Redundancy of Lossless Data Compression for Known Sources by Analytic Methods

Michael Drmota
TU Wien
michael.drmota@tuwien.ac.at

Wojciech Szpankowski
Purdue University
szpan@purdue.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Drmota and W. Szpankowski. *Redundancy of Lossless Data Compression for Known Sources by Analytic Methods*. Foundations and Trends® in Communications and Information Theory, vol. 13, no. 4, pp. 277–417, 2016.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-284-6

Full text available at: http://dx.doi.org/10.1561/0100000090

Foundations and Trends® in Communications and Information Theory
Volume 13, Issue 4, 2016

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2016, Volume 13, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Foundations and Trends® in Communications and
Information Theory
Vol. 13, No. 4 (2016) 277–417

DOI: 10.1561/0100000090

the essence of knowledge

Redundancy of Lossless Data Compression for Known Sources by Analytic Methods

Michael Drmota
TU Wien
michael.drmota@tuwien.ac.at

Wojciech Szpankowski
Purdue University
szpan@purdue.edu

Contents

Abstract

Lossless data compression is a facet of source coding and a well studied problem of information theory. Its goal is to find a shortest possible code that can be unambiguously recovered. Here, we focus on rigorous analysis of code redundancy for *known sources*. The redundancy rate problem determines by how much the actual code length exceeds the optimal code length. We present precise analyses of three types of lossless data compression schemes, namely fixed-to-variable (FV) length codes, variable-to-fixed (VF) length codes, and variable-to-variable (VV) length codes. In particular, we investigate the average redundancy of Shannon, Huffman, Tunstall, Khodak and Boncelet codes. These codes have succinct representations as trees, either as coding or parsing trees, and we analyze here some of their parameters (e.g., the average path from the root to a leaf). Such trees are precisely analyzed by analytic methods, known also as analytic combinatorics, in which complex analysis plays decisive role. These tools include generating functions, Mellin transform, Fourier series, saddle point method, analytic poissonization and depoissonization, Tauberian theorems, and singularity analysis. The term *analytic information theory* has been coined to describe problems of information theory studied by analytic tools. This approach lies on the crossroad of information theory, analysis of algorithms, and combinatorics.

M. Drmota and W. Szpankowski. *Redundancy of Lossless Data Compression for Known Sources by Analytic Methods*. Foundations and Trends® in Communications and Information Theory, vol. 13, no. 4, pp. 277–417, 2016.
DOI: 10.1561/0100000090.

1

Introduction

The basic problem of *source coding* better known as (lossless) *data compression* is to find a binary code that can be unambiguously recovered with shortest possible description either on average or for individual sequences. Thanks to Shannon's work we know that on average the number of bits per source symbol cannot be smaller than the source entropy rate. There are many codes asymptotically achieving the entropy rate, therefore one turns attention to *redundancy*. The average redundancy of a source code is the amount by which the expected number of binary digits per source symbol for that code exceeds entropy. One of the goals in designing source coding algorithms is to minimize the redundancy. In this survey, we discuss various classes of source coding and their corresponding redundancy. It turns out that such analyses often resort to studying certain intriguing trees such as Huffman, Tunstall, Khodak and Boncelet trees, as well as various algorithms such as divide-and-conquer approach. We study them using tools from the analysis of algorithms and analytic combinatorics[1] to discover precise and minute behavior of lossless compression codes.

[1] Andrew Odlyzko has argued that: "analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision."

Lossless data compression comes in three flavors: fixed-to-variable (FV) length codes, variable-to-fixed (VF) length codes, and finally variable-to-variable (VV) length codes. The latter includes the previous two families of codes and is the least studied among all data compression schemes. Over years we have seen a resurgence of interest in redundancy rate for *fixed-to-variable* coding (cf. [25, 28, 29, 30, 66, 90, 91, 92, 101, 103, 124, 126, 130, 132, 131, 139, 140, 151, 152, 164, 173, 180, 176, 177]). Surprisingly there are only a handful of results for variable-to-fixed codes (cf. [77, 97, 112, 133, 131, 134, 156, 161, 185]) and an almost non-existing literature on variable-to-variable codes (cf. [42, 50, 80, 97]). While there is some work on universal VF codes [156, 161, 185], to the best of our knowledge redundancy for universal VF and VV codes were not studied with the exception of some work of the Russian school [97, 96] (cf. also [99]).

