapping information for redundancy in representation. As discussed earlier in Section 2.2, this is enough spatial information for word spotting because

Table 4.1: Bag-of-Symbols Conjunctive Query in the Annotated ARQMath Collection for Equations 4.1 and 4.2

	Bag-of-Symbols	
Rank	$f(x) = x + \frac{1}{x}$	$f(x) = \frac{x}{1+x}$
1	$\boxed{f(x) = x + \frac{1}{x}}$	$f(x) = x + \frac{1}{x}$
2	$f(x) = 1 + \frac{1}{x}$	$f(x) = 1 + \frac{1}{x}$
3	$f(x) = \frac{x+1}{x}$	$f(x) = \frac{x+1}{x}$
4	$f(x) = \frac{x}{1+x}$	$\boxed{f(x) = \frac{x}{1+x}}$
5	$f(x) = \frac{x}{x+1}$	$f(x) = \frac{x}{x+1}$

a word is a sequence that progresses from left to right. Math formulas on the other hand are tree like structures where symbols can be above and below the baseline for various operations such as fractions and exponentiation. An example of the improvement in search results when using the PHOC embedding can be seen in Table 4.2. Here the same queries made previously are run again, with the PHOC embedding the retrieval results reflect the horizontal placement of symbols in the formula, and not just the presence of specific symbols.

PHOC vs. XY-PHOC. Experimentation with PHOC can help understand if the difference in vertical layout that is seen in math formulas is important to be captured or if there is enough information in capturing the horizontal layout of the symbols. If we look at the equation $f(x) = \frac{x+1}{x}$ which appears at rank 3 result of the query for Equation 4.1 and $f(x) = \frac{x}{x+1}$ which appears at rank 2 of the query for Equation 4.2 both would have identical PHOC embeddings since the symbols are in the same positions in the horizontal direction but in XY-PHOC would have different embeddings because of the vertical differences.

4.1.3 Tangent

The three Tangent systems that will be used as baselines represent the graph based models being used for math formula retrieval, with all three capable of similarity search and Tangent-V being also used for spatial autocompletion. The Tangent models all use relational data between the symbols in the formula in order to perform math formula retrieval. There is a difference in the types of data each uses, with Tangent-V using visual data, the symbols in space, and Tangent-S and Tangent-CFTED work on the relational trees that represent the formulas. We will be comparing against these systems to gain a better understanding of how the similar XY-PHOC embeddings against state-of-the-art systems.

4.1.4 Tangent-V

To better understand how XY-PHOC compares against another system using only visual data we will be running Tangent-V. Tangent-V [5] generates line-of-sight graphs representing what symbols can see other symbols, storing these symbol pairs. Generating the entire graph representation is very time

Table 4.2: PHOC Conjunctive Query in the Annotated ARQMath Collection for Equations 4.1 and 4.2

	PHOC	
Rank	$f(x) = x + \frac{1}{x}$	$f(x) = \frac{x}{1+x}$
1	$f(x) = x + \frac{1}{x}$	$f(x) = \frac{x}{1+x}$
2	$f(x) = x^2 + \frac{1}{x}$	$f(x) = \frac{x}{x+1}$
3	$f(x) = \frac{x+1}{x}$	$f(x) = \frac{x}{1+\lvert x\rvert}$
4	$f(x) = x + \frac{16}{x}$	$f(x) = x + \frac{1}{1+x}$
5	$f(x) = 1 + \frac{1}{x}$	$f_n(x) = n + \frac{x}{1+x}$

consuming, as determining if two symbols can see one another through line of sight is a nontrivial task. During the graph generation the line-of-sight between symbols must be tested, determining if two symbols can see each other and if the distance between those symbols falls within a threshold along the line-of-sight.

With in the index generated by Tangent-V and inverted index is generated over the symbol pairs from the graph edges. That means that any symbol in the query that doesn't see a symbol it has a line-of-sight to in the target it will never be able to match successfully, making it not as useful for autocomplete, such as with $^2 + ^2 = ^2$ the exponents never see the variables a, b, or c and the symbol pairs from the inverted index would not be returned. The reliance on pairs of symbols also restricts how good the system will be at autocompletion when only the first symbol is input, formulas with three or fewer symbols in the index will be returned when a single symbol is in the query, due to special case code in Tangent-V to index smaller formulas with self pairs.

When ranking, a two layer system is used. The first layer gets all postings containing the symbol pairs from the inverted index. Then formulas with large differences in displacement angles and/or symbol size ratios relative to the query edge are filtered out. Then an edge-based ranking metric is applied, keeping the top 1000 formulas. Then a re-ranking layer is applied which focuses on mapping line-of-sight nodes one-to-one with the query. Structural alignment is used for matching the graphs.

4.1.5 Tangent-S

To better understand how XY-PHOC compares against the state-of-the-art, a comparison against Tangent-S the top performing system on the ARQMath similarity search task will be performed. Tangent-S [4] utilizes the SLT and OPT in order to generate tuples which are stored in the index. The SLT representation is built around the writing lines, which leads to deep trees with few branches. The OPT represents the formula as a hierarchy of operations, which makes a shallow tree with many branches. The index is then made by making an inverted index on symbol pair tuples. The symbol pair tuples are made up of the ancestor and descendant symbols, as well as the sequence of edge labels in the path from the ancestor to the descendant. The tuples from the SLT and OPT contains a lot of relational data to help better match the

formulas on.

When performing retrieval, first candidates are selected by matching the query symbol tuples in the index. A harmonic mean of precision and recall of matched symbol tuples is used to assign an initial score. Then a structural matching score is made by finding the the largest connected match between the query and candidates, which is obtained by a greedy algorithm, evaluating pairwise alignments between trees. The output of the structural matching is a subtree of the candidate formula that has been successfully aligned to the query. A re-ranking is performed on the Maximum Subtree Similarity (MSS), negative count of candidate nodes matched with unification, and negative count of candidate nodes matched without unification. A linear regressor was trained using the relevance numbers for query matches to learn how to best combine these metrics in [4] and a learn to rank model was later applied in [14].

4.1.6 Tangent-CFTED

A further improvement on the Tangent-S system, Tangent-CFT(ED) [15]. Tangent-CFT as presented in [15] uses the output of the Tangent-S model to generate the tuple embeddings for the SLT, OPT, and SLT-Type (each node representing only the type and not the value) which are embedded into a vector which is combined together. To better handle partial matching, tree-edit distance is added to the model. Tree-edit distance (TED) is the minimum cost of converting one tree to another. The tree-edit distance is used to re-rank the retrieval results.

All three of these systems will be able to help show if a graph structure is required for math formula retrieval. While embedding of the graph data are able to hold a large amount of data about the relationships between symbols this also take a large amount of time to compute and in the case of SLT, in Equation1.1 we show a possible query formula this is a disjoint subgraph which can not be handled by tree based systems. Additionally the TED used for re-ranking in Tangent-CFTED adds a penalty on similarity between graphs that require new nodes and edges be added, but in auto-completion adding symbols will be necessary as the query being matched will be incomplete. The XY-PHOC model, also prefers shorter matched formulas, with the redundant capture of symbols in relative space there is less of a penalty for adding more symbols if they already appeared once in the query compared to needing

unique symbols.

4.2 Datasets

The ARQMath formula dataset will be used in order evaluate the XY-PHOC embedding. ARQMath dataset [26] contains 9,340,034 visually distinct formulas, comprised of formulas from Math Stack exchange. Of the entire dataset there is a subset of 5,676 annotated formulas, on 45 test topics. 10 of these formulas were determined to not be valid LaTeX, and were not included in the index. All 10 of the invalid formulas had a relevance of 0. The total number of formulas in the annotated index is $5,666$ and $9,326,795$ formulas in the full index. The annotated index size on disk is 1.6 MB for the symbol index and 836 KB for the formula index, and the full index size on disk is 2.3 GB for the symbol index and 814 MB for the formula index. The annotated index was generated for Tangent-V where in addition to the 10 formulas determined to not be valid latex were not generated as well as 51 formulas that failed to render during the index. All of these formulas have a 0 relevance rating leaving the Tangent-V index with 5615 formulas.

In order to render the LaTeX string the node.js library MathJax was used to render the formulas to a SVG image file format. From the SVG the locations and labels were able to be extracted directly for use. To make sure as many formulas as possible were able to be rendered certain tags and commands were removed from the strings such as the MathJax specific \toggle and \endtoggle commands which only work in HTML rendering. Along with the latex commands for tag and label. The label command has no visual impact on the formulas, removing them would not effect the output files and the tag commands while effecting the visual output do not effect the actual math content of the rendered formula, it simply allows the user to override what reference number is used for the formula in display math mode.

4.3 Formula Search and Autocompletion Experiments

Since retrieval is based on similarity matching and autocompletion is based on exact-matching, the conditions that work the best for one of the systems will not necessarily be the best for both systems. Several parameters of the

XY-PHOC system must be tested and tuned so that the best system using this model can be compared to the state of the art.

Across tests of the system different variables can be adjusted. The first test run will compare the the different individual levels of the embedding function, the space split into N regions for level N, on their own. Starting with level 1 and running with just level-x for all 5 levels. Using just level 1 is equivalent to the bag-of-symbols. The idea for this experiment is to demonstrate how the embedding works without the redundant information from the previous levels. Similar to how XY-Trees only compare at the leaf level of the tree there will be no redundant information to improve robustness. Additionally this experiment will give a sense of the importance of any specific level has in retrieval. An experiment will be run on just the just the horizontal cuts, which is equivalent to the original PHOC embedding, to determine if the added vertical information provides an improvement for both retrieval tasks.

The next test will demonstrate the impact of more levels of the embedding make starting with level 1-2 and increasing in levels till we get to 5 with both vertical and horizontal cuts. These test will help to show how additional levels giving both redundant information and better separation of symbols helps with the robustness of the representation. The deeper the range of levels, the better the XY-PHOC model performs at both similarity search and autocompletion.

The condition to include a symbol inside a split of the PHOC can be experimented with to provider better distinguishing embeddings. The original criteria is that any pixel of the symbol is with in the bounds of the given split. Several additional criteria will be used including the top left corner of the symbol, the centroid of the symbol, and projections of the symbols onto a single axis. As well as test to test these conditions separately on their impact on the vertical and horizontal cuts of the XY-PHOC. The condition used to determine if a symbol is within a split helps to get better discrimination of the location of symbol in the formulas. There needs to be a balance between discrimination of symbols locations, to help better match exact matches, while keeping enough redundancy of the symbol that is helpful for finding similar formulas with different symbol spacing.

4.3.1 Formula Similarity Search

The math formula retrieval task consists of a complete query formula and relevant formulas are returned. To evaluate retrieval using XY-PHOC we use the ARQMath formula retrieval task (45 topics). The three metrics used for comparison on the ARQMath dataset are the normalized discounted cumulative gain (nDCG), mean average precision (MAP), and precision (P). These scores are measured using `trec_eval`, a standard tool in the information retrieval space, which provides the ability to read in results files and generate the desired scores.

The prime version of these metrics are use as defined in [26] because they are only calculated on the dataset for the formulas with annotated relevance scores. These are the standard metrics used for comparing systems on ARQ-Math Task 2. nDCG is defined as follows:

$$nDCG@p = \frac{rel_1 + \sum_{i=1}^{p} \frac{rel_i}{log_2 i}}{rel_{max} + \sum_{i=1}^{p} \frac{rel_{max}}{log_2 i}} \tag{4.3}$$

Where rel_i represents the relevance score for the retrieved result at rank i, and rel_{max} represents the max relevance for this collection. The p represents the cut off for how many results to look at when calculating the score, and if no p is given then all retrieval results are used. The nDCG measure is the discounted cumulative again divided by the ideal discounted cumulative gain which gives a score which is ≤ 1 at any rank position.

The formula for MAP is defined as follows:

$$\text{MAP} = \frac{\sum_{q=1}^{|Q|} AveragePrecision(Q_q)}{|Q|} \tag{4.4}$$

Where Q is a set of queries, Q_i represents the ith query in the set and $|Q|$ represents the number of queries in the set. This is the mean of the average precision for all queries in the set. Precision is defined as follows:

$$Precision(q) = \frac{|rel_q \cap ret_q|}{|rel_q|} \tag{4.5}$$

Where rel_q represents the relevant results for query q and ret_q represents the retrieved results for query q. This gets the ratio of retrieved relevant results

to the total number of results retrieved. This metric is measured with cuts represent as P@k where k represents the top-k documents retrieved to look at.

We run all test topics and generate nDCG′, nDCG′@5, MAP′ (mean average precision), P′@10, and P′@5 scores that can be compared directly to previous systems, which are evaluated using only assessed hits. For our conjunctive model, to avoid errors with `trec_eval`, we generate an irrelevant hit for queries that return no result (e.g., because some symbols in a test formula are not in the indexed collection).

Tests will be done to see if a different number of vertical and horizontal splits is helpful in improving the efficacy on the formula retrieval. The redundant information given by extra levels could cause noise and so having different resolution of cuts on the x and y axis may prove beneficial. The insight that will be gained from these experiments is if the more information packed into the embedding helps to improve results. In order to keep the embedding efficient a minimal representation required to get the best results can be used so no unneeded information needs to be stored. It is also possible additional layers cause extra noise which will decrease performance, removing any levels that add too much noise to the embedding will help to improve the efficacy in retrieval.

The highest performing system from the experiments will be compared against the baseline and state-of-the-art systems. The default configuration will also be compared to the state-of-the-art systems on both the annotated and full ARQMath collection.

Tangent-V being another system using visual information for retrieval, will help to show how the XY-PHOC embedding compares to another system using visual data for math formula retrieval. In order to establish a comparable baselines with Tangent-V the annotated ARQMath data will be indexed and task 2 queries will be run.

4.3.2 Formula Autocompletion

To test autocompletion, we used a selection of formulas from the annotated ARQMath index as targets for the completion. In a series of experiments we will test the effect of the input order for the symbols in the query. In order to evaluate the different input orders a new metric is presented rsaved, along with the use of number of symbols needed to obtain a top-k rank for the target

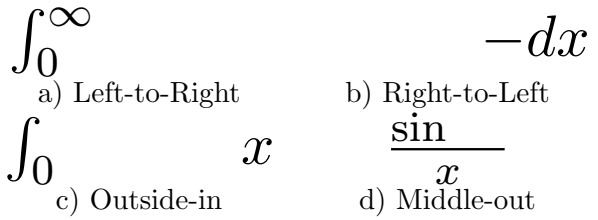


Figure 4.1: Example of inputting the Equation $\int_0^\infty \frac{\sin(x)}{x}\,dx$ with 3 symbols when the order entered is a) Left-to-Right, b) Right-to-left, c) Alternating from Left-Right from the Outside-in, d) Alternating Left-Right from the Middle-Out.

query.

The queries used for the autocompletion testing are taken from the annotated collection of ARQMath, as any query used for the autocomplete metrics needs to be present in the index. A stratified sample of the queries is made, where the queries are binned by the number of symbols they contain. Bins spanning 4 sizes are generated starting with 2 symbols (e.g. 2-5, 6-9...) with no queries of length greater than 200 include as formulas with this many symbols are rare in the collection and by just being that long their autocomplete result would be guaranteed after enough symbols are entered do to the minimum length restriction included in the XY-PHOC autocompletion model. From each bin 10 formulas are selected randomly and for some bins where there were not 10 formulas all the formulas in that bin were selected giving a total of 422 queries. All the formulas included in the autocomplete query set are presented in Appendix A.

The orders that will be tested are left-to-right, right-to-left, middle-out, and outside-in. The left-to-right order is equivalent to the standard use of autocompletion on strings, where the symbols are input ordered on position from left-to-right and top-to-bottom. Right-to-left uses the reverse order but still breaking ties top to bottom. The middle-out order will start at the middle an alternate between the adding symbols to the left and right until the full string is used. For the Outside-in the left symbols is added to start followed by the right-most symbols alternating left and right until the whole formula is input. An example of these four input orders is shown in Figure 4.1, where the equation $\int_0^\infty \frac{\sin(x)}{x}\,dx$ is input with 3 symbols for each order.

The reason to test the input order in this way is to help to understand the effect of giving users a new way of inputting symbols will impact there ability to quick find the formula they are looking for. While we recognize that these strict orders do not necessarily match how a user would actually go about inputting a formula, these conditions let us easily automate the test for large portions of data. While getting an accurate user test for the actual orders a user might enter the symbols in would require a large user-study that is beyond the scope of this paper and will be done as future work.

After identifying the optimal configuration on the annotated ARQMath index, the same queries will be run again using the full ARQMath index, again measuring with the same metrics. With the full ARQMath index we will be able to identify how the system works at scale for both its efficiency with large data and the impact of additional data impacts the results.

To establish a baseline for autocompletion the Tangent-V system will be run with the same queries as the XY-PHOC System. For ease of comparison the standard left to right input order will be used for the symbols. The Tangent-V System being not designed for the same types of sparse queries as XY-PHOC, the other orders error on some cases were the symbols are not with in the threshold for line-of-sight, leaving disjoint nodes in the graph. This will help better understand the efficiency of the XY-PHOC query model and to understand how the scores compare against another visual system. To score autocompletion a new metric based on esaved from [12] is proposed.

Previously the effort saved [12] metric was proposed for measuring the amount of effort saved while entering each character in a word for autocompletion, which weights a probability of user satisfaction for the rank of the target formula in the results by the ratio of characters that have not been typed in the word:

$$esaved(q) = \sum_{i=1}^{|q|} \left(1 - \frac{i}{|q|}\right) \sum_{j} P(S_{ij} = 1) \tag{4.6}$$

We have opted for a simpler, more intuitive measure to characterize autocompletion performance as each symbol is entered, the *rsaved* metric, which is the average reciprocal rank of the target after each symbol is added to the query:

$$rsaved(q, k) = \frac{1}{|q|} \sum_{i=1}^{|q|} rr(q, k) \tag{4.7}$$

This metric has a close relationship to the number of alternatives at and above the rank of the target formula after entering each character.

Additionally for each query, we observe the number of symbols that needs to be entered before the target appears in the top-k (where k is 5). Our expectation is that the outside-in ordering of adding symbols will be most efficient, as it constrains the formula structure from both ends of the query, this is helpful for the XY-PHOC embedding because it stabilizes were the symbols are in space based on the full span of the formula. When adding symbols just from left-to-right, the spacing of the cuts based on the width of the bounding box formed by all the symbols keeps growing, adding variance as the embedding space grows.

4.4 Summary

In order to test the XY-PHOC model the ARQMath dataset [26] will be used with a series of experiments. To understand how useful the XY-PHOC model is, it will be compared against several baselines. These baselines include two systems that are structurally similar with with different levels of data, that being the BoS and PHOC models. Then a state-of-the-art visual system, Tangent-V [5], will be run to understand how the simpler visual embedding of XY-PHOC compares to embedding the line-of-sight graph for both similarity search and autocompletion. Then the top performing systems on the ARQMath Task 2 will be compared against to understand how XY-PHOC, compares to the state-of-the-art systems for the similarity search on math formula task.

Chapter 5

Results

In this section we will present the results for the XY-PHOC Model on the ARQMath formula retrieval task [26], and an analysis of our XY-PHOC autocompletion model for formulas using symbols entered in different orders (e.g., left-right and outside-in). Several experimental conditions were tested on both tasks to get a better understanding of the impact of redundant information in the embedding. Each task may favor different conditions for embedding as ARQMath formula retrieval is based on similarity and autocompletion is based on exact matching.

5.1 Formula Similarity Search

To gain an understanding of the strength of each individual level of the XY-PHOC embedding we present the retrieval metrics for each single level in Table 5.1. It is possible that having multiple levels can make the embedding noisy by adding too much redundant information, so it is important to look at these single levels to see if they out perform the full embedding. An important pattern to recognize is that for levels 2 and 3 disjunctive queries the horizontal levels have higher scores than their respective vertical levels and this pattern inverts for levels 4 and 5. This makes sense, because in vertical spacing the symbols are much closer together than in the horizontal spacing. So with just vertical splits there are fewer distinguishing cuts which will matter more with fewer cuts. In Levels 4 and 5 there are more splits which can provide more information about the relative positions.

Table 5.1: Single Level of XY-PHOC using Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection. Prime ($'$) levels are split vertically.

	$\mu(\sigma)$					
	Disjunctive			**Conjunctive**		
Level	**nDCG$'$**	**MAP$'$**	**P'@10**	**nDCG$'$**	**MAP$'$**	**P$'$@10**
1 (BoS)	0.697 (0.166)	0.397 (0.238)	0.369 (0.279)	0.244 (0.269)	0.151 (0.218)	0.218 (0.277)
2	0.738 (0.145)	0.450 (0.214)	0.424 (0.240)	0.259 (0.278)	0.164 (0.217)	0.224 (0.266)
$2'$	0.736 (0.158)	0.439 (0.230)	0.416 (0.286)	0.263 (0.282)	0.170 (0.229)	0.236 (0.288)
3	0.738 (0.158)	0.447 (0.225)	0.420 (0.266)	0.252 (0.270)	0.153 (0.205)	0.204 (0.250)
$3'$	0.731 (0.164)	0.441 (0.236)	0.393 (0.296)	0.266 (0.285)	0.174 (0.232)	0.236 (0.281)
4	0.726 (0.169)	0.455 (0.241)	0.415 (0.253)	0.259 (0.276)	0.164 (0.215)	0.200 (0.238)
$4'$	0.745 (0.160)	0.456 (0.233)	0.411 (0.290)	0.266 (0.281)	0.174 (0.228)	**0.247 (0.291)**
5	0.738 (0.162)	**0.462 (0.228)**	**0.440 (0.259)**	0.260 (0.279)	0.163 (0.215)	0.191 (0.2284)
$5'$	**0.741 (0.156)**	0.446 (0.231)	0.422 (0.277)	**0.269 (0.283)**	**0.179 (0.235)**	0.240 (0.283)

As the depth increases for the single levels, for disjunctive queries, there is a trend of increasing scores with level $5'$ having the best nDCG$'$ and level 5 having the best MAP$'$ and P$'$@10. This is due to the increase of splits giving more distinguishing information about the relative position of the symbols. A surprising insight is that level 1, which is equivalent to a bag-of-symbols show strong results. This is a very simple approach, just capturing if a symbol is present at all in the formula and works well for this task. In the conjunctive queries, $5'$ is the best for nDCG$'$, MAP$'$, and $4'$ has the best P$'$@10. This shows a similar trend as the disjunctive that the more cuts in the embedding help to improve the similarity search results.

Next we look to understand the impact of packing more information into the embedding. We present the retrieval results for increasing the range of level used in the embedding for both disjunctive and conjunctive queries in Table 5.2. The expectation is that as we pack more information into the embedding that the scores will improve, as we will have more information about the placements of the symbols. This hypothesis holds true as the range increases depth the scores improve across the board, for both disjunctive and conjunctive queries. The best scores for disjunctive queries, with levels 1-5, has an nDCG$'$ of 0.778 which while only slightly higher than the previous 1-4 with an nDCG$'$ of 0.775 and a bit better than just the single level $5'$ with an nDCG$'$ of 0.741 it is worth the over head of using all the levels because the embedding is packed into bit-strings, and modern computer systems are

Table 5.2: Increase Depth of XY-PHOC on Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection. The Ranges Include Both the Horizontal and Vertical levels (e.g 1-2 includes 1, 2, and 2′)

	$\mu(\sigma)$					
	Disjunctive			**Conjunctive**		
Range	**nDCG′**	**MAP′**	**P′@10**	**nDCG′**	**MAP′**	**P′@10**
1 (BoS)	0.697 (0.166)	0.397 (0.238)	0.369 (0.279)	0.244 (0.269)	0.151 (0.218)	0.218 (0.277)
1-2	0.761 (0.156)	0.481 (0.237)	0.460 (0.268)	0.266 (0.284)	0.173 (0.232)	0.236 (0.288)
1-3	0.770 (0.153)	0.490 (0.235)	0.453 (0.276)	0.267 (0.285)	0.172 (0.227)	0.233 (0.274)
1-4	0.775 (0.153)	0.505 (0.240)	0.458 (0.270)	0.269 (0.285)	0.176 (0.233)	0.231 (0.266)
1-5	**0.778 (0.152)**	**0.511 (0.240)**	**0.469 (0.268)**	**0.270 (0.285)**	**0.176 (0.229)**	**0.231 (0.264)**

Table 5.3: Different Membership conditions using Disjunctive and Conjunctive Queries for Similarity Search in Annotated ARQMath collection.

	$\mu(\sigma)$					
	Disjunctive			**Conjunctive**		
Condition	**nDCG′**	**MAP′**	**P′@10**	**nDCG′**	**MAP′**	**P'@10**
Default	0.778 (0.152)	0.511 (0.240)	0.469 (0.268)	**0.270 (0.285)**	0.176 (0.229)	0.231 (0.264)
Top-Left	0.793 (0.136)	0.533 (0.233)	0.469 (0.261)	0.269 (0.285)	0.176 (0.232)	0.227 (0.263)
Centroid	0.791 (0.139)	0.535 (0.230)	0.473 (0.268)	0.269 (0.285)	0.178 (0.236)	**0.236 (0.267)**
Vert. Center	**0.794 (0.143)**	**0.540 (0.235)**	**0.473 (0.267)**	0.269 (0.284)	**0.179 (0.236)**	0.229 (0.259)

designed to work fastest on 64-bit integers so there is no in memory waste for the extra level.

The next condition we look to study the effect of is changing the condition for a symbol to be a member of a split. In the default conditions, if any part of the symbols bounding box is with in the split the symbols is included in that split, this gives a lot of redundancy in the embedding. This may also have the effect of making it so there is less unique information in the embedding. We hypothesised that changing this condition to make it so that the splits are more discriminating with improve the score, as it will reduce redundancy that makes the embedding less unique. Three alternative membership conditions along with the default condition are presented for both disjunctive and conjunctive queries in Table 5.3.

The Top-Left and Centroid conditions reduce each symbol to a single point for determining membership and the Vertical Center condition reduces each symbol to a line in the center of the symbol. Reducing the symbols membership helps make each cut either in the horizontal or vertical space more

discriminating. In the disjunctive queries, using the vertical center condition shows the strongest results across the board, and is the best system for the XY-PHOC Model. This makes sense because of the difference in vertical and horizontal spacing as mentioned previously. Since symbols are already spread out more in the horizontal axis the redundant information given by including it in any split along the horizontal axis more discriminating information. The vertical spacing between symbols being smaller, benefits from the reduced redundancy of only using the middle of the symbol rather than the entire span of the symbol.

For conjunctive queries all the conditions provide very similar results, this can be explained by the limited number for formulas that pass the initial conjunctive filtering. With only a limited number of formulas to even consider there is much less impact on score from having more or less discriminating information in the embedding.

We present the current top systems for the ARQMath Task 2 retrieval, along with our baseline systems in Table 5.4. The best disjunctive and conjunctive systems for XY-PHOC are presented, along with Tangent-V as a comparison against another spatial system. The PHOC system utilizes the same model as XY-PHOC but only uses the horizontal cuts, to help understand if the added vertical information is helpful. The current top system Tangent-S and Tangent-CFTED are presented twice, once with the original results for the systems and once re-ranked with a learn to rank model [14]. An interesting result present is how strong both the XY-PHOC mode and the Tangent-V systems are on this task. The XY-PHOC system uses a much simpler representation than the Tangent-S and Tangent-CFTED systems and before re-ranking XY-PHOC out performs both of them. Even with re-ranking, XY-PHOC has the strongest MAP' (0.54). A learn to rank model being trained on XY-PHOC could further show the competitive results this embedding can provide, additionally it could be used in conjunction with the other features used in the Tangent systems.

In order to confirm that the results seen on the annotated set applied to the full collection, we ran the disjunctive query model with the default membership condition on the full index. These results in Table 5.4 show that on the full collection the metrics stay stable other than the MAP$'$. The lower MAP$'$ can be explained by the larger collection having more formulas that are matched in the top 1000 before filtering down to only the annotated formulas

Table 5.4: ARQMath Task 2 Formula Similarity Search Benchmark

System	Collection	nDCG′@5	$\mu(\sigma)$ MAP′	P′@5	Q. Time (s)
XY-PHOC-Dis. Vert. Center	Annotated	0.63 (0.26)	**0.54 (0.24)**	0.56 (0.32)	0.25 (0.06)
XY-PHOC-Dis. Vert. Center	Full	0.61 (0.24)	0.44 (0.23)	0.55 (0.29)	565.8 (162.5)
XY-PHOC-Con. Vert. Center	Annotated	0.35 (0.33)	0.18 (0.24)	0.29 (0.32)	**0.02 (0.01)**
XY-PHOC-Dis. Default	Annotated	0.61 (0.27)	0.51 (0.24)	0.56 (0.31)	0.26 (0.07)
XY-PHOC-Dis. Default	Full	0.61 (0.26)	0.42 (0.23)	0.56 (0.30)	565.7 (170.2)
PHOC - Disjunctive	Annotated	0.59 (0.27)	0.49 (0.24)	0.53 (0.30)	**0.23 (0.05)**
Tangent-V	Annotated	0.62 (0.29)	0.53 (0.27)	0.59 (0.33)	0.59 (0.51)
Tangent-S	Full	0.58 (0.27)	0.45 (0.22)	0.51 (0.30)	3.75 (5.53)
Tangent-CFTED	Full	0.56 (0.26)	0.52 (0.27)	0.53 (0.36)	1.75 (1.62)
Re-ranked Runs					
Tangent-S	Full	**0.69 (0.23)**	0.52 (0.22)	**0.62 (0.30)**	-
Tangent-CFTED	Full	0.65 (0.26)	0.42 (0.27)	0.58 (0.32)	-

for the ′ (prime) metrics. The query time on the full collection is much longer than on the annotated collection and than the other systems working with the full collection. With the disjunctive queries there is a wider net cast so more formulas need to be scored, there is also room for optimization in the code base for improving efficiency. With how well the XY-PHOC model performed at similarity search we next will show how it performed at autocompletion.

5.2 Formula Autocompletion

In this section we will look at the results of running our autocompletion experiments. We hope to gain an understanding of how the levels of the embedding impact the exact retrieval of formulas in the index. We additionally will be studying the impact of 4 different input orders on the autocompletion metrics, to simulate the benefit of being unrestricted in which order you can input symbols for queries. The orders test are left to right (standard input order for strings), right to left, outside-in and middle-out.

To better understand the impact of the single levels have in the entire embedding the autocompletion queries are run for all 4 orders with just a single level of the embedding presented in Table 5.5. The mean *rsaved* is the new metric proposed to help capture amount of ranks saved by entering more symbols for a given query, and the in top-5 score represents the number of symbols

Table 5.5: Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With a single level where the number represents the number of sections in that level. Prime ($'$) levels are split vertically.

	$\mu(\sigma)$							
	Left to Right		**Right to Left**		**Outside In**		**Middle Out**	
Level	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**
1-(BoS)	0.774 (0.239)	10.8 (15.1)	0.778 (0.225)	11.2 (15.6)	0.819 (0.216)	8.19 (11.5)	0.760 (0.237)	12.0 (16.7)
2	0.805 (0.202)	8.87 (11.1)	0.807 (0.192)	8.55 (9.47)	0.868 (0.150)	6.31 (5.77)	0.824 (0.181)	9.11 (11.7)
$2'$	0.806 (0.205)	9.47 (13.2)	0.813 (0.196)	9.97 (14.9)	0.846 (0.182)	7.44 (11.0)	0.802 (0.205)	10.4 (15.0)
3	0.812 (0.197)	7.54 (8.53)	0.808 (0.198)	7.65 (7.4)	0.886 (0.130)	5.48 (4.73)	0.815 (0.187)	8.67 (10.9)
$3'$	0.816 (0.202)	9.07 (13.1)	0.822 (0.193)	9.28 (14.5)	0.853 (0.176)	7.08 (10.9)	0.810 (0.205)	9.28 (14.1)
4	0.811 (0.200)	7.34 (7.62)	0.810 (0.198)	7.58 (7.37)	0.898 (0.116)	4.99 (3.51)	**0.823 (0.186)**	**7.84 (8.76)**
$4'$	0.821 (0.196)	8.86 (12.9)	0.825 (0.187)	9.09 (14.2)	0.858 (0.171)	6.93 (10.7)	0.818 (0.199)	9.22 (14.3)
5	0.807 (0.203)	**7.15 (7.16)**	0.801 (0.209)	**7.48 (7.44)**	**0.905 (0.110)**	**4.71 (3.29)**	0.810 (0.193)	7.91 (8.85)
$5'$	**0.825 (0.200)**	8.43 (11.9)	**0.828 (0.190)**	8.52 (12.0)	0.863 (0.170)	6.71 (10.6)	0.819 (0.204)	8.70 (12.7)

required to get the target result in the top-5 results and the whole length is used if no number of symbols ever gets a rank with in the top-5 because this represents the full query being put in with no benefit of autocompletion for that query.

