Zeckendorf Decomposition and Conditioned Markov Chains

Iddo Ben-Ari, University of Connecticut
Joint with Steven J. Miller, Williams College

(to appear Siam Journal of Discrete Math
http://dx.doi.org/10.1137/140996859)

Introduction

This is about

- Application of standard tools in theory of Markov processes in certain enumeration systems, which we refer to as *generalized Zeckendorf decompositions*.
- The tools include Doob's h-transform and ergodicity of Markov chains.
- Our work provides a unified approach to the study of statistical properties of the model.

Recurrence relations and decompositions of integers

Consider a linear recurrence relation of length $L \geq 1$ with integer coefficients $c_1, \ldots, c_L \in \mathbb{Z}_+$, $c_1 c_L > 0$:

$$\begin{cases} G_1 = 1 \\ G_{n+1} = c_1 G_n + \cdots + c_n G_1 + 1 & n = 1, \ldots, L-1 \\ G_{n+1} = c_1 G_n + \cdots + c_L G_{n+1-L} & n \geq L. \end{cases}$$

Facts

- Given the sequence $(G_n : n \in \mathbb{N})$, each natural number can be written as a linear combination with nonnegative integer coefficients its elements. The coefficients are called *digits*.
- The decomposition is unique if we require a consistency condition generalizing the greedy algorithm used to obtain base k-decomposition. We call the decomposition the *generalized Zeckendorf decomposition*.

Zeckendorf decomposition

Idea

- Obtain digits by "exhausting" the recurrence relation, not individual digits each step.

Definition

- We fix a linear recurrence relation

$$G_{n+1} = c_1 G_n + \cdots + c_L G_{n+1-L}, n \geq L.$$

- $N \in \mathbb{N}$ has a *legal* decomposition of length $n \in \mathbb{N}$ if there exist *digits* $a_1 \in \mathbb{N}, a_2, \ldots, a_n \in \mathbb{Z}_+$, such that

$$N = \sum_{i=1}^{n} a_i G_{n+1-i},$$

  and

  - $n < L$ and $a_i = c_i$ for $1 \leq i \leq n$; or
  - there exists some $s \in \{1, \ldots, L\}$ such that

$$\left.\begin{array}{l} a_1 = c_1,\ a_2 = c_2,\ \ldots,\ a_{s-1} = c_{s-1}, \text{ and } a_s < c_s, \\ a_{s+1}, \ldots, a_{s+\ell} = 0 \text{ for some } \ell \geq 0, \\ \{b_i\}_{i=1}^{n-s-\ell} \text{ with } b_i = a_{s+\ell+i}, \text{ is either legal or empty.} \end{array}\right\}$$

- The legal decomposition is also called the *generlized Zeckendorf decomposition*, and can be written as

$$[N]_G = a_1 \ldots a_n.$$

Existence and Uniqueness of a Legal Decompositions

Theorem 1

Under above assumptions, every $N \in \mathbb{N}$ has a unique legal decomposition.

For a reference see [MW1].

Properties

- The legal decomposition of N is of length n iff $G_n \leq N < G_{n+1}$.
- The digits in the legal decompositions are in $\{0, \ldots, \max_i c_i\}$, but there are certain constraints. Not all words of of length n are legal.

Examples

1. $G_{n+1} = kG_n, \ k = 2, 3, \ldots$
   $\Rightarrow [N]_G$ is base-k expansion.
   Digits are in $\{0, 1, \ldots, k-1\}$, $a_1 \neq 0$ and all other digits are free.
2. $G_{n+1} = G_n + G_{n-1}, \ G_1 = 1, G_2 = 2$ (Fibonacci sequence)
   $\Rightarrow$ Zeckendorf decomposition.
   - Digits are in $\{0, 1\}$
   - A sequence if legal iff $a_1 = 1$ and no two adjacent digits are 1.

j	9	8	7	6	5	4	3	2	1
G_j	45	34	21	13	8	5	3	2	1
$[16]_G$				1	0	0	1	0	0

   The sequence 11100 also represents 16, but is not legal.
3. $G_{n+1} = 3G_n + 4G_{n-1} + 2G_{n-2}$

j	6	5	4	3	2	1
G_j	1159	283	69	17	4	1
$[280]_G$			3	4	1	1

   If we were trying to maximize the digit at each step we would get a different, not legal, decomposition:
   $$280 = 4 * G_4 + 1 * G_2$$
   Another example of a legal sequence $[624]_G = 2\ 0\ 31\ 3$.