In the fixed-to-variable code, discussed in Chapter 3, the encoder maps fixed length blocks of source symbols into variable-length binary code strings. Two important fixed-to-variable length coding schemes are the Shannon code and the Huffman code. In this survey we follow [152, 114]. We first discuss precise analyses of Shannon code redundancy for memoryless and Markov sources. We show that the average redundancy either converges to an explicitly computable constant, as the block length increases, or it exhibits a very erratic behavior fluctuating between 0 and 1. We also observe a similar behavior for the worst case or maximal redundancy. Then we move to the Huffman code. Despite the fact that Huffman codes have been so well known for so long, it was only relatively recently that their redundancy was fully understood. In [1] Abrahams summarizes much of the vast literature on fixed-to-variable length codes. Here, we present a precise analysis from our work [152] of the Huffman average redundancy for memoryless sources. We show that the average redundancy either converges to an explicitly computable constant, as the block length increases, or it exhibits a very erratic behavior fluctuating between 0 and 1. Following [114] we also present similar results for Markov sources.

Next, in Chapter 4 we study variable-to-fixed codes. A VF encoder partitions the source string into variable-length phrases that belong to a given dictionary $\mathcal{D}$. Often a dictionary is represented by a complete tree (i.e., a tree in which every node has maximum degree), also known as the *parsing tree*. The code assigns a fixed-length word to each dictionary entry. An important example of a variable-to-fixed code is the Tunstall code [157]. Savari and Gallager [131] present an analysis of the dominant term in the asymptotic expansion of the Tunstall code redundancy. In this survey, following [34], we describe a precise analysis of the phrase length (i.e., path from the root to a terminal node in the corresponding parsing tree) for such a code and its average redundancy. We also discuss a variant of Tunstall code known as VF Khodak code.

In the next Chapter 5 we continue analyzing VF codes due to Boncelet [15] who used the *divide-and-conquer principle* to design a practical encoding. Boncelet's algorithm is computationally fast and its practicality stems from the divide and conquer strategy: It splits the input (e.g., parsing tree) into several smaller subproblems, solving each subproblem separately, and then knitting together to solve the original problem. We use this occasion to present a careful analysis of a divide-and conquer recurrence which is at foundation of several divide-and-conquer algorithms such as heapsort, mergesort, discrete Fourier transform, queues, sorting networks, compression algorithms, and so forth [47, 86, 153].

In Chapter 6 we consider variable-to-variable codes. A variable-to-variable (VV) code is a concatenation of variable-to-fixed and fixed-to-variable codes. A variable-to-variable length encoder consists of a *parser* and a *string encoder*. The parser, as in VF codes, segments the source sequence into a concatenation of phrases from a predetermined dictionary $\mathcal{D}$. Next, the string encoder in a variable-to-variable scheme takes the sequence of dictionary strings and maps each one into its corresponding binary codeword of variable length. Aside from the special cases where either the dictionary strings or the codewords have a fixed length, very little is known about variable-to-variable length codes, even in the case of memoryless sources. In 1972 Khodak [80] described a VV scheme with small average redundancy that decreases with the growth of phrase length. He did not offer, however, an explicit VV code construction. We will remedy this situation and follow [16].

Finally, in Chapter 7 we discuss redundancy of one-to-one codes that are not necessarily prefix or even uniquely decodable. Recall that non-prefix codes are such codes which are not prefix free and do not satisfy Kraft's inequality. In particular, we analyze binary and non-binary one-to-one codes whose average lengths are smaller than the source entropy in defiance of the Shannon lower bound.

Throughout this survey, we study various intriguing trees describing Huffman, Tunstall, Khodak and Boncelet codes. These trees are studied by analytic techniques of analysis of algorithms [47, 85, 86, 87, 153]. The program of applying tools from analysis of algorithms to problems of source coding and in general to information theory lies at the crossroad of computer science and information theory. It is also known as *analytic information theory*. In fact, the interplay between information theory and computer science dates back to the founding father of information theory, Claude E. Shannon. His landmark paper "A Mathematical Theory of Communication" is hailed as the foundation for information theory. Shannon also worked on problems in computer science such as chess-playing machines and computability of different Turing machines. Ever since Shannon's work on both information theory and computer science, the research at the interplay between these two fields has continued and expanded in many exciting ways. In the late 1960s and early 1970s, there were tremendous interdisciplinary research activities, exemplified by the work of Kolmogorov, Chaitin, and Solomonoff, with the aim of establishing algorithmic information theory. Motivated by approaching Kolmogorov complexity algorithmically, A. Lempel (a computer scientist), and J. Ziv (an information theorist) worked together in the late 1970s to develop compression algorithms that are now widely referred to as Lempel-Ziv algorithms. Analytic information theory is a continuation of these efforts.

Finally, we point out that this survey deals only with source coding for *known sources*. The more practical *universal source coding* (in which the source distribution is unknown) is left for our future book *Analytic Information Theory*. However, at the end of this survey we provide an extensive bibliography on the redundancy rate problem, including universal source coding.