A consistent pattern is that for each level the outside-in input order is the best for rsaved and number of symbols for a rank in top-5. This is caused by the span of the query adding symbols to both sides gives. XY-PHOC uses splits that are based the bounding box formed by the symbols in the query, giving symbols on both ends of the formula gives the same split locations as the full formula. Outside-in also show a steady decrease in variance in both scores as the levels go deeper. This is the most stable ordering so heading additional splits to this order improves the scores and reduces variance between the scores. A surprising result is how strong the autocompletion results are for level 5, especially with the outside-in order with the highest rsaved ($\mu 0.905, \sigma 0.110$) and lowest number of symbols required for rank in top-5 ($\mu 4.71, \sigma 3.29$).

Another important insight looking at the impact of the single levels is that as the depth increases there is an improvement in the scores which is constant across most orders. Depending on the order there is also an improvement on using the vertical or horizontal split for the respective level.

To observe the impact of increasing the amount of data in the embedding we present the 4 input orders withe increasing depth of the embedding represented as the range of levels included, presented in Table 5.6. The expected pattern is observed as the depth in the embedding increases the autocomplete

Table 5.6: Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With Increasing Depth Represent As A Range Of Levels Included. The Ranges Include Both the Horizontal and Vertical levels (e.g 1-2 includes 1, 2, and 2′)

	$\mu(\sigma)$							
	Left to Right		**Right to Left**		**Outside In**		**Middle Out**	
Level	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**
1-(BoS)	0.774 (0.239)	10.8 (15.1)	0.778 (0.225)	11.2 (15.6)	0.819 (0.216)	8.19 (11.5)	0.760 (0.237)	12.0 (16.7)
1-2	0.815 (0.191)	9.04 (11.6)	0.822 (0.179)	9.26(12.6)	0.864 (0.154)	6.88 (9.93)	0.820 (0.182)	10.0 (14.4)
1-3	0.828 (0.177)	7.87 (7.9)	0.833 (0.169)	8.19(9.74)	0.877 (0.138)	6.16 (6.98)	0.835 (0.169)	8.6 (9.69)
1-4	0.838 (0.165)	7.54 (7.3)	0.842 (0.159)	7.85(9.84)	0.888 (0.119)	5.75 (5.0)	0.846 (0.157)	7.92 (9.02)
1-5	**0.841 (0.163)**	**7.19 (6.93)**	**0.844 (0.159)**	**7.32 (7.08)**	**0.895 (0.107)**	**5.5 (4.67)**	**0.849 (0.157)**	**7.63 (8.77)**

Table 5.7: Conjunctive Queries on Annotated ARQMath Data For Autocompletion Result For All 4 Input Orders With Different Membership Conditions.

	$\mu(\sigma)$							
	Left to Right		**Right to Left**		**Outside In**		**Middle Out**	
Condition	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**
Default	0.841(0.163)	7.19(6.93)	0.844(0.159)	7.32(7.08)	0.895(0.107)	5.5(4.67)	0.849(0.157)	7.63(8.77)
Top-Left	0.846(0.161)	7.06(6.88)	0.846(0.156)	7.19(6.79)	0.901(0.104)	5.23(4.37)	0.850(0.160)	7.42(7.32)
Centroid	**0.847(0.161)**	7.0(6.75)	**0.852(0.153)**	6.98(6.66)	**0.903(0.101)**	**5.1(4.25)**	0.853(0.155)	**7.08(8.25)**
Vert. Center	0.845(0.164)	**6.86(6.8)**	0.851(0.157)	**6.95(6.77)**	0.902(0.102)	5.12(4.29)	**0.855(0.154)**	7.09(8.33)

results improve. The outside-in condition is the strongest for all ranges. A surprising result is that, while the trend of more information improves the embedding applies in this test, compared to the single levels just the level 5 is better than the range of levels 1-5. This could mean that there is too much redundant information in the full range, making it less discriminating of an embedding.

To help get a better understanding of making the embedding more discriminating, the 3 previously proposed split membership conditions are tested on all 4 input orders. The results on this experimentation are presented in Table 5.7. All alternative conditions show improvement over the default condition, which may have too much redundant information for the exact retrieval task. While it is possible this redundant information is better at retrieving information if there is more variance in the layout of symbols, when the symbol placement is exactly the same the more discriminating conditions produce better results. The centroid condition produced the best results, also beating the single level 5 results, but the other conditions show very similar improvements with the differences between the alliterative membership conditions being small.

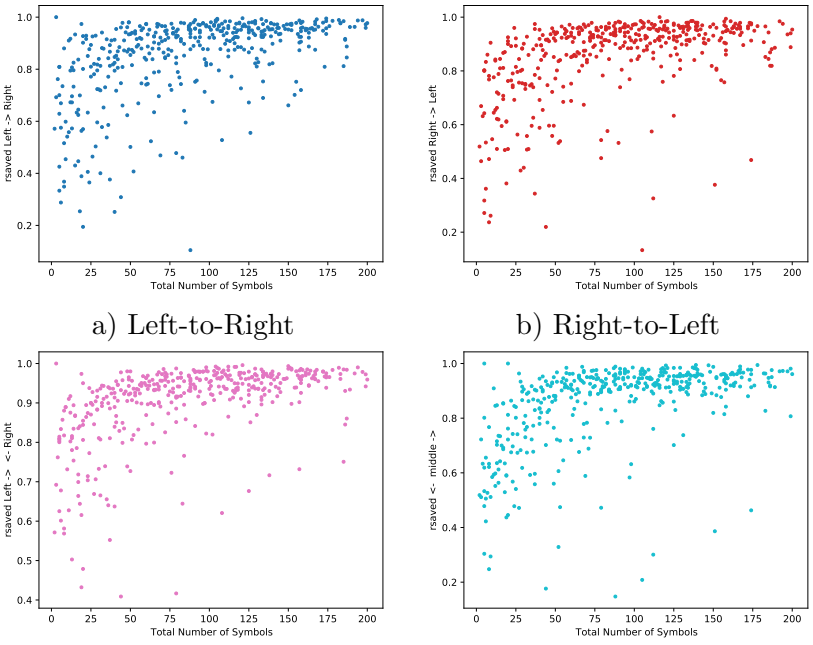
Table 5.8: Autocompletion Benchmark on ARQMath Annotated.

System	$\mu(\sigma)$ Left to Right rsaved	In Top-5	Query Time (s)
XY-PHOC	0.847 (0.161)	**7.0 (6.75)**	0.038 (0.015)
PHOC	0.844 (0.159)	7.29 (7.09)	**0.033 (0.012)**
Tangent-V	**0.863 (0.240)**	9.91 (28.9)	1.75 (2.06)

To get a sense of how well the autocompletion works we compare the results of the best XY-PHOC System for autocompletion with Tangent-V as a comparison against another visual based system. Additionally, to get a sense of how the addition of vertical splits improves the results we compare the autocompletion of the PHOC system with just the horizontal splits. These benchmarks are presented in Table 5.8. While the best XY-PHOC system has the top scores, it is only marginally better than the baseline PHOC System. The baseline PHOC System with conjunctive queries also have the fastest query time ($\mu 0.033, \sigma 0.012$), this could show that adding the vertical cuts is not as useful in the autocompletion task. We believe since the additional vertical splits showed stronger results in the similarity retrieval task, that they will still provide an important advantage in autocompletion with more variation in the input data. The query times for all three models (Tangent-V, XY-PHOC, and PHOC) were gathered using a Python implementation running on a desktop Linux system with and Intel i7-8700K CPU (3.7GHz) with 32GB RAM.

To better understand the trend of the rsaved metric we graphed the rsaved values based on the number of symbols in the target formula in Figure 5.1. With the larger formulas the rsaved approaches 1.0 because there are less formulas with those lengths in the collection and even less with all the same symbols. What is important to recognize is the speed at which convergence happens. The Outside-in as the tightest curve in the scatter plot that is almost logarithmic. There are also significantly fewer outliers in the Outside-in condition which accounts for the low variance observed for rsaved with the outside-in input order.

To understand how the number of symbols entered and how the percent of



a) Left-to-Right b) Right-to-Left

c) Outside-in d) Middle-out

Figure 5.1: `rsaved` Values By Target Formula Size Using a) Left-to-Right b) Right-to-Left c) Outside-in d) Middle-out Input Order for the Centroid Membership condition. The Quicker Convergence to 1 and lower number of outliers in the Outside-in condition shows how its scores across the board are better.

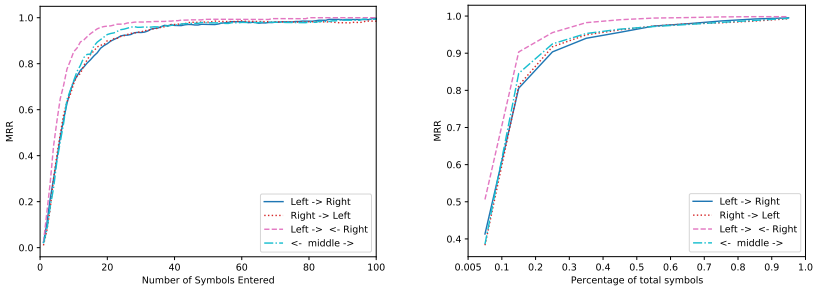


Figure 5.2: Mean Reciprocal Rank (MRR) of all 4 input orders by the number of symbols entered and the % of target formula symbols entered for the Centroid Membership condition. The Quicker Convergence to 1 in the Outside-in condition shows how its scores across the board are better.

total symbols from the target effects the retrieval we graphed the Mean Reciprocal Rank (MRR) to this conditions in Figure 5.2. Both of these line graphs show all 4 input conditions. In Figure 5.2.a the graph ends at 100 symbols input because beyond this point the MRR was just continuing to approach 1. In Figure 5.2.b the x-axis represents bins of 10%, where all percents are plotted in the center of the range. Graphing over both symbols entered and percent of target enter, the Outside-in ($left \rightarrow\leftarrow right$) condition starts off better and approaches 1 the fastest. This shows how outside-in is the most stable input order when using the XY-PHOC model. The graphs in both Figure 5.1 and Figure 5.2 for all other experimental conditions are presented in Appendix B, C.

To get a better understanding of the behavior of the autocompletion we present Table5.9 and Table 5.10. In Table5.9 there are 2 queries, the left is given in a outside-in order and the right shown the Pythagorean theorem example proposed in 1.1 which would not generate a valid SLT. With just 4 and 5 symbols respectively the target formula is at rank 1. These cases work well for autocompletion because the left and right sides are bounded. Looking at the other returned formulas the symbols have a similar placement of the given symbols and in all retrieved results all the symbols in the query are contained.

Table 5.9: XY-PHOC Conjunctive Query in the Annotated ARQMath Collection, Where the Target is Returned at Rank 1

	XY-PHOC		
Rank	1^3	6^3	$^2 + ^2 = ^2$
1	$1^3 + 5^3 + 5^3 = 2^3 + 3^3 + 6^3$		$a^2 + b^2 = c^2$
2	1 1 2 3 3 4 5 6 6 7		$A^2 + B^2 = a^2$
3	$(-36)(-1) = 36$		$A^2 + B^2 = C^2$
4	$1 - 0.39764$		$a^2 + b^2 = c^2.$
5	$(-12)(-3) = 36$		$2^{2n+1} = 2$

In Table 5.10 there are 2 queries, both using a left to right input order. For both of these queries, the left having 4 and Right having 5 symbols entered their target is at rank 5. This helps to identify a problem case when the index is full of too many similar formulas. With the queries given there is no way to better identify the exact formula we want to match. While being in the top 5 is still helpful to users, this pattern can also push the exact formula we want to match out of the top 5, since there are more than 4 formulas in the collection that start of with $\int_0^\infty$ and $\frac{\partial f}{\partial x}$.

To verify that the model holds up on the entire ARQMath collection we have run a smaller number of the autocomplete queries to get a sense of the impact of the full collection scale. In the current implementation query times dramatically increase; changes in the system architecture could yield faster searches. To see how the model compares on the whole collection the first 19 queries from the autocomplete collection and compare it to the first 19 queries run on the annotated collection. In Table 5.11 we see all 4 input orders for both models. The scores are lower when using the much larger collection because there are more formulas that are candidates. To get an understanding of what is happening we look at the Pythagorean theory query from the annotated set and the full collection.

In Table 5.11 we can see a query that worked well for the annotated set

Table 5.10: XY-PHOC Conjunctive Query in the Annotated ARQMath Collection, Where the Target is Returned at Rank 5

	XY-PHOC	
Rank	$\int_0^\infty$	$\frac{\partial f}{\partial x}$
1	$\int_0^\infty$	$\frac{\partial f}{\partial x^i}$
2	$\int_0^\infty \frac{dx}{x^2}$	$\frac{\partial f}{\partial x^1}$
3	$\int_0^\infty \frac{\sin t}{t}$	$\frac{\partial f}{\partial x_2}$
4	$\int_{x=0}^\infty \frac{dx}{xe^x}.$	$\frac{\partial f}{\partial x_1}$
5	$\int_0^\infty \frac{\sin x}{x}$	$\frac{\partial f}{\partial x_3}$

Table 5.11: Comparing Conjunctive Queries on Annotated ARQMath and Full ARQMath collection of the first 19 queries of the Autocomplete Benchmark.

	$\mu(\sigma)$							
	Left to Right		**Right to Left**		**Outside In**		**Middle Out**	
Condition	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**	**rsaved**	**In Top-5**
XY-PHOC-Annotated	**0.595(0.177)**	**2.7(1.35)**	**0.598(0.234)**	**2.95(1.47)**	**0.732(0.116)**	**2.15(0.726)**	**0.591(0.194)**	**3.05(1.69)**
XY-PHOC-Full	0.281(0.130)	4.95(1.8)	0.294(0.150)	5.0(1.67)	0.346(0.150)	4.3(1.19)	0.290(0.129)	4.9(1.7)

Table 5.12: XY-PHOC Conjunctive Query on the Annotated ARQMath Collection and Full ARQMath Collection

	XY-PHOC-Annotated	XY-PHOC-Full
Rank	$^2 + ^2 = ^2$	$^2 + ^2 = ^2$
1	$\boxed{a^2 + b^2 = c^2}$	$2 + 2 = 2^2$
2	$A^2 + B^2 = a^2$	$2 + 2 + 2 = 2$
3	$A^2 + B^2 = C^2$	$2 + 2 = 2$
4	$a^2 + b^2 = c^2.$	$2 + 2 = 22$
5	$2^{2n+1} = 2$	$+ 2 = 2$

is not able to get the target formula in full collection. This is because in the full collection there are several orders of magnitude more formulas. With so many more formulas there are many more formulas with the same XY-PHOC embedding. This pushes the correct formula out of the range of the top 5.

5.3 Summary

The XY-PHOC Model is able to produce competitive results in the similarity search of ARQMath Task 2 and strong results in the autocomplete task. From our experimentation the condition that performed the best similarity search is not the same as the strongest system for autocompletion. In similarity search there is an improvement of score when using all 5 levels of the embedding but for autocompletion the best configuration was the single level 5. Changing the membership condition for being within a split improves scores in both tasks and more experiments can be performed to find the best condition. With a single level performing the best for autocompleteion, it is possible that all the redundant information is not as useful for autocompletion, but additional experiments could be run with different combinations of levels to see if a combination of levels (e.g., levels 5 and $5'$) is better than the single level.

Chapter 6

Conclusion

In this thesis we present an expansion on the PHOC Embedding, XY-PHOC, for the task for math formula retrieval. XY-PHOC is a spatial embedding which captures the labels and relative positions of symbols in space. We proposed an efficient way to index XY-PHOC, with an efficient way to score candidate formulas. In this chapter we will summarize our findings and discuss future work.

6.1 Summary of Findings

The XY-PHOC model using our proposed inverted index be easily and efficiently searched with conjunctive and disjunctive queries. Using the reduced bitstring representation, storage of the index is reduced in size and allows for fast computation of scores.

The strong results on the ARQMath retrieval task 2 show that a simple spatial representation is strong for similarity retrieval. With additional reranking or used in conjunction with other systems can produce even better results than the current state-of-the-art systems.

Across all autocompletion experiments the outside-in input order showed the best results. This could mean that visual based queries are more powerful for finding exact formulas faster. This could be a product of how putting symbols on the left and right of the query helps fix the width of the query which helps for the relative positioning captured by XY-PHOC and future experiments should be run with a fixed width canvas that queries can be

normalized to fit.

Through our experiments we showed that based on which task you are doing, a different balance must be met between redundant information making the embedding more generalized versus improving uniqueness to improve distinguishability. Using an increased depth improved the results for similarity retrieval but for autocompletion using a single level was the best before studying the impact of the different split membership conditions.

The XY-PHOC model is an efficient and effective system for math formula retrieval, for both similarity retrieval and autocompletion. The benefit of using the XY-PHOC model is the ease of use that it will have for novice users as they will not need to know anything about LaTeX to effectively use the search system nor will the need to be able to provide enough of the formula to produce a valid SLT or OPT. The added power visual systems give for math formula retrieval in combination with the efficient scoring this embedding gives makes it useful for at scale math information retrieval systems.

6.2 Future Work

In future work, we believe that using the Inverse Document Frequency (IDF) for symbols as a term weight during scoring may be beneficial, as well as extending the PHOC encoding to include n-grams [22]. To help speed up our conjunctive queries the use of skip lists would reduce the number of postings that have to be looped through.

A user interface needs to be made and integrated with to better utilize the embedding. This includes an easy way for symbols to be inserted in space. Making it so the matching location can be a parameter controlled at run time with the use of radio buttons or a drop down to select where to anchor the symbols in the formula. As well a user test should be run to verify how useful the added input methods XY-PHOC is capable of.

Additionally, a collection should be gathered for math autocompletion. Currently using the entire ARQMath collection for autocomplete does not make sense. Instead a collection generated from user logs of the most common queries would make more sense as using common queries for autocompletion would help to reduce the effort for users without having a collection full of formulas that might not be relevant to the average user.

An area that can be explored in future work, is how handwritten queries

perform in the XY-PHOC model. The configuration that worked best on typeset queries might not work the best for handwritten queries. Handwritten queries have more variation and the added redundancy from some of the experimentation conditions will work better for these queries. With handwritten queries it will be easier to also to take advantage of the alternative input order for the math formulas.

Finally, our queries are simply symbols in space, and there is nothing math-specific about our approach. We expect that this technique may be beneficial for retrieving other graphical objects by appearance (e.g., tables, figures, plots, etc.). It is also possible that you could embed formula trees in XY-PHOC embeddings, where splits are made at node depths and between siblings rather than in space.

Bibliography

[1] Mohammad Reza Ameri, Michael Stauffer, Kaspar Riesen, Tien D. Bui, and Andreas Fischer. Graph-based keyword spotting in historical manuscripts using Hausdorff edit distance. *Pattern Recognition Letters*, 121:61–67, April 2019.

[2] Dena Bazazian, Dimosthenis Karatzas, and Andrew D. Bagdanov. Word Spotting in Scene Images Based on Character Recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1953–19532, June 2018. ISSN: 2160-7516.

[3] Kenny Davila, Ritvik Joshi, Srirangaraj Setlur, Venu Govindaraju, and Richard Zanibbi. Tangent-V: Math Formula Image Search Using Line-of-Sight Graphs. In Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra, editors, *Advances in Information Retrieval*, volume 11437, pages 681–695. Springer International Publishing, Cham, 2019. Series Title: Lecture Notes in Computer Science.

[4] Kenny Davila and Richard Zanibbi. Layout and Semantics: Combining Representations for Mathematical Formula Search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1165–1168, Shinjuku Tokyo Japan, August 2017. ACM.

[5] Kenny Davila and Richard Zanibbi. Visual Search Engine for Handwritten and Typeset Math in Lecture Videos and LATEX Notes. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 50–55, August 2018.

[6] David Fernández, Simone Marinai, Josep Lladós, and Alicia Fornés. Contextual word spotting in historical manuscripts using Markov logic networks. In *Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing*, HIP '13, pages 36–43, New York, NY, USA, August 2013. Association for Computing Machinery.

[7] Andreas Fischer, Andreas Keller, Volkmar Frinken, and Horst Bunke. HMM-based Word Spotting in Handwritten Documents Using Subword Models. In *2010 20th International Conference on Pattern Recognition*, pages 3416–3419, August 2010. ISSN: 1051-4651.

[8] Volkmar Frinken, Andreas Fischer, R. Manmatha, and Horst Bunke. A Novel Word Spotting Method Based on Recurrent Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):211–224, February 2012. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[9] Suman K. Ghosh and Ernest Valveny. R-PHOC: Segmentation-Free Word Spotting Using CNN. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 801–806, November 2017. ISSN: 2379-2140.

[10] Angelos P. Giotis, Giorgos Sfikas, Basilis Gatos, and Christophoros Nikou. A survey of document image word spotting techniques. *Pattern Recognition*, 68:310–332, August 2017.

[11] Marti A. Hearst. *Search User Interfaces*. Cambridge University Press, USA, 1st edition, 2009.

[12] Eugene Kharitonov, Craig Macdonald, Pavel Serdyukov, and Iadh Ounis. User model-based metrics for offline query suggestion evaluation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '13, pages 633–642, New York, NY, USA, July 2013. Association for Computing Machinery.

[13] Parag Mali, Puneeth Kukkadapu, Mahshad Mahdavi, and Richard Zanibbi. ScanSSD: Scanning Single Shot Detector for Mathematical Formulas in PDF Document Images. *arXiv:2003.08005 [cs]*, March 2020. arXiv: 2003.08005.

[14] Behrooz Mansouri. Learning to Rank for Mathematical Formula Retrieval. SIGIR '21, page 10, 2021.

[15] Behrooz Mansouri, Douglas W Oard, and Richard Zanibbi. DPRL Systems in the CLEF 2020 ARQMath Lab. page 12.

[16] Behrooz Mansouri, Shaurya Rohatgi, Douglas W. Oard, Jian Wu, C. Lee Giles, and Richard Zanibbi. Tangent-CFT: An Embedding Model for Mathematical Formulas. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '19, pages 11–18, New York, NY, USA, September 2019. Association for Computing Machinery.

[17] Mohamed Mhiri, Christian Desrosiers, and Mohamed Cheriet. Convolutional pyramid of bidirectional character sequences for the recognition of handwritten words. *Pattern Recognition Letters*, 111:87–93, August 2018.

[18] M. Okamoto and Bin Miao. Recognition Of Mathematical Expressions by Using the Layout Structures of Symbols. *Proc. First Int'l Conf. Document Analysis and Recognition*, 1:242–250, 1991.

[19] Tony M. Rath and R. Manmatha. Word spotting for historical documents. *International Journal of Document Analysis and Recognition (IJDAR)*, 9(2):139–152, April 2007.

[20] Shaurya Rohatgi, Wei Zhong, Richard Zanibbi, Jian Wu, and C. Lee Giles. Query Auto Completion for Math Formula Search. *arXiv:1912.04115 [cs]*, December 2019. arXiv: 1912.04115.

[21] Yang Song, Dengyong Zhou, and Li-wei He. Post-ranking query suggestion by diversifying search results. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, page 815–824, New York, NY, USA, 2011. Association for Computing Machinery.

[22] Sebastian Sudholt and Gernot A. Fink. PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 277–282, Shenzhen, China, October 2016. IEEE.

[23] Sebastian Sudholt and Gernot A. Fink. Evaluating Word String Embeddings and Loss Functions for CNN-Based Word Spotting. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 493–498, November 2017. ISSN: 2379-2140.

[24] Sebastian Sudholt and Gernot A. Fink. Attribute CNNs for word spotting in handwritten documents. *International Journal on Document Analysis and Recognition (IJDAR)*, 21(3):199–218, September 2018.

[25] Shunyi Yao, Ying Wen, and Yue Lu. HoG based two-directional Dynamic Time Warping for handwritten word spotting. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 161–165, August 2015.

[26] Richard Zanibbi, Douglas W. Oard, Anurag Agarwal, and Behrooz Mansouri. Overview of ARQMath 2020: CLEF Lab on Answer Retrieval for Questions on Math. In Avi Arampatzis, Evangelos Kanoulas, Theodora Tsikrika, Stefanos Vrochidis, Hideo Joho, Christina Lioma, Carsten Eickhoff, Aurélie Névéol, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, volume 12260, pages 169–193. Springer International Publishing, Cham, 2020. Series Title: Lecture Notes in Computer Science.

[27] Richard Zanibbi and Li Yu. Math Spotting: Retrieving Math in Technical Documents Using Handwritten Query Images. In *2011 International Conference on Document Analysis and Recognition*, pages 446–451, Beijing, China, September 2011. IEEE.

[28] Liang Zheng, Yi Yang, and Qi Tian. SIFT Meets CNN: A Decade Survey of Instance Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5):1224–1244, May 2018. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Appendices

Appendix A

Autocompletion Benchmark Query Set

In this Appendix the we have included the render formulas used for autocompletion.On the following pages are the visual ids and rendered formulas for the formulas used for the autocomplete experiments. The formulas were rendered using the MathJax API.

580216 $(-2)^+$

19179 $\sqrt{-1} = i$

5400 $h \to 0$

2798007 $\alpha = \sqrt{-1}$

5545460 $union$

68247 $2'$

2956795 $x \Rightarrow 1 = 1$

1166077 $\frac{t^n}{n!}$

4247472 $X^+ \geq 0$

111745 $b + 1$

2178452 $S : A \to P(B)$

8458002 $\sum_{k=0}^{\infty} a_k = L$

1148475 $n \equiv 3 \mod 6$

2329990 $1 - \frac{\sqrt{3}}{2}$

2369233 $-1 \mod 2 \equiv 1$

5004786 $\sqrt{3}(1 + 1/2)$

277577 $\sum_{n=0} a_n x^n$

1996187 $w = e^{2i\pi/3}$

62113 $n = dq + r$

2320413 $\frac{\partial f}{\partial x_3}$

2711889 $y = xp + \frac{1}{2}p^2$

3012963 $\pi : z = ax + by + c$

8230323 $2^{n-1} \cdot \frac{1}{2^n} = \frac{1}{2}$

6249673 $\lim_{x \uparrow \infty} \frac{x^m}{e^x} = 0$

596332 $\sum_{n=1}^{\infty} \frac{\sin(n)}{n}$

8259101 $\sum_a \frac{\chi(a)}{\mathrm{N}(a)^s}$

2213348 $x^2 + (\dot{x})^2 = a^2$

1535348 $\frac{\pi(\sqrt{2}-1)}{2\sqrt{2}}.$

3070821 $\sum_{i=0}^{x} \sum_{j=0}^{i+1} 1$

7743106 $\int_0^{\infty} \frac{|\sin x|}{x^2}$

5169991 $\frac{1}{2}(y')^2 = -\frac{1}{y} + C$

6180034 $\sqrt{(-1)(-1)} \neq \sqrt{-1} \cdot \sqrt{-1}$

947595 $P(z) = \sum_{n=0}^{N} a_n z^n$

4227963 $y' = \frac{x}{2y\ln(y)+y}$

7097065 $1^3 + 5^3 + 5^3 = 2^3 + 3^3 + 6^3$

5604479 $\frac{D't^2+F'}{(t+1)^2},$

1412688 $\sum_{k=0}^{n} a_k < \sum_{k=0}^{n+1} a_k$

8246812 $$\int_0^{\pi} \sin 2x \sin x \, dx$$

808021 $\forall x, y(\neg(xRy \to yRx))$

6191071 $2x - 1 = 0 \Rightarrow 2x = 1 \Rightarrow x = \frac{1}{2}$

8322352 $\sum_{i=0}^{0} i = 0 = \frac{1}{2}0(0+1)$

2971158 $A^2B = AAB = -ABA = BAA = BA^2$

1477029 $(x^T Ax)^T = x^T A^T x = -x^T Ax$

4809176 $f(x) = \frac{x^2+ax+1}{x^2+x+1}$

4118079 $\binom{n}{s} + \binom{n}{s+1} = \binom{n+1}{s+1}$

4199743 $\sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^w$

6007670 $\int_0^1 \frac{\ln(x)\ln^2(1-x)}{x} dx$

8691009 $\sum_{k=0}^{n} (-1)^k \binom{n}{k} B_{n-k}$

1812552 $(-1-\xi)(-1-\xi) = 1 + 2\xi + \xi^2$

1188924 $\mathcal{I}_{\mathrm{n}} := \int_{-\infty}^{\infty} \frac{\sin(x)}{x^n} \, \mathrm{d}x$

8861763 $21269 * 37469 - 15190 * 52464 = 1$

9270748 $n \times 6 = \operatorname{lcm}(6,8) = \frac{6 \times 8}{\gcd(6,8)}$

5597696 $\sigma(\Sigma^n) \subset \sigma\left(\bigcup_{i=1}^{n} pr_i^{-1}(\Sigma)\right)$

6367323 $I_1 = \int_0^1 \frac{1}{t} \ln\left(\frac{1+\sqrt{1-t^2}}{2}\right) dt,$

4632826 $x = \frac{1}{2}(1 + \sqrt{101} - \sqrt{98 - 2\sqrt{101}})$

7441598 $2^{n+2} \geq 2^{n+1} + 2^n + 2 \geq 2^{n+1} + 2^n + 1$

3388583 $v(n) \Sigma_{k=1}^{m-r} (-1)^k \binom{m-r}{k}$

2081962 $3(x^2 + y^2 + z^2) \geq (x + y + z)^2 \geq k^2$

5686258 $\sum_{k=1}^{n} (-1)^k \binom{n}{k} \log(1 + 2k)$

9013464 $x^T Ay = (x^T Ay)^T = y^T A^T x = y^T Ax;$

6171115 $0 \times \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{1}{3} = \frac{2-1}{4-1}.$

4723412 $\sum_{k=0}^{n} (-1)^{k+1} k^2 = (-1)^{n+1} \binom{n+1}{2}$

3001953 $f(x) = \sum_{j=0}^{k-1} \cos\left(\frac{(2j+1)x}{2k}\right)$

3554913 $77976^2 = 1026^3 + 1710^3 = 228^3 + 1824^3$

1986267 $\frac{(a+b+c)^3 - 27abc}{a^3+b^3+c^3-3abc}$

1488793 $1 = \frac{2}{2} = \frac{1+1}{1+1} \stackrel{*}{=} \frac{1}{1} + \frac{1}{1} = 1 + 1 = 2.$

168807 $\frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{5}{6} = 3\frac{1}{3}.$

4308484 $224312161 + 372170768 = 24423^2$

6274264 $x - \frac{x^3}{3 \cdot 1!} + \frac{x^5}{5 \cdot 2!} - \frac{x^7}{7 \cdot 3!} + \cdots$

3442126 $\frac{1}{(1-3^{1/3}+3^{1/6})(1+3^{1/3}-3^{1/6})},$

4341285 $\frac{\sin\left(n+\frac{1}{2}\right)x}{\sin\frac{x}{2}} = 1 + 2\sum_{k=1}^{n} \cos kx$

7560292 $x^3 + x - x^2(x-1) = x^3 + x - (x^3 - x^2) = x^2 + x,$

4976234 $\lim_{n\to\infty} a_n^{1/n} = \left(\lim_{n\to\infty} a_n\right)^{\lim_{n\to\infty} \frac{1}{n}}$

1063943 $\int_0^\infty \frac{1-\cos(x)}{x^2} dx = \int_0^\infty \frac{\sin(x)}{x} dx$

