Theoretical Computer Science 537 (2014) 105–124

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Towards a generic view of primality through multiset decompositions of natural numbers

Paul Tarau

Dept. of Computer Science and Engineering, University of North Texas, USA

ARTICLE INFO

Article history:
Received 26 February 2012
Received in revised form 24 March 2014
Accepted 18 April 2014

Keywords:
Bijective datatype transformations
Multiset encodings and prime numbers
Möbius and Mertens functions
Experimental mathematics and functional programming
Automorphisms of $\mathbb{N}$

ABSTRACT

Factorization results in multisets of primes and this mapping can be turned into a bijection between multisets of natural numbers and natural numbers. At the same time, simpler and more efficient bijections exist that share some interesting properties with the bijection derived from factorization.

This paper describes mechanisms to emulate properties of *prime numbers* through isomorphisms connecting them to computationally simpler representations involving bijections from natural numbers to *multisets* of natural numbers.

As a result, interesting automorphisms of $\mathbb{N}$ and emulations of the rad, Möbius and Mertens functions emerge in the world of our much simpler multiset representations.

Finally we generalize the multiset decomposition mechanism derived from the Diophantine equation $2^x(2y+1)=z$ and study emulations of the behavior of the ω function that counts the number of distinct prime factors of a number as well as the equivalent of *b-smoothness* (a property characterizing natural numbers having all their prime factors smaller than b).

The paper is organized as a self-contained *literate Haskell program*. The code extracted from the paper is available as a standalone program at http://logic.cse.unt.edu/tarau/research/2012/jprimes.hs.

1. Introduction

This paper is an extended version of [1], with new material concentrated in Sections 2, 6 and 7.

Paul Erdös's statement, shortly before he died, that "*It will be another million years at least, before we understand the primes*" is indicative of the difficulty of the field as perceived by number theorists. The growing number of conjectures [2] and the large number of still unsolved problems involving prime numbers [3] shows that the field is still open to surprises, after thousands of years of effort by some of the brightest human minds.

Interestingly, some significant progress on prime numbers correlates with unexpected paradigm shifts, the prototypical example being Riemann's paper [4] connecting primality and complex analysis, all evolving around the still unsolved *Riemann Hypothesis* [5–8]. The genuine difficulty of the problems and the seemingly deeper and deeper connections with fields ranging from cryptography to quantum physics suggest that unusual venues might be worth trying out.

A number of breakthroughs in various sciences involve small scale emulation of complex phenomena. Common sense analogies thrive on our ability to extrapolate from simpler (or, at least, more frequently occurring and better understood) mechanisms to infer surprising properties in a more distant ontology.

E-mail address: tarau@cs.unt.edu.

http://dx.doi.org/10.1016/j.tcs.2014.04.025
0304-3975/

Prime numbers exhibit a number of fundamental properties of natural phenomena and human artifacts in an unusually pure form. For instance, *reversibility* is present as the ability to recover the operands of a product of distinct primes. This relates to the information-theoretical view of multiplication [9] and it suggests investigating connections between combinatorial properties of multisets and operations on multisets and multiplicative number theory.

With such methodological hints in mind, this paper will explore mappings between multiset encodings and prime numbers. It is based on our *data type transformation framework* connecting most of the fundamental data types used in computer science with a *groupoid of isomorphisms* [10–12].

The paper is organized as follows. In Section 2 we generalize the mechanism described in our ICTAC'11 paper to a family of multiset decompositions. Section 3 revisits the well-known connection between multisets and primes using a variant of Gödel's encoding [13]. Section 4 describes our computationally efficient multiset encoding, seen as an instance of the general mechanism described in Section 2. Based on this encoding, Section 5 explores the analogy between multiset decompositions and factoring and describes a multiplicative monoid structure on multisets that "emulates" properties of the monoid induced by ordinary multiplication as well as generic definitions in terms of multiset encodings of the `rad`, Möbius and Mertens functions.

In Section 6 we introduce a generic framework for multiset-based emulations of primality. Using instances of this framework, in Section 7 we study emulations of the behavior of the ω function (counting the number of distinct prime factors of a number) and the distribution of *smooth* numbers (numbers that are the product of only "small" prime factors).

Section 8 describes automorphisms of $\mathbb{N}$ derived from alternative multiset encodings. Section 9 overviews some related work and Section 10 concludes the paper.

We organize our literate programming code as a Haskell module, relying only on the List library module.

2. A generalized multiset decomposition mechanism

We will start by describing a mechanism for deriving bijections between $\mathbb{N}$ and finite sequences of natural numbers from one-solution Diophantine equations. Let's observe that

Proposition 1. $\forall z \in \mathbb{N} - \{0\}$ *the Diophantine equation*

$$2^x(2y+1) = z \tag{1}$$

has exactly one solution $x, y \in \mathbb{N}$.

This follows immediately from the unicity of the decomposition of a natural number as a multiset of prime factors.

We will generalize this observation to obtain a family of bijections from finite sequences to natural numbers by choosing an arbitrary base b instead of 2.

Definition 1. Given a number $n \in \mathbb{N},\ n > 1$, the n-adic evaluation of a natural number m is the largest exponent k of n, such that n^k divides m. It is denoted $\nu_n(m)$.

Note that the solution x of Eq. (1) is actually $\nu_2(z)$. This suggest deriving a similar Diophantine equation for an arbitrary n-adic valuation using the well known division theorem that, when restricted to $\mathbb{N}$, states that for every pair $y, b \in \mathbb{N}$ where $b > 0$, there exist unique $q, m \in \mathbb{N}$ such that $y = qb + m$ and $0 \le m < b$.

We start by observing that the following holds:

Proposition 2. *For all* $b, y \in \mathbb{N},\ b > 1$ *such that* b *does not divide* y, *let* m *be the remainder of the division of* y *by* b, (q, m) *be the unique pair such that* $b > m > 0,\ q \ge 0,\ y = bq + m$. *Then there's a unique pair* (y', m'), $b - 1 > m' \ge 0$ *such that* $y' = (b-1)q + m'$ *and the function associating* (y', m') *to* (y, m) *is a bijection.*

Proof. $y = bq + m, b > m > 0$ can be rewritten as $y - q - 1 = bq - q + m - 1,\ b > m > 0$, or equivalently $y - q - 1 = (b-1)q + (m-1),\ b > m > 0$ from where, given the existence and unicity of (q, m), it follows that setting $y' = y - q - 1$ and $m' = m - 1$ ensures the existence and unicity of y' and m' such that $y' = (b-1)q + m'$ and $b - 1 > m' > 0$. We can therefore define a function f that transforms a pair (y, m) such that $y = bq + m$ with $b > m > 0$ into a pair (y', m') such that $y' = q(b-1) + m'$ with $b - 1 > m' \ge 0$. Note that the transformation works also in the opposite direction with $y' = y - q - 1$ giving $y = y' + q + 1$, and $m' = m - 1$ giving $m = m' + 1$, and therefore f is a bijection. □

Proposition 3. $\forall b \in \mathbb{N}, b > 1,\ \forall z \in \mathbb{N},\ z > 0$ *if* (q, m') *are the quotient and the remainder of the division of* y' *by* $b - 1$, *the system of Diophantine equations and inequations*

$$b^x\left(y' + q + 1\right) = z \tag{2}$$

$$y' = (b-1)q + m' \tag{3}$$

$$b > m' \ge 0 \tag{4}$$