This survey is organized as follows. In the next chapter, we present some preliminary results such as Kraft's inequality, Shannon's lower bound, and Barron's lemma. In Section 3 we analyze Shannon and Huffman codes. Then we turn our attention in Section 4 to the Tunstall and VF Khodak codes. Finally, in Section 6 we discuss the VV code of Khodak and its interesting analysis. We conclude this survey with a chapter concerning the average redundancy for non-prefix codes such as one-to-one codes.

References

[1] J. Abrahams. Code and parse trees for lossless source encoding. *Communications in Information and Systems*, 1:113–146, 2001.

[2] M. Akra and L. Bazzi. On the solution of linear recurrence equations. *Computational Optimization and Applications*, (10):195–201, 1998.

[3] J. Allouche and J. Shallit. *Automatic Sequences*. Cambridge University Press, 2008.

[4] N. Alon and A. Orlitsky. A lower bound on the expected length of one-to one codes. *IEEE Trans. Information Theory*, (40):1670–1672, 1994.

[5] T. Apostol. *Introduction to analytic number theory*. Springer, 1976.

[6] K. Atteson. The asymptotic redundancy of bayes rules for markov chains. *IEEE Trans. on Information Theory*, (45):2104–2109, 1999.

[7] R. C. Baker. Dirichlet's theorem on diophantine approximation. *Cambridge Philos.*, (83):37–59, 1978.

[8] A. Barron. Logically smooth density estimation. *Ph.D. Thesis*, (Stanford University):Stanford, CA, 1985.

[9] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Information Theory*, (44):2743–2760, 1998.

[10] J. Bernardo. Reference posterior distributions for bayesian inference. *J. Roy. Stat. Soc. B.*, (41):113–147, 1979.

[11] V. Bernik and M.M. Dodson. *Metric Diophantine approximation on manifolds*. Cambridge University Press, 1999.

[12] P. Billingsley. Statistical methods in markov chains. *Ann. Math. Statistics*, (32):12–40, 1961.

[13] P. Billingsley. Convergence of probability measures. *John Wiley and Sons*, page New York, 1968.

[14] L. Biza. Asymptotically optimal tests for finite markov chains. *Ann. Math. Statistics*, (42):1992–2007, 1971.

[15] C. Boncelet. Block arithmetic coding for source compression. *IEEE Trans. Information Theory*, (39):1546–1554, 1993.

[16] Y. Bugeaud, M. Drmota, and W. Szpankowski. On the construction of (explicit) Khodak's code and its analysis. *IEEE Trans. Information Theory*, (54), 2008.

[17] J. W. S. Cassels. An introduction to diophantine approximation. *Cambridge University Press*, 1957.

[18] V. Choi and M. J. Golin. Lopsided trees. *I. Analyses.*, (31):240–290, 2001.

[19] B. Clarke and A. Barron. Information-theoretic asymptotics of bayes methods. *IEEE Trans. Informational Theory*, (36):453–471, 1990.

[20] B. Clarke and A. Barron. Jeffrey's prior is asymptotically least favorable under entropy risk. *J. Stat. Planning Inference*, (41):37–61, 1994.

[21] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth. On the lambert w function. *Adv. Computational Mathematics*, (5):329–359, 1996.

[22] T. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Trans. Information Theory*, (42):348–363, 1996.

[23] T. M. Cover and J. A. Thomas. Elements of information theory. *John Wiley and Sons*, (New York), 1991.

[24] I. Csiszar and J. Korner. Information theory: Coding theorems for discrete memoryless systems. *Academic Press*, (New York), 1981.

[25] I. Csiszar and P. Shields. Redundancy rates for renewal and other processes. *IEEE Trans. Information Theory*, (42):2065–2072, 1996.

[26] L. Davisson. Universal noiseless coding. *IEEE Trans. Inform. Theory*, (19):783–795, 1973.

[27] L. Davisson and A. Leon-Garcia. A source matching approach to finding minimax codes. *IEEE Trans. Inform. Theory*, (26):166–174, 1980.

[28] A. Dembo and I. Kontoyiannis. The asymptotics of waiting times between stationary processes allowing distortion. *Annals of Applied Probability*, (9):413–429, 1999.

[29] A. Dembo and I. Kontoyiannis. Critical behavior in lossy coding. *IEEE Trans. Inform. Theory*, (47):1230–1236, 2001.

[30] A. Dembo and I. Kontoyiannis. Source coding large deviations and approximate pattern matching. *IEEE Trans. Information*, (48):1590–1615, 2002.

[31] H. Dickinson and M. M. Dodson. Extremal manifolds and hausdorff dimension. *Duke Math*, (101):271–281, 2000.

[32] M. Drmota. A bivariate asymptotic expansion of coefficients of powers of generating. *Europ. J. Combinatorics*, (15):139–152, 1994.

[33] M. Drmota, H. K. Hwang, and W. Szpankowski. Precise average redundancy of an idealized arithmetic coding. *Proc. Data Compression Conference*, (Snowbird):222–231, 2002.