102550 $(1+y)^{(k+1)n} \geq (1+ny)^{k+1} \geq n^{k+1} y^{k+1}$

764453 $Y = Y^1 \frac{\partial}{\partial x^1} + Y^2 \frac{\partial}{\partial x^2} + Y^3 \frac{\partial}{\partial y^1} + Y^4 \frac{\partial}{\partial y^2}$

2117894 $\sqrt[3]{7 + 5\sqrt{2}} + \sqrt[3]{7 - 5\sqrt{2}} = (1 + \sqrt{2}) + (1 - \sqrt{2}) = 2.$

2235267 $2017W(2017) - 1 = 2016 \cdot 2015 \cdot \cdots \cdot 1 = 2016!$

8286760 $(x_1, y_1) + (x_2, y_2) = (2x_1 - 3x_2, y_1 - y_2)$

2593904 $0, \frac{1}{2}, 1, \frac{2}{3}, \frac{1}{3}, 0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1, \ldots$

1331042 $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots = \sum_{i=0}^{\infty} \frac{x^i}{i!}.$

2045376 $\int_0^1 \frac{\sum_{j=0}^{2n-1} x^{2j}}{(1+x^{2n})(-\ln x)^{\frac{1}{2}}} dx \ldots ?$

1563316 $[1], [4], [9], [1, 4], [1, 9], [4, 9], [1, 4, 9]$

9058562 $10(a^4 + b^4 + c^4 - 3abc) \geq 17(a^3 + b^3 + c^3 - 3abc)$

2861323 $1, \sin(x), \sin(2x), \sin(3x), \ldots, \sin(nx)$

7712246 $\int_0^\infty \frac{\sin x}{x} dx = \int_0^\pi \frac{\sin x}{x} dx + \int_\pi^\infty \frac{\sin x}{x} dx$

2536825 $\sum_{k=0}^{n-1} \cos(\frac{2k\pi}{n}) = 0 = \sum_{k=0}^{n-1} \sin(\frac{2k\pi}{n})$

6092507 $\sum_{i=1}^{n+1} \frac{1}{i(i+1)} = \frac{n+1}{(n+1)+1} = \frac{n+1}{n+2}.$

5651048 $\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \ldots + \binom{n}{k} = \binom{n+1}{k+1}$

3777978 $\sum_{n=1}^{N} \sin nx = \frac{\cos\frac{x}{2} - \cos(N+\frac{1}{2})x}{2\sin\frac{x}{2}}$

6715164 $n \geq 2 \text{ and } n \not\equiv 0 \pmod{10} \text{ and } S(n) = S\left(n^2\right) = \cdots = S\left(n^k\right)$

6519069 $\sum_{k=1}^{n} n^{1/k} = n + H_n \log(n) + \sum_{m\geq 2} \frac{H_n^{(m)}}{m!} (\log n)^m.$

1511104 $\left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^k = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n} \neq 1$

579643 $T_n = \sum_{k=n}^{n^2} \frac{1}{k} = \frac{1}{n^2} + \sum_{j=1}^{n-1} \sum_{k=jn}^{(j+1)n-1} \frac{1}{k}$

5604323 $\frac{F(x) - a_0 - a_1}{x^2} = a_2 + a_3 x + a_4 x^2 + \ldots = \sum_{n=0}^{\infty} a_{n+2} x^n$

100900 $a^3 + b^3 + c^3 = (a+b+c)(a^2 + b^2 + c^2 - ab - bc - ca) + 3abc$

6498907 $g(x) = \sum_{n=0}^{\infty} b_n x^n = 0 + x + 0 - \frac{1}{6}x^3 + 0 + \frac{1}{120}x^5 + \ldots$

8357719 $50 \to 75 \to 87.5 \to 93.75 \to 96.875 \to 98.4375 \to 97.65625$

8569443 $x_n = (n \sin\frac{1}{n}, n \sin\frac{1}{n}, \ldots, n \sin\frac{1}{n}, 0, 0, \ldots)$

21847 $(1 - 0.32)(1 - 0.64)(1 - 0.81)(1 - 0.89)(1 - 0.94).$

8260680 $\oint P(x,y)dx = -\iint \frac{\partial P}{\partial x} dxdy, \oint Q(x,y)dy = \iint \frac{\partial Q}{\partial y} dxdy$

3027868 $I_{t-}^{1-\alpha} \eta_t(x) = I_{t-}^{1-\alpha} D_{t-}^{1-\alpha} (f - f(t))(x) = f(x) - f(t)$

8266677 $\{\}, \{1\}, \{\{2\}, \{3\}\}, \{\{1, 2\}, \{1, 3\}\}, \{2, 3\}, \{1, 2, 3\}$

1704028 $p_{5,1} : p_{5,2} : p_{5,3} : p_{5,4} : p_{5,5} = 1 : 105 : 1050 : 2100 : 840$

1672024 $\int_0^1 \frac{\ln\left(1+x^{2+\sqrt{3}}\right)}{1+x} dx = \frac{\pi^2}{12}\left(1 - \sqrt{3}\right) + \ln\left(1 + \sqrt{3}\right)\ln 2.$

6115691 $\frac{\partial g}{\partial x} = yf(x,y) + xy\frac{\partial f}{\partial x}, \frac{\partial g}{\partial y} = xf(x,y) + xy\frac{\partial f}{\partial y}$

6710287 $\frac{1}{(x^2-1)^2} = \frac{1}{4}\left(\frac{1}{(x-1)^2} - \frac{2}{(x-1)(x+1)} + \frac{1}{(x+1)^2}\right)$

8983775 $\sum_{n=1}^{\infty} a_n \cdot x^n = (1 + x + x^2 \ldots) \cdot \ldots \cdot (1 + x + x^2 + \ldots) = \frac{1}{(1-x)^k}$

8974123 $\frac{1}{a} \sum_{n=1}^{\infty} a^n \int_0^{2\pi} x^2 \sin(nx) dx = \frac{4\pi^2 \ln(1-a)}{a}, \ |a| < 1$

6729544 $\lim_{N\to\infty} \int_{-N}^{N} f(u) e^{-2\pi i n u} du = \int_{-\infty}^{\infty} f(u) e^{-2\pi i n u} du.$

8628194 $x \succeq y \iff x_i \geq y_i \quad (i \in \{1, \ldots, n\}) X \succeq Y \iff X - Y \text{ is positive semidefinite}$

6023403 $J = \{x \in (a,b) : f(x+0), f(x-0) \text{ both exist but are not equal }\}$

4785775 $000:0, 001:1, 010:1, 011:2, 100:1, 101:2, 110:2, 111:3.$

6968671 $\int\limits_E \varliminf\limits_{n\to\infty} f_n = \varliminf\limits_{n\to\infty} \int\limits_E g_n = \varliminf\limits_{n\to\infty} inf \int\limits_E g_n \leq \varliminf\limits_{n\to\infty} \int\limits_E f_n$

8988418 $[S_k = \sin(x) + \sin(2x) + \sin(3x) + \cdots + \sin(nx) = \sum_n \sin(n(x))]$

8887322 $I = \int_0^1 \frac{\ln(1-x)\ln^2(1+x)}{x} dx \qquad J = \int_0^1 \frac{\ln^2(1-x)\ln(1+x)}{x} dx$

2768536 $(\frac{\partial x}{\partial y})^2 \frac{\partial^3 y}{\partial x^3} + 3\frac{\partial^2 y}{\partial x^2}\frac{\partial^2 x}{\partial y^2} + (\frac{\partial y}{\partial x})^2 \frac{\partial^3 x}{\partial y^3} = 0$

1878544 $I = \int \frac{2t}{(t^2+1)^2} \log(t)\, dt = -\frac{\log(t)}{t^2+1} + \int \frac{dt}{t(t^2+1)}$

8880653 $(1) \qquad \log \varphi_n(\lambda) = \sum_{k=1}^n \log\left(1 - \frac{1}{k}\left(1 - \cos\frac{\lambda k^{3/2}}{a_n^{1/2}}\right)\right).$

8685026 $1 \cdot (5/6)^3 - 1 \cdot 3(1/6)(5/6)^2 - 2 \cdot 3(1/6)^2(5/6) - x \cdot (1/6)^3$

4290458 $\frac{1^3}{1^4+4} - \frac{3^3}{3^4+4} + \frac{5^3}{5^4+4} - \ldots + \frac{(-1)^n(2n+1)^3}{(2n+1)^4+4}$

6379853 $\int\limits_0^1 \frac{\ln(1+x)}{1+x^2}dx - \frac{\pi}{8}\ln 2 = \int\limits_0^{\frac{\pi}{4}} \ln(1+\tan\theta)d\theta - \frac{\pi}{8}\ln 2 =$

7559841 $\frac{\partial^3 Q}{\partial x \partial y^2} = -\frac{1}{D}\left(-(A+B)\frac{\partial^3 Q}{\partial x^3} + (E+B)\frac{\partial^4 P}{\partial x^4} - F\frac{\partial^2 P}{\partial y^2}\right)$

8446688 $\operatorname{Spec}(S) = \{\text{prime ideals } P \subset S\} \xrightarrow{\phi^*} \{\text{prime ideals } Q \subset R\} = \operatorname{Spec}(R)$

8333907 $\frac{d(C_1,C_2)}{n_1 n_2} = \frac{d(C_1,C_1)}{n_1^2} + \frac{d(C_2,C_2)}{n_2^2} + 2||m_1 - m_2||^2.$

4813742 $\sum_{k=0}^{n-1}(4k+1) = 4\sum_{k=0}^{n-1} k + \sum_{k=0}^{n-1} 1 = 4\frac{n(n-1)}{2} + n = 2n^2 - n.$

968811 $\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$

2205691 $\sum_{m=k}^{n+1} \binom{m}{k} = \sum_{m=k}^{n} \binom{m}{k} + \binom{n+1}{k} = \binom{n+1}{k+1} + \binom{n+1}{k} = \binom{n+2}{k+1}$

8973652 $\emptyset, \{0\}, \{1\}, \{0,1\}, \{2\}, \{0,2\}, \{3\}, \{0,3\}, \{1,2\}, \{0,1,2\}, \ldots$

8421579 $\frac{1}{4\sin^2 18^\circ} - \frac{1}{2\sin 18^\circ} - 1 = \frac{1 - 2\sin 18^\circ - 4\sin^2 18^\circ}{4\sin^2 18^\circ}$

1271185 $= \frac{1}{1-x}\left(\sum_{a=1}^{s+t-1} 1 - \sum_{a=1}^{s+t-1} x^{a+1}\right) = \frac{1}{1-x}\left(s + t - 1 + x - \sum_{a=1}^{s+t} x^a\right)$

3250205 $\frac{\partial v_1}{\partial x_{j_1}} = \frac{\partial v_1}{\partial \varphi_1}\frac{\partial \varphi_1}{\partial x_{j_1}} + \frac{\partial v_1}{\partial \varphi_2}\frac{\partial \varphi_2}{\partial x_{j_1}} + \frac{\partial v_3}{\partial \varphi_1}\frac{\partial \varphi_3}{\partial x_{j_1}}.$

3351018 $s = \big(s_1(\vec{y}_1) \wedge \neg s_2(\vec{y}_2)\big) \vee \big(\neg s_1(\vec{y}_1) \wedge s_2(\vec{y}_2)\big) \vee \big(s_1(\vec{y}_1) \wedge s_2(\vec{y}_2)\big)$

8232280 $\sin x := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \quad ; \quad \cos x := \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$

7923130 $\frac{1}{2}bc\sin 120^\circ = \frac{1}{2}ab\sin 60^\circ + \frac{1}{2}ac\sin 60^\circ = a(b+c)\cdot\frac{1}{2}\sin 60^\circ.$

191729 $I(\alpha,\lambda) = \int\limits_0^1 x^{\lambda-\alpha}(1-x)^{\alpha-1}\,\mathrm{d}x = \mathrm{B}(1+\lambda-\alpha,\alpha) = \frac{\Gamma(\alpha)\Gamma(1+\lambda-\alpha)}{\Gamma(1+\lambda)}.$

4160774 $\frac{\partial^2 u_i}{\partial x_j \partial x_j} = \sum_{j=1}^3 \frac{\partial^2 u_i}{\partial x_j^2} = \frac{\partial^2 u_i}{\partial x_1^2} + \frac{\partial^2 u_i}{\partial x_2^2} + \frac{\partial^2 u_i}{\partial x_3^2}$

320393 $\frac{\partial g}{\partial x} = \frac{\partial g}{\partial \varphi_1}\frac{\partial \varphi_1}{\partial x} + \frac{\partial g}{\partial \varphi_2}\frac{\partial \varphi_2}{\partial x} = \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial x}.$

2537458 $\frac{\sqrt{2+\sqrt{3}}}{2} = \frac{\sqrt{3}+1}{2\sqrt{2}} = \cos 30^\circ \cos 45^\circ + \sin 30^\circ \sin 45^\circ = \cos(45^\circ - 30^\circ)$

142635 $\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \frac{z^9}{9!} + \ldots = \sum_{k=0} \frac{(-1)^k}{(2k+1)!} z^{2k+1}$

2485925 $\varphi_{n,m} : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{nm+n+m}, ((p_0 : \ldots : p_n)(q_0 : \ldots : q_m)) \mapsto (p_0q_0 : p_0q_1 : \ldots : p_nq_m)$

4919636 $\sum_{n=1}^{\infty} \frac{a_n}{1+n^\alpha a_n} = \sum_{n=1}^{\infty} \frac{n^{\lambda\alpha}}{1+n^{\alpha+\lambda\alpha}} = \sum_{n=1}^{\infty} \frac{1}{n^{-\lambda\alpha}+n^\alpha} \leqslant \sum_{n=1}^{\infty} \frac{1}{n^{-\lambda\alpha}}.$

8448692 $\frac{\sqrt{2+\sqrt{3}}}{-\sqrt{2-\sqrt{3}}} = -\sqrt{\frac{2+\sqrt{3}}{2-\sqrt{3}}} = -\sqrt{\frac{(2+\sqrt{3})(2+\sqrt{3})}{(2-\sqrt{3})(2+\sqrt{3})}} = -(2+\sqrt{3})$

4195341 $(\lambda x + (1-\lambda)y)^\alpha = x^\alpha(\lambda + (1-\lambda)\frac{y}{x})^\alpha \leq x^\alpha(\lambda^\alpha + (1-\lambda)^\alpha \frac{y^\alpha}{x^\alpha}) = \lambda^\alpha x^\alpha + (1-\lambda)^\alpha y^\alpha$

1088695 $\int_0^\infty \frac{\ln^2 x}{1+x^2}dx = \int_0^1 \frac{\ln^2 x}{1+x^2}dx + \int_1^\infty \frac{\ln^2 x}{1+x^2}dx = 2\int_0^1 \frac{\ln^2 x}{1+x^2}dx$

7556515 $(1-1) + \left(\frac{1}{2} + \frac{1}{2} - \frac{1}{2} - \frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4}\right) + \ldots$

7031540 $\int_0^\pi \frac{\sin(nx)}{nx}dx = \frac{1}{n}\int_0^{n\pi} \frac{\sin y}{y} = \frac{1}{n}\sum_{j=0}^{n-1}\int_0^\pi (-1)^j \frac{\sin y}{y+j\pi}dy,$

410986 $\mu_n = \mathbb{E}(\sum_{k=1}^n X_i) = n\frac{\alpha}{\alpha+1} \qquad \sigma_n^2 = \mathbb{V}\mathrm{ar}(\sum_{k=1}^n X_i) = \sum_{k=1}^n \mathbb{V}\mathrm{ar}(X_i) = \frac{n\alpha}{\alpha+2}\frac{1}{(\alpha+1)^2}$

5144032 $ghkh^{-1}k^{-1}g^{-1} = (gh^{-1}g^{-1})(gkg^{-1})(gh^{-1}g^{-1})(gk^{-1}g^{-1}) = h_2k_2h_2^{-1}k_2^{-1}$

8650621 $\tan^{-1}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} = \sum_{n=0}^{\infty} (-1)^n \cdot \frac{x^{2n+1}}{2n+1}$

381546 〃 〃

8988417 $\sin(x) + \sin(2x) + \sin(3x) + \cdots + \sin(nx) = \frac{cos(1/2) - cos(n-1/2)}{2\sin(1/2)2}$

5949152 $x \in (A_1 \triangle \cdots \triangle A_{n-1}) \triangle A_n \iff x \text{ is in exactly an odd number of } A_i$
$\iff \Big|\{i \mid 1 \leq i \leq n, x \in A_i\}\Big| \text{ is odd.}$

5906285 $$g(x) = \sum_{i=a+1}^{x} \Delta(i) = \left(\sum_{i=a+1}^{x} \lambda(i) \right) \frac{\sum_{x=a+1}^{b} \mu(x)}{\sum_{x=a+1}^{b} \lambda(x)} - \sum_{i=a+1}^{x} \mu(i)$$

7075140 $$\sin(105°) = \sin(90° + 15°) = \cos(15°) = \sqrt{\frac{1+\cos(30°)}{2}} = \sqrt{\frac{1+\frac{1}{2}\sqrt{3}}{2}} = \sqrt{\frac{2+\sqrt{3}}{4}}.$$

5087123 $$10515389996560312960 = 3022993637 \times 6003631993 + 9069920719 \times 9592692301$$

9153002 $$\sum_{i=1}^{n} \frac{i^2}{n^2} = \frac{1}{n^2} \sum_{i=1}^{n} i^2 = \frac{1}{n^2} \frac{n(n+1)(2n+1)}{6} = \frac{n^2}{n^2} \frac{(1+\frac{1}{n})(2n+1)}{6}$$

6940955 $$\int_0^2 f(x) = \int_{-1}^{1} f(t+1)dt \sim A_1 f(t_1+1) + A_2 f(t_2+1) = f(1-1/\sqrt{(3)} + f(1+1/\sqrt{(3)}$$

8532653 $$\text{''} \sum_n \lambda \alpha_n = \lambda x = (S + S^*)x = \sum_{n=1}^{\infty} \alpha_n e_{n+1} + \sum_{n=2}^{\infty} \alpha_n e_{n-1} = \alpha_2 e_1 + \sum_{n=2}^{\infty} (\alpha_{n-1} + \alpha_{n+1}) e_n \text{.''}$$

6385056 $$I = \iint_R f(x,y)\,dx\,dy = \int_{x=-r}^{0} \int_{y=-r}^{0} f(x,y)\,dx\,dy + \int_{x=0}^{r} \int_{y=0}^{r} f(x,y)\,dx\,dy = A + B$$

2205186 $$\lim_{j\to\infty} x_j \overset{(2)}{=} \lim_{j\to\infty} \left(\lim_{n\to\infty} x_j^{(n)} \right) \overset{(3)}{=} \lim_{n\to\infty} \left(\lim_{j\to\infty} x_j^{(n)} \right) \overset{(1)}{=} \lim_{n\to\infty} 0 = 0$$

8828733 $$\sum_{n=1}^{\infty} \frac{1}{n(n+3)} = \frac{1}{3} \sum_{n=1}^{\infty} \frac{1}{n} - \frac{1}{n+3} = \frac{1}{3}\left(1 - \frac{1}{4} + \frac{1}{2} - \frac{1}{5} + \frac{1}{3} - \frac{1}{6} + \frac{1}{4} - \frac{1}{7} + \cdots\right)$$

4963196 $$\sum_{i=1}^{n} \binom{n}{i} i = \sum_{i=1}^{n} \frac{n}{i} \cdot \binom{n-1}{i-1} i = n \cdot \sum_{i=1}^{n} \binom{n-1}{i-1} = n \sum_{j=0}^{n-1} \cdot \binom{n-1}{j} = n 2^{n-1}.$$

7105197 $$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1+1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\lambda^{a+1}\Gamma(a+1)}{\lambda^{a+1}\Gamma(a+1)} \int_0^\infty x^{\alpha-1+1} e^{-\lambda x} dx =$$

7004390 $$\sum_{n=1}^{\infty} \frac{t^n}{n} = \sum_{n=0}^{\infty} \frac{t^{n+1}}{n+1} = \sum_{n=0}^{\infty} \int_0^t s^n ds = \int_0^t \sum_{n=0}^{\infty} s^n ds = \int_0^t \frac{1}{1-s} ds = -\ln(1-t)$$

7266541 $$= \sum_{n=0}^{\infty} \frac{k(k-1)(k-2)\ldots(k-n+1)(k-n)(k-n-1)\ldots 2\times 1}{n!(k-n)(k-n-1)\ldots 2\times 1} x^n$$

3153003 $$0 < \left\langle df\left(\frac{\partial \varphi_\beta}{\partial r}\right) \wedge df\left(\frac{\partial \varphi_\beta}{\partial s}\right), N_2 \right\rangle = \frac{\partial(u,v)}{\partial(r,s)} \left\langle df\left(\frac{\partial \varphi_\alpha}{\partial u}\right) \wedge df\left(\frac{\partial \varphi_\alpha}{\partial v}\right), N_2 \right\rangle$$

6607662 $$x \in f^{-1}(C \cup D) \Leftrightarrow f(x) \in C \cup D \Leftrightarrow f(x) \in C \vee f(x) \in D \Leftrightarrow x \in f^{-1}(C) \vee x \in f^{-1}(D) \Leftrightarrow x \in f^{-1}(C) \cup f^{-1}(D)$$

4867926 $$\frac{\beta}{\alpha-1} = \mathrm{E}[X] = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{\mu}, \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} = \mathrm{Var}[X] = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \hat{\sigma}^2.$$

6981740 $$x_n \to x \text{ weakly } \Leftrightarrow f(x_n) \to f(x)\ \forall\, f \in X^* \Leftrightarrow \lim_{m\to\infty} d_m(x_n) \to \lim_{m\to\infty} d_m(x)\ \forall\, (d_m) \subset D \text{ s.t. } d_m \to f$$

857944 $$\int_0^\infty \frac{\sin t}{t^\alpha}\, dt = \int_0^\pi \frac{\sin t}{t^\alpha}\, dt + \sum_{n=1}^{\infty} \int_{n\pi}^{(n+1)\pi} \frac{\sin t}{t^\alpha}\, dt := \int_0^\pi \frac{\sin t}{t^\alpha}\, dt + \sum_{n=1}^{\infty} a_n'$$

701781 $$\tan^2\theta \sin^2\theta = (\sec^2\theta - 1)\sin^2\theta = \sec^2\theta \sin^2\theta - \sin^2\theta = \frac{\sin^2\theta}{\cos^2\theta} - \sin^2\theta = \tan^2\theta - \sin^2\theta$$

2632213 $$\sum_{m=0}^{n-1} \frac{1}{(2n+1)(z_m)^{2n}} \frac{(-1)^n}{4^n} \sum_{k=0}^{n} (-1)^k \binom{2n+1}{k} \sum_{j=0}^{2n} \frac{[\mathrm{i}(2n+1-2k)z_m]^j}{j!} ?$$

6881546 $$s_1' = 0 + 1 + 2x + 3x^3 + 4x^3 + 5x^4 + \ldots + nx^{n-1} + \ldots = (\tfrac{1}{1-x})'1 + 2x + 3x^2 + 4x^3 + \ldots = \tfrac{1}{(1-x)^2}$$

3989242 $$1 = \sum_{x=1}^{3} \sum_{y=1}^{x} c(x+y) = c\sum_{x=1}^{3} (x^2 + \sum_{y=1}^{x} y) = c\sum_{x=1}^{3} (x^2 + x(x+1)/2) = \tfrac{c}{2} \sum_{x=1}^{3} (3x^2 + x) \vdots$$

6843982 $$S = \int \left(\frac{\alpha - x}{\beta - x}\right)^\lambda dx = (\alpha - \beta) \int \frac{t^\lambda}{(1-t)^2}\, dt = \frac{(\alpha-\beta)}{\lambda - 1} \frac{t^{\lambda+1}}{(1-t)^2}\, {}_2F_1\left(1, 2; 2-\lambda; \frac{1}{1-t}\right)$$

340683 $$\frac{\partial f}{\partial x_1} = 19x_2x_3x_4 - 3(x_2 + x_3 + x_4), \frac{\partial^2 f}{\partial^2 x_1} = 0, \frac{\partial^2 f}{\partial x_1 \partial x_2} = 19x_3x_4 - 3 = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 16$$

6917607 $$\sum_j p_j a_{ij}^{\frac{1}{1+r}} \le \left(\sum_j p_j a_{ij}\right)^{\frac{1}{1+r}} =: x_i^{\frac{1}{1+r}}, \quad \sum_k q_k b_{ik}^{\frac{1}{1+r}} \le \left(\sum_j q_k b_{ik}\right)^{\frac{1}{1+r}} =: y_i^{\frac{1}{1+r}}$$

2248865 $$\begin{aligned} \mathcal{F}_s\left(\frac{\partial^2 u}{\partial y^2}\right) &= \int_0^\infty \frac{\partial^2 u}{\partial x^2} \sin(kx) dx \\ &= \frac{\partial^2}{\partial y^2} \int_0^\infty u(x,y) \sin(kx) dx \\ &= \frac{\partial^2}{\partial y^2} \hat{u}(k,y) \end{aligned}$$

7399755 $$\iint_0^1 \lfloor \mathrm{e}^x \mathrm{e}^{-y} \rfloor dxdy = \int_0^1 \int_x^1 0\, dydx + \int_0^1 \int_0^x 1\, dydx + \int_{\ln 2}^1 \int_0^{x-\ln 2} 1\, dydx = 1 - \ln 2 + (\ln 2)^2/2.$$

3234731 $$\lim_{N\to\infty} \tfrac{1}{N} \sum_{n\le N} \lim_{k\to\infty} f(n,k) = \lim_{N\to\infty} \lim_{k\to\infty} \tfrac{1}{N} \sum_{n\le N} f(n,k) = \lim_{k\to\infty} \lim_{N\to\infty} \tfrac{1}{N} \sum_{n\le N} f(n,k)$$

3148754 $$\sum_{n=1}^{N} \cos(2^n) = \mathcal{O}(1) \iff \sum_{n=1}^{\infty} \frac{\cos(2^n) - \cos(2^{n+N})}{n} = 2\sum_{n=1}^{\infty} \frac{\cos^2(2^{n-1}) - \cos^2(2^{n-1+N})}{n}$$

8276013 $$f(x_n + c_2h, y_n + c_2hk_1) = f(x_n, y_n) + \tfrac{\partial}{\partial x} f(x_n, y_n) \tfrac{(c_2h)}{1!} + \tfrac{\partial}{\partial y} f(x_n, y_n) \tfrac{(c_2hk_1)}{1!} + O(h^2)$$

1599195 $$\sum_{i=0}^{m-1} \binom{n-1+i}{i} x^n y^i = \sum_{i=0}^{m-1} \binom{n-1+i}{i} x^n (1-x)^i = \sum_{i=0}^{m-1} \binom{n-1+i}{i} x^n \sum_{k=0}^{i} (-1)^k \binom{i}{k} x^k$$

$$\frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots + \frac{1}{x^2} <$$

6010851 $$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{x-1} - \frac{1}{x}\right) =$$

$$= 1 - \frac{1}{x} = \frac{x-1}{x}$$

2668306 $$x_1 = \arccos\tfrac{3}{5} + \arccos\tfrac{2\sqrt{2}}{3} = \arcsin\sqrt{1 - \tfrac{3}{5}^2} + \arcsin\sqrt{1 - \tfrac{2\sqrt{2}}{3}^2} = \arcsin\tfrac{4}{5} + \arcsin\tfrac{1}{3}$$

3477896 $$(10011101001)_2 = 1\times 2^{10} + 0\times 2^9 + 0\times 2^8 + 1\times 2^7 + 1\times 2^6 + 1\times 2^5 + 0\times 2^4 + 1\times 2^3 + 0\times 2^2 + 0\times 2^1 + 1\times 2^0 = 1257$$

3513170 $$\int_0^\infty \tfrac{1}{x^2} dx = \lim_{\epsilon\to 0} \lim_{N\to\infty} \int_\epsilon^N \tfrac{1}{x^2} dx = \lim_{\epsilon\to 0} \lim_{N\to\infty} -\tfrac{1}{x}\Big|_\epsilon^N = \lim_{\epsilon\to 0} \lim_{N\to\infty} \left(-\tfrac{1}{N} + \tfrac{1}{\epsilon}\right) = \infty.$$

4032405 $$n^2 \equiv 1 \mod p \implies (2k+1)^2 \equiv 1 \mod p \implies 4k(k+1) \equiv 0 \mod p \implies 4k \equiv 0 \mod p \text{ or } k+1 \equiv 0 \mod p \implies k \equiv \pm 1 \mod p \implies n \equiv :$$

$$\begin{cases} 6x \equiv 2\,[8] \\ 5x \equiv 5\,[6] \end{cases} \Leftrightarrow \begin{cases} 3x \equiv 1\,[4] \\ x \equiv 1\,[6] \end{cases}$$

9150972 $$\Rightarrow \begin{cases} x \equiv -1\,[4] \\ x \equiv 1\,[6] \end{cases} \Rightarrow \begin{cases} x \equiv 3\,[4] \\ x \equiv 1\,[6] \end{cases}$$

$$4k + 3 = 6m + 1 \Leftrightarrow 1 = 3m - 2k$$

2451289 $$0000 \to 1000 \to 1100 \to 0100 \to 0110 \to 1110 \to 1010 \to 0010 \to 0001 \to 1001 \to 1101 \to 0101 \to 0111 \to 1111 \to 1011 \to 0011$$

2884618 $$\tan(x) \cdot y' + y = \frac{\sin(x)\cdot y' + \cos(x)\cdot y}{\cos x} = \frac{\sin(x)\cdot y' + (\sin(x))'\cdot y}{\cos x} = \frac{(\sin(x)\cdot y)'}{\cos x}$$

4386575 $\sum_{k=2001}^{m} a_k \sin(kx) \leq 2ma_m + \sum_{k=2001}^{m-1} 2k(a_k - a_{k+1}) \leq 2 + 2\left(\sum_{k=2001}^{m-1} ka_k - \sum_{k=2001}^{m-1} ka_{k+1}\right),$

8866996 $\operatorname{Cov}(Q^1_{N_1}, Q^2_{N_2}) = \frac{\operatorname{Cov}(f^1(X), f^2(X))}{\max(N_1, N_2)} = \frac{\sqrt{\operatorname{Var}(f^1(X))\operatorname{Var}(f^2(X))}}{\max(N_1, N_2)}\rho_{12},$

3679687 $\sum_{n=1}^{\infty} \frac{1}{n^2+a^2} = \frac{\pi}{2a}\coth(\pi a) - \frac{1}{2a^2} \sum_{n=1}^{\infty} \frac{1}{[n^2+a^2]^2} = \frac{\pi}{4a^3}\coth(\pi a) + \frac{\pi^2}{4a^2}\operatorname{csch}(\pi a) - \frac{1}{2a^4}$

6906793 $k(\sqrt{1-\cos^2 x} + \sqrt{1+\sin^2 x}) = (\sqrt{1-\cos^2 x} - \sqrt{1+\sin^2 x})(\sqrt{1-\cos^2 x} + \sqrt{1+\sin^2 x}) = 1 - \cos^2 x - (1+\sin^2 x) = -1$

9326890 $\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = \sum_{n=-1}^{\infty} \frac{(-1)^{n+1}}{(2n+2)!} x^{2n+2} = 1 + \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(2n+2)!} x^{2n+2}$

6371749 $\sum_{j=0}^{n+1} \binom{m+j}{j} = \underbrace{\sum_{j=0}^{n} \binom{m+j}{j}}_{\binom{m+n+1}{n}} + \binom{m+n+1}{n+1} = \binom{m+n+1}{n} + \binom{m+n+1}{n+1} = \binom{m+n+2}{n+1}$

2930140 $\det(A+I) = \prod_{i=1}^{n}(1+\lambda_i) \geq \prod_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} \lambda_i \geq \det(A) + \operatorname{tr}(A) \geq \det(A) + \operatorname{tr}(I_n) = \det(A) + n\det(I_n)$