Statistical Properties

Probabilistic Structure

- Statistics under sequence of uniform probability measure on $\{G_n, G_n + 1, \ldots, G_{n+1} - 1\}$, that is, the numbers whose decomposition is of length n.
- Natural and convenient choice for combinatorial manipulation, which is how the model has been studied.
- Work extends to uniform measure on initial segments $\{1, \ldots, N\}$ (under some additional assumptions).

Questions

- Number of nonzero digits: Moments, LLN, CLT, LIL, LDP.
- Gaps between nonzero digits.

Previous and current work based primarily on nontrivial combinatorial analysis.

What we do

- Express the decompositions in terms of Markov chain dynamics.
- This framework provides immediate access to standard results on statistical properties.

Probabilistic construction

Idea

- Start with uniform measure on all sequences of digits of length n.
- Obtain uniform on legal decompositions through conditioning ... on the set of legal sequences.
- Associate the procedure to a Markov chain conditioned not to hit a set and apply the classical Doob's h-processes.

Construction

- $Z = \big(Z_n = (X_n, Y_n) : n \in \mathbb{N}\big)$ canonical process on state space $\{0, \dots, \max_i c_i\} \times \{1, \dots, L\}$.
- P, probability measure under which the Z is an IID, Z_1 uniformly distributed over state space.
- A (infinite) realization of Z is *legal* if
  1. (nontrivial first digit) $X_1 \in \{1, \dots, c_1\}$, $Y_1 = 1$
  2. (finite) There exists $J = J(Z) \in \mathbb{N}$ such that $X_J > 0$ and $Z_n = (0, 1)$ for $n > J$
  3. (legality) For all $n \in \mathbb{N}$, either
     - $X_n = c_{Y_n}$ and $Y_{n+1} = Y_n + 1$; or
     - $X_n < c_{Y_n}$ and $Y_{n+1} = 1$.
- *Legality test*

$$\tau = \inf\{n \geq 1 : (Z_1, \dots, Z_n) \text{ does not extend to an infinite legal realization}\}.$$

Finally, define:

$$Q_n((Z_1, \dots, Z_n) \in \cdot) = P((Z_1, \dots, Z_n) \in \cdot | \tau > n).$$

As a result: Q_n inherits uniformity from P.

Completing the construction

- Recall $Q_n(\cdot) = P(\cdot|\tau > n)$.
- $\Rightarrow$ the distribution of $(X_1, \ldots, X_n)$ under Q_n is uniform over all legal decompositions of length n. Specifically the random variable

$$N = \sum_{j=1}^{n} X_j G_{n+1-j}$$

is uniform on $\{G_n, \ldots, G_{n+1} - 1\}$ under Q_n.
- **Goal.** Identify conditioned measure as MC conditioned not to hit a set.
- Problem: τ is not a hitting time for Z.
- Letting $Z'_n = (X_n, Y_n, Y_{n+1})$, then Z' is a MC under P and it follows that

$$\tau = \inf\{n : Z'_n \in A\},$$

where

$$A^c = \{(x, j, j') : (x < c_j \text{ and } j' = 1) \text{ or } (x = c_j \text{ and } j' = j+1) \text{ or } \big((x, j) = (c_L, L) \text{ and } j' = 1\big)\}.$$

Example

Consider the recurrence relation

$$G_{n+1} = 3G_n + 4G_{n-1} + 2G_{n-2}.$$

We have $L = 3$, $c_1 = 3, c_2 = 4, c_3 = 2$.