[34] M. Drmota, Y. Reznik, S. Savari, and W. Szpankowski. Precise asymptotic analysis of the tunstall code. *2006 International Symposium on Information Theory*, pages 2334–2337, 2006.

[35] M. Drmota, Y. Reznik, and W. Szpankowski. Tunstall code, Khodak variations, and random walks. *IEEE Trans. Inform. Theory*, (56):2928–2937, 2010.

[36] M. Drmota and W. Szpankowski. Precise minimax redundancy and regrets. *IEEE Trans. Information Theory*, (50):2686–2707, 2004.

[37] M. Drmota and W. Szpankowski. Variations on Khodak's variable-to-variable codes. *42nd Annual Allerton Conference on Communication Control Computing*, page Urbana, 2004.

[38] M. Drmota and W. Szpankowski. On the exit time of a random walk with the positive drift. *2007 Conference on Analysis of Algorithms Juan-les-Pins France and Proc. Discrete Mathematics and Theoretical Computer Science*, (291-302), 2007.

[39] M. Drmota and W. Szpankowski. A master theorem for discrete divid and conquer recurrences. *J. of the ACM*, (60):16:1–16:49, 2013.

[40] M. Drmota and R. Tichy. Sequences discrepancies and applications. *Springer Verlag Berlin Heidelberg*, 1997.

[41] Y. Ephraim and N. Merhav. Hidden markov processes. *IEEE Trans. Information Theory*, (48):1518–1569, 2002.

[42] F. Fabris. Variable-length-to-variable-length source coding: A greedy step-by-step algorithm. *IEEE Trans. Information Theory*, (38):1609–1617, 1992.

[43] J. Fan, T. Poo, and B. Marcus. Constraint gain. *IEEE Trans. Information Theory*, (50):1989–2001, 2004.

[44] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Trans. Information Theory*, (38):1258–1270, 1992.

[45] P. Flajolet. Singularity analysis and asymptotics of bernoulli sums. *Theoretical Computer Science*, (215):371–381, 1999.

[46] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Special Volume on Mathematical Analysis of Algorithms*, (144):3–58, 1995.

[47] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2008.

[48] P. Flajolet and W. Szpankowski. Analytic variations on redundancy rates of renewal processes. *IEEE Trans. Information Theory*, (48):2911–2921, 2002.

[49] P. H. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *Discrete Math.*, (3):216–240, 1990.

[50] G. H. Freeman. Divergence and the construction of variable-to-variable-length lossless codes by source-word extensions. *Data Comnpression Conference 1993*, pages 79–88, 1993.

[51] R. Gallager. Information theory and reliable communications. *New York Wiley*, 1968.

[52] R. Gallager. Variations on the theme by huffman. *IEEE Trans. Information Theory*, (24):668–674, 1978.

[53] R. Gallager and D. van Voorhis. Optimal source codes for geometrically distributed integer alphabets. *IEEE Trans. Information Theory*, (21):228–230, 1975.

[54] S. Golomb. Run-length coding. *IEEE Trans. Information Theory*, (12):399–401, 1996.

[55] G.Park, H. Hwang, P. Nicodeme, and W. Szpankowski. Profile in tries. *SIAM J. Computing*, (38):1821–1880, 2009.

[56] P. Grabner and J. Thuswaldner. Analytic continuation of a class of dirichlet series. *Abh. Math. Sem. Univ. Hamburg*, (66):281–287, 1996.

[57] R. M. Gray. Optimization noise spectra. *IEEE Trans. Information Theory*, (36):1220–1244, 1990.

[58] D. K. He and E. H. Yang. Performance anaylsis of grammer-based codes revisited. *IEEE Trans. Information Theory*, (50):1524–1535, 2004.

[59] P. Howard and J. Vitter. Analysis of arithmetic coding for data compression. *Proc. Data Compression Conference*, pages 3–12, 1991.

[60] K. H. Hwang. Large deviations for combinatorial distributions i: Central limit theorems. *Ann. Appl. Probab.*, (6):297–319, 1996.

[61] P. Jacquet, C. Knessl, and W. Szpankowski. Counting markov types, balanced matrices, and eulerian graphs. *IEEE Trans. Information Theory*, (58):4261–4272, 2012.

[62] P. Jacquet, G. Seroussi, and W. Szpankowski. On the entropy of a hidden markov process. *Data Compression Conference*, pages 362–371, 2004.

[63] P. Jacquet, G. Seroussi, and W. Szpankowski. On the entropy of a hidden markov process. *Theoretical Computer Science*, (395):203–219, 2008.

[64] P. Jacquet and W. Szpankowski. Analysis of digital tries with markovian dependency. *IEEE Trans. Information Theory*, (37):1470–1475, 1991.