822662 $\int_0^1 \frac{dx}{x^2-x+1} = \int_{-1/2}^{1/2} \frac{dx}{x^2+3/4} = \frac{\sqrt{3}}{2} \cdot \frac{4}{3} \int_{-1/\sqrt{3}}^{1/\sqrt{3}} \frac{dx}{x^2+1} = \frac{4\sqrt{3}}{3} \arctan \frac{1}{\sqrt{3}} = \frac{2\pi}{3\sqrt{3}}.$

6781362 $\{\hat{1}\}, \{\hat{2}\}, \{\hat{1}, 2\}, \{1, \hat{2}\}, \{\hat{3}\}, \{\hat{1}, 3\}, \{1, \hat{3}\}, \{\hat{2}, 3\}, \{2, \hat{3}\}, \{\hat{1}, 2, 3\}, \{1, \hat{2}, 3\}, \{1, 2, \hat{3}\}.$

3072169 $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-a^2}}{1+e^{-(x+a)}} dx = \frac{e^{-a^2}}{2}\left[\sum_{n=0}(-1)^n \operatorname{erfcx}(-a+n/2) + \sum_{n=1}^{\infty}(-1)^{n+1}\operatorname{erfcx}(a+n/2)\right],$

84771 $(y-1)^2 = 4(y+2) \Rightarrow y^2 - 2y + 1 - 4y - 8 = 0 \Rightarrow y^2 - 6y - 7 = 0 \Rightarrow (y+1)(y-7) = 0 \Rightarrow y = -1, 7 \Rightarrow A = \int_{-1}^{7} \left(y + 1 - \left(\frac{(y-1)^2}{4} - 1\right)\right) dy$

4865288 $(\partial^2 F_1/\partial x^2 + \partial^2 F_1/\partial y^2 + \partial^2 F_1/\partial z^2, \partial^2 F_2/\partial x^2 + \partial^2 F_2/\partial y^2 + \partial^2 F_2/\partial z^2, \partial^2 F_3/\partial x^2 + \partial^2 F_3/\partial y^2 + \partial^2 F_3/\partial z^2)$

5659929 $\{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,3,4\}, \{1,2,4\}, \{2,3,4\}$

5773917 $\int_0^\infty \frac{(\sin x)^2}{x^2} dx = -\int_0^\infty (\sin x)^2 d\frac{1}{x} = -\frac{(\sin x)^2}{x}\Big|_0^\infty + \int_0^\infty \frac{(2\sin x \cos x)}{x} dx = \int_0^\infty \frac{(\sin 2x)}{2x} d2x = \dots$

5352283 $\int_0^1 \frac{\ln^2(1-x)}{2-x^2} dx \quad \int_0^1 \frac{\ln^2(1+x)}{2-x^2} dx \quad \int_0^1 \frac{\ln(1+x)\ln(1-x)}{\sqrt{2}-x} dx \quad \int_0^1 \frac{\ln(1+x)\ln(1-x)}{\sqrt{2}+x} dx$

5445589 $\frac{\sum_{n=1}^{2^k} a_n + \sum_{n=2^k+1}^{2^{k+1}} a_n}{2^{k+1}} \geq \frac{1}{2}\left(\sqrt[2^k]{\prod_{n=1}^{2^k} a_n} + \sqrt[2^k]{\prod_{n=2^k+1}^{2^{k+1}} a_n}\right) \geq \sqrt[2^{k+1}]{\prod_{n=1}^{2^{k+1}} a_n}$

6524517 $\sum_{k=1}^{N}\left(\cos\frac{\pi}{2k} - \cos\frac{\pi}{2(k+2)}\right) = \sum_{k=1}^{N}\left(\cos\frac{\pi}{2k} - \cos\frac{\pi}{2(k+1)}\right) + \sum_{k=1}^{N}\left(\cos\frac{\pi}{2(k+1)} - \cos\frac{\pi}{2(k+2)}\right)$

6115808 $= \dfrac{\sin 50^\circ + \dfrac{1}{2}}{\dfrac{\sqrt{3}}{2} + 2 \cdot \sin 50^\circ \cdot \cos 30^\circ} = \dfrac{\sin 50^\circ + \dfrac{1}{2}}{\dfrac{\sqrt{3}}{2} + \sqrt{3} \cdot \sin 50^\circ} = \dfrac{\sin 50^\circ + \dfrac{1}{2}}{\sqrt{3}\left(\dfrac{1}{2} + \sin 50^\circ\right)}$

5133523 $u_{x_i} = (2-n)\left(\frac{1}{1+|x|^2}\right)^{\frac{n-2}{2}}\left(\frac{1}{1+|x|^2}\right)|x|x_i \text{ and } (u_\lambda)_{x_i} = (2-n)\left(\frac{\lambda}{\lambda^2+|x|^2}\right)^{\frac{n-2}{2}}\left(\frac{1}{\lambda^2+|x|^2}\right)|x|x_i,$

9016545 $128^{128^{128}} \mod 7 \equiv 128^{128^{128} \bmod 3} \mod 7 \equiv 128^{128^{128 \bmod 2} \bmod 3} \mod 7 \equiv 128^{128^0 \bmod 3} \mod 7 \equiv 128^1 \mod 7 \equiv 128 \mod 7 \equiv 2$

7617396 $\{\{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, N\}.$

5254228 $a_2 = \csc(x) - \frac{1}{2}\csc(x/2) a_3 = \frac{1}{2}\csc(x/2) - \frac{1}{4}\csc(x/4) \vdots a_n = \frac{1}{2^{n-2}}\csc(x/2^{n-2}) - \frac{1}{2^{n-1}}\csc(x/2^{n-1}).$

5220392 $\left(e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}\right)' = e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2} \cdot \left(-\frac{1}{2}\right)\sum_{i=1}^{n}(2\theta - 2x_i) = -e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}\sum_{i=1}^{n}(\theta - x_i)$

8306674 $\sum_{a=1}^{x}\sum_{b=1}^{y}(x-a+1) \times (y-b+1) \times ab = \sum_{a=1}^{x}\left(a(x-a+1)\sum_{b=1}^{y} b(y-b+1)\right) = \left(\sum_{a=1}^{x} a(x-a+1)\right)\left(\sum_{b=1}^{y} b(y-b+1)\right)$

5914751 $f((s,x,y)) \star f((s',x',y')) = (s + \frac{1}{2}xy, x, y) \star (s' + \frac{1}{2}x'y', x', y') = (s + s' + \frac{1}{2}xy + \frac{1}{2}x'y' + \frac{1}{2}xy', x)$

448459 $\int_0^{\pi/2} \frac{\cos^2 x \cos x dx}{\sqrt{\sin x}(1+\sin^2 x)} = \int_0^{\pi/2} \frac{(1-\sin^2 x)\cos x dx}{\sqrt{\sin x}(1+\sin^2 x)} = \int_0^1 \frac{(1-u^2)du}{\sqrt{u}(1+u^3)}$

4147561 $\zeta(2)^2 = \sum_{(x,y)\in\mathbb{N}^2} \frac{1}{x^2y^2} = \sum_{\substack{(a,b)\in A \\ g\in\mathbb{N}}} \frac{1}{(ga)^2(gb)^2} = \sum_{g\in\mathbb{N}} \frac{1}{g^4} \sum_{(a,b)\in A} \frac{1}{a^2b^2} = \zeta(4)\sum_{(a,b)\in A} \frac{1}{a^2b^2}$

1445496 $= \int_\alpha^\beta \int_{h(x)}^{g(x)} \left|\left(\frac{\partial\varphi_1(f)}{\partial x}, \frac{\partial\varphi_2(f)}{\partial x}, \frac{\partial\varphi_3(f)}{\partial x}, \right) \times \left(\frac{\partial\varphi_1(f)}{\partial y}, \frac{\partial\varphi_2(f)}{\partial y}, \frac{\partial\varphi_3(f)}{\partial y}, \right)\right| \cdot |J_f| \, dy \, dx.$

2345403 *Let $(x_n) \subset S$ and $x \in S$. Show that $x_n \to x$ iff for all $\epsilon > 0$, the open ball $B(\epsilon; x)$ contains all but finitely many terms of x_n*

1287640 $\frac{1}{2^n\sqrt{\prod_{i=1}^{n} x_i}} + \sum_{k=1}^{n} \frac{x^k}{\prod_{i=1}^{k}(1+x_i)} \geq 1 \Leftrightarrow \frac{1}{2^n\sqrt{\prod_{i=1}^{n} x_i}} + \sum_{k=1}^{n} \frac{1+x_k}{\prod_{i=1}^{k}(1+x_i)} \geq 1 + \sum_{k=1}^{n} \frac{1}{\prod_{i=1}^{k}(1+x_i)} \Leftrightarrow$

551619 $\int_{-3}^{4} e^{-\lambda x^2} \log(1+x^2) dx = \int_{-3}^{-1+\delta} e^{-\lambda x^2} \log(1+x^2) dx + \int_{-1+\delta}^{1-\delta} e^{-\lambda x^2} \log(1+x^2) dx + \int_{1-\delta}^{4} e^{-\lambda x^2} \log(1+x^2) dx$

9042733 $\implies \sin\frac{3\pi}{8} = \sin 67.5^\circ = \sin(90^\circ - 22.5^\circ) = \cos 22.5^\circ = -\sqrt{\frac{1+\cos 45^\circ}{2}} = \sqrt{\frac{1+\frac{1}{\sqrt{2}}}{2}} = \sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}} = \sqrt{\frac{2+\sqrt{2}}{4}} = \frac{\sqrt{2+\sqrt{2}}}{2}$

6395610 $g \in \ker f \iff f(g) = \hat{e} \iff p(\phi(g)) = \hat{e} \iff \phi(g) \in \phi(N) \iff \exists n \in N \text{ such that } \phi(g) = \phi(n) \iff \phi(gn^{-1}) = e \iff gn^{-1} \in \ker\phi \subseteq N \iff$

6500517 $\sum_{i=1}^{10}(-2)^{-i} = \sum_{i=1}^{10}\left(-\frac{1}{2}\right)^i = \frac{1-(-\frac{1}{2})^{11}}{1-(-\frac{1}{2})} = \frac{2^{11}+1}{2^{11}} \cdot \frac{2}{3} = \frac{2^{11}+1}{3\cdot 2^{10}} = \frac{2049}{3072} = \frac{683}{1024}.$

2219431 $(-1)^{\varphi(n)}\Phi_n(-1) = \prod_{\substack{1\leq k\leq n \\ \gcd(k,n)=1}}\left(1 + e^{\frac{2\pi i k}{n}}\right) = 2^{\varphi(n)} \prod_{\substack{1\leq k\leq n \\ \gcd(k,n)=1}} e^{\frac{\pi i k}{n}} \prod_{\substack{1\leq k\leq n \\ \gcd(k,n)=1}} \cos\frac{\pi k}{n}$

6225402 $1 \leqslant A = \lim_{n\to\infty}(\lim_{k\to\infty}(\frac{1}{1+2^{n-k}})) \leqslant \lim_{n\to\infty}(\lim_{k\to\infty}(\frac{1}{2^{n-k}})) \leqslant \lim_{n\to\infty}(\lim_{k\to\infty}(\frac{2^k}{2^n})) = \dfrac{\lim_{k\to\infty} 2^k}{\lim_{n\to\infty} 2^n}$

253072 $(1-\epsilon)\sum_{k=1}^{n}\|v_k\|_{V'}^2 \leq \epsilon + v(x) \leq \epsilon + \|v\|_{L^2(V)'}\left(\sum_{k=1}^{n}\|v_k\|_{V'}^2\right)^{1/2} \leq \epsilon + \frac{1}{2}\|v\|_{L^2(V)'}^2 + \frac{1}{2}\left(\sum_{k=1}^{n}\|v_k\|_{V'}^2\right)^{1/2},$

4408909 $\frac{\pi}{2} = \int_0^\infty \frac{\sin x}{x} dx = \int_0^\infty \frac{\sin 2u}{2u} d(2u) = \int_0^\infty \frac{\sin 2u}{u} du = \underbrace{\left[\frac{\sin^2 u}{u}\right]_0^\infty}_{=0} + \int_0^\infty \frac{\sin^2 u}{u^2} du = \int_0^\infty \frac{\sin^2 u}{u^2} du$

4747000 $\lim_{n\to+\infty}\left(1 - \frac{1}{n}\right)^{n^2} = \lim_{n\to+\infty} e^{n^2\ln\left(1-\frac{1}{n}\right)} = e^{\lim_{n\to+\infty} n^2\ln\left(1-\frac{1}{n}\right)} = e^{\lim_{n\to+\infty} n^2 \lim_{n\to+\infty}\ln\left(1-\frac{1}{n}\right)} = e^{\lim_{n\to+\infty} -n} = 0$

8633090 $\lambda(2) = 1\lambda(4) = 2\lambda(2^k) = 2^{k-2}\lambda(p^k) = p^{k-1}(p-1)\lambda(n) = lcm[\lambda(p_1^{\alpha_1}), \lambda(p_2^{\alpha_2}), \dots, \lambda(p_r^{\alpha_r})] \text{ for } n = p_1^{\alpha_1}p_2^{\alpha_2}\dots p_r^{\alpha_r}$

2059379 $A(z,u) - \sum_{n\geq 0}(n+1)z^{n+1} = \sum_{n\geq 0}\sum_{q\geq 1} u^q z^{n+1} \sum_{m=1}^{n}[z^m][u^{q-1}]A(z,u) + \sum_{n\geq 0}\sum_{q\geq 1} u^q z^{n+1}\sum_{m=1}^{n}\binom{m-1}{q-1}10^q \times (n+1).$

3960467 $\{\{\}, \{\}\}, \{\{1\}, \{1\}\}, \{\{2\}, \{2\}\}, \{\{3\}, \{3\}\}, \{\{1,2\}, \{1,2\}\}, \{\{1,3\}, \{1,3\}\}, \{\{2,3\}, \{2,3\}\}, \{\{1,2,3\}, \{1,2,3\}\}$

5337683 $\int_0^1 \frac{\ln(a-x)}{1+x}dx = \int_1^2 \frac{\ln(a+1-x)}{x}dx = \ln 2\ln(a+1) + \int_1^2 \frac{\ln(1-\frac{x}{a+1})}{x}dx = \ln 2\ln(a+1) + \operatorname{Li}_2\left(\frac{1}{a+1}\right) - \operatorname{Li}_2\left(\frac{2}{a+1}\right).$

2382504 $J_1 = \int_0^1 \frac{1}{(a^2(t^2+2)+1)(3t^2+5)}dt = \frac{\sqrt{15}\tan^{-1}\left(\frac{\sqrt{15}}{5}\right)}{5}\int_0^1 \frac{1}{a^2+3}da - \int_0^1 \frac{a\tan^{-1}\left(\frac{a}{\sqrt{2a^2+1}}\right)}{\sqrt{2a^2+1}(a^2+3)}da.$

5996309 $\frac{P}{2^{29}} = \frac{(\sqrt{3}\cos(1°)+\sin(1°))}{2\cos(1°)}\frac{(\sqrt{3}\cos(2°)+\sin(2°))}{2\cos(2°)}\cdots\frac{(\sqrt{3}\cos(29°)+\sin(29°))}{2\cos(29°)}$

4012489 $\lim_{x\to 0}\frac{f(x)-f(0)}{x} = \begin{cases} \lim\limits_{x\to 0,\, x\neq\frac{1}{2^n}} \dfrac{0-0}{x} = 0 \\ \lim\limits_{x=\frac{1}{2^n},x\to 0 \iff n\to\infty} \dfrac{\frac{1}{2^{2n}}-0}{\frac{1}{2^n}} = \lim\limits_{x=\frac{1}{2^n},x\to 0 \iff n\to\infty} \dfrac{1}{2^n} = 0 \end{cases}$

3210903 $b_{k+1}^2\sum_{i=1}^k a_i^2 + a_{k+1}^2\sum_{i=1}^k b_i^2 \geq 2a_{k+1}b_{k+1}\sum_{i=1}^k a_ib_i \implies a_{k+1}^2\sum_{i=1}^k b_i^2 - 2a_{k+1}b_{k+1}\sum_{i=1}^k a_ib_i + b_{k+1}^2\sum_{i=1}^k a_i^2 \geq 0.$

4403309 $\sqrt[3]{\frac{74}{43}+\sum_{k=0}^1\cos(\frac{2\pi}{7}\cdot 3^{3k})} + \sqrt[3]{\frac{74}{43}+\sum_{k=0}^1\cos(\frac{2\pi}{7}\cdot 3^{3k+1})} + \sqrt[3]{\frac{74}{43}+\sum_{k=0}^1\cos(\frac{2\pi}{7}\cdot 3^{3k+2})} = \sqrt[3]{\frac{392}{43}}$

7045786 $e \overset{def}{=} \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \lim_{n\to\infty}\lim_{N\to\infty}\sum_{k=0}^N \frac{\alpha_{n,k}}{k!} = \lim_{N\to\infty}\lim_{n\to\infty}\sum_{k=0}^N \frac{\alpha_{n,k}}{k!} = \lim_{N\to\infty}\sum_{k=0}^N \frac{1}{k!} \overset{def}{=} \sum_{k=0}^\infty \frac{1}{k!}$

5324440 $I = \sum_{n=1}^\infty \frac{2^{2n}}{n^3(2n+1)\binom{2n}{n}} = \sum_{n=1}^\infty \frac{2^{2n}}{n^3\binom{2n}{n}} - \sum_{n=1}^\infty \frac{2^{2n+1}}{n^2(2n+1)\binom{2n}{n}} = 4\int_0^1 \frac{\arcsin^2 x}{x}\,dx - 4\int_0^1 \arcsin^2 x\,dx$

200206 $\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}\left(-e^{2i\pi\frac{k}{n}}\right) = \left(e^{\frac{2i\pi}{n}\sum_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}} k}\right)\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}(-1) = \left(e^{\frac{2i\pi}{n}\frac{n}{2}\varphi(n)}\right)(-1)^{\varphi(n)} = +1,$

8450768 $1\cdot 3 = (2-1)(2+1) = 2^2-1^2 < 2^2 1\cdot 3\cdot 5 = (3-2)3(3+1) = (3^2-1^2)3 < 3^3 1\cdot 3\cdot 5\cdot 7 = (4-3)(4-1)(4+1)(4+3) = (4^2-3^3)(4^2-1^2) < 4$

5562454 $\frac{1}{2}\sum_{n=1}^{99}\frac{1}{n^2-n+1} - \frac{1}{n^2+n+1} = \frac{1}{2}[\sum_{n=1}^{99}\frac{1}{n^2-n+1} - \sum_{n=2}^{100}\frac{1}{n^2-n+1}] = \frac{1}{2}[\frac{1}{1^2-1+1} - \frac{1}{100^2-100+1}] = \frac{1}{2} - \frac{1}{18802}$

$\sum_{k=1}^n k^{\overline{2}} = \sum_{k=1}^n k(k+1) = \frac{1}{3}n^{\overline{3}}$

4990568 $\sum_{k=1}^{n-1} k^{\underline{-3}} = \sum_{k=2}^n \frac{1}{k(k+1)(k+2)} =$

$= \frac{1}{1-3}\left(\frac{1}{(n+1)^{\underline{2}}} - \frac{1}{2^{\underline{2}}}\right) = \frac{1}{2}\left(\frac{1}{6} - \frac{1}{(n+1)(n+2)}\right)$

7952392 $\sum_{|\lambda|=n} q^{\binom{n}{2}+\sum_{j=1}^r\binom{b_j}{2}}\frac{\prod_{k=1}^n(q^k-1)}{\prod_{j=1}^r\prod_{i=1}^{b_j}(q^i-1)} = q^{\binom{n}{2}}\sum_{|\lambda|=n} q^{\sum_{j=1}^r\binom{b_j}{2}}\begin{bmatrix} a \\ b_1,b_2,\ldots,b_r\end{bmatrix}_q\prod_{k=a+1}^n(q^k-1)$

4292205 $0 \leq \sum_{j=1}^n (a_{1j}-\lambda x_j)^2 = \sum_{j=1}^n a_{1j}^2 - 2\lambda\sum_{j=1}^n a_{1j}x_j + \lambda^2\sum_{j=1}^n x_j^2 = f(\lambda), \forall\lambda\in\mathbb{R} \Rightarrow \triangle' \leq 0 \Rightarrow \left(\sum_{j=1}^n a_{1j}x_j\right)^2 - \sum_{j=1}^n a_{1j}^2\sum_{j=1}^n x_j^2 \leq 0$

6789156 $\int_0^1 \frac{\ln(1+x)}{x^2+1}dx = \begin{bmatrix} t=\frac{1}{x} \\ dt = -\frac{1}{x^2}dx\end{bmatrix} = \int_1^\infty \frac{\ln(t+1)-\ln t}{1+t^2}dt = \begin{bmatrix} u=t-1 \\ du=dt\end{bmatrix} = \int_0^\infty \frac{\ln(u+2)-\ln(u+1)}{1+(u+1)^2}du.$

5383271 $J = -2\cdot 2^{1/3}t^{-2/3}F\left(2^{1/3}at^{1/3}\right) + 2\cdot(-2)^{1/3}t^{-2/3}F\left(2^{1/3}e^{-2\pi i/3}at^{1/3}\right) - 2\cdot(-1)^{2/3}2^{1/3}t^{-2/3}F\left(2^{1/3}e^{2\pi i/3}at^{1/3}\right)$

6468617 $\frac{\langle x,y\rangle}{\|x\|\|y\|} = \frac{1}{\sqrt{2}} \implies \frac{-1/2+\alpha^2}{1+\alpha^2} = \frac{1}{\sqrt{2}} \implies -1/\sqrt{2}+\sqrt{2}\alpha^2 = 1+\alpha^2 \implies (\sqrt{2}-1)\alpha^2 = \frac{1+\sqrt{2}}{\sqrt{2}} \implies \alpha^2 = \frac{1+\sqrt{2}}{\sqrt{2}(\sqrt{2}-1)} = \frac{4+3\sqrt{2}}{2} \implies \alpha = \sqrt{\frac{4+3\sqrt{2}}{2}}$

2870 $\begin{cases} 20 = \alpha(\frac{1+\sqrt{-3}}{2})^{16} + \beta(\frac{1-\sqrt{-3}}{2})^{16} = \alpha(\frac{-1-\sqrt{-3}}{2}) + \beta(\frac{-1+\sqrt{-3}}{2}) \\ 16 = \alpha(\frac{1+\sqrt{-3}}{2})^{20} + \beta(\frac{1-\sqrt{-3}}{2})^{20} = \alpha(\frac{1-\sqrt{-3}}{2}) + \beta(\frac{1+\sqrt{-3}}{2}) \end{cases}$

6799405 $\therefore 3\left(\frac{\partial^2 u}{\partial\eta^2} - 6\frac{\partial^2 u}{\partial\zeta\partial\eta} + 9\frac{\partial^2 u}{\partial\zeta^2}\right) + 10\left(-3\frac{\partial^2 u}{\partial\zeta^2} + 10\frac{\partial^2 u}{\partial\zeta\partial\eta} - 3\frac{\partial^2 u}{\partial\eta^2}\right) + 3\left(\frac{\partial^2 u}{\partial\zeta^2} - 6\frac{\partial^2 u}{\partial\zeta\partial\eta} + 9\frac{\partial^2 u}{\partial\eta^2}\right) = \sin\frac{\zeta+\eta}{-2}$

3297691 $[\{2,1,1\}] = \{\{\{1,2\},\{3\},\{4\}\},\{\{2,3\},\{1\},\{4\}\},\{\{3,4\},\{1\},\{2\}\},\{\{1,4\},\{2\},\{3\}\}\,\{\{1,3\},\{2\},\{4\}\}\{\{2,4\},\{1\},\{3\}\}\}$

8640511

5571052 $\Delta f = \frac{1}{3}\left[\frac{\partial^2 f}{\partial x^2} + \underbrace{\left(\cos\tfrac{2\pi}{3}\cdot\frac{\partial^2 f}{\partial x^2} + \sin\tfrac{2\pi}{3}\cdot\frac{\partial^2 f}{\partial y^2}\right)}_{\phi=2\pi/3} + \underbrace{\left(\cos\tfrac{4\pi}{3}\cdot\frac{\partial^2 f}{\partial x^2} + \sin\tfrac{4\pi}{3}\cdot\frac{\partial^2 f}{\partial y^2}\right)}_{\phi=4\pi/3}\right]$

7062721 $x = [4.105 - \ln(\sqrt{y})]^2 \Rightarrow \pm\sqrt{x} = 4.105 - \ln(\sqrt{y}) \Rightarrow \pm\sqrt{x} = 4.105 - \ln(y^{\frac{1}{2}}) \Rightarrow \pm\sqrt{x} = 4.105 - \frac{1}{2}\ln(y) \Rightarrow \ln(y) = 2\cdot 4.105 \pm 2\sqrt{x} \Rightarrow y = e^{2\cdot 4.105 \pm 2\sqrt{x}}$

2516293 $\Psi^\star\frac{\partial^2\Psi}{\partial x\partial t} - \frac{\partial^2\Psi^\star}{\partial x\partial t}\Psi = [\frac{i\hbar}{2m}\Psi^\star\frac{\partial^3\Psi}{\partial x^3} + \frac{i\hbar}{2m}\Psi\frac{\partial^3\Psi^\star}{\partial x^3}] - [\frac{i}{\hbar}V\Psi^\star\frac{\partial\Psi}{\partial x} + \frac{i}{\hbar}V\Psi\frac{\partial\Psi^\star}{\partial x}] - \frac{i}{\hbar}\frac{\partial V}{\partial x}|\Psi|^2 - \frac{i}{\hbar}\frac{\partial V}{\partial x}|\Psi|^2$

5233448 $(2+5)+\left(2^2+5^2\right)+\left(2^3+5^3\right)+\left(2^4+5^4\right) = \left(\sum_{i=1}^4 2^i\right)+\left(\sum_{i=1}^4 5^i\right) = \frac{2-2^5}{1-2} + \frac{5-5^5}{1-5} = 30+5(5^4-1)/4 = 30+5\times 624/4 = 30+780 = 810$

6199256 $\int_0^\infty\int_0^\infty e^{-xy}\sin x\,dy\,dx = \int_0^\infty \frac{\sin x}{x}dx = \frac{\pi}{2}, \int_0^\infty\int_0^\infty e^{-xy}\sin x\,dx\,dy = \int_0^\infty \operatorname{Im}\int_0^\infty e^{-xy+ix}\,dx\,dy = \int_0^\infty \operatorname{Im}\frac{1}{y-i}\,dy = \int_0^\infty \frac{1}{y^2+1}\,dy = \frac{\pi}{2}.$

307286 $PA^2 = PB^2 \Rightarrow (x+1)^2+(y-4)^2 = (x-3)^2+(y+2)^2 \Rightarrow x^2+2x+1+y^2-8y+16 = x^2-6x+9+y^2+4y+4 \Rightarrow 2x-8y+17 = -6x+4y+1$

3648070 $\psi_1 = 1, \quad \psi_2 = 2y\psi_3 = 3x^4+6x^2-1\psi_4 = 4y(x^6+5x^4-5x^2-1)\psi_{2n+1} = \psi_{n+2}\psi_n^3 - \psi_{n-1}\psi_{n+1}^3 \quad \text{for } n\geq 2 2y\psi_{2n} = \psi_n(\psi_{n+2}\psi_{n-1}^2 - \psi_{n-2}\psi_{n+1}^2) \quad \text{for } n$

6387213 $\frac{1+\sqrt{1+y}}{1+\sqrt{1-y}} = \frac{1+\sqrt{2}\cos u}{1+\sqrt{2}\sin u} = \frac{\frac{1}{\sqrt{2}}+\cos u}{\frac{1}{\sqrt{2}}+\sin u} = \frac{2\cos\frac{45°+u}{2}\cos\frac{45°-u}{2}}{2\sin\frac{45°+u}{2}\cos\frac{45°-u}{2}} = \cot\frac{45°+u}{2}$

5282088 $I = \int_0^\infty \frac{\tanh^2 z}{z^2}dz = \int_0^\infty\int_0^\infty\int_0^\infty \frac{\sin(xx)\sin(xt)}{x^2\sinh(\frac{\pi}{2}x)\sinh(\frac{\pi}{2}t)}dxdtdz = \int_0^\infty\int_0^\infty \frac{f(x,t)}{\sinh(\frac{\pi}{2}x)\sinh(\frac{\pi}{2}t)}dxdt$

470495 $(D-2)^2(D^2+1)(D-3)Jte^{3t} = ((D-3)+1)^2((D-3+3)^2+1)(D-3)Jte^{3t} = ((D-3)+1)^2((D-3)^2+6(D-3)+1)(D-3)Jte^{3t} = (D-3)Jt$

3475520

$$F(t) = \sum_{k=1}^{\max\left(\frac{k_1}{\gcd(|k_1|,|k_2|)}, \frac{k_2}{\gcd(|k_1|,|k_2|)}, 0\right) - \min\left(\frac{k_1}{\gcd(|k_1|,|k_2|)}, \frac{k_2}{\gcd(|k_1|,|k_2|)}, 0\right)} \Theta_k(t)\lambda_k^t + F_p(t)\,,$$

9186119 $$\frac{\partial E(a)}{\partial a} = \sum_{k=0}^{r} \frac{2\sin(x_k)(-y_k + a\sin(x_k))}{\log(1+x_k^2)} = -\left(\sum_{k=0}^{r} \frac{2\sin(x_k)y_k}{\log(1+x_k^2)}\right) + a\left(\sum_{k=0}^{r} \frac{2\sin^2(x_k)}{\log(1+x_k^2)}\right)$$

8053122 $$\begin{aligned}&\left(\textstyle\sum_{i=1}^{k}(-1)^{i-1}\alpha^i(v)\alpha^1\wedge\cdots\wedge\alpha^{i-1}\wedge\alpha^{i+1}\wedge\cdots\wedge\alpha^k\right)\wedge\alpha^{k+1}\wedge\cdots\wedge\alpha^{k+l}\\ +\ &\alpha^1\wedge\cdots\wedge\alpha^k\left(\textstyle\sum_{i=1}^{l}(-1)^{k+i-1}\alpha^{k+i}(v)\alpha^{k+1}\wedge\cdots\wedge\alpha^{k+i-1}\wedge\alpha^{k+i+1}\wedge\cdots\wedge\alpha^{k+l}\right)\end{aligned}$$

8152468 $$E[2^Y] = E[2^{x-2}] = \textstyle\sum_x g(x)\cdot p_x(x) = \sum_{i=2}^{\infty} 2^i\cdot\binom{i-1}{r-1}\cdot 0.4^{i-2}\cdot 0.6^2 = \sum_{i=2}^{\infty} 2^i\cdot\binom{i-1}{1}\cdot 0.4^{i-2}\cdot 0.6^2 = \sum_{i=2}^{\infty} 2^i\cdot(i-1)\cdot 0.4^{i-2}\cdot 0.6^2$$

1956862 $$\frac{1}{(1+x^2)\cdot(1+x^{2015})} = \frac{1}{(2i)\cdot\prod_{n=0}^{2014}(i-\lambda_n)} + \frac{1}{(2i)\cdot\prod_{n=0}^{2014}(i+\lambda_n)} + \sum_{k=0}^{2014}\lim_{x\to\lambda_k}\frac{\lambda_k-\lambda_n}{(1+\lambda_k^2)\cdot\prod_{n=0}^{2014}(\lambda_k-\lambda_n)}$$

4040519 $$\begin{aligned}\frac{\frac{\sqrt{3}+1}{\sqrt{2}} - \frac{\sqrt{3}-1}{\sqrt{2}}}{\frac{\sqrt{3}+1}{\sqrt{2}} + \frac{\sqrt{3}-1}{\sqrt{2}}} &= \frac{\tan\frac{\theta-(90°-\theta)}{2}}{\tan\frac{\theta+(90°-\theta)}{2}} \Rightarrow\\ \frac{1}{\sqrt{3}} &= \frac{\tan(\theta-45°)}{1} \Rightarrow\\ \tan(\theta-45°) &= \frac{1}{\sqrt{3}} \Rightarrow\\ \theta-45° &= 30° \Rightarrow\\ \theta &= 75°.\end{aligned}$$

