A NEW PROOF OF THE PROUHET-TARRY-ESCOTT PROBLEM

Hieu D. Nguyen
Department of Mathematics, Rowan University, Glassboro, New Jersey
nguyen@rowan.edu

Received: 1/6/15, Revised: 12/19/15, Accepted: 1/2/16, Published: 1/21/16

Abstract

The famous Prouhet-Tarry-Escott problem seeks collections of mutually disjoint sets of non-negative integers that have equal sums of like powers. In this paper we present a new proof of the solution to this problem by deriving a generalization of the product generating function formula for the classical Prouhet-Thue-Morse sequence.

1. Introduction

The well-known Prouhet-Tarry-Escott (PTE) problem [3, 9] seeks $p \geq 2$ sets of non-negative integers S_0, S_1, ..., S_{p-1} that have equal sums of like powers (ESP) up to degree $M \geq 1$, i.e.,

$$\sum_{n \in S_0} n^m = \sum_{n \in S_1} n^m = \cdots = \sum_{n \in S_{p-1}} n^m$$

for all $m = 0, 1, \ldots, M$. In 1851, E. Prouhet [6] announced a solution; however, a full proof of the solution was never published. Prouhet's solution involved partitioning the first p^{M+1} non-negative integers into the sets $S_0, S_1, \ldots, S_{p-1}$ according to the rule

$$n \in S_{u_p(n)}.$$

Here, $u_p(n)$ is the generalized Prouhet-Thue-Morse sequence defined by computing the residue of the sum of digits of n (base p):

$$u_p(n) = \sum_{j=0}^{d} n_j \mod p,$$

where $n = n_d p^d + \cdots + n_0 p^0$ is the base-p expansion of n. When $p = 2$, $u(n) := u_2(n)$ generates the classical Prouhet-Thue-Morse sequence: $0, 1, 1, 0, 1, 0, 0, 1, \ldots$. For

example, the two sets

$$S_0 = \{0, 3, 5, 6, 9, 10, 12, 15\}$$
$$S_1 = \{1, 2, 4, 7, 8, 11, 13, 14\},$$

defined by the assignment $n \in S_{u(n)}$, solve the PTE problem with $p = 2$ and $M = 3$ since

$$\begin{aligned}
8 &= 0^0 + 3^0 + 5^0 + 6^0 + 9^0 + 10^0 + 12^0 + 15^0 \\
&= 1^0 + 2^0 + 4^0 + 7^0 + 8^0 + 11^0 + 13^0 + 14^0 \\
60 &= 0 + 3 + 5 + 6 + 9 + 10 + 12 + 15 \\
&= 1 