[65] P. Jacquet and W. Szpankowski. Autocorrelation on words and its applicaitons. analysis of suffix trees by string-ruler approach. *Combinatorial Theory Ser A.*, (66):237–269, 1994.

[66] P. Jacquet and W. Szpankowski. Asymptotic behavior ofhte lempel-ziv parsing scheme and digital search trees. *Theoretical Computer Science*, (144):161–197, 1995.

[67] P. Jacquet and W. Szpankowski. Anaylytical depoissonization and its applications. *Theoretical Computer Science*, (201):1–62, 1998.

[68] P. Jacquet and W. Szpankowski. Entropy computations via analytic depoissonization. *IEEE Trans. Information Theory*, (45):1072–1081, 1999.

[69] P. Jacquet and W. Szpankowski. A combinatorial problem arising in information theory: Precise minimax redundancy for markov sources. In *Proc. Colloquium on Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities*, pages 311–328. Birkhauser, 2002.

[70] P. Jacquet and W. Szpankowski. Analytic approach to pattern matching applied combinatorics on words. *Cambridge University Press*, page Chapter 7, 2004.

[71] P. Jacquet and W. Szpankowski. Markov types and minimax redundancy for markov sources. *IEEE Trans. Information Theory*, (50):1393–1402, 2004.

[72] P. Jacquet and W. Szpankowski. Joint string complexity for markov sources. In *23rd International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, pages 1–12, 2012.

[73] P. Jacquet and W. Szpankowski. *Analytic Pattern Matching: From DNA to Twitter*. Cambridge University Press, Cambridge, 2015.

[74] P. Jacquet, W. Szpankowski, and I. Apostol. A universal predictor based on pattern matching. *IEEE Trans. Information Theory*, (48):1462–1472, 2002.

[75] P. Jacquet, W. Szpankowski, and J. Tang. Average profile of hte lempel-ziv parsing scheme for a markovian source. *Algorithmica*, (31):318–360, 2001.

[76] S. Janson. Moments for first passage and last exit times the minimum and related quantities for random walks with positive drift. *Adv. Appl. Probab.*, (18):865–879, 1986.

[77] F. Jelinek and K. S. Schneider. On variable-length-to-block coding. *Trans. Information Theory*, (18):765–774, 1972.

[78] G. Katona and G. Tusnady. The principle of conservation of entropy in a noiseless channel. *Studia Sci. Math.*, (2):20–35, 1967.

[79] G. L. Khodak. Connection between redundancy and average delay of fixed-length coding. *All-Union Conference on Problems of Theoretical Cybernetics*, (Novosibirsk USSR), 1969.

[80] G. L. Khodak. Bounds of redundancy estimates for word-based encoding of sequences produced by a bernoulli source. *Problemy Peredachi Informacii*, (8):21–32, 1972.

[81] J. C. Kieffer. A unified approach to weak universal source coding. *IEEE Trans. Information Theory*, (24):340–360, 1978.

[82] J. C. Kieffer. Sample converses in source coding theory. *IEEE Trans. Information Theory*, (37):263–268, 1991.

[83] J. C. Kieffer. Strong converses in source coding relative to a fidelity criterion. *IEEE Trans. Information Theory*, (37):257–262, 1991.

[84] C. Knessl and W. Szpankowski. Enumeration of binary trees lempel-ziv '78 parsings and universal types. *Proc. the Second workshop on Analytic Algorithmics and Combinatorics*, page Vancouver, 2005.

[85] D. E. Knuth. The art of computer programming. fundmental algorithms. *Addison-Wesley Reading*, (Vol 1):Third Edition, 1997.

[86] D. E. Knuth. The art of computer programming. seminumerical algorithms. *Addison Esley Reading*, (Vol 2):Third Edition, 1998.

[87] D. E. Knuth. The art of computer programming sorting and searching. *Addison-Wesley Reading*, (Vol 3):Second Edition, 1998.

[88] D. E. Knuth. Linear probing and graphs. *Algorithmica*, (22):561–568, 1998.

[89] D. E. Knuth. Selected papers on the analysis of algorithms. *Cambridge University Press*, 2000.

[90] I. Kontoyiannis. An implementable lossy version of the lempel-ziv algorithm-part i: Optimality for memoryless sources. *IEEE Trans. Information Theory*, (45):2285–2292, 1999.

[91] I. Kontoyiannis. Pointwise redundancy in lossy data compression and universal lossy data compression. *IEEE Trans. Information Theory*, (46):136–152, 2000.

[92] I. Kontoyiannis. Sphere-covering measure concentration and source coding. *IEEE Trans. Information Theory*, (47):1544–1552, 2001.

[93] I. Kontoyiannis and S. Verdu. Optimal lossless data compression: Non-asymptotics and asymptotics. *IEEE Trans. Information Theory*, (60):777–795, 2014.