4192504 $$\sqrt{\log_x(\sqrt{3x})}\cdot\log_3 x = -1 \Rightarrow \sqrt{\tfrac{1}{2}\left(\tfrac{1}{\log_3 x}+1\right)}\cdot\log_3 x = -1 \Rightarrow \sqrt{\tfrac{\log_3^2 x}{2}\left(\tfrac{1}{\log_3 x}+1\right)} = -1 \Rightarrow \sqrt{\tfrac{1}{2}\left(\log_3 x + \log_3^2 x\right)} = \sqrt{i} \Rightarrow \tfrac{1}{2}\left(\log_3 x + \log_3^2 x\right) = i$$

5795722 $$I = \int_{-\pi/2}^{\pi/2}\sum_{k=0}^{\infty}\frac{\sin^k(\theta)}{k!}d\theta = \sum_{k=0}^{\infty}\int_{-\pi/2}^{\pi/2}\frac{\sin^k(\theta)}{k!}d\theta = \sum_{k=0}^{\infty}\int_{-\pi/2}^{\pi/2}\frac{\sin^{2k}(\theta)}{(2k)!}d\theta = 2\sum_{k=0}^{\infty}\int_{0}^{\pi/2}\frac{\sin^{2k}(\theta)}{(2k)!}d\theta$$

7584851 $$\textstyle\sum_{k=1}^{n}\frac{k^2}{n}\binom{2n-k-1}{n-1} \overset{k\to n+1-k}{=} \sum_{k=1}^{n}\frac{(n+1-k)^2}{n}\binom{n+k-2}{n-1} \overset{\binom{a}{b}=\binom{a}{a-b}}{=} \sum_{k=1}^{n}\frac{(n+1-k)^2}{n}\binom{n+k-2}{k-1} = \sum_{k=0}^{n}\frac{(n-k)^2}{n}\binom{n+k-1}{k}$$

4403310 $$\sqrt[3]{\tfrac{5105}{11349} + \textstyle\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k})} + \sqrt[3]{\tfrac{5105}{11349} + \textstyle\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k+1})} + \sqrt[3]{\tfrac{5105}{11349} + \textstyle\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k+2})} = \sqrt[3]{\tfrac{21}{1261}}$$

7706090 $$\begin{aligned}t_{11}:&\quad (10-2i-6\sqrt{6i}) - 32ix + 32y = 0\\ t_{12}:&\quad (10-2i+6\sqrt{6i}) - 32ix + 32y = 0\\ t_{21}:&\quad (10+2i-6\sqrt{-6i}) + 32ix + 32y = 0\\ t_{22}:&\quad (10+2i+6\sqrt{-6i}) + 32ix + 32y = 0\end{aligned}$$

7922591 $$\mathcal{P}(B) = \{\{\},\{1\},\{2\},\{3\},\{4\},\{5\},\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{2,2\},\{2,3\},\{2,4\},\{2,5\},\{3,3\},\{3,4\},\{3,5\},\{4,4\},\{4,5\},\{5,5\}$$

8053123 $$\begin{aligned}&\left(\textstyle\sum_{i=1}^{k}(-1)^{i-1}\alpha^i(v)\alpha^1\wedge\cdots\wedge\alpha^{i-1}\wedge\alpha^{i+1}\wedge\cdots\wedge\alpha^k\right)\wedge\alpha^{k+1}\wedge\cdots\wedge\alpha^{k+l}\\ +\ &(-1)^k\alpha^1\wedge\cdots\wedge\alpha^k\left(\textstyle\sum_{i=1}^{l}(-1)^{i-1}\alpha^{k+i}(v)\alpha^{k+1}\wedge\cdots\wedge\alpha^{k+i-1}\wedge\alpha^{k+i+1}\wedge\cdots\wedge\alpha^{k+l}\right)\end{aligned}$$

1167668 $$\frac{1000}{\pi\cdot(\frac{500}{\pi})^{\frac{2}{3}}} = \frac{1000}{\pi^{\frac{1}{3}}\cdot 2^{-\frac{2}{3}}\cdot 2^{\frac{2}{3}}\cdot 500^{\frac{2}{3}}} = \frac{1000}{\pi^{\frac{1}{3}}\cdot 2^{-\frac{2}{3}}\cdot 1000^{\frac{2}{3}}} = \frac{2\cdot 1000^{\frac{1}{3}}}{(2\pi)^{\frac{1}{3}}} = \frac{20}{\sqrt[3]{2\pi}} = \frac{10\sqrt[3]{4\pi^2}}{\pi} = 10\sqrt[3]{\frac{4}{\pi}}$$

6578256 $$\begin{aligned}\int_{\pi}^{\infty}\frac{\sin(x)}{x+\sin(x)}\,\mathrm{d}x &= -\frac{\cos(x)}{x+\sin(x)}\Big]_{\pi}^{\infty} - \int_{\pi}^{\infty}\frac{\cos(x)(1+\cos(x))}{(x+\sin(x))^2}\,\mathrm{d}x\\ &= -\frac{1}{\pi} - \int_{\pi}^{\infty}\frac{\cos(x)(1+\cos(x))}{(x+\sin(x))^2}\,\mathrm{d}x\end{aligned}$$

3628244 $$\begin{aligned}\alpha_I:\beta_I:\gamma_I &= \sin A:\sin B:\sin C &&= t_Ac_A:t_Bc_B:t_Cc_C\\ \alpha_O:\beta_O:\gamma_O &= \sin 2A:\sin 2B:\sin 2C &&= 2t_Ac_A^2:2t_Bc_B^2:2t_Cc_C^2\\ \alpha_H:\beta_H:\gamma_H &= \tan A:\tan B:\tan C &&= t_A:t_B:t_C\end{aligned}$$

2101643 $$x^{x+y} = y^{y-x} \iff (xy)^x = \left(\tfrac{y}{x}\right)^y \Rightarrow y = kx(kx^2)^x = k^{kx} \iff x^{2x} = k^{x(k-1)} \Rightarrow k-1 = 2nx^{2x} = k^{x(k-1)} \Rightarrow x^{2x} = (2n+1)^{2nx} \Rightarrow x = (2n+1)^n \Rightarrow (x,$$

7679466 $$8\sin^2 80° - 2\sqrt{3}\sin 40° - 2\cos 40° = 8\sin^2 80° - 4(\tfrac{\sqrt{3}}{2}\sin 40° + \tfrac{1}{2}\cos 40°) = 8\sin^2 80° - 4\cos 20° = 8\cos^2 10° - 4\cos 20° = 4(\cos 20°+1) - 4\cos 20°$$

1049853 $$\textstyle\sum_{i=0}^{n}\sum_{k=i+1}^{n} i^2 = \sum_{k=1}^{n}\sum_{i=0}^{k-1} i^2 = \sum_{k=1}^{n}\sum_{i=0}^{k-1}\binom{i}{2} + \binom{i+1}{2} = \sum_{k=n}^{n}\binom{k}{3} + \binom{k+1}{3} = \binom{n+1}{4} + \binom{n+2}{4} = \frac{2n\cdot(n+1)n(n-1)}{4!} = \frac{1}{12}n^2(n^2-1)\quad\blacksquare$$

2124971 $$\sum_{n=0}^{\infty}\frac{1}{(24n+5)(24n+19)} = \frac{1}{14}\sum_{n=0}^{\infty}\left(\frac{1}{24n+5} - \frac{1}{24n+19}\right) = \frac{1}{14}\sum_{n=0}^{\infty}\left(\int_0^1\left(x^{24n+4} - x^{24n+18}\right)dx\right) = \frac{1}{14}\int_0^1\left(\sum_{n=0}^{\infty}\left(x^{24n+4} - x^{24n+18}\right)\right)dx$$

4727951 $$\lim_{x\to a}\frac{1}{f(x)} = \frac{1}{\lim_{x\to a} f(x)} = \begin{cases}\frac{1}{L},\ L\neq 0,\\ \begin{cases}+\infty, L = \lim_{x\to 0} x^2 = 0,\\ -\infty, L = \lim_{x\to 0} -x^2 = 0,\\ undefined, L = \lim_{x\to 0} x\sin\frac{1}{x} = 0\end{cases}\\ undefined,\ \ f(x) = 0\end{cases},$$

2095505 $$\cos 36^\circ = \sin 54^\circ\cos 2\cdot 18^\circ = \sin 3\cdot 18^\circ\theta = 18^\circ\cos 2\theta = \sin 3\theta 1 - 2\sin^2\theta = 3\sin\theta - 4\sin^3\theta 4\sin^3\theta - 2\sin^2\theta - 3\sin\theta + 1 = 0\,(\sin\theta - 1)\left(4\sin^2\theta + 2s\right.$$

6313498 $$\frac{1}{2-x} = \frac{1}{1-(x-1)} = \sum_{n=0}^{\infty}(x-1)^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}x^k \overset{*}{=} \lim_{m\to\infty}\textstyle\sum_{n=0}^{m}\binom{n}{0}(-1)^n + \sum_{n=1}^{m}\binom{n}{1}(-1)^{n-1}x + \ldots + \sum_{n=k}^{m}\binom{n}{k}(-1)^{n-k}x^k + \ldots$$

6876563 $$\int_0^1\frac{1}{1+(1-\frac{1}{x})^{2015}} = \int_0^1\frac{1}{1+(\frac{x-1}{x})^{2015}} = \int_0^1\frac{1}{1+(\frac{(x-1)^{2015}}{x^{2015}})} = \int_0^1\frac{1}{\frac{x^{2015}+(x-1)^{2015}}{x^{2015}}} = \int_0^1\frac{x^{2015}}{x^{2015}+(x-1)^{2015}}dx$$

1952940 $$\frac{\sqrt{x^3+7x}}{\sqrt{4x^3+5}} = \frac{(x^3+7x)^{\frac{1}{2}}}{(4x^3+5)^{\frac{1}{2}}} = \frac{\left(x^3(1+\frac{7}{x^2})\right)^{\frac{1}{2}}}{\left(x^3(4+\frac{5}{x^3})\right)^{\frac{1}{2}}} = \frac{x^{\frac{3}{2}}(1+\frac{7}{x^2})^{\frac{1}{2}}}{x^{\frac{3}{2}}(4+\frac{5}{x^3})^{\frac{1}{2}}} = \frac{\sqrt{1+7x^{-2}}}{\sqrt{4+5x^{-3}}}$$

6688369 $$\mathrm{Var}(S_n/n) = \frac{1}{n^2}Var(S_n) = \frac{1}{n^2}\textstyle\sum_{i\neq j}\mathrm{Cov}(X_i,X_j) = \frac{1}{n^2}(\sum_{k=1}^{n}\mathrm{Cov}(X_i,X_i) + \sum_{i=1}^{n-1}\sum_{j=i}^{n}\mathrm{Cov}(X_i,X_j)) \le \frac{1}{n^2}(n\cdot\epsilon_0 + \sum_{i=1}^{n-1}\sum_{j=i}^{n}\mathrm{Cov}(X_i,X_j))$$

1599730 $$T(n) \ge \tfrac{1}{2}\textstyle\sum_{k=0}^{\lfloor\log_2 n\rfloor} 2^k(\lfloor\log_2 n\rfloor + 1 - k)2^{\lfloor\log_2 n\rfloor - k} = \frac{1}{2}\times 2^{\lfloor\log_2 n\rfloor}\sum_{k=0}^{\lfloor\log_2 n\rfloor}(\lfloor\log_2 n\rfloor + 1 - k) = \frac{1}{2}\times 2^{\lfloor\log_2 n\rfloor}\left(\frac{1}{2}\lfloor\log_2 n\rfloor^2 + \frac{3}{2}\lfloor\log_2 n\rfloor + 1\right).$$

4018339 $$\iiint_V R^2\sin\theta dRd\theta d\phi=\int_0^{\pi}\int_0^{2\pi}\int_0^{\sqrt{\sin^2\theta\sin^2(2\phi)}}R^2\sin\theta dRd\phi d\theta=\int_0^{\pi}\int_0^{2\pi}\frac{R^3}{3}\Big|_0^{\sqrt{\sin^2\theta\sin^2(2\phi)}}\sin\theta d\phi d\theta=\int_0^{\pi}\int_0^{2\pi}\frac{|\sin\theta|^3|\sin(2\phi)|^3}{3}\sin\theta d\phi d\theta$$

$$\begin{aligned}\lim_{x\to\infty}-x^{n+1}e^{-x}&=-(2n+2)^{n+1}\lim_{x\to\infty}\left(\sqrt{\frac{x}{2n+2}}e^{-\frac{x}{2n+2}}\right)^{2n+2}\\ 5456325\qquad&=-(2n+2)^{n+1}\left(\lim_{x\to\infty}\sqrt{\frac{x}{2n+2}}e^{-\frac{x}{2n+2}}\right)^{2n+2}\\&=-(2n+2)^{n+1}\left(\lim_{x\to\infty}\frac{\sqrt{u}}{e^u}\right)^{2n+2}\end{aligned}$$

3341753 $$3^x+4^x=5^x\Leftrightarrow\frac{3^x}{3^x}+\frac{4^x}{3^x}=\frac{5^x}{3^x}\Leftrightarrow1+\left(\frac{4}{3}\right)^x=\left(\frac{5}{3}\right)^x\Leftrightarrow1+\left(\frac{4}{3}\right)^{\frac{x}{2}\cdot2}=\left(\frac{5}{3}\right)^{\frac{x}{2}\cdot2}\Leftrightarrow\left(\frac{4}{3}\right)^{\frac{x}{2}\cdot2}-\left(\frac{5}{3}\right)^{\frac{x}{2}\cdot2}=-1\Leftrightarrow\left(\left(\frac{4}{3}\right)^{\frac{x}{2}}\right)^2-\left(\left(\frac{5}{3}\right)^{\frac{x}{2}}\right)^2=-1$$

108722 $$\int\arctan(x)^2dx=x\arctan(x)^2-\log(1+x^2)\arctan(x)+\frac{1}{(2i)}\sum_{\xi\in\{-1,1\}}\xi\left(\log(i\xi+x)\log(\frac{1}{2}(1+i\xi x))+Li_2(\frac{1}{2}(1-i\xi x))-\frac{1}{2}\log(x+i\xi)^2\right)\quad(i)$$

3041752 $$|T|-\sum_{j=1}^{q}(-1)^{j-1}\sum_{X\in\binom{[p+q]}{j}}|\bigcap_{y\in X}A_y|=\binom{p+q}{i}-\sum_{j=1}^{q}(-1)^{j-1}\sum_{X\in\binom{[p+q]}{j}}\binom{p+q-j}{i-j}=\binom{p+q}{i}-\sum_{j=1}^{q}(-1)^{j-1}\binom{q}{j}\binom{p+q-j}{i-j},$$

5301956 $$\int\frac{1}{\cos\psi}d\psi=\ln\sqrt{\frac{1+\sin\psi}{1-\sin\psi}}+C=\ln\sqrt{\frac{(1+\sin\psi)(1-\sin\psi)}{(1-\sin\psi)^2}}+C=\ln\sqrt{\frac{1-\sin^2\psi}{(1-\sin\psi)^2}}+C=\ln\sqrt{\frac{\cos^2\psi}{(1-\sin\psi)^2}}+C=\ln\frac{\cos\psi}{1-\sin\psi}+C$$

797447 $$\lim_{x\to0}\frac{\ln\cos x}{\ln\cos3x}=\lim_{x\to0}\frac{\ln\sqrt{1-\sin^2x}}{\ln\sqrt{1-\sin^23x}}=\lim_{x\to0}\frac{\frac{1}{2}\cdot\ln(1-\sin^2x)}{\frac{1}{2}\cdot\ln(1-\sin^23x)}=\lim_{x\to0}\frac{\ln(1-\sin^2x)}{\ln(1-\sin^23x)}=$$

5936514 $$\frac{1}{1+x^n}=\sum_{k=0}^{n/2-1}\frac{a_kx+b_k}{x^2-2\cos(x_k)x+1}=\sum_{k=0}^{n/2-1}\frac{(a_kx+b_k)\prod_{j\neq k}^{j\neq n-1-k}(x-\alpha_j)}{1+x^n}=\sum_{k=0}^{n/2-1}\frac{\frac{a_kx+b_k}{x-\alpha_{n-1-k}}\prod_{j\neq k}(x-\alpha_j)}{1+x^n}.$$

4983615 $$\prod_{n=1}^{N}\sqrt[n]{\cos(nx)}=\prod_{n=1}^{N}\sqrt[n]{1+\frac{(nx)^2}{2}+o(x^2)}=\prod_{n=1}^{N}\left(1+\frac{(nx)^2}{n2}+o(x^2)\right)=\prod_{n=1}^{N}\left(1+\frac{nx^2}{2}+o(x^2)\right)=1+\frac{x^2}{2}\sum_{n=1}^{N}n+o(x^2)=1+\frac{N(N+1)}{4}\cdot x^2+o($$

4867822 $$\operatorname{Tr}\left((A(xA+I))^{-1}-\frac{I}{2+x}\right)=\operatorname{Tr}\left(P\left((A(xA+I))^{-1}-\frac{I}{2+x}\right)P^{-1}\right)=\operatorname{Tr}\left((xD^2+D)^{-1}-\frac{I}{2+x}\right)=\sum_{i=1}^{n}\frac{1}{x\lambda_i^2+\lambda_i}-\frac{1}{2+x}=\sum_{i=1}^{n}\frac{x(1-\lambda_i)^2+2-\lambda_i}{(2+x)\lambda_i(x\lambda_i+1)}.$$

7018260 $$\lim_{n\to\infty}[-(f(b)-f(a))]\frac{1}{n}\leq\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(nx)\leq\lim_{n\to\infty}(f(b)-f(a))\frac{1}{n}\Rightarrow0\leq\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(nx)\leq0\Rightarrow\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(n$$

7018263 $$\lim_{n\to\infty}(f(b)-f(a))\frac{1}{n}\leq\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(nx)\leq\lim_{n\to\infty}[-(f(b)-f(a))]\frac{1}{n}\Rightarrow0\leq\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(nx)\leq0\Rightarrow\lim_{n\to\infty}\frac{1}{n}\int_a^bf'(x)\cos(n$$

3683737 $$\sum_{n=1}^{\infty}\frac{\cot^{-1}(n)}{n}=\sum_{n=1}^{\infty}\frac{\int_0^{\infty}\frac{\exp(-nx)\sin(x)}{x}dx}{n}=\int_0^{\infty}\frac{\sin(x)\sum_{n=1}^{\infty}\frac{\exp(-nx)}{n}}{x}dx=\int_0^{\infty}\frac{\sin(x)(-\log(1-e^{-x}))}{x}dx=\int_0^{\infty}-\frac{\sin(x)\log(1-e^{-x})}{x}dx$$

1964627 $$\lim_{n\to\infty}n^2x(1-x^2)^n=x\lim_{n\to\infty}n^2a^n\leq\lim_{n\to\infty}n^2a^n=\lim_{n\to\infty}\frac{n^2}{(1/a)^n}=\lim_{n\to\infty}\frac{2n}{\ln(1/a)(1/a)^n}=\lim_{n\to\infty}\frac{2}{\ln^2(1/a)(1/a)^n}=\frac{2}{\ln^2(1/a)}\lim_{n\to\infty}a^n=0$$

2519965 $$\sum_{n=0}^{\infty}\left(Y_n''-n^2\pi^2Y_n(y)\right)\int_0^1\cos(m\pi x)\cos(n\pi x)=\sum_{n=0}^{\infty}\left(Y_n''-n^2\pi^2Y_n(y)\right)\frac{1}{2}\int_0^1\cos(2m\pi x)+1dx=\sum_{n=0}^{\infty}\left(Y_n''-n^2\pi^2Y_n(y)\right)\frac{1}{2}\left[\frac{sin(nπmx)}{2mπ}+x\right|_0^1\Big]$$

3011982 $$\begin{cases}\text{square of length } C_1C_3: & (x_3-3)^2+y_3^2 &=& (3+x)^2 & (a)\\ \text{square of length } C_2C_3: & (x_3-8)^2+y_3^2 &=& (2+x)^2 & (b)\\ \text{square of length } C_0C_3: & (x_3-5)^2+y_3^2 &=& (5-x)^2 & (c)\end{cases}$$

4092194 $$\sum_{i=1}^{n}(x_i-\mu)^2\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\sum_{i=1}^{n}(x_i-\bar{x})^2+2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x})+\sum_{i=1}^{n}(\bar{x}-\mu)^2\sum_{i=1}^{n}(x_i-\bar{x})=0\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2$$

4010093 $$x_{n+2}=2018^2(x_1\cdot x_2\cdot\cdots\cdot x_nx_{n+1})^{-\frac{1}{n+2}}=2018^2\left(\left(\frac{2018^2}{x_{n+1}}\right)^{n+1}x_{n+1}\right)^{-\frac{1}{n+2}}=\left(2018^2\right)^{1-\frac{n+1}{n+2}}\left(\left(x_{n+1}\right)^{-n-1}x_{n+1}\right)^{-\frac{1}{n+2}}=2018^{\frac{2}{n+2}}\left(x_{n+1}\right)^{\frac{n}{n+2}}$$

8990145 $$((1+x^2)^2)^2=(x^4+2x^2+1)^2=(ax^3+bx)^2\leq(a^2+b^2)(x^6+x^2)=(a^2+b^2)x^2(x^4+1)\implies a^2+b^2\geq\frac{(1+x^2)^4}{x^2(x^4+1)}\geq8\iff(1+x^2)^4\geq8x^2(1+$$

6375186 $$-x^2y'y+xy''-y'=0\overset{\cdot\frac{1}{x^2}}{\iff}\frac{y''}{x}-\frac{y'}{x^2}-y'y=0\overset{\int dx}{\iff}\frac{y'}{x}-\frac{y^2}{2}=c\iff\frac{2y'}{2c+y^2}=x\overset{\int dx,\ s=\frac{y}{\sqrt{2c}}}{\iff}\frac{\sqrt{2}\tan^{-1}\left(\frac{y}{\sqrt{2c}}\right)}{\sqrt{c}}=\frac{x^2}{2}+c_1\iff y=\sqrt{2c}\tan\left(\frac{\sqrt{c}(x^2+c_1)}{2\sqrt{2}}\right.$$

6936678 $$\int_0^1\frac{\ln x\ln(1+x)}{(1+x)^2}dx=-\frac{\ln x\ln(1+x)}{1+x}\Big|_0^1+\int_0^1\frac{1}{1+x}\left(\frac{\ln(1+x)}{x}+\frac{\ln x}{1+x}\right)dx=-\ln2+\int_0^1\frac{\ln(1+x)}{x}dx-\int_0^1\frac{\ln(1+x)}{1+x}dx=\frac{\pi^2}{12}-\frac{1}{2}\ln^22-\ln2.$$

6789162 $$\int_0^R\frac{\log^2(x+i\epsilon+2)}{(1+x+i\epsilon)^2+1}dx+\int_R^0\frac{\log^2(x-i\epsilon+2)}{(1+x-i\epsilon)^2+1}dx\to\int_0^{\infty}\frac{\ln^2(x+2)-(\ln(x+2)+2i\pi)^2}{(1+x)^2+1}dx=\int_0^{\infty}\frac{4\pi^2-4i\pi\ln(x+2)}{(1+x)^2+1}dx,$$

8552650 $$=\alpha_n^n(x_1+\alpha_nx_2+\alpha_n^2x_3+\cdots+\alpha_n^{n-1}x_n)^n=(x_n+\alpha_nx_1+\alpha_n^2x_2+\cdots+\alpha_n^{n-1}x_{n-1})^n=\alpha_n^n(x_n+\alpha_nx_1+\alpha_n^2x_2+\cdots+\alpha_n^{n-1}x_{n-1})^n=(x_{n-1}+\alpha_nx_n+a$$

5092375 $$||U_n||_{\infty}=\sup_{x\geq a}\left(\frac{x}{(1+(n-1)x)(1+nx)}\right)\geq\sup_{x\geq a}\left(\frac{x}{(1+(n-1)a)(1+na)}\right)\geq\sup_{x\geq a}\left(\frac{x}{(1+na)^2}\right)\geq\frac{1}{(1+a)^2}\sup_{x\geq a}\left(\frac{x}{n^2}\right)\geq\frac{1}{(1+a)^2}\times\frac{n}{n^2}=\frac{1}{(1+a)^2}\frac{1}{n}=a_n$$

5852903 $$\frac{\sin(B/2)}{\sin(C/2)}=\frac{\sin(24^\circ)}{\sin(18^\circ)}\cdot\frac{CE}{BD}=\frac{\sin(24^\circ)}{\sin(18^\circ)}\cdot\frac{\frac{ab}{a+c}}{\frac{ac}{a+b}}=\frac{\sin(24^\circ)}{\sin(18^\circ)}\cdot\frac{\sin B}{\sin C}\cdot\frac{\sin(A)+\sin(B)}{\sin(A)+\sin(C)}$$

7819258 $$U_{r+3}-6U_{r+2}+11U_{r+1}-6U_r=0,\ U_0=1,\ U_1=-1,\ U_2=0\therefore L=4\therefore p(x)=\frac{\sum_{k=0}^{2}\sum_{r=0}^{2-k}a_kU_rx^r}{\sum_{r=0}^{3}a_rx^r}=\frac{1-7x+17x^2}{1-6x+11x^2-6x^3}=\frac{1-7x+17x^2}{(1-x)(1-2x)(1-3x)}$$

4269433 $$\ln P=(1+x)\ln(1+x)+(1-x)\ln(1-x)=x(\ln(1+x)-\ln(1-x))+\ln(1+x)+\ln(1-x)=2x\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots\right)-2\left(\frac{x^2}{2}+\frac{x^4}{4}+\frac{x^6}{6}+\cdots\right)=$$

4028973 $$\sin(\arcsin\tfrac{x}{3})=\tfrac{x}{3}\sin^2(\arcsin\tfrac{x}{3})=\tfrac{x^2}{9}1-\sin^2(\arcsin\tfrac{x}{3})=1-\tfrac{x^2}{9}\cos^2(\arcsin\tfrac{x}{3})=1-\tfrac{x^2}{9}\cos(\arcsin\tfrac{x}{3})=\sqrt{1-\tfrac{x^2}{9}}\cos(\arcsin\tfrac{x}{3})=\tfrac{\sqrt{9-x^2}}{3}$$

5334898 $$b\sum_{n=a}^{b}\binom{n}{a}-\left(\sum_{n=a}^{b-1}\binom{n}{a}+\sum_{n=a}^{b-2}\binom{n}{a}+\cdots+\sum_{n=a}^{a}\binom{n}{a}\right)=b\binom{b+1}{a+1}-\left(\binom{b}{a+1}+\binom{b-1}{a+1}+\cdots+\binom{a+1}{a+1}\right)=b\binom{b+1}{a+1}-\sum_{i=a+1}^{b}\binom{i}{a+1}=b\binom{b+1}{a+1}-\binom{b+1}{a+2}$$

32977 $$\frac{x^5+x^3+(-1)^{5/6}\sqrt{3\sqrt[3]{-1}x^2+3}\sqrt{1-(-1)^{2/3}x^2}F(i\sinh^{-1}((-1)^{5/6}x)|(-1)^{2/3})+\sqrt[3]{-1}\sqrt{\sqrt[3]{-1}x^2+1}\sqrt{1-(-1)^{2/3}x^2}E(i\sinh^{-1}((-1)^{5/6}x)|(-1)^{2/3})+x}{3\sqrt{x^4+x^2+1}},$$

8131484 $$\lim_{n\to\infty}a_n=\lim_{n\to\infty}\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{n+n}==\lim_{n\to\infty}\frac{1}{n}(\frac{1}{1+\frac{1}{n}}+\frac{1}{1+\frac{2}{n}}+\frac{1}{1+\frac{3}{n}}+\ldots+\frac{1}{1+\frac{n}{n}})=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\frac{i}{n}}=\int_0^1\frac{1}{1+x}dx=ln(1+x)=ln(2)$$

4644947 $$\sin54^\circ\cos108^\circ=\cos36^\circ\cos(2\cdot54^\circ)=\cos36^\circ(\cos^254^\circ-\sin^236^\circ)=\cos36^\circ\cos^254^\circ-\cos^354^\circ=\tfrac{1}{2}\left[\cos90^\circ+\cos18^\circ\right]\cdot\cos54^\circ-\cos^336^\circ=\tfrac{1}{2}\cos18^\circ$$

8158550 $$=\ln314\sum_{n=0}^{\infty}\frac{x}{4^n(n!)^2}+\ln314\sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{((k-1)!)^2\sin x\cos^{2k-1}x}{4^{n-k+1}(n!)^2(2k-1)!}+\ln314\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{(-1)^kn!\sin^{2k+1}x}{(2n+1)!k!(n-k)!(2k+1)}+C$$

1865517 $$\lim_{t\to0}(1+t^2)=1,\qquad\lim_{t\to0}\frac{1}{t^2}=\infty\qquad\lim_{t\to0}(1+t^2)^{1/t^2}=e\lim_{t\to0}(1+t)=1,\qquad\lim_{t\to0}\frac{1}{t^2}=\infty\qquad\lim_{t\to0}(1+t)^{1/t^2}=\infty\lim_{t\to0}(1+t^2)=1,$$

6598436 $$n_1=120250=(68^2+1)(5^2+1)=(57^2+1)(6^2+1)=(43^2+1)(8^2+1)n_3=505723158638933050=(98048^2+1)(7253^2+1)=(82137^2+1)(8658^2$$

4935286 $$\frac{2\left(x^5+x^2-3^{3/4}\sqrt{-\sqrt[6]{-1}(x+(-1)^{2/3})}\sqrt{(-1)^{2/3}x^2+\sqrt[3]{-1}x+1}\left((-1)^{5/6}F\left(\sin^{-1}\left(\frac{\sqrt{-(-1)^{5/6}(x+1)}}{\sqrt[6]{3}}\right)|\sqrt[3]{-1}\right)+\sqrt{3}E\left(\sin^{-1}\left(\frac{\sqrt{-(-1)^{5/6}(x+1)}}{\sqrt[6]{3}}\right)|\sqrt[3]{-1}\right)\right)\right)}{7\sqrt{x^3+1}}$$

4273512 \()$. It follows that $k_1,\ldots,k_n$ is a permutation of $1,2,\ldots,n$.