Here is a legal sequence

X	340	2	33	0	0	2
Y	123	1	12	1	1	1

- Length is 9, of which 6 nonzero digits.
- There are 6 "blocks".

  Note that the block length varies.
- Why Y ?
  Since $c_2 = 4$, first "3" digit in a block can be followed by any of the digits $\{0, \ldots, 4\}$. Since $L = 3$ and $c_3 = 2$, if the digit "3" appears in the second position in a block, then it can be only followed by the digits 0,1.
- Note that for a legal decomposition, the values of Y are determined by history of X, but Y is useful for keeping bookkeeping. s of size zero.

The Resulting Structure I

Corollary 1

The restriction of TF for Z' induced by P to A^c is substochastic, irreducible and aperiodic. Let λ_c denote its Perron eigenvalue, φ_c corresponding, ℓ^1-normalized positive eigenvector.
Define a TF Q on A^c through

$$Q(z, z') = \frac{1}{\lambda_c \varphi_c(z)} P(z, z') \varphi_c(z').$$

Then Z' is an ergodic MC under Q.

Proposition 1

Let $f = f(Z'_1, \ldots, Z'_n)$. Then

$$E^{Q_n}(f) = \frac{E^Q_{\tilde{\varphi}_c}\left(\frac{f}{\varphi_c(Z'_n)}\right)}{E^Q_{\tilde{\varphi}_c}\left(\frac{1}{\varphi_c(Z'_n)}\right)},$$

where $\tilde{\varphi}_c$ is the distribution φ_c conditioned on $\{(z, 1, j'),\ z \in \{1, \ldots, c_1\}\}$.

- All Q_n are treated through dynamics according to the TF Q.

The Resulting Structure II

- If $f_n = f_n(Z_1, \dots, Z_n')$ do not depend too much on last coordinates of Z', then ergodicity of Z' allows to decouple Z_n' from f in numerator to obtain

$$E^{Q_n}(f_n) \sim E^{Q}_{\tilde{\varphi}_c}(f_n).$$

Here is a sketch of a proof:

Proof.

Suppose $z_1, \dots, z_n \in A^c$. Then

$$\begin{aligned}
P(Z_1' = z_1, \dots, Z_n' = z_n, \tau > n) &= P(Z_1' = z_1, \dots, Z_n' = z_n) \\
&= \prod_{j=1}^{n-1} P(z_j, z_{j+1}) \\
&= \lambda_c \varphi_c(z_1) \frac{Q(z_1, z_2)}{\varphi_c(z_2)} \times \lambda_c \varphi_c(z_2) \frac{Q(z_2, z_3)}{\varphi_c(z_3)} \times \cdots \times \lambda_c \varphi_c(z_{n-1}) \frac{Q(z_{n-1}, z_n)}{\varphi_c(z_n)} \\
&= \lambda_c^n \frac{\varphi_c(z_1)}{\varphi_c(z_n)} \prod_{j=1}^{n-1} Q(z_j, z_{j+1}).
\end{aligned}$$

From here we see that

$$E^P_{z_1}[f(Z'), \tau > n] = \lambda_c^n \varphi_c(z_1) E^Q_{z_1}[\frac{f(Z')}{\varphi_c(Z_n')}],$$

and the result follows after some manipulation (sum over initial configurations and condition) □

Additive functionals

The results in this section are not new, but proofs and approach is new.

Let π^Q denote the stationary distribution for Q.

Group inverse, $Q^{\#}$

The restriction of $I - Q$ to the Q-invariant subspace $\{g : \sum \pi^Q(z)g(z) = 0\}$ is invertible.

Denote inverse by $Q^{\#}$, and extend to all functions by letting $Q^{\#}\mathbf{1} = 0$.

Let $S_n = \sum_{j=1}^{n} g(Z_j')$, for some fixed g.

Proposition 2

$$E^{Q_n}[S_n] = nE_{\pi^Q} g + E_{\tilde{\varphi}_c}(Q^{\#} g) + o(1).$$

This could be used for computing expectation number of summands under Q_n.

Recall that Zeckendorf decomposition corresponds to the Fibonacci sequence.