[94] J. Korevaar. A century of complex tauberian theory. *Bull. Amer. Soc.*, (39):475–531, 2002.

[95] C. Krattenthaler and P. Slater. Asymptotic redundancies for univeral quantum coding. *IEEE Trans. Information Theory*, (46):801–819, 2000.

[96] R. Krichevsky. Universal compression and retrieval. *Kluwer Dordrecht*, 1994.

[97] R. Krichevsky and V. Trofimov. The performance of universal coding. *IEEE Trans. Information Theory*, (27):199–207, 1981.

[98] L. Kuipers and H. Niederreiter. Uniform distribution of sequences. *John Wiley and Sons*, 1974.

[99] J. Lawrence. A new universal coding scheme for the biary memoryless source. *IEEE Trans. Information Theory*, (23):466–472, 1977.

[100] A. Lempel and J. Ziv. On the cojmplexity of finite sequences. *IEEE Information Theory*, (22):75–81, 1976.

[101] T. Linder, G. Lugosi, and K. Zeger. Fixed-rate universal lossy source coding and rates of convergence for memoryless sources. *IEEE Information Theory*, (41):665–676, 1995.

[102] S. Lonardi, W. Szpankowski, and M. Ward. Error resilient lz'77 data compression: Algorithms analysis and experiments. *IEEE Trans. Information Theory*, (53):1799–1813, 2007.

[103] G. Louchard and W. Szpankowski. Average profile and limiting distribution for a phrase size in the lempel-ziv parsing algorithm. *IEEE Trans. Information Theory*, (41):478–488, 1995.

[104] G. Louchard and W. Szpankowski. On the average redundancy rate of the lempel-ziv code. *IEEE Trans. Information Theory*, (43):2–8, 1997.

[105] G. Louchard, W. Szpankowski, and J. Tang. Average profile for the generalized digital search trees and the generalized lempel-ziv algorithms. *SIAM J. Computing*, (28):935–954, 1999.

[106] T. Luczak and W. Szpankowski. A suboptimal lossy data compression based in approximate pattern matching. *IEEE Trans. Information Theory*, (43):1439–1451, 1997.

[107] H. Mahmoud. Evolution of random search trees. *John Wiley and Sons*, 1992.

[108] K. Marton and P. Shields. The positive-divergence and blowing-up properties. *Israel J. Math*, (80):331–348, 1994.

[109] J. Massey. The entropy of a rooted tree with probabilities. In *International Symposium on Information Theory*, 1983.

[110] N. Merhav and M. Feder. A strong version of the redundancy-capacity theory of universal coding. *IEEE Trans. Information Theory*, (41):714–722, 1995.

[111] N. Merhav, M. Feder, and M. Gutman. Some properties of sequential predictors for binary markov sources. *IEEE Trans. Information Theory*, (39):887–892, 1993.

[112] N. Merhav and D. Neuhoff. Variable-to-fixed length codes provided better large deviations performance than fixed-to-variable codes. *IEEE Trans. Information Theory*, (38):135–140, 1992.

[113] N. Merhav, G. Seroussi, and M. Weinberger. Optimal prefix codes for sources with two-sided geometric distributions. *IEEE Trans. Information Theory*, (46):121–135, 2000.

[114] N. Merhav and W. Szpankowski. Average redundancy of the shannon code for markov sources. *IEEE Trans. Information Theory*, (59):7186–7193, 2013.

[115] N. Merhav and J. Ziv. On the amount of statistical side information required for lossy data compression. *IEEE Trans. Information Theory*, (43):1112–1121, 1997.

[116] R. Noble and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

[117] A. Odlyzk. Asymptotic enumeration. *Handbook of Cominatorics*, (II):1063–1229, 1995.

[118] A. Orlitsky and P. Santhanam. Speaking of infinity. *IEEE Trans. Information Theory*, (50):2215–2230, 2004.

[119] A. Orlitsky, Prasad Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Trans. Information Theory*, (50):1469–1481, 2004.

[120] D. Ornstein and P. Shields. Universal almost sure data compression. *Ann. Probab.*, (18):441–452, 1990.

[121] D. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Informaiton Theory*, (39):78–83, 1993.

[122] E. Plotnik, M. J. Weinberger, and J. Ziv. Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the lempel-ziv algorithm. *IEEE Trans. Information Theory*, (38):66–72, 1992.

[123] Y. Reznik and W. Szpankowski. On average redundancy rate of the lempel-ziv codes with k-error protocol. *Information Sciences*, (135):57–70, 2001.

[124] J. Rissanen. Complexity of strings in the class of markov sources. *IEEE Trans. Information Theory*, (30):526–532, 1984.