Conversely, if A has $\epsilon_i = \pm1$'s at entries $(i,\pi(i))$ $($for a fixed permutation $\pi\in S_n\)

4440774 $\frac{(2n)!!}{(2n+3)!!} = \prod_{k=1}^{n}\left(1-\frac{3}{2k+3}\right) \le \exp\left(-\sum_{k=1}^{n}\frac{3}{2k+3}\right) \le \exp\left(-\frac{3}{2}\sum_{k=3}^{n+2}\frac{1}{k}\right) \le \exp\left(-\frac{3}{2}\sum_{k=3}^{n+2}\log\frac{k+1}{k}\right) = \left(\prod_{k=3}^{n+2}\frac{k+1}{k}\right)^{-3/2} = \left(\frac{n+3}{3}\right)^{-3/2} < 6\cdot n^{-3/2}$

2076828 $\sum_{n\ge1}\frac{\cos n}{n} = \int_0^{+\infty}\frac{e^s\cos(1)-1}{e^{2s}-2e^s\cos(1)+1}\,s\,ds = \int_0^{+\infty}\left(s-\log\sqrt{e^{2s}-2e^s\cos(1)+1}\right)ds = -\frac{1}{2}\int_0^{+\infty}\log\left(1+e^{-2s}-2e^{-s}\cos(1)\right)ds = -\int_0^1\frac{\log(1+t^2-2t\cos(1))}{2t}\,dt$

5326795

6926751 $\frac{I}{\pi}\Sigma_k\frac{1}{k}[2\sin(k\theta)+\sin(k(\theta-\frac{\pi}{3}))-\sin(k(\theta-\frac{2\pi}{3}))-2\sin(k(\theta\pm\pi))-\sin(k(\theta+\frac{2\pi}{3}))+\sin(k(\theta+\frac{\pi}{3}))] = \frac{I}{\pi}\Sigma_k\frac{\sin(k\theta)}{k}[1+\cos(\frac{k\pi}{3})-\cos(\frac{2k\pi}{3})-\cos(k\pi)]$

5449454 $2017^{13164589}\equiv 2017^{13164589 \mod \phi(6)}\equiv 2017^{13164589 \mod 2}\equiv 2017^1\equiv 1 \mod 62099^{2017^{13164589}}\equiv 2099^{2017^{13164589 \mod \phi(9)}}\equiv 2099^{2017^{13164589 \mod 6}}\equiv 2099^1\equiv 2$

1212316 $f(h,1)=\sum_{n=1}^{h}n=\frac{h(h+1)}{2}f(h,2)=\sum_{n=1}^{h}n^2=\frac{f(h,1)(2h+1)}{3}f(h,3)=\sum_{n=1}^{h}n^3=f(h,1)^2f(h,4)=\sum_{n=1}^{h}n^4=\frac{f(h,2)(3h^2+3h-1)}{5}f(h,5)=\sum_{n=1}^{h}n^5=2f(h$

6677118 $I=\int\sqrt{\frac{-\cos^2 t}{-\sin^2 t}}\sin 2tdt=\int 2\cos^2 tdt=\int(1+\cos 2t)dt=t+\frac{1}{2}\sin 2t+cI=\underbrace{\cos^{-1}\sqrt{3-x}}_{\pi/2-\sin^{-1}\sqrt{3-x}}+\sqrt{x-2}\sqrt{3-x}+cI=\underbrace{\sqrt{x-2}\sqrt{3-x}}_{\sqrt{5x-x^2-6}}-\sin^{-1}\sqrt{3-x}+c$

7219582 $\sum_{k=0}^{2n}\binom{2n}{k}\sin((n-k)x)=\sum_{i=1}^{n}\binom{2n}{n-i}\sin((n-(n-i))x)+\sum_{i=1}^{n}\binom{2n}{n+i}\sin((n-(n+i))x)\overset{\binom{2n}{n-i}=\binom{2n}{n+i}}{=}\sum_{i=1}^{n}\binom{2n}{n-i}\sin(ix)-\sum_{i=1}^{n}\binom{2n}{n-i}\sin(ix)=0$

2248393 $x=2\sin u\Rightarrow dx=2\cos u-1=2\sin u_1, 1=2\sin u_2 u_1=\arcsin(-1/2), u_2=\arcsin(1/2)\Rightarrow\int\frac{1}{\sqrt{4-x^2}}dx=\int\frac{1}{\sqrt{4-4\sin^2 u}}2\cos u du=\int\frac{1}{2\sqrt{1-\sin^2 u}}2\cos ud$

4345069 $u(x,t)=1+\sum_{n=0}^{\infty}\frac{(-1)^n x^{4n}\sin t}{(4n)!}+\sum_{n=0}^{\infty}\frac{(-1)^n x^{4n+2}\cos t}{(4n+2)!}+\sum_{n=0}^{\infty}\frac{(-1)^n x^{4n+1}(\sin t+\cos t)}{\sqrt{2}(4n+1)!}+\sum_{n=0}^{\infty}\frac{(-1)^n x^{4n+3}(\sin t-\cos t)}{\sqrt{2}(4n+3)!}$

9173048 $\sum_{j=1}^{n}|(1-\lambda)x_j+\lambda y_j|^2\le\sum_{j=1}^{n}((1-\lambda)|x_j|+\lambda|y_j|)^2=\sum_{j=1}^{n}\left((1-\lambda)^2|x_j|^2+2\lambda(1-\lambda)|x_j||y_j|+\lambda^2|y_j|^2\right)=(1-\lambda)^2\sum_{j=1}^{n}|x_j|^2+2\lambda(1-\lambda)\sum_{j=1}^{n}|x_$

643860 $t_{\alpha/2,df=n-k}=t_{0.025,20}=2.086; MSW=\frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{n-k}=\frac{920.5}{20}=46.025; LSD=2.086\sqrt{46.025\left(\frac{1}{6}+\frac{1}{6}\right)}=8.17; |x_1-x_2|=|47.17-15.67|=31.5>8.17$

x	$2x:x^2-1:x^2+1$
8	16 : 63 : 65
12	24 : 143 : 145
18	36 : 323 : 325
22	44 : 483 : 485
6633755 28	56 : 783 : 785
30	60 : 899 : 901
32	64 : 1023 : 1025
34	68 : 1155 : 1157
38	76 : 1443 : 1445
42	84 : 1763 : 1765

1613064 $$\int_0^1\frac{\ln(1-x)\ln(1+x)}{1+x^2}dx=\int_0^1\frac{(\ln(1+x))^2}{1+x^2}dx-\int_0^1\frac{\ln(1+x)\ln x}{1+x^2}dx+\ln 2\int_0^1\frac{\ln\left(\frac{1-x}{1+x}\right)}{1+x^2}dx$$
$$=\int_0^1\frac{(\ln(1+x))^2}{1+x^2}dx-\int_0^1\frac{\ln(1+x)\ln x}{1+x^2}dx-G\ln 2$$

$$(\underbrace{n}_{A}+\underbrace{1}_{B})+\underbrace{\sum_{k=1}^{n}k\binom{n}{k}}_{C}+\underbrace{\sum_{k=1}^{n}(k-1)\binom{n}{k-1}}_{D}+\underbrace{\sum_{k=1}^{n}\binom{n}{k-1}}_{E}$$

2863193

2977124 $I=\int\sqrt{\frac{\cos x-\cos^3 x}{1-\cos^3 x}}dx=\int\sqrt{\frac{\cos x(1-\cos^2 x)}{1-\cos^3 x}}dx=\int\sqrt{\frac{\cos x\sin^2 x}{1-\cos^3 x}}dx=\int\sin x\sqrt{\frac{\cos x}{1-\cos^3 x}}dx=\int\sqrt{\frac{\cos x}{1-\cos^3 x}}(\sin x)dx=-\int\sqrt{\frac{\cos x}{1-\cos^3 x}}(-\sin x)dx$

771141 $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{4m\cos\left(\frac{2\pi mg}{T_x}\right)\cos\left(\frac{2\pi ng}{T_y}\right)\sin\left(\frac{\pi ml_x}{T_x}\right)\sin\left(\frac{\pi nl_y}{T_y}\right)\left(1-\exp\left(-\pi h\sqrt{\left(\frac{m}{T_x}\right)^2+\left(\frac{n}{T_y}\right)^2}\right)\cosh\left(2\pi z\sqrt{\left(\frac{m}{T_x}\right)^2+\left(\frac{n}{T_y}\right)^2}\right)\right)}{\pi^2\left(nT_x^2\left(\left(\frac{m}{T_x}\right)^2+\left(\frac{n}{T_y}\right)^2\right)\right)}$

4957972 $\int_0^1\frac{(x-\frac{1}{2})+\frac{3}{2}}{3(x-\frac{1}{2})^2+\frac{9}{4}}dx=\int_{-1/2}^{1/2}\frac{u+\alpha}{3u^2+\alpha^2}du=\frac{1}{3}\int_{-1/2}^{1/2}\frac{\alpha}{u^2+\frac{\alpha^2}{3}}du=\frac{2\alpha}{3}\int_0^{1/2}\frac{1}{u^2+\beta^2}du=\frac{1}{\beta}\arctan\frac{u}{\beta}\Big|_0^{1/2}=\frac{2}{\sqrt{3}}\arctan\frac{1}{\sqrt{3}}=\frac{2}{\sqrt{3}}\frac{\pi}{6}=\frac{\pi}{3\sqrt{3}},$

9284719 $2^n=\sum_{k=0}^{n}\binom{n}{k}=\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor-1}\binom{n}{k}+\binom{n}{\lfloor\frac{n}{2}\rfloor}+\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{k}=\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor-1}\binom{n}{n-k}+\binom{n}{\lfloor\frac{n}{2}\rfloor}+\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{k}=\sum_{j=n-\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{j}+\binom{n}{\lfloor\frac{n}{2}\rfloor}+\sum_{k=\lfloor\frac{n}{2}\rfloor+1}^{n}\binom{n}{k}.$

7028735 $P(n,x)=\frac{1}{365^{x-1}}+365\sum_{m=2}^{n}\sum_{y=2}^{x}\left(\frac{m-2}{365}\right)^{x-y}\left(\frac{1}{365}\right)^{y}\binom{x}{y}(2^y-2)=\frac{1}{365^{x-1}}+365\sum_{m=2}^{n}\left[\sum_{y=2}^{x}\left(\frac{m-2}{365}\right)^{x-y}\left(\frac{2}{365}\right)^{y}\binom{x}{y}-2\sum_{y=2}^{x}\left(\frac{m-2}{365}\right)^{x-y}\left(\frac{1}{365}\right)^{y}\binom{x}{y}\right].$

4312643 $f(a)=\log(1+ax^2)-\ln a\Rightarrow f'(a)=\frac{x^2}{1+ax^2}-\frac{1}{a}<0\Rightarrow \text{ since } b=\left(2+\frac{1}{\sqrt{2}}\right)^2>\left(1-\frac{1}{\sqrt{2}}\right)^2=a\Rightarrow f(a)<f(b) \text{ and } \ln\left(\frac{b}{a}\right)\le\frac{b}{a}-1\Rightarrow \text{the in}$

8699392 $A=\frac{1}{r^2+1+2r\cos(\theta-k\omega)}=\frac{1}{r^2+1}\left(1-a\cos(\theta-k\omega)+a^2\cos^2(\theta-k\omega)+\cdots\right),\ a=\frac{2r}{r^2+1}=\frac{1}{r^2+1}\sum_{l=0}^{\infty}(-1)^l a^l\cos^l(\theta-k\omega)=\frac{1}{r^2+1}\sum_{l=0}^{\infty}\sum_{m=0}^{\lfloor l/2\rfloor}\frac{a^l}{2^l}\binom{l}{2m}\cos((l-$

45579 $\sin(x)\prod_{i=0}^{n}\cos(2^i x)=\sin(x)\cos(x)\cos(2x)\cos(4x)\ldots\cos(2^n x)==\frac{\sin(2x)}{2}\cos(2x)\cos(4x)\ldots\cos(2^n x)==\frac{\sin(4x)}{4}\cos(4x)\ldots\cos(2^n x)=\ldots==\frac{\sin(2^{n+1}x)}{2^{n+1}}$

5326813

$$a=9\Rightarrow 5+4+3+2+1 \text{ possible numbers}$$
$$a=7\Rightarrow 4+3+2+1 \text{ possible numbers}$$
$$a=6\Rightarrow 3+2+1 \text{ possible numbers}$$
6164551 $$a=5\Rightarrow 2+1 \text{ possible numbers}$$
$$a=4\Rightarrow 1 \text{ possible number: } 432$$
$$a=3\Rightarrow \text{no possible numbers}$$

6084032 $$y(x)=C_1\,\mathrm{e}^{\frac{1}{6}\,ix\left(2\,x^2+3\right)}\,HeunT\left(-\frac{3^{2/3}\sqrt[3]{2}(4\,c-1)}{8},\frac{3\,i}{2},-\frac{2^{2/3}\sqrt[3]{3}}{2},\frac{i\sqrt[3]{2}\,3^{2/3}}{3}\,x\right)+C_2\mathrm{e}^{-\frac{1}{6}\,ix\left(2\,x^2+3\right)}\,HeunT\left(-\frac{3^{2/3}\sqrt[3]{2}(4\,c-1)}{8},-\frac{3\,i}{2},-\frac{2^{2/3}\sqrt[3]{3}}{2},-\frac{i\sqrt[3]{2}\,3^{2/3}}{3}\,x\right).$$

7630297

2109759 $$\sum_{n=0}^{\infty}\frac{\partial A(n,t)}{\partial t}\sin\frac{(2n+1)\pi\ln x}{2}-\sum_{n=0}^{\infty}\frac{(2n+1)\pi A(n,t)}{2}\cos\frac{(2n+1)\pi\ln x}{2}+\sum_{n=0}^{\infty}\frac{(2n+1)^2\pi^2A(n,t)}{4}\sin\frac{(2n+1)\pi\ln x}{2}+\sum_{n=0}^{\infty}\frac{(2n+1)\pi A(n,t)}{2}\cos\frac{(2n-}$$

$$T_1=(1,0)$$
$$T_2=(-1,0)$$
$$T_3=\left(\tfrac{1}{8}(-5+\sqrt{73}),-\tfrac{1}{4}\sqrt{\tfrac{1}{2}(7+13\sqrt{73})}\right)$$
5707816 $$T_4=\left(-\tfrac{1}{8}(-5+\sqrt{73}),\tfrac{1}{4}\sqrt{\tfrac{1}{2}(7+13\sqrt{73})}\right)$$
$$T_5=\left(\tfrac{1}{4}(3+\sqrt{73}),\tfrac{1}{2}\sqrt{\tfrac{1}{2}(-5+\sqrt{73})}\right)$$
$$T_6=\left(-\tfrac{1}{4}(3+\sqrt{73}),-\tfrac{1}{2}\sqrt{\tfrac{1}{2}(-5+\sqrt{73})}\right)$$

1928775 $$f(n)=\begin{cases}\overbrace{\sum_{k=0}^{(n-1)/2}\binom{n}{k}\sum_{l=0}^{n+1}\binom{n+1}{l}}^{=2^{n-1}\qquad\qquad=2^{n+1}}, & \text{if } n \text{ is odd}\\ \underbrace{\sum_{k=0}^{\frac{n}{2}-1}\binom{n}{k}}_{=2^{n+1}}\underbrace{\sum_{l=0}^{n+1}\binom{n+1}{l}}+\binom{n}{n/2}\underbrace{\sum_{l=\frac{n}{2}+1}^{n+1}\binom{n+1}{l}}_{=2^{n+1}/2=2^n}, & \text{if } n \text{ is even.}\end{cases}$$

4167907 $$I=\int_0^1\frac{\log(1-u)}{u-1}(\log u)^2\,du=\int_0^1\sum_{n=0}^{\infty}H_nu^n(\log u)^2\,du=\sum_{n=0}^{\infty}H_n\int_0^1u^n(\log u)^2\,du=\sum_{n=0}^{\infty}H_n\frac{2}{(n+1)^3}=2\sum_{n=1}^{\infty}\frac{H_{n-1}}{n^3}=2\left(\sum_{n=1}^{\infty}\frac{H_n}{n^3}-\sum_{n=1}^{\infty}\frac{1}{n^4}\right)=$$

6438104 $$|S|=|\bigcup_{i=1}^{5}A_i|=\sum_{i=1}^{5}|A_i|=\sum_{i=1}^{5}\frac{(6+i)!}{i!6!}=\frac{7!}{1!6!}+\frac{8!}{2!6!}+\frac{9!}{3!6!}+\frac{10!}{4!6!}+\frac{11!}{5!6!}=7+28+\frac{9\times8\times7}{3\times2}+\frac{10\times9\times8\times7}{4\times3\times2}+\frac{11\times10\times9\times8\times7}{5\times4\times3\times2}=7+28+84+210+462=791$$

n	Fat subsets
0	$\varnothing$
1	$\varnothing,\{1\}$
3803587 2	$\varnothing,\{1\},\{2\}$
3	$\varnothing,\{1\},\{2\},\{3\},\{2,3\}$
4	$\varnothing,\{1\},\{2\},\{3\},\{2,3\},\{4\},\{2,4\},\{3,4\}$
5	$\varnothing,\{1\},\{2\},\{3\},\{2,3\},\{4\},\{2,4\},\{3,4\},\{5\},\{2,5\},\{3,5\},\{4,5\},\{3,4,5\}$

$$A=-\frac{1}{2}\cos^2(\zeta)\mathrm{d}\varphi_1\wedge\mathrm{d}\varphi_2-\left(\frac{\partial P}{\partial\varphi_1}\mathrm{d}\varphi_1+\frac{\partial P}{\partial\varphi_2}\mathrm{d}\varphi_2\right)\wedge\mathrm{d}\zeta$$
8659507 $$A=-\frac{1}{2}\cos^2(\zeta)\mathrm{d}\varphi_1\wedge\mathrm{d}\varphi_2-\left(\frac{\partial P}{\partial\zeta}\mathrm{d}\zeta+\frac{\partial P}{\partial\varphi_1}\mathrm{d}\varphi_1+\frac{\partial P}{\partial\varphi_2}\mathrm{d}\varphi_2\right)\wedge\mathrm{d}\zeta$$
$$A=-\frac{1}{2}\cos^2(\zeta)\mathrm{d}\varphi_1\wedge\mathrm{d}\varphi_2-\mathrm{d}P\wedge\mathrm{d}\zeta$$
$$A=-\frac{1}{2}\cos^2(\zeta)\mathrm{d}\varphi_1\wedge\mathrm{d}\varphi_2+\mathrm{d}\zeta\wedge\mathrm{d}P$$

2011531 $$I/2=\int\frac{\sin^2t+2}{\sqrt{4\sin^2t-\sin^4t}}\sin t\cos t dtI/2=\int\frac{\sin^2t+2}{\sin t\sqrt{4-\sin^2t}}\sin t\cos t dt=\int\frac{\sin^2t+2}{\sqrt{4-\sin^2t}}\cos t dt=1/2\int\frac{\sin^2t+2}{\sqrt{1-\frac{\sin^2t}{4}}}\cos t dt\Rightarrow I=\int\frac{\sin^2t+2}{\sqrt{1-\frac{\sin^2t}{4}}}\cos t dt$$

7780785 $$\int_{\theta=0}^{\theta=2\pi}\int_{\varphi=0}^{\varphi=\pi/2}(8\sin^3\varphi\cos^3\theta,2\cos\varphi,2\sin\varphi\sin\theta)\cdot(4\sin^2\varphi\cos\theta,4\sin^2\varphi\sin\theta,4\sin\varphi\cos\varphi)d\varphi d\theta=\int_{\theta=0}^{\theta=2\pi}\int_{\varphi=0}^{\varphi=\pi/2}(32\sin^5\varphi\cos^4\theta+16\sin^2\varphi\cos\varphi\sin\theta)d$$

5834667 $$0=\lim_{k\to\infty}\sum_{l=0}^{k}0=\overbrace{\lim_{k\to\infty}\sum_{l=0}^{k}\left(\lim_{n\to\infty}\frac{1}{10^n}\right)=\lim_{n\to\infty}\left(\lim_{k\to\infty}\sum_{l=0}^{k}\frac{1}{10^n}\right)}^{\text{You cannot interchange limits arbitrarily}}=\lim_{n\to\infty}\lim_{k\to\infty}\frac{k}{10^n}=\lim_{n\to\infty}\infty=\infty$$
This is incorrect

6878060 $= \frac{1}{2\sqrt{3}}(2\sin^2(A) + 2\sin^2(B) + 4 + 2\cos(B) + 2\cos(A)) = \frac{1}{\sqrt{3}}(1 - \cos^2(A) + 1 - \cos^2(B) + 2 + \cos(B) + \cos(A)) = \frac{1}{\sqrt{3}}\left(1 + \frac{1}{4} - \left(\cos(A) - \frac{1}{2}\right)^2 + 1 +$

2109758 $$\sum_{n=0}^{\infty} \frac{\partial A(n,t)}{\partial t} \sin\frac{(2n+1)\pi \ln x}{2} - x\sum_{n=0}^{\infty} \frac{(2n+1)\pi A(n,t)}{2x} \cos\frac{(2n+1)\pi \ln x}{2} - x^2\left(-\sum_{n=0}^{\infty} \frac{(2n+1)^2\pi^2 A(n,t)}{4x^2} \sin\frac{(2n+1)\pi \ln x}{2} - \sum_{n=0}^{\infty} \frac{(2n+1)\pi A(n,t)}{2x^2}\right.$$

$$\sum_{n=0}^{N-1} e^{-j2\pi n(k-1)/N} = \frac{1 - e^{-j2\pi(k-1)}}{1 - e^{-j2\pi(k-1)/N}}$$

8719789
$$= \frac{e^{-j\pi(k-1)}\left(e^{j\pi(k-1)} - e^{-j\pi(k-1)}\right)}{e^{-j\pi(k-1)/N}\left(e^{j\pi(k-1)/N} - e^{-j\pi(k-1)/N}\right)}$$
$$= e^{-j\pi(k-1)(N-1)/N} \frac{\sin(\pi(k-1))}{\sin(\pi(k-1)/N)}$$

6246174 $\{\emptyset\}, \{\emptyset, \{1\}\}, \{\emptyset, \{2\}\}, \{\emptyset, \{3\}\}, \{\emptyset, \{1\}, \{2\}\}, \{\emptyset, \{1\}, \{3\}\}, \{\emptyset, \{2\}, \{3\}\}, \{\emptyset, \{1\}, \{2\}, \{1,2\}\}, \{\emptyset, \{1\}, \{3\}, \{1,3\}\}, \{\emptyset, \{2\}, \{3\}, \{2,3\}\}, \{\emptyset, \{1\}, \{2\}, \{3\}$

6729543 $\sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(u)e^{-2\pi imu}du = \sum_{m\in\mathbb{Z}} \lim_{N\to\infty}\left[\sum_{n=0}^{N} \int_n^{n+1} f(u)e^{-2\pi imu}du + \sum_{n=-1}^{-N} \int_n^{n+1} f(u)e^{-2\pi imu}du\right] = \sum_{m\in\mathbb{Z}} \lim_{N\to\infty}\left[\int_0^{N+1} f(u)e^{-2\pi imu}du +\right.$

5097647 $\log 2 + \log\left(1 + \frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{12}x^3 + \ldots\right) = \log 2 + \left(\frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{12}x^3 + \ldots\right) - \frac{1}{2}\left(\frac{1}{2}x + \frac{1}{4}x^2 + \ldots\right)^2 + \frac{1}{3}\left(\frac{1}{2}x + \ldots\right)^3 + \ldots = \log 2 + \frac{1}{2}x + \left(\frac{1}{4} - \frac{1}{8}\right)x^$

$$E\left(\frac{X_{(k)}}{\sum_{i=1}^n X_i}\right) - E\left(\frac{X_{(k-1)}}{\sum_{i=1}^n X_i}\right) = \frac{1}{n(n-k+1)}$$
$$E\left(\frac{X_{(k-1)}}{\sum_{i=1}^n X_i}\right) - E\left(\frac{X_{(k-2)}}{\sum_{i=1}^n X_i}\right) = \frac{1}{n(n-(k-1)+1)}$$

3039149 $\vdots$

$$E\left(\frac{X_{(2)}}{\sum_{i=1}^n X_i}\right) - E\left(\frac{X_{(1)}}{\sum_{i=1}^n X_i}\right) = \frac{1}{n(n-1)}$$
$$E\left(\frac{X_{(1)}}{\sum_{i=1}^n X_i}\right) = \frac{1}{n^2}$$

7959701 $\sum_{n=1}^{\infty} \frac{1}{n^2}\left(\frac{1}{2n+1} + \frac{1}{2n+2} + \frac{1}{2n+3} + \frac{1}{2n+4} + \frac{1}{2n+5}\right) = \sum_{n=1}^{\infty} \frac{137}{60n^2} - \frac{5269}{1800\,n} + \frac{1}{2(n+1)} + \frac{1}{8(n+2)} + \frac{4}{2n+1} + \frac{4}{9(2n+3)} + \frac{4}{25(2n+5)} = \frac{1036}{225}\ln(2) + \frac{137}{60}\zeta(2) - \frac{298373}{54000}$

6133444 $y_1(x) = e^x\begin{pmatrix}1\\1\\1\end{pmatrix} y_2(x) = e^{\frac{1}{2}x}\begin{pmatrix}\frac{1}{2}(-\cos(\frac{\sqrt{3}x}{2}) + \sqrt{3}\sin(\frac{\sqrt{3}x}{2}))\\ \frac{1}{2}(\cos(\frac{\sqrt{3}x}{2}) + \sqrt{3}\sin(\frac{\sqrt{3}x}{2}))\\ 1\end{pmatrix} y_3(x) = e^{\frac{1}{2}x}\begin{pmatrix}\frac{1}{2}(-\cos(\frac{\sqrt{3}x}{2}) + \sqrt{3}\sin(\frac{\sqrt{3}x}{2}))\\ \frac{1}{2}(\cos(\frac{\sqrt{3}x}{2}) + \sqrt{3}\sin(\frac{\sqrt{3}x}{2}))\\ 1\end{pmatrix}$

863247 $(1+\cos x)(2a'y_1' + a''y_1) + \sin x(a'y_1) = 0, a''y_1(1+\cos x) + a'\{2y_1'(1+\cos x) + y_1\sin x\} = 0, \frac{a''}{a'} + \left(\frac{2y_1'}{y_1} + \frac{\sin x}{1+\cos x}\right) = 0, \ln|a'| + \ln y_1^2 - \ln|1+\cos x|$

3100678
$$\begin{cases} \text{If} \quad x < -1: & -(x-1) - (x+1) = x - 3 \quad \text{sol: } x = 1 \quad \text{not compatible with } x < -1 \\ \text{If} \quad -1 \le x \le 1: & -(x-1) + (x+1) = x - 3 \quad \text{sol: } x = 5 \quad \text{not compatible with } -1 \le x \le 1 \\ \text{If} \quad x > 1: & (x-1) + (x+1) = x - 3: \quad \text{sol: } x = -3 \quad \text{not compatible with } x > 1 \end{cases}$$

8997593 $\frac{3-\tan^2\frac{\pi}{7}}{1-\tan^2\frac{\pi}{7}} = 1 + \frac{2}{1-\tan^2\frac{\pi}{7}} = 1 + \frac{2\tan\frac{\pi}{7}}{1-\tan^2\frac{\pi}{7}}\frac{1}{\tan\frac{\pi}{7}} = 1 + \frac{\tan\frac{2\pi}{7}}{\tan\frac{\pi}{7}} = 1 + \frac{\sin\frac{2\pi}{7}\cos\frac{\pi}{7}}{\cos\frac{2\pi}{7}\sin\frac{\pi}{7}} = \frac{\sin\frac{3\pi}{7}}{\cos\frac{2\pi}{7}\sin\frac{\pi}{7}} = \frac{\sin\frac{4\pi}{7}}{\cos\frac{2\pi}{7}\sin\frac{\pi}{7}} = 4\cos\frac{\pi}{7}.$

945139 $f(x) = \int_0^x \frac{\sin t dt}{t} = \int_0^x \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}t^{2n} = \sum_{n=0}^{\infty} \int_o^x \frac{(-1)^n}{(2n+1)!}t^{2n}dt = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)(2n+1)!}t^{2n+1}\Big|_0^x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)(2n+1)!}x^{2n+1} = x - \frac{x^3}{3*3!} + \frac{x^5}{5*5!} - \frac{x^7}{7*7!} + \cdots$

4742648 $\sum_{i=0}^{20}(|A_i| \cdot |X_i|) = \sum_{i=0}^{20}\binom{i+3-1}{i}\binom{10-i+3-1}{10-1}$ and from (2) $\sum_{i=0}^{20}\binom{i+3-1}{i}\binom{10-i+3-1}{10-1} = \sum_{i=0}^{10}\binom{2i+3-1}{2i}\binom{10-i+3-1}{10-i} =$

3105896 $T_1(n) = \sum_{i=0}^{\log_2(n)-1} 3^i \frac{n}{2^i} = n\sum_{i=0}^{\log_2(n)-1}\left(\frac{3}{2}\right)^i = n\frac{\left(\frac{3}{2}\right)^{\log_2(n)}-1}{\frac{3}{2}-1} = \frac{n((\frac{3}{2})^{\log_{3/2}(n)})^{\frac{1}{\log_{3/2}(2)}}-n}{0.5} = \frac{nn^{\frac{1}{\log_{3/2}2}}-n}{0.5} = \frac{n^{1.5849}-n}{0.5} = 2n^{1.5849} - 2n = \Theta(n^{1.5849})$

4022105 $\frac{\arctan(\log(1+\sqrt{x}))\cdot\sin^3(x^{\frac{3}{4}})}{(e^{\tan x}-1)\cdot(1-\sin^2 x)} = \frac{\arctan(\log(1+\sqrt{x}))}{\log(1+\sqrt{x})} \cdot \frac{\log(1+\sqrt{x})}{\sqrt{x}} \cdot \frac{\sin^3(x^{\frac{3}{4}})}{x^{\frac{9}{4}}} \cdot \frac{\tan x}{e^{\tan x}-1} \cdot \frac{x}{\tan x} \cdot \frac{\sqrt{x}\cdot x^{\frac{9}{4}}}{x(1-\sin^2 x)} \to 1\cdot 1\cdot 1\cdot 1\cdot 1\cdot 0 = 0$

$$f(j, \alpha(\lambda^{(1)}) + (1-\alpha)(\lambda^{(2)})) = \langle y^{(j)}, \alpha\sum_{i=1}^{m}\lambda_i^{(1)}x^{(i)} + (1-\alpha)\sum_{i=1}^{m}\lambda_i^{(2)}x^{(i)} - (1-\alpha)z - \alpha z\rangle$$

8336400
$$= (1-\alpha)\langle y^{(j)}, \sum_{i=1}^{m}\lambda_i^{(1)}x^{(i)} - z\rangle + \alpha\langle y^{(j)}, \sum_{i=1}^{m}\lambda_i^{(2)}x^{(i)} - z\rangle$$
$$= \alpha f(j, \lambda^{(1)}) + (1-\alpha)f(j, \lambda^{(2)}),$$

3248789 $\int_0^1 x^m(1-x)^n dx = \int_0^1 x^m\left[\sum_{k=0}^n \binom{n}{k}(-1)^k x^k\right]dx = \int_0^1 \sum_{k=0}^n \binom{n}{k}(-1)^k x^{k+m}dx = \sum_{k=0}^n \int_0^1 \binom{n}{k}(-1)^k x^{k+m}dx$ assuming I can interchange the int

7562248 $y_{n+1} - y_n = hf(t_{n+1}, y_{n+1})y_{n+2} - \frac{4}{3}y_{n+1} + \frac{1}{3}y_n = \frac{2}{3}hf(t_{n+2}, y_{n+2})y_{n+3} - \frac{18}{11}y_{n+2} + \frac{9}{11}y_{n+1} - \frac{2}{11}y_n = \frac{6}{11}hf(t_{n+3}, y_{n+3})y_{n+4} - \frac{48}{25}y_{n+3} + \frac{36}{25}y_{n+2} - \frac{16}{25}y_$

$$\frac{2}{5-\sqrt{2}+\sqrt{3}}\left(\frac{5+\sqrt{2}-\sqrt{3}}{5+\sqrt{2}-\sqrt{3}}\right) = \frac{10+2\sqrt{2}-2\sqrt{3}}{20+2\sqrt{6}}$$
$$= \frac{10+2\sqrt{2}-2\sqrt{3}}{20+2\sqrt{6}}\left(\frac{20-2\sqrt{6}}{20-2\sqrt{6}}\right)$$

705817
$$= \frac{200-20\sqrt{6}+40\sqrt{2}-4\sqrt{12}-40\sqrt{3}+4\sqrt{18}}{400-24}$$
$$= \frac{200-20\sqrt{6}+52\sqrt{2}-48\sqrt{3}}{376}$$
$$= \frac{50-5\sqrt{6}+13\sqrt{2}-12\sqrt{3}}{94}$$
$$\approx .37609$$

3178136 $f(x) = (1-x) + (1-2x) + \cdots + (1-nx) + ((n+1)x - 1) + \cdots + (118x-1) + (119x-1) = \sum_{k=1}^{n}(1-kx) + \sum_{k=n+1}^{119}(kx-1) = 2\sum_{k=1}^{n}(1-kx) +$

6665375 $438976\left(\left(\sqrt[3]{77} - \sqrt[3]{75}\right)^{-1} - \sqrt[3]{5775}\right)^{-3} + 17328\frac{1}{\left(\left(\sqrt[3]{77}-\sqrt[3]{75}\right)^{-1} - \sqrt[3]{5775}\right)^2\left(76\left(\sqrt[3]{77}+\sqrt[3]{75}\right)^{-1} + \sqrt[3]{5775}\right)} + 228\frac{1}{\left(\left(\sqrt[3]{77}-\sqrt[3]{75}\right)^{-1} - \sqrt[3]{5775}\right)\left(76\left(\sqrt[3]{77}+\sqrt[3]{75}\right)^{-1} + \sqrt[3]{5775}\right)^2} + \left(76\left(\sqrt[3]{\right.\right.$