Example 2 (Lekkerkerker)

For Zeckendorf,

$$E^{Q_n}[\#\text{summands}] = \frac{n}{\phi^2 + 1} + \frac{\phi^2}{5} + o(1),$$

where ϕ is the Golden ratio, $\phi = \frac{1+\sqrt{5}}{2}$.

LLN and CLT for additive functionals

Proposition 3

1. *LLN.* $S_n/n \to E_{\pi Q}\, g$ *in* Q_n*-probability.*
2. *CLT.* $Q_n(\frac{S_n - nE_{\pi Q}\, g}{\sqrt{n}} \in \cdot) \Rightarrow N(0, \sigma^2)$, *where* $\sigma^2 = E_{\pi Q}\tilde{g}(2Q^{\#} - I)\tilde{g}$, *and* $\tilde{g} = g - E_{\pi Q}\, g$.

Previously obtained by detailed combinatorial analysis of decompositions.

Example 3

For Zeckendorf

$$\frac{\#\ \text{summands} - n/(\phi^2 + 1)}{\sqrt{n}} \Rightarrow N(0, \frac{\sqrt{5}}{25}).$$

Gaps

Recall the identification:

$$N = \sum_{j=1}^{n} X_j G_{n+1-j}.$$

Now if $K = K(N) = \#\{j : X_j > 0\}$ and $n = j_1 > j_2 > \cdots > j_K$ are the indices $X_1 = R_1, \ldots, R_K$ are the respective digits, then

$$N = \underbrace{G_{j_1} + \cdots + G_{j_1}}_{R_1 \text{ times}} + \underbrace{G_{j_2} + \cdots + G_{j_2}}_{R_2 \text{ times}} + \cdots + \underbrace{G_{j_K} + \cdots + G_{j_K}}_{R_K \text{ times}}$$

This leads the definition of *gaps* (in indices)

1. Repetition of a summand G_j exactly R_j times yields $j-1$ gaps of size 0.
2. Change of index from j_k to j_{k+1} yields a gap of size $j_k - j_{k+1} \geq 1$.

Let

$$N_n(k) = \text{number of gaps of size } k,$$

and define the empirical gap-size distribution:

$$\mu_n(k) = \frac{N_n(k)}{\sum_k N_n(k)}$$

Also let

$$M_n = \max\{k : N_n(k) > 0\}.$$

Gap distribution

Let π_1^Q marginal of the first component (X) under π^Q, M denote the corresponding mean, and let λ_C denote the largest characteristic root for the recurrence relation,

$$\lambda_C = \lim_{n\to\infty} G_{n+1}/G_n.$$

$n_k \nearrow \infty$ satisfies the *spacing condition* with respect to α and $q \in (0,1)$ if

$$\liminf_{k\to\infty} \inf_{z\in\mathbb{Z}_+} \left| \frac{\ln(n_k\alpha)}{\ln \frac{1}{q}} - z \right| > 0.$$

This means that $n_k\alpha$ is eventually uniformly far from integer powers of $1/q$.

Proposition 4

1. *Let* $k \in \mathbb{Z}_+$. *Then* $\mu_n(k) \to \begin{cases} 1 - \frac{1-\pi_1^Q(0)}{M} & k = 0; \\ \frac{1-\pi_1^Q(0)-\pi_1^Q(0)(1-\lambda_C^{-1})}{M} & k = 1; \\ \frac{\pi_1^Q(0)(\lambda_C-1)^2\lambda_C^{-k}}{M} & k \geq 2. \end{cases}$

   where the convergence is in Q_n*-probability.*
2. *Let* $k \in \mathbb{Z}$. *Then*

   $$Q_n(M_n \leq \lfloor \frac{\ln n\alpha}{\ln \lambda_C} \rfloor + k) \to e^{-\lambda_C^{-(k-2)}},$$

   along sequences satisfying spacing with respect to $\alpha = \pi_1(0)(1-\frac{1}{\lambda_C})$ *and* $q = \frac{1}{\lambda_C}$.