[125] J. Rissanen. Universal coding and information and prediction and estimation. *IEEE Trans. Information Theory*, (30):629–636, 1984.

[126] J. Rissanen. Fisher information and stochastic complexity. *IEEE Trans. Information Theory*, (42):40–47, 1996.

[127] B. Ryabko. Twice-universal coding. *Problems of Information Transmission*, pages 173–177, 1984.

[128] B. Ryabko. Prediction of random sequences and universal coding. *Problems of Information Transmission*, (24):3–14, 1988.

[129] B. Ryabko. The complexity and effectiveness of prediction algorithms. *J. Complexity*, (10):281–295, 1994.

[130] S. Savari. Redundancy of the lempel-ziv incremental parsing rule. *IEEE Trans. Information Theory*, (43):9–21, 1997.

[131] S. Savari and R. Gallager. Generalized tunstall codes for sources with memory. *IEEE Trans. Information Theory*, (43):658–668, 1997.

[132] S. A. Savari. Variable-to-fixed length codes for predictable sources. *Proc Ieee Data Compresion Conference*, pages 481–490, 1998.

[133] S. A. Savari. Variable-to-fixed length codes and the conservation of enthropy. *Trans. Information Theory*, (45):1612–1620, 1999.

[134] J. Schalkwijk. An algorithm for source coding. *IEEE Information Theory*, (18):395–399, 1972.

[135] W. M. Schmidt. *Diophantine Approximation*. Springer, 1980.

[136] R. Sedgewick and P. Flajolet. An introduction to the analysis of algorithms. *Addison-Wesley Publishing Company Reading Mass.*, 1995.

[137] G. Seroussi. On universal types. *IEEE Transactions on Informatoin Theory*, (52):171–189, 2006.

[138] P. Shields. Universal redundancy rates do not exist. *IEEE Information Theory*, (39):520–524, 1993.

[139] P. Shields. The ergodic theory of discrete sample path. *American Mathematical Society*, 1996.

[140] Y. Shtarkov. Universal sequential coding of single messages. *Problems of Information Transmission*, (23):175–186, 1987.

[141] Y. Shtarkov, T. Tjalkens, and F. M. Willems. Multi-alphabet universal coding of memoryless sources. *Problems of Information Transmission*, (31):114–127, 1995.

[142] V. G. Sprindzuk. *Metric Theory of Diophantine Approximations*. Wiley, 1979.

[143] R. Stanley. Enumerative combinatorics. *Watsworth Monterey*, 1986.

[144] R. Stanley. Enumerative combinatorics. *Cambridge University Press*, (vol. 2), 1999.

[145] Y. Steinberg and M. Gutman. An algorithm for source coding subject to a fidelity criterion based on string matching. *IEEE Trans. Information Theory*, (39):877–886, 1993.

[146] P. Stubley. On the redundancy of optimum fixed-to=variable length codes. *Proc. Data Compression Conference*, pages 90–97, 1994.

[147] B. Sury. Weierstrass's theorem - leaving no stone unturned. In *Workshop on Linear Algebra and Analysis*, 2006.

[148] W. Szpankowski. Asymptotic properties of data compress and suffix trees. *IEEE Trans. Information Theory*, (39):1647–1659, 1993.

[149] W. Szpankowski. A generalized suffix tree and its (un) expected asymptotic behaviors. *SIAM J. Compt.*, (22):1176–1198, 1993.

[150] W. Szpankowski. On asymptotics of certain sums arising in coding theory. *IEEE Trnas. Information Theory*, (41):2087–2090, 1995.

[151] W. Szpankowski. On asymptotics of certain recurrences arising in universal coding. *roblems of Information Transmission*, (34):55–61, 1998.

[152] W. Szpankowski. Asymptotic redundancy of huffman (and other) block codes. *IEEE Trans. Information Theory*, (46):2434–2443, 2000.

[153] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* Wiley New York, New York, 2001.

[154] W. Szpankowski. A one-to-one code and its anti-redundancy. *IEEE Trans. Information Theory*, (54):4762–4766, 2008.

[155] W. Szpankowski and S. Verdu. Minimum expected length of fixed-to-variable lossless compression without prefix constraints. *IEEE Trans. Information Theory*, (57):4017–4025, 2011.

[156] T. Tjalkens and F. Willems. A universal variable-to-fixed length source code based on lawrence's algorithm. *IEEE Trans. Information Theory*, (38):247–253, 1992.

[157] B. P. Tunstall. Synthesis of noiseless compression codes. *Ph.D. dissertation*, (Georgia Inst. Technology), 1967.

[158] J. D. Vaaler. Some extremal functions in fourier analysis. *Bull. Amer. Math. Soc.*, (12):183–216, 1985.

[159] B. Vall. Dynamics of the binary euclidean algorithm functional analysis and operators. *Algorithmica*, (22):660–685, 1998.