2135286

$$\exists x\ (D(x) \to Q)$$
$$\Updownarrow \text{(implication equivalence)}$$
$$\exists x\ (\neg D(x) \vee Q)$$
$$\Updownarrow \text{(distribuition } \exists \text{ over } \vee)$$
$$\exists x\ \neg D(x)\ \vee\ \exists x\ Q$$
$$\Updownarrow (x \text{ is not free in } Q)$$
$$\exists x\ \neg D(x)\ \vee\ Q$$
$$\Updownarrow \text{(dual negation)}$$
$$\neg \forall x\ D(x)\ \vee\ Q$$
$$\Updownarrow \text{(implication equivalence)}$$
$$\big(\forall x\ D(x)\big)\ \to\ Q$$

6439142
$$\begin{array}{l|l|l} 56\times 65 & 5^2+6^2=25+36\to 6:1 & \overline{(30+6)(30+10)}=3640 \\ 43\times 34 & 3^2+4^2=9+16\to 2:5 & \overline{(12+2)(12+50)}=1462 \\ 73\times 37 & 3^2+7^2=9+49\to 5:8 & \overline{(21+5)(21+80)}=2701 \quad \textit{carry} \\ 88\times 88 & 8^2+8^2=64+64\to 12:8 & \overline{(64+12)(64+80)}=7744 \quad \textit{carry} \end{array}$$

8619638 $f(a)=\int_0^\infty e^{-at}\frac{\sin t}{t}dtf'(a)=\int_0^\infty -e^{-at}\sin tdtf'(a)=\frac{e^{-at}(\cos t-a\sin t)}{1+a^2}|_0^\infty f'(a)=-\frac{1}{1+a^2}f(\infty)-f(0)=-\int_0^\infty \frac{1}{1+a^2}daf(\infty)-f(0)=-\arctan\infty+\arctan$

8182998 $\alpha=2\tan^{-1}(2\sqrt{2}-1)=2\tan^{-1}(1.82)>2\tan^{-1}\sqrt{3}=2.\frac{\pi}{3}\implies \boxed{\alpha>\frac{2\pi}{3}=0.67\pi}\beta=3\sin^{-1}\frac{1}{3}+\sin^{-1}\frac{3}{5}=3\sin^{-1}(0.33)+\sin^{-1}(0.6)=\sin^{-1}(0.8)$

4151565 $\therefore \int_{\sin^{-1}y_1}^{\sin^{-1}y_2}\sum_{n=0}^\infty\frac{\alpha^{2n}\sin^{2n+1}x}{(2n)!}dx-\int_{\sin^{-1}y_1}^{\sin^{-1}y_2}\sum_{n=0}^\infty\frac{\alpha^{2n+1}\sin^{2n+2}x}{(2n+1)!}dx-\int_{\sin^{-1}y_1}^{\sin^{-1}y_2}\sum_{n=0}^\infty\frac{\alpha^{2n}\sin^{2n+3}x}{(2n)!}dx+\int_{\sin^{-1}y_1}^{\sin^{-1}y_2}\sum_{n=0}^\infty\frac{\alpha^{2n+1}\sin^{2n+4}x}{(2n+1)!}dx$

4916483 $y(x)=\frac{\sqrt[4]{\alpha}c_2\cdot{}_1F_3\left(\frac{\beta}{4\alpha}+\frac{1}{4};\frac{1}{4},\frac{3}{4},\frac{5}{4};-\frac{x^4\alpha}{64}\right)}{\sqrt{2}}+c_1\cdot{}_1F_3\left(\frac{\beta}{4\alpha};\frac{1}{4},\frac{1}{2},\frac{3}{4};-\frac{x^4\alpha}{64}\right)+\frac{\alpha^{3/4}c_4x^2\cdot{}_1F_3\left(\frac{\beta}{4\alpha}+\frac{3}{4};\frac{3}{4},\frac{3}{2},\frac{7}{4};-\frac{x^4\alpha}{64}\right)}{\sqrt{2}}+\sqrt{\alpha}c_3x^2{}_1F_3\left(\frac{\beta}{4\alpha}+\frac{1}{2};\frac{3}{4},\frac{5}{4},\frac{3}{2};-\frac{x^4\alpha}{64}\right)$

1479939 $\int_{-1}^1 f(x)dx\approx\int_{-1}^1\left(f(0)+f'(0)x+\frac{1}{2!}f''(0)x^2+\frac{1}{3!}f'''(0)x^3+\frac{1}{4!}f''''(0)x^4\right)dx=\left(f(0)x+\frac{1}{2!}f'(0)x^2+\frac{1}{3!}f''(0)x^3+\frac{1}{4!}f'''(0)x^4+\frac{1}{5!}f''''(0)x^5\right)\Big|_{-1}^1=2f($

5393938 $\sum_{r=493701}^{506199}\sum_{k=0}^{100}(-1)^k\frac{\binom{100}{k}\binom{r-10001k+99}{99}}{\binom{r+99}{r}}=\sum_{k=0}^{100}(-1)^k\binom{100}{k}\sum_{r=493701}^{506199}\frac{\binom{r-10001k+99}{99}}{\binom{r+99}{r}}=\sum_{k=0}^{100}(-1)^k\binom{100}{k}\sum_{r=493701}^{506199}\prod_{s=1}^{99}\left(1-\frac{10001k}{r+s}\right)$

4150833 $=\left[\sum_{n=0}^\infty\sum_{k=0}^n\frac{(-1)^{k+1}n!\alpha^{2n}\cos^{2k+1}x}{2^{2n+1}(2n)!k!(n-k)!(2k+1)}\right]_{\cos^{-1}(2y-1)}^{\pi}-\left[\sum_{n=0}^\infty\frac{\alpha^{2n+1}x}{2^{4n+3}n!(n+1)!}-\sum_{n=0}^\infty\sum_{k=0}^n\frac{(k!)^2\alpha^{2n+1}\sin^{2k+1}x\cos x}{2^{4n-2k+3}n!(n+1)!(2k+1)!}\right]_{\cos^{-1}(2y-1)}^{\pi}$

4063102 $(x-a)(x-12)+2=(x+b)(x+c)\Rightarrow x^2-(a+12)x+12a+2=x^2+(b+c)x+bc\Rightarrow\begin{cases}-a-12=b+c\\12a+2=bc\end{cases}\overset{12R_1+R_2}{\Rightarrow}-142=12b+12c+bc\Rightarrow(b+12$

9096632 $\lambda=\frac{\left(3^{\frac{1}{3}}2^{\frac{2}{3}}\left(9\cdot3^{\frac{1}{3}}2^{\frac{2}{3}}-\left(\frac{3}{12013156}\sqrt{3}\sqrt{-113452970512451555823}+\frac{427215759480560}{15614099511}\right)^{\frac{2}{3}}\right)\right)}{6\left(\frac{3}{12013156}\sqrt{3}\sqrt{-113452970512451555823}+\frac{427215759480560}{15614099511}\right)^{\frac{1}{3}}}-\frac{91979}{5199}$

3475519 $\lambda^{-\min\left(\frac{k_1}{\gcd(|k_1|,|k_2|)},\frac{k_2}{\gcd(|k_1|,|k_2|)},0\right)}-a\lambda^{\frac{k_1}{\gcd(|k_1|,|k_2|)}-\min\left(\frac{k_1}{\gcd(|k_1|,|k_2|)},\frac{k_2}{\gcd(|k_1|,|k_2|)},0\right)}-b\lambda^{\frac{k_2}{\gcd(|k_1|,|k_2|)}-\min\left(\frac{k_1}{\gcd(|k_1|,|k_2|)},\frac{k_2}{\gcd(|k_1|,|k_2|)},0\right)}=0$

1693574 $\therefore\int_0^{\sin^{-1}\Omega}\left(1+\sum_{n=1}^\infty\frac{\sin^{2n}x}{(2n)!}\right)dx+\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{2(n+1)\sin^{2n+1}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=1}^\infty\frac{2n\sin^{2n+2}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{2(n+1)\sin^{2n+3}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{\sin^{2n+4}}{(2n+1}$

$$f=\frac{(A+2)^n}{A\cdot(A+1)^{n-K-1}}$$
$$=\frac{\sum_{k=0}^K\binom{n}{k}(A+1)^{n-k}+\sum_{k=K+1}^n\binom{n}{k}(A+1)^{n-k}}{A\cdot(A+1)^{n-K-1}}$$
5968293
$$=\frac{\sum_{k=0}^K\binom{n}{k}(A+1)^{K+1-k}}{A}+\frac{\sum_{k=K+1}^n\binom{n}{k}(A+1)^{K+1-k}}{A}$$
$$=\sum_{k=0}^K\binom{n}{k}+\frac{\sum_{k=0}^K\binom{n}{k}}{A}+\frac{\sum_{k=K+1}^n\binom{n}{k}(A+1)^{K+1-k}}{A}=\sum_{k=0}^K\binom{n}{k}+0$$

1693565 $=\int_0^{\sin^{-1}\Omega}\left(1+\sum_{n=1}^\infty\frac{\sin^{2n}x}{(2n)!}\right)dx+\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{2(n+1)\sin^{2n+1}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=1}^\infty\frac{2n\sin^{2n+2}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{2(n+1)\sin^{2n+3}x}{(2n+1)!}dx-\int_0^{\sin^{-1}\Omega}\sum_{n=0}^\infty\frac{\sin^{2n+4}}{(2n+1}$

$$xy^2\frac{dy}{dx}-x^3-y^3=0\Rightarrow\frac{dy}{dx}-\frac{x^2}{y^3}-\frac{y}{x}=0$$
$$u=\frac{y}{x}\Rightarrow xu=y\Rightarrow udx+xdu=dy$$
$$\frac{xdu}{dx}-\frac{1}{u^2}=0\Rightarrow u^2du=\frac{dx}{x}$$
2162727 $$\frac{u^3}{3}=\ln(x)+\ln(c^{-3})\Rightarrow u^3=3\ln(c^{-3}x)=\ln(cx^3)$$
$$u^3=\ln(cx^3)$$
$$\left(\frac{y}{x}\right)^3=\ln(cx^3)$$
$$x=1\wedge y=1\Rightarrow\ln(c)=1\Rightarrow c=e$$
$$\left(\frac{y}{x}\right)^3=\ln(ex^3)$$

3833443 $J=-\sum_{n=1}^\infty\frac{(-1)^n}{n}\sum_{m=0}^\infty\binom{n}{m}(-1)^{n-m}\int_0^\infty\frac{x^{m+3}}{e^x-1}dx=-\sum_{n=1}^\infty\frac{(-1)^n}{n}\sum_{m=0}^\infty\binom{n}{m}(-1)^{n-m}\Gamma(m+4)\zeta(m+4)=-\sum_{n=1}^\infty\frac{1}{n}\sum_{m=0}^\infty\binom{n}{m}(-1)^m(m+3)!\zeta(m+$

4620828

$$\begin{array}{l}\left|\begin{array}{l}A\vee B\\\hline\left|\begin{array}{l}A\\\hline B\vee\neg B\\\left|\begin{array}{l}B\\\hline A\wedge B\end{array}\right.\\(A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B))\end{array}\right.\\B\to((A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B)))\\\quad\vdots\\\neg B\to((A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B)))\\(A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B))\end{array}\right.\\A\to((A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B)))\\\quad\vdots\\B\to((A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B)))\\(A\wedge B)\vee((\neg A\wedge B)\vee(A\wedge\neg B))\end{array}$$

3775542 $T(n)\leq\frac{2}{\sqrt{13}}\sum_{j=0}^{\lfloor\log_2 n\rfloor}\left(\rho_1^{j+1}-\rho_2^{j+1}\right)\sum_{k=j}^{\lfloor\log_2 n\rfloor}2^k=\frac{2}{\sqrt{13}}\sum_{j=0}^{\lfloor\log_2 n\rfloor}\left(\rho_1^{j+1}-\rho_2^{j+1}\right)\left(2^{\lfloor\log_2 n\rfloor+1}-2^j\right)=\frac{2^{\lfloor\log_2 n\rfloor+2}}{\sqrt{13}}\left(\frac{\rho_1^{\lfloor\log_2 n\rfloor+2}-1}{\rho_1-1}-\frac{\rho_2^{\lfloor\log_2 n\rfloor+2}-1}{\rho_2-1}\right)-\frac{1}{\sqrt{13}}\sum_{j=0}^{\lfloor\log_2 n\rfloor}\left((2\rho\right.$

5326815 $\int_0^1\frac{\ln(1+x^2)}{(1+x^2)^2}(1-x^2)dx=-2\int_0^{\frac{\pi}{4}}\cos^2(\theta)(1-\tan^2\theta)\ln\cos\theta\,d\theta=-2\int_0^{\frac{\pi}{4}}\cos(2\theta)\ln\cos\theta\,d\theta=2\ln 2\int_0^{\frac{\pi}{4}}\cos(2\theta)d\theta+\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\int_0^{\frac{\pi}{2}}\cos\theta\cos(n\theta)d\theta=\ln 2-$

4403311 $\sqrt[3]{-\frac{2306997866696}{1047656140569}+\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k})}+\sqrt[3]{-\frac{2306997866696}{1047656140569}+\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k+1})}+\sqrt[3]{-\frac{2306997866696}{1047656140569}+\sum_{k=0}^{1}\cos(\frac{2\pi}{7}\cdot 3^{3k+2})}=-\sqrt[3]{\frac{118149192000}{1908298981}}$

3985113 $\left|\bigcup_{i=1}^{k}A_i\right|=\left|\bigcup_{i=1}^{k-1}A_i\right|+|A_k|-\left|\left(\bigcup_{i=1}^{k-1}A_i\right)\bigcap A_k\right|=\left|\bigcup_{i=1}^{k-1}A_i\right|+|A_k|-\left|\bigcup_{i=1}^{k-1}(A_i\bigcap A_k)\right|\overset{(1)}{\geq}\left|\bigcup_{i=1}^{k-1}A_i\right|+|A_k|-\sum_{i=1}^{k-1}|A_i\bigcap A_k|\geq\left|\bigcup_{i=1}^{k-1}A_i\right|+2n-(k-1)\geq\left|\bigcup_{i=1}^{k-2}A_i\right|-$

8554107 $/\!/Q=\frac{L(\alpha,\hat{\sigma}^2)}{L(\hat{\alpha},\hat{\sigma}^2)}=\frac{(2\pi\hat{\sigma}^2)^{-n/2}\exp\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\alpha x_i)^2\right)}{(2\pi\hat{\sigma}^2)^{-n/2}\exp\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\hat{\alpha}x_i)^2\right)}=\left(\frac{\sum_{i=1}^{n}(y_i-\alpha x_i)^2}{\sum_{i=1}^{n}(y_i-\hat{\alpha}x_i)^2}\right)^{-n/2}\frac{e^{-n/2}}{e^{-n/2}}=\left(1+\frac{(\hat{\alpha}-\alpha)^2\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(y_i-\hat{\alpha}x_i)^2}\right)^{-n/2}/\!/$

6064602 $\lim_{x\to 0}\frac{\sqrt{\cos x}-\sqrt[3]{\cos x}}{\sin^2 x}=\lim_{x\to 0}\frac{\sqrt{\cos x-1+1}-\sqrt[3]{\cos x}}{\sin^2 x}=\lim_{x\to 0}\frac{\sqrt{1+\cos x-1}-1-(\sqrt[3]{1+\cos x-1}-1)}{\cos x-1}\cdot\frac{\cos x-1}{\sin^2 x}=\lim_{x\to 0}[\frac{(1+\cos x-1)^{\frac{1}{2}}-1}{\cos x-1}-\frac{(1+\cos x-1)^{\frac{1}{3}}-1}{\cos x-1}]\cdot\frac{-1}{2}=(\frac{1}{2}-\frac{1}{3})\cdot-$

2275092
$$\begin{array}{ll}k=0: & 3^0\equiv 1\pmod{11} \quad x^2+4\equiv 5\pmod{11}\\ k=1: & 3^1\equiv 3\pmod{11} \quad x^2+4\equiv 13\equiv 2\pmod{11}\\ k=2: & 3^2\equiv 9\pmod{11} \quad x^2+4\equiv 85\equiv 8\pmod{11}\\ k=3: & 3^3\equiv 5\pmod{11} \quad x^2+4\equiv 29\equiv 7\pmod{11}\\ k=4: & 3^3\equiv 4\pmod{11} \quad x^2+4\equiv 20\equiv 9\pmod{11}\\ k=5: & 3^5\equiv 1\pmod{11} \quad \text{and it cycles there...}\end{array}$$

1909260 $6+12+18+24=\sum_{i=1}^{4}6ix_1-x_2+x_3-x_4+\ldots+x_7-x_8=\sum_{i=1}^{8}(-1)^{i-1}x_i 2*3*4*9*8*27=\prod_{i=1}^{3}2^i 3^i 1+2x+3x^2+4x^3+\ldots+10x^9=\sum_{i=1}^{10}i$

3680362 $\sum_{n=1}^{\infty}\frac{\cos(n)}{n}=\sum_{n=1}^{\infty}\cos(n)\int_0^{\infty}\exp(-xn)\,dx=\int_0^{\infty}\left(\sum_{n=1}^{\infty}\cos(n)\exp(-xn)\right)dx=\int_0^{\infty}\Re\left(\sum_{n=1}^{\infty}\exp(in)\exp(-xn)\right)dx=\Re\left(\int_0^{\infty}\left(\sum_{n=1}^{\infty}e^{-n(-i+x)}\right)dx\right)=$

759486 $(e^x)'=e^x\iff\ln' x=\frac{1}{e^{\ln x}}=\frac{1}{x}\sin' x=+\cos x=+\sqrt{1-\sin^2 x}\iff\arcsin' x=+\frac{1}{\sqrt{1-\sin^2(\arcsin x)}}=+\frac{1}{\sqrt{1-x^2}}\cos' x=-\sin x=-\sqrt{1-c}$

$$\begin{aligned}0&=\left(\frac{\partial L[x,y+hv,y'+hv']}{\partial y}-\frac{d}{dx}\frac{\partial L[x,y+hv,y'+hv']}{\partial y'}\right)-\left(\frac{\partial L[x,y,y']}{\partial y}-\frac{d}{dx}\frac{\partial L[x,y,y']}{\partial y'}\right)\\ &=h\left(v\frac{\partial^2 L}{\partial y^2}+v'\frac{\partial^2 L}{\partial y\partial y'}-\frac{d}{dx}\left(v\frac{\partial^2 L}{\partial y\partial y'}+v'\frac{\partial^2 L}{\partial y'^2}\right)\right)+o(h)\\ &=h\left(v\left(\frac{\partial^2 L}{\partial y^2}-\frac{d}{dx}\frac{\partial^2 L}{\partial y\partial y'}\right)-\frac{d}{dx}\left(v'\frac{\partial^2 L}{\partial y'^2}\right)\right)+o(h)\end{aligned}$$

1626431 $\Phi_n(-x)=\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}(-x-e^{2\pi ik/n})=(-1)^{\varphi(n)}\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}(x+e^{2\pi ik/n})=(-1)^{\varphi(n)}\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}(x-e^{\pi i+2\pi ik/n})=(-1)^{\varphi(n)}\prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}}(x-e^{2\pi i(k/n+1/2)})=(-1)^{\varphi(n}$

5373740 $LHC=\frac{\sin(\theta)}{\left(\frac{\cos(\theta)}{\sin(\theta)}\right)\left(\frac{1}{\sin(\theta)}\right)-\cos(\theta)}=\frac{\sin(\theta)}{\left(\frac{\cos(\theta)}{\sin^2(\theta)}\right)-\cos(\theta)}=\frac{\sin(\theta)}{\frac{\cos(\theta)-\cos(\theta)\sin(\theta)}{\sin^2(\theta)}}=\frac{\sin^3(\theta)}{\cos(\theta)-\cos(\theta)\sin^2(\theta)}=\frac{\sin^3(\theta)}{\cos(\theta)(1-\sin^2(\theta))}=\frac{\sin^3(\theta)}{\cos(\theta)\cos^2(\theta)}=\tan^3(\theta)$

5531925 $I=\int_{23\pi}^{71\pi/2}\ln(1+\cos x)dxI=\int_0^{25\pi/2}\ln(1+\cos x)dxI=12\int_0^{\pi}\ln(1+\cos x)dx+\int_0^{\pi/2}\ln(1+\cos x)dxI=12\int_0^{\pi}\ln(1-\cos x)dx+\int_0^{\pi/2}\ln(1+\sin x)dx2I$

7501924 $(x\in A\wedge x\notin B\wedge x\in C)\vee(x\in C\wedge x\in A\wedge x\notin B)\vee(x\in A\wedge x\notin B\wedge x\notin C)\vee(x\in A\wedge x\in B\wedge x\notin C)\vee(x\in B\wedge x\notin C\wedge x\in A)\vee(x\in B\wedge x\notin$

4520916 $\sinh^{-1}(x):=\ln(x+\sqrt{x^2+1})\implies\sinh^{-1}(-x)=\ln(-x+\sqrt{x^2+1})=\ln(-(x-\sqrt{x^2+1}))=\ln\left(\frac{x-\sqrt{1+x^2}}{x^2-1-x^2}\right)=\ln\left(\frac{1}{x-\sqrt{1+x^2}}\frac{x+\sqrt{1+x^2}}{x+\sqrt{1+x^2}}\right)$

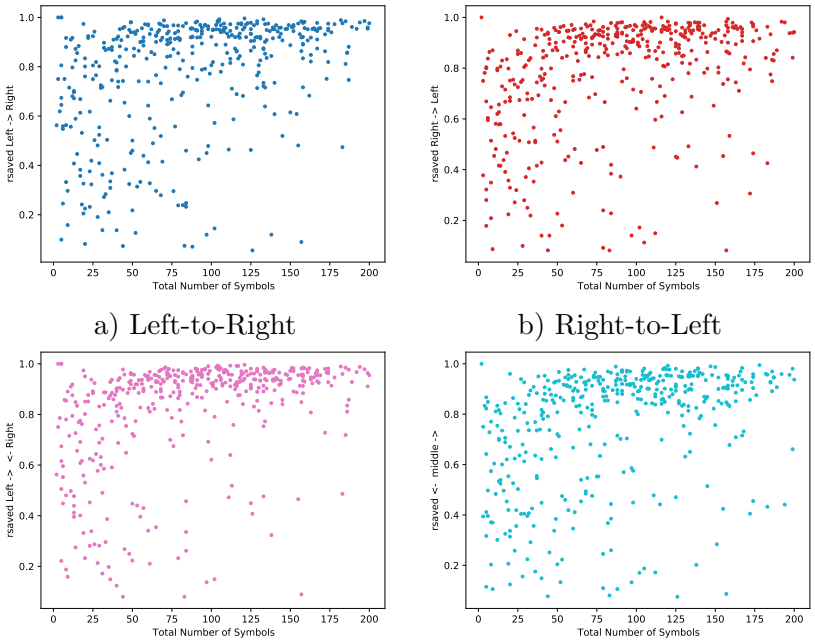
4520916 $\sinh^{-1}(x):=\ln(x+\sqrt{x^2+1})\implies\sinh^{-1}(-x)=\ln(-x+\sqrt{x^2+1})=\ln(-(x-\sqrt{x^2+1}))=\ln\left(\frac{x-\sqrt{1+x^2}}{x^2-1-x^2}\right)=\ln\left(\frac{1}{x-\sqrt{1+x^2}}\frac{x+\sqrt{1+x^2}}{x+\sqrt{1+x^2}}\right)$

4520916 $\sinh^{-1}(x):=\ln(x+\sqrt{x^2+1})\implies\sinh^{-1}(-x)=\ln(-x+\sqrt{x^2+1})=\ln(-(x-\sqrt{x^2+1}))=\ln\left(\frac{x-\sqrt{1+x^2}}{x^2-1-x^2}\right)=\ln\left(\frac{1}{x-\sqrt{1+x^2}}\frac{x+\sqrt{1+x^2}}{x+\sqrt{1+x^2}}\right)$

Appendix B

Rsaved Graphs

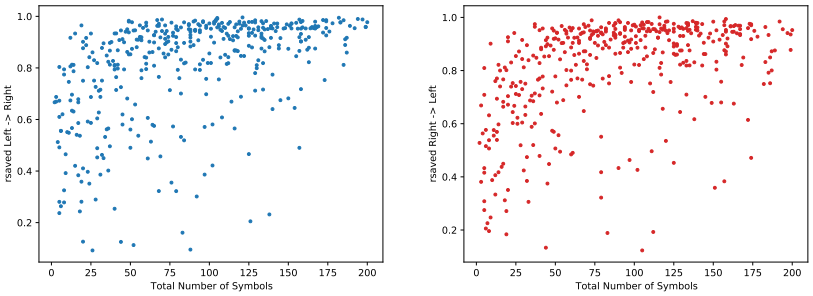
The following graphs are the `rsaved` score by the target formula size using all 4 input orders for all the experimental runs.



a) Left-to-Right b) Right-to-Left

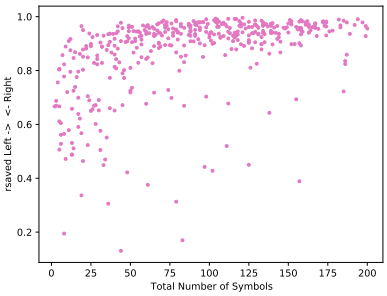
c) Outside-in d) Middle-out

Figure B.1: Level 1 (BoS)

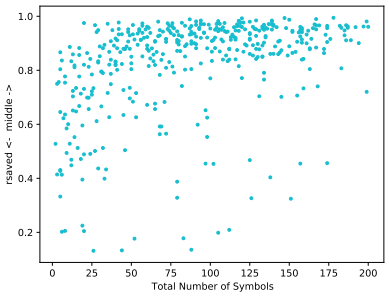


a) Left-to-Right

Right-to-Left

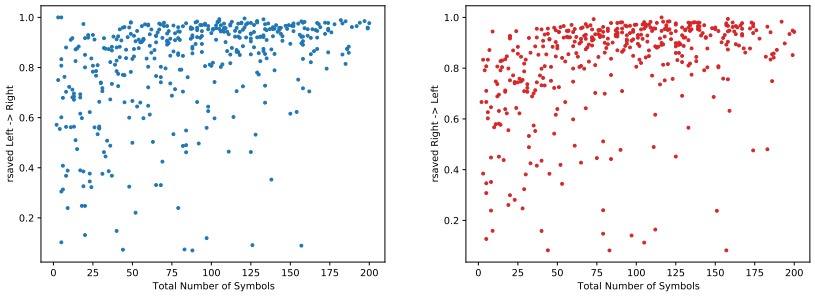


c) Outside-in



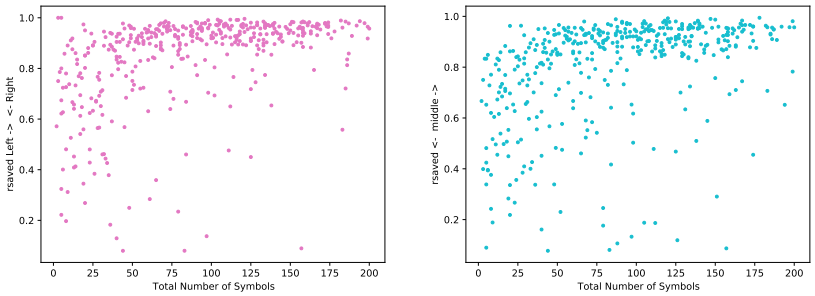
d) Middle-out

Figure B.2: Level 2



a) Left-to-Right

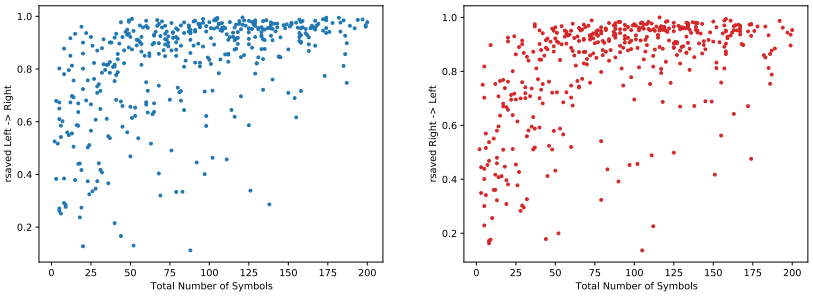
Right-to-Left



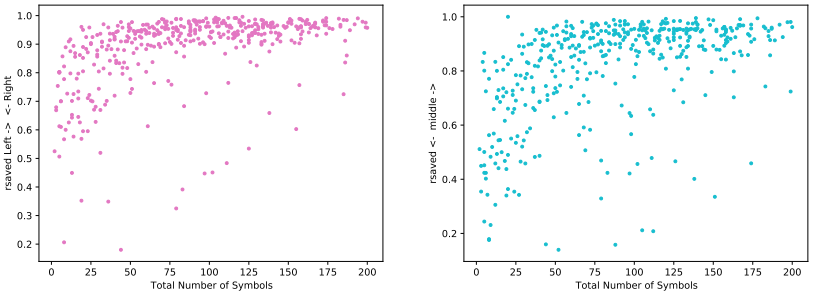
c) Outside-in

d) Middle-out

Figure B.3: 2'

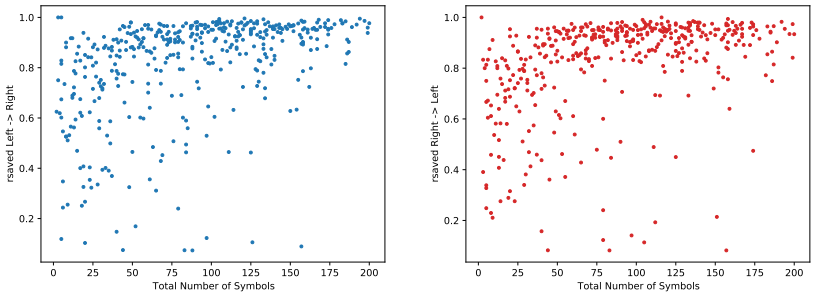


a) Left-to-Right Right-to-Left



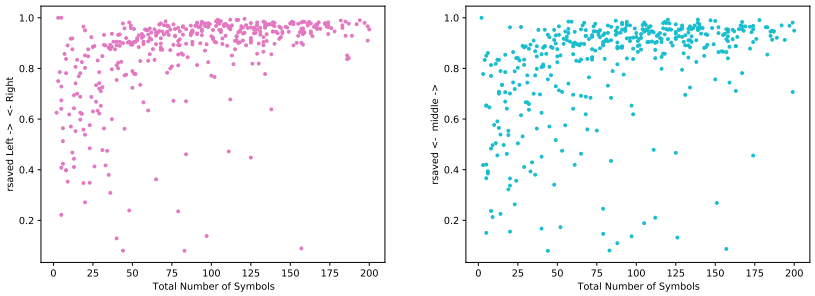
c) Outside-in d) Middle-out

Figure B.4: Level 3



a) Left-to-Right

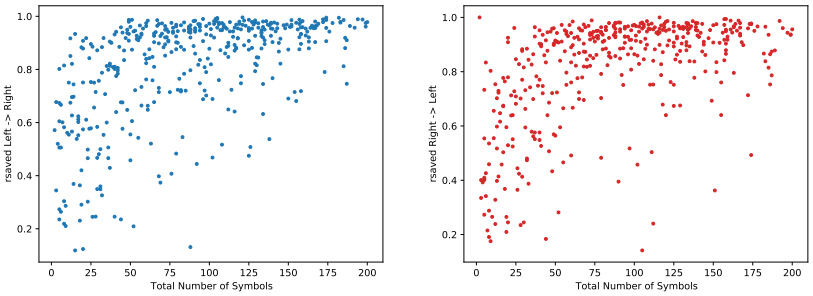
Right-to-Left



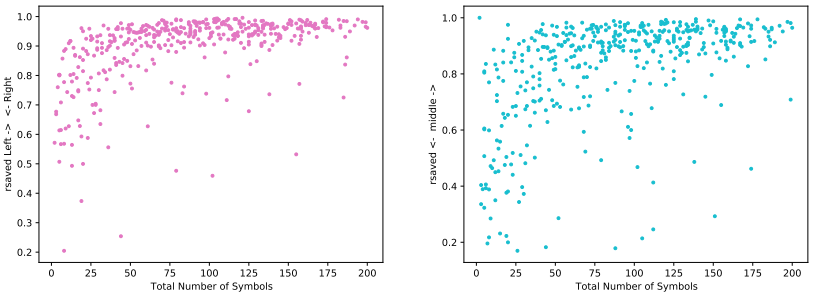
c) Outside-in

d) Middle-out

Figure B.5: Level 3'