No surprises: Maximal gap distribution is asymptotically same as maximum of the average number of gaps of independent geometric random variables with parameter $1 - \frac{1}{\lambda_C}$.

Gap Distribution - Example

Example 4

For Zeckendorf

1. $\mu_n(k) \underset{n\to\infty}{\to} \begin{cases} 0 & k = 0, 1; \\ \phi^{-k} & k \geq 2. \end{cases}$

2. $$Q_n(M_n \leq \lfloor \frac{\ln n - \ln(\phi^2+1)}{\ln \phi} \rfloor + k) \to e^{-\phi^{-(k-2)}},$$

   along sequences satisfying spacing with respect to $\alpha = \frac{1}{\phi^2+1}$ and $q = \frac{1}{\phi}$.

Thank you !

Bibliography I

H. Alpert, *Differences of multiple Fibonacci numbers*, Integers: Electronic Journal of Combinatorial Number Theory **9** (2009), 745–749.

O. Beckwith, A. Bower, L. Gaudet, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *The Average Gap Distribution for Generalized Zeckendorf Decompositions*, Fibonacci Quarterly **51** (2013), 13–27.

Iddo Ben-Ari and Michael Neumann, *Probabilistic approach to Perron root, the group inverse, and applications*, Linear Multilinear Algebra **60** (2012), no. 1, 39–63.

A. Bower, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *The Distribution of Gaps between Summands in Generalized Zeckendorf Decompositions* (and an appendix on *Extensions to Initial Segments* with Iddo Ben-Ari), Journal of Combinatorial Theory, Series A **135** (2015), 130–160.

E. Burger, D. C. Clyde, C. H. Colbert, G. H. Shin and Z. Wang, *A Generalization of a Theorem of Lekkerkerker to Ostrowski's Decomposition of Natural Numbers*, Acta Arith. **153** (2012), 217–249.

L. H. Y. Chen, H.-K. Hwang and V. Zacharovas, *Distribution of the sum-of-digits function of random integers: A survey*, Probability Surveys **11** (2014), 177–237.

Bibliography II

M. Catral, P. Ford, P. E. Harris, S. J. Miller, D. Nelson, *Generalizing Zeckendorf's Theorem: The Kentucky Sequence*, Proceedings of the Sixteenth International Conference on Fibonacci Numbers and Their Applications, Volume 52, Number 5, pp. 69–91.

M. Catral, P. Ford, P. Harris, S. J. Miller and D. Nelson, *Legal Decompositions Arising from Non-positive Linear Recurrences*, preprint.

M. Catral, P. Ford, P. E. Harris, S. J. Miller and D. Nelson, *New Behavior in Legal Decompositions Arising from Non-positive Linear Recurrences*, preprint.

D. E. Daykin, *Representation of Natural Numbers as Sums of Generalized Fibonacci Numbers*, J. London Mathematical Society **35** (1960), 143–160.

P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, D. Moon and U. Varma, *Generalizing Zeckendorf's Theorem to f-decompositions*, Journal of Number Theory **141** (2014), 136–158.

P. Demontigny, T. Do, A. Kulkarni, S. J. Miller and U. Varma, *A Generalization of Fibonacci Far-Difference Representations and Gaussian Behavior*, to appear in the Fibonacci Quarterly.
http://arxiv.org/pdf/1309.5600v2.

R. Dorward, P. Ford, E. Fourakis, P. E. Harris, S. J. Miller, E. Palsson and H. Paugh, *A Generalization of Zeckendorf's Theorem via Circumscribed m-gons*, preprint.

Bibliography III

M. Drmota and J. Gajdosik, *The distribution of the sum-of-digits function*, J. Théor. Nombrés Bordeaux **10** (1998), no. 1, 17–32.

J.-M. Dumont, *Formules sommatoires et systèmes de numération lies aux substitutions*, Séminaires de Théorie des Nombres, Bordeaux (1987/88).

J.-M. Dumont and A. Thomas, *Systèmes de numération et fonctions fractales relatifs aux substitutions*, Theor. Comput. Sci. **65** (1989), 153–169.