[160] B. Vall. Dynamical sources in information theory: Fundamental intervals and word prefixes. *Algorithmica*, (29):262–306, 2001.

[161] K. Visweswariah, S. Kulkurani, and S. Verdu. Universal variable-to-fixed length source codes. *IEEE Trans. Information Theory*, (47):1461–1472, 2001.

[162] J. Vitter and P. Krishnan. Optimal prefetching via data compression. *ACM*, (43):771–793, 1996.

[163] M. Ward and W. Szpankowski. Analysis of a randomized selection algorithm motivated by the lz'77 shceme. *the First Workshop on Analytic Algorithmics and Combinatorics*, (New Orleans):153–160, 2004.

[164] M. Weinberger, N. Merhav, and M. Feder. Optimal sequential probability assignments for individual sequences. *IEEE Trans. Information Theory*, (40):384–396, 1994.

[165] M. Weinberger, J. Rissanen, and R. Arps. Applications of universal context modeling to lossless compression of gray-scale images. *IEEE Trans. Image Processing*, (5):575–586, 1996.

[166] M. Weinberger, J. Rissanen, and M. Feder. A universal finite memory sources. *IEEE Trans. Information Theory*, (41):643–652, 1995.

[167] M. Weinberger, G. Seroussi, and G. Sapiro. Loco-i: A low complexity context-based lossless image compression algorithms. *Proc. Data Compression Conference*, pages 140–149, Snowbird 1996.

[168] P. Whittle. Some distribution and moment formulae for markov chain. *J. Roy. Stat. Soc.*, (17):235–242, 1955.

[169] F. M. Willems. the context-tree weighting method: Extensions. *IEEe Trans. Information Theory*, to appear.

[170] F. M. Willems, Y. Shtarkov, and T. Tjalkens. Context weighting for general finite context sources. *IEEE Trans. Information Theory*, (42):1514–1520, 1996.

[171] F. M. Williams, Y. Shtarkov, and T. Tjalkens. The context-tree weighting method: Basic properties. *IEEE Trans. Information Theory*, (41):653–664, 1995.

[172] A. Wyner and J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Information Theory*, (35):1250–1258, 1989.

[173] A. D. Wyner. An upper bound on the entropy series. *Information and Control*, (20):176–181, 1972.

[174] A. D. Wyner. An upper bound on the entropy series. *Informiationa nd Control*, (20):176–181, 1972.

[175] A. J. Wyner. The redundancy and distribution of the phrase lengths of the fixed-database lempel-ziv algorithm. *IEEE Trans. Information Theory*, (43):1439–1465, 1997.

[176] Q. Xie and A. Barron. Minimax redundancy for the class of memoryless souces. *IEEE Trans. Information Theory*, (43):647–657, 1997.

[177] Q. Xie and A. Barron. Asymptotic minimax regret for data compression and gambling and prediction. *IEEE Trans. Information Theory*, (46):431–445, 2000.

[178] E. H. Yang and J. Kieffer. Simple universal lossy data compression schemees derived from lempel-ziv algorithm. *IEEE Trans. Information Theory*, (42):239–245, 1996.

[179] E. H. Yang and J. Kieffer. On the redundancy of the fixed-database lempel-ziv algorithm for mixing sources. *IEEE Trans. Information Theory*, (43):1101–1111, 1997.

[180] E. H. Yang and J. Kieffer. On the performance of data compression algorithms based upon string matching. *IEEE Trans. Information Theory*, page 44, 1998.

[181] E. H. Yang and Z. Zhang. The shortest common superstring problem: Average case analysis for both exact matching and approximate matching. *IEEE Trans. Information Theory*, (45):1867–1886, 1999.

[182] Y. Yang and A. Barron. Informatoin-theoretic determination of minimax rates of convergence. *The Ann. Stat.*, (27):1564–1599, 1999.

[183] Z. Zhang and V. Wei. An on-line universal lossy data compressoin algorithm via continuous codebook reinement part i: Basic results. *IEEE Trans. Information Theory*, (2):803–821, 1996.

[184] J. Ziv. Coding of source with unknown statistics part ii: Distortion relative to a fidelity criterion. *IEEE Trans. Information Theory*, (18):389–394, 1972.

[185] J. Ziv. Variable-to-fixed length codes are better than fixed-to-variable length codes for markov sources. *IEEE Trans. Information Theory*, (36):861–863, 1990.

[186] J. Ziv. Back from infinity: A constrained resouces approach to information theory. *IEEE Information Theory Society Newsletter*, (48):30–33, 1998.

[187] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Information Theory*, (23):337–343, 1977.

[188] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Information Theory*, (24):530–536, 1978.

[189] A. Zygmund. *Trigonometric Series*. Cambridge University Press, New York, 1959.