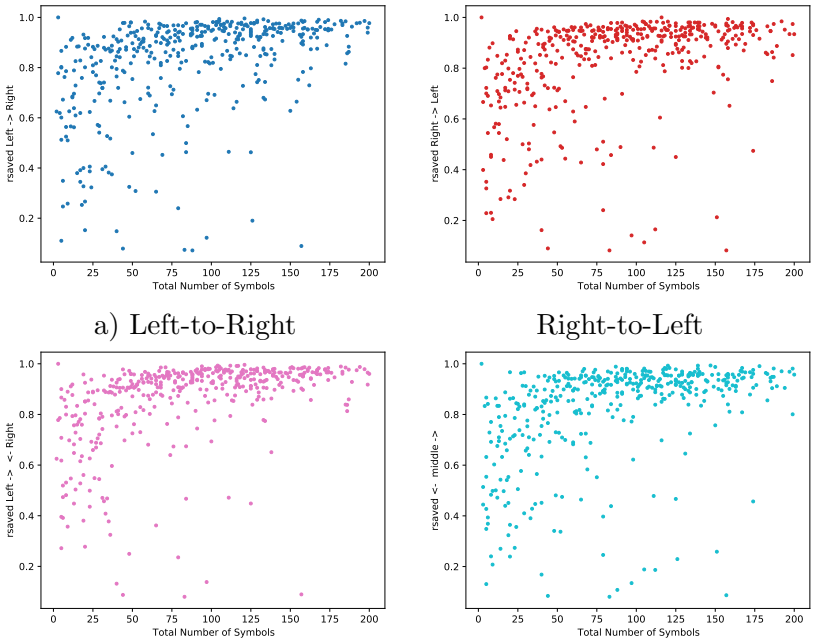


a) Left-to-Right Right-to-Left



c) Outside-in d) Middle-out

Figure B.6: Level 4



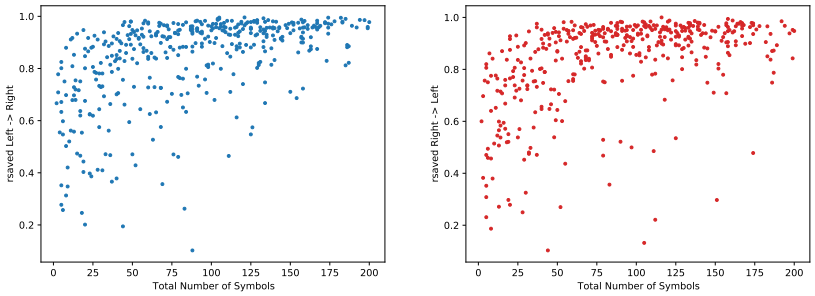
a) Left-to-Right

Right-to-Left

c) Outside-in

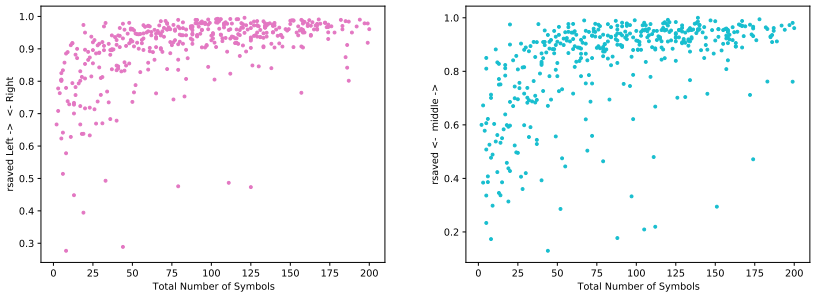
d) Middle-out

Figure B.7: Level 4'



a) Left-to-Right

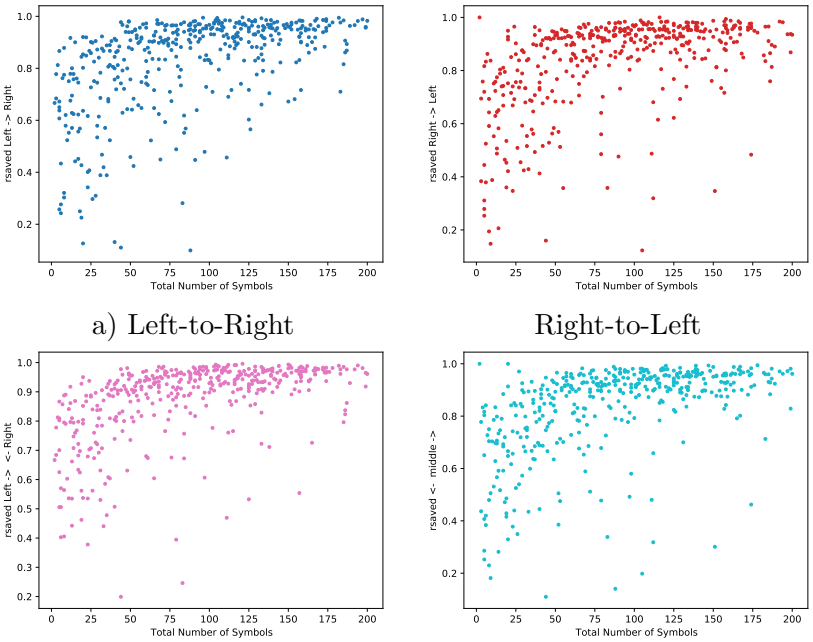
Right-to-Left



c) Outside-in

d) Middle-out

Figure B.8: Level 5



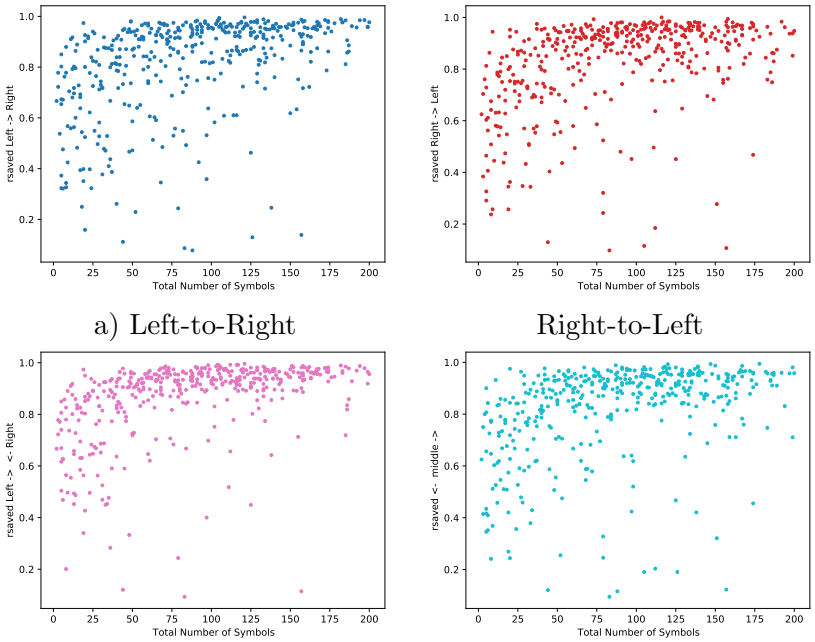
a) Left-to-Right

Right-to-Left

c) Outside-in

d) Middle-out

Figure B.9: 5'



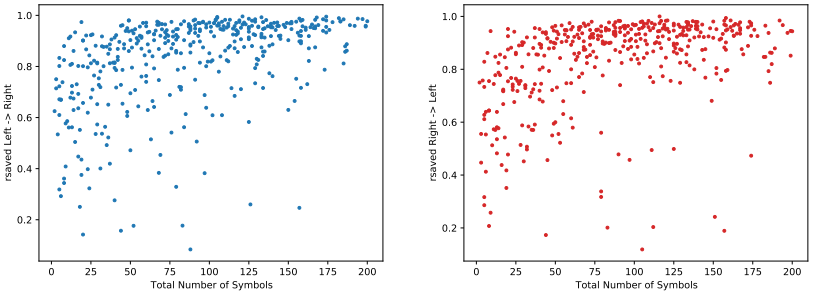
a) Left-to-Right

Right-to-Left

c) Outside-in

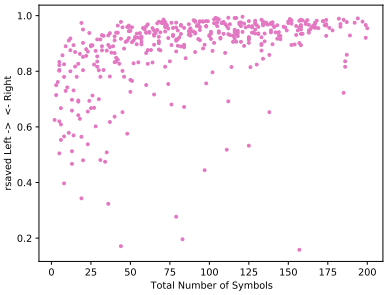
d) Middle-out

Figure B.10: Levels 1-2

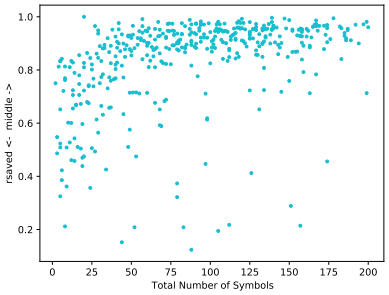


a) Left-to-Right

Right-to-Left

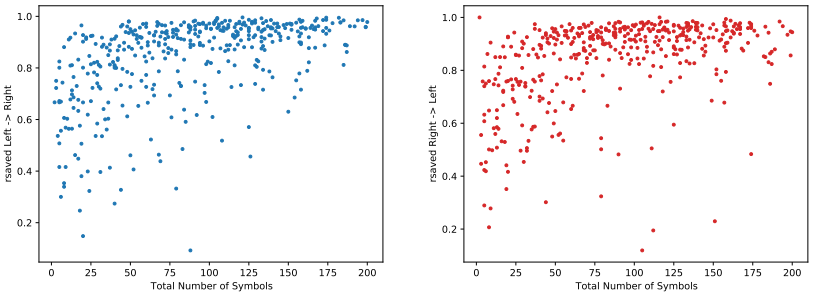


c) Outside-in

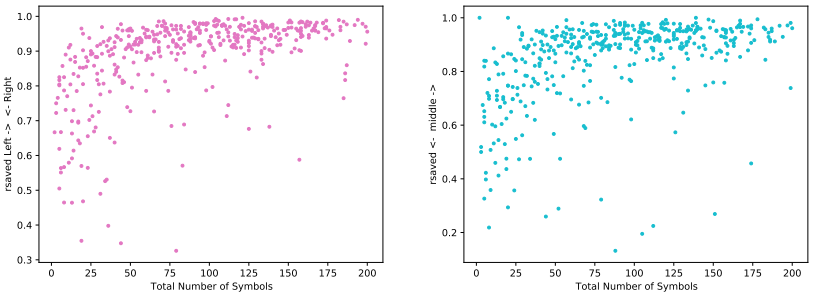


d) Middle-out

Figure B.11: Levels 1-3

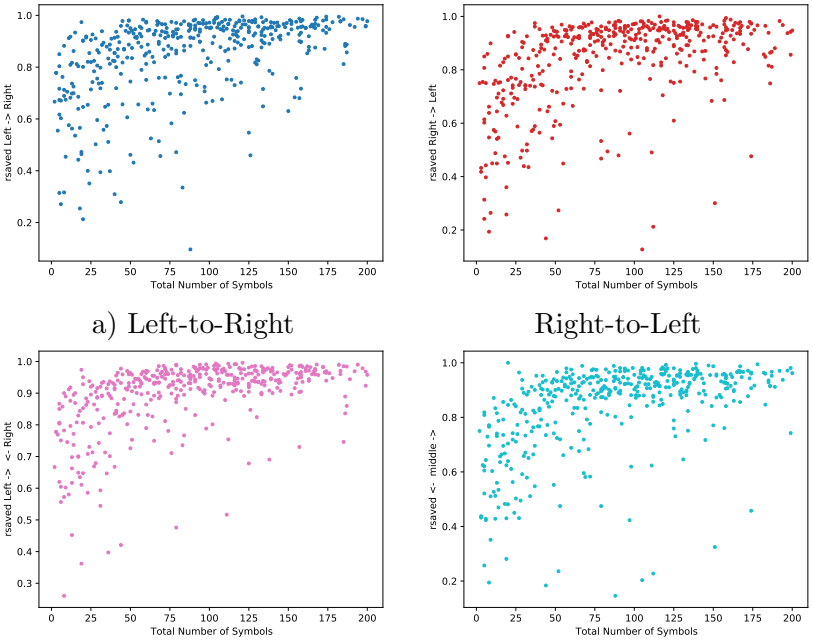


a) Left-to-Right Right-to-Left



c) Outside-in d) Middle-out

Figure B.12: Levels 1-4



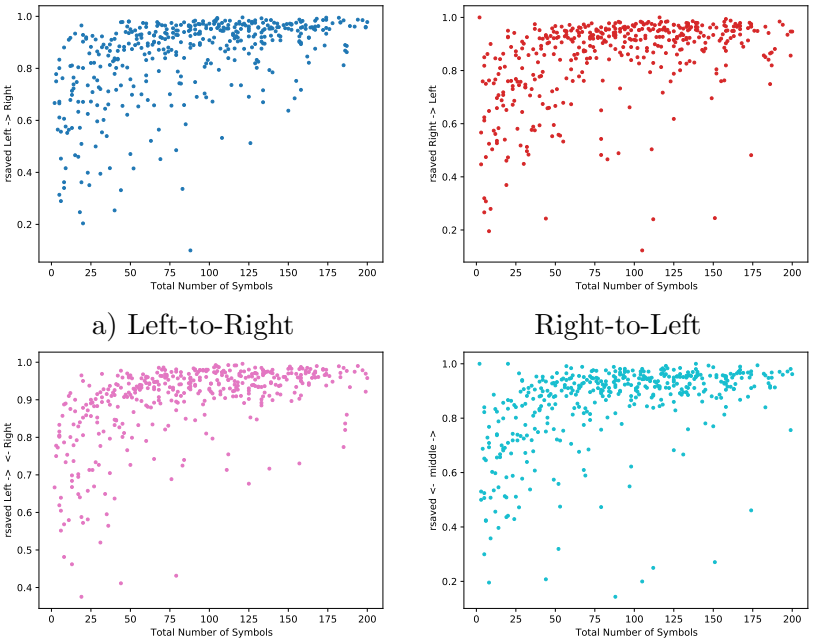
a) Left-to-Right

Right-to-Left

c) Outside-in

d) Middle-out

Figure B.13: PHOC Embedding



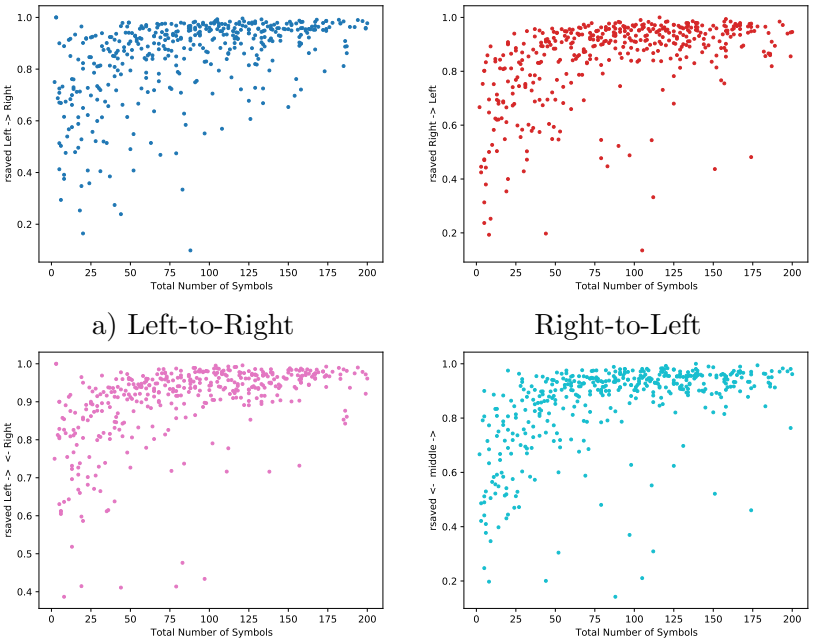
a) Left-to-Right

Right-to-Left

c) Outside-in

d) Middle-out

Figure B.14: XY-PHOC Embedding



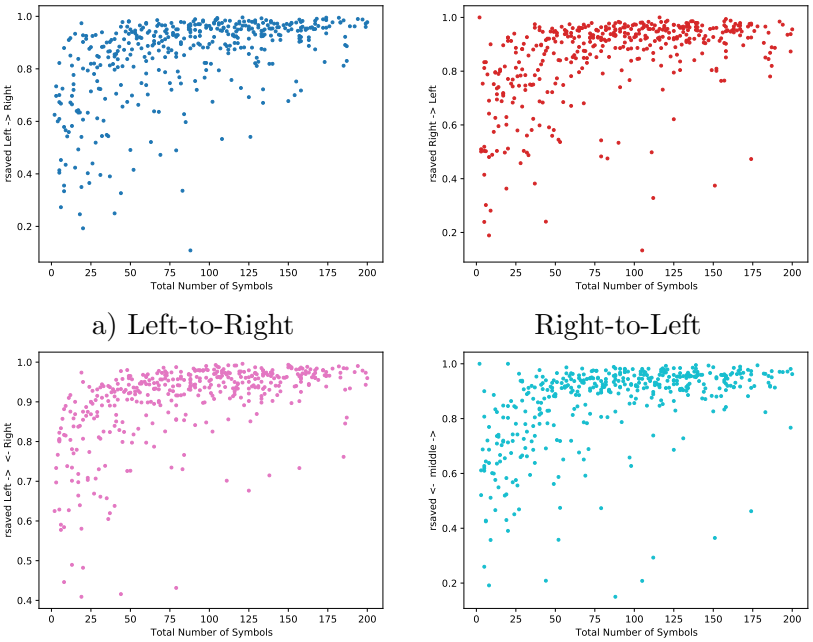
a) Left-to-Right

Right-to-Left

c) Outside-in

d) Middle-out

Figure B.15: Top-Left membership condition



a) Left-to-Right

Right-to-Left

c) Outside-in

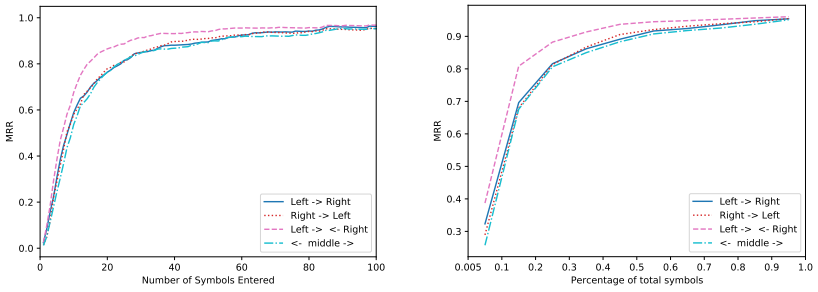
d) Middle-out

Figure B.16: Vertical Center membership condition

Appendix C

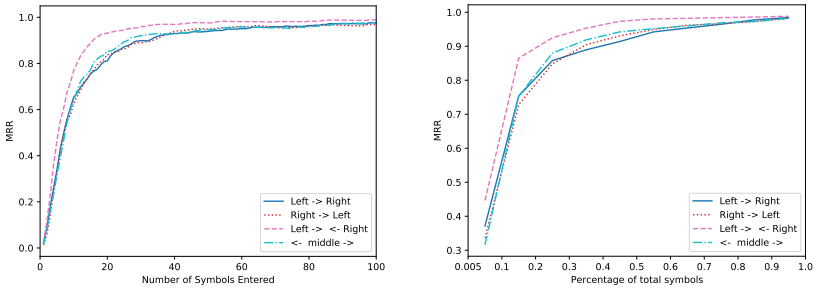
Mean Reciprocal Rank Graphs

Mean Reciprocal Rank of all 4 input orders by the number of symbols entered and the % of target formula symbols entered for all experimental conditions.



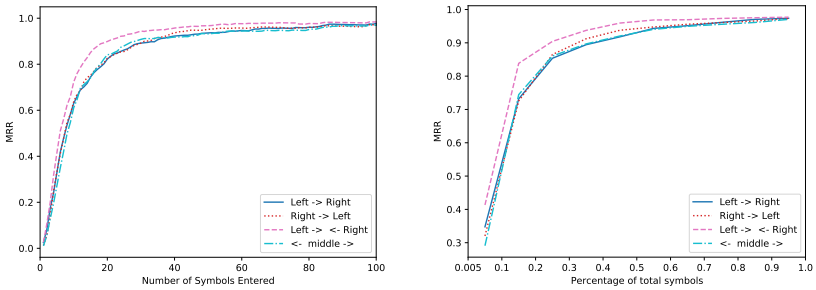
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.1: Level 1 (BoS)



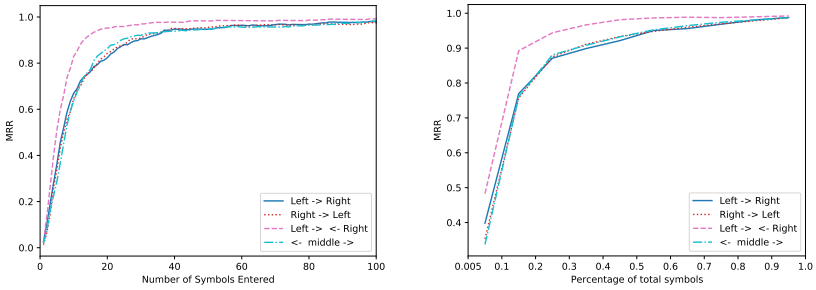
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.2: Level 2. The Quicker Convergence to 1 in the Outside-in condition shows how its scores across the board are better



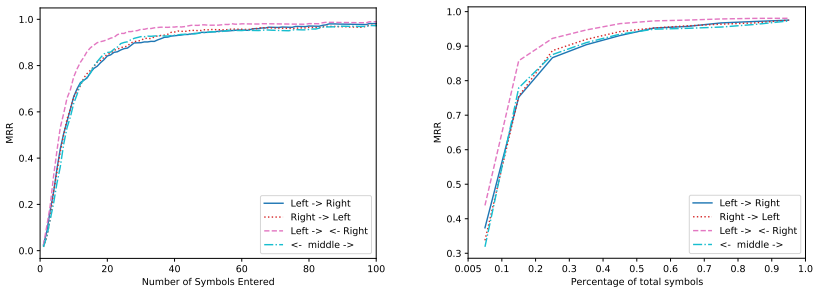
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.3: Level 2'



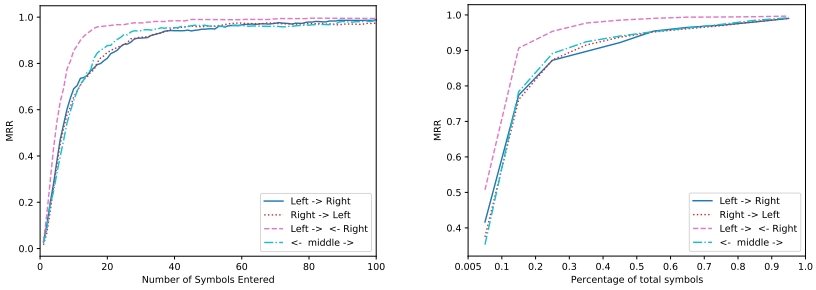
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.4: Level 3



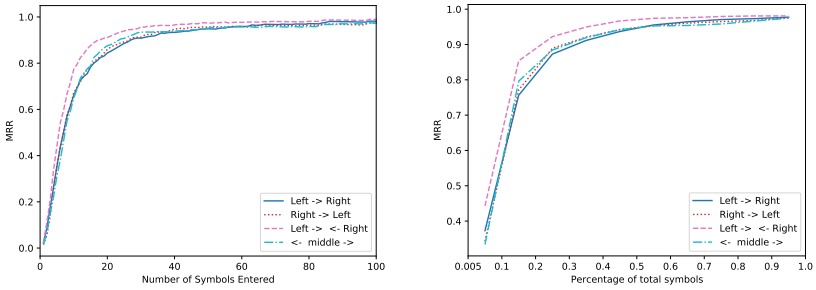
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.5: Level 3'



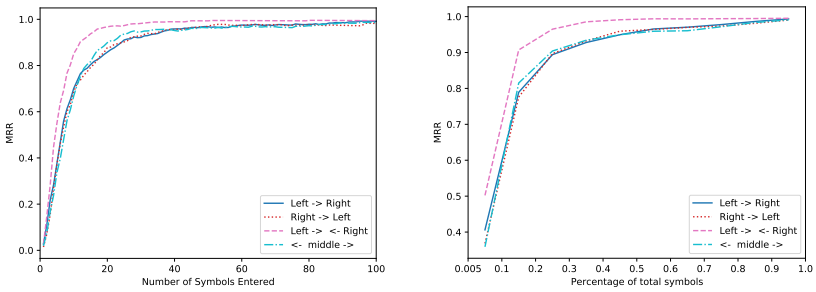
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.6: Level 4



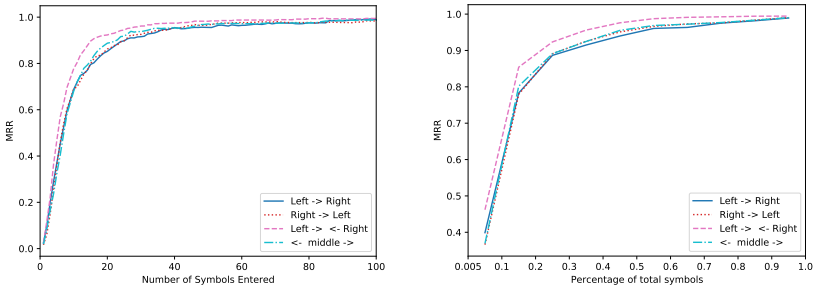
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.7: Level 4'



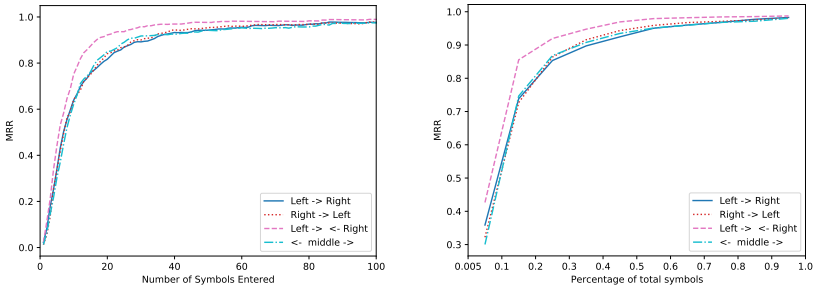
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.8: Level 5



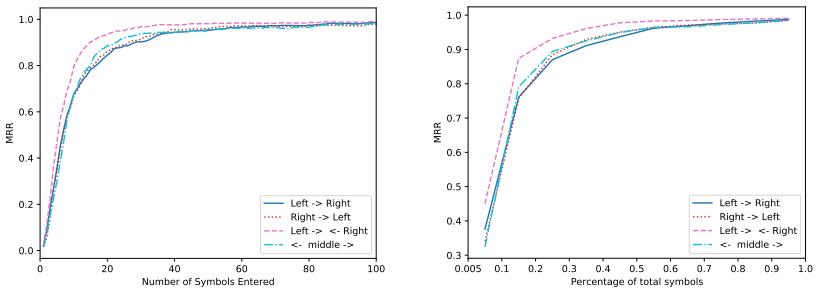
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.9: Level 5'



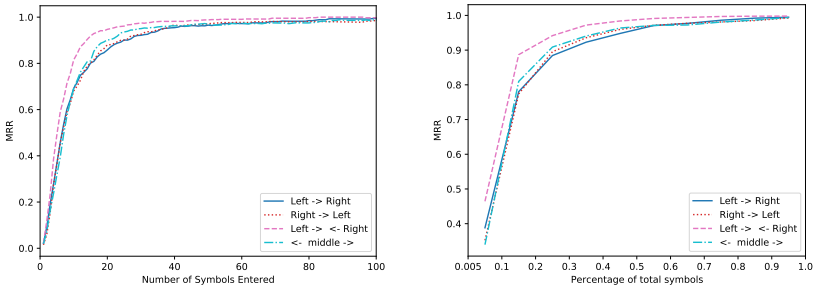
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.10: Levels 1-2



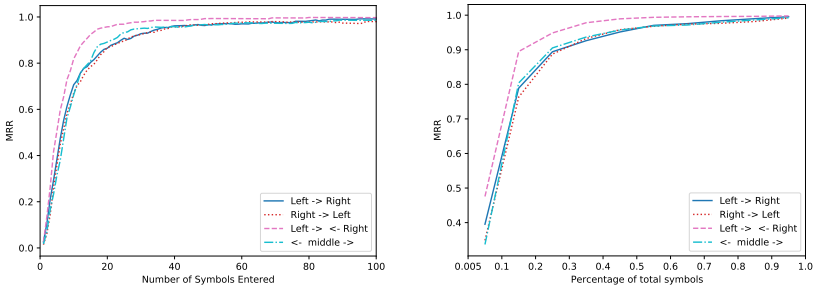
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.11: Levels 1-3



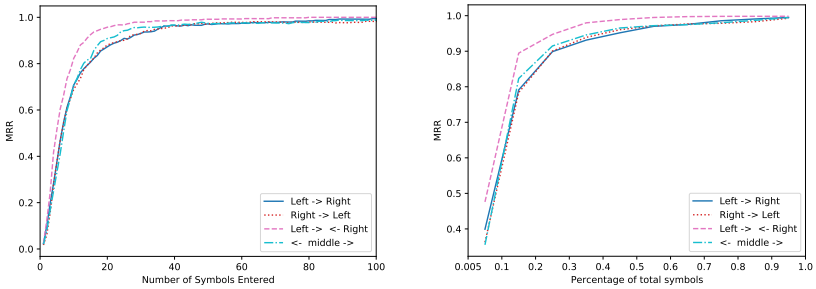
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.12: Levels 1-4



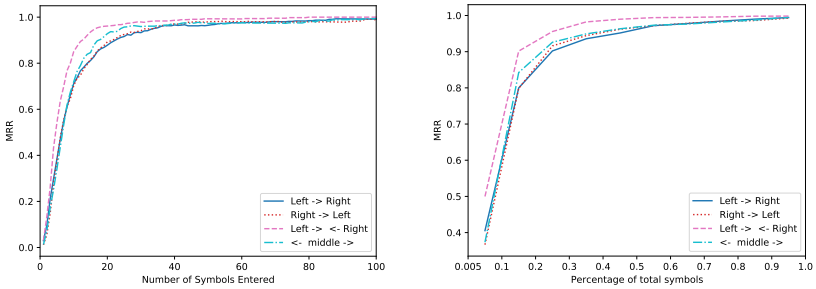
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.13: PHOC Embedding



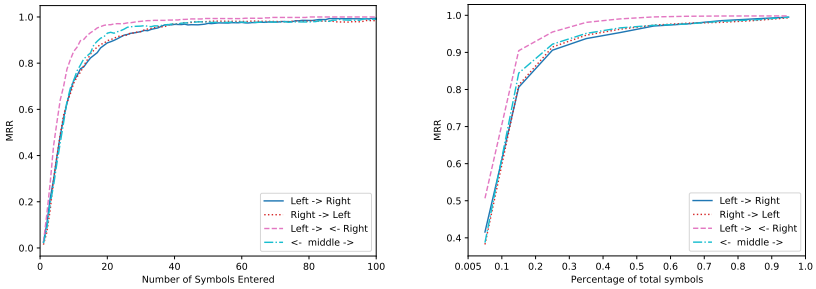
a) MRR for symbols entered b) MRR for % symbols entered

Figure C.14: XY-PHOC Embedding



a) MRR for symbols entered b) MRR for % symbols entered

Figure C.15: Top-Left membership condition



a) MRR for symbols entered b) MRR for % symbols entered

Figure C.16: Vertical Center membership condition