J.-M. Dumont and A. Thomas, *Gaussian asymptotic properties of the sum-of-digits function*, J. Number Th. **62** (1997), 19–38.

Rick Durrett, *Probability: theory and examples*, fourth ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.

P. Filipponi, P. J. Grabner, I. Nemes, A. Pethö, and R. F. Tichy, *Corrigendum to: "Generalized Zeckendorf expansions"*, Appl. Math. Lett., **7** (1994), no. 6, 25–26.

A. S. Fraenkel, *Systems of numeration*, Amer. Math. Monthly **92** (1985), no. 2, 105–114.

P. J. Grabner and M. Rigo, *Distribution of additive functions with respect to numeration systems on regular languages*, Theory Comput. Syst. **40** (2007), 205–223.

Bibliography IV

P. J. Grabner and R. F. Tichy, *Contributions to digit expansions with respect to linear recurrences*, J. Number Theory **36** (1990), no. 2, 160–169.

P. J. Grabner, R. F. Tichy, I. Nemes, and A. Pethö, *Generalized Zeckendorf expansions*, Appl. Math. Lett. **7** (1994), no. 2, 25–28.

N. Hamlin, *Representing Positive Integers as a Sum of Linear Recurrence Sequences*, Abstracts of Talks, Fourteenth International Conference on Fibonacci Numbers and Their Applications (2010), pages 2–3.

V. E. Hoggatt, *Generalized Zeckendorf theorem*, Fibonacci Quarterly **10** (1972), no. 1 (special issue on representations), pages 89–93.

M. Hollander, *Greedy numeration systems and regularity*, Theory Comput. Syst. **31** (1998), no. 2, 111–133.

T. J. Keller, *Generalizations of Zeckendorf's theorem*, Fibonacci Quarterly **10** (1972), no. 1 (special issue on representations), pages 95–102.

M. Lamberger and J. M. Thuswaldner, *Distribution properties of digital expansions arising from linear recurrences*, Math. Slovaca **53** (2003), no. 1, 1–20.

P. B. A. Lecomte and M. Rigo, *Numeration systems on a regular language*, Theory Comput. Syst. **34** (2001), no. 1, 27–44.

Bibliography V

T. Lengyel, *A Counting Based Proof of the Generalized Zeckendorf's Theorem*, Fibonacci Quarterly **44** (2006), no. 4, 324–325.

C. G. Lekkerkerker, *Voorstelling van natuurlyke getallen door een som van getallen van Fibonacci*, Simon Stevin **29** (1951-1952), 190–195.

M. Koloğlu, G. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.

T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley-Interscience, New York, 2001.

E. Manstavičius, *Probabilistic theory of additive functions related to systems of numeration*, New trends in probability and statistics, Vol. **4** (Palanga, 1996), VSP, Utrecht, 1997, pp. 413–429.

Michael Maxwell and Michael Woodroofe, *Central limit theorems for additive functionals of Markov chains*, Ann. Probab. **28** (2000), no. 2, 713–724.

S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, Journal of Combinatorial Theory, Series A **119** (2012), no. 7, 1398–1413.

S. J. Miller and Y. Wang, *Gaussian Behavior in Generalized Zeckendorf Decompositions*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 159–173.

Bibliography VI

Jeffrey Shallit, *Numeration systems, linear recurrences, and regular sets*, Inform. and Comput. **113** (1994), no. 2, 331–347.

W. Steiner, *Parry expansions of polynomial sequences*, Integers **2** (2002), Paper A14.

W. Steiner, *The Joint Distribution of Greedy and Lazy Fibonacci Expansions*, Fibonacci Quarterly **43** (2005), 60–69.

K. B. Stolarsky, *Power and exponential sums of digital sums related to binomial coefficient parity*, SIAM J. Appl. Math. **32** (1977), no. 4, 717–730.

E. Zeckendorf, *Représentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas*, Bulletin de la Société Royale des Sciences de Liége **41** (1972), pages 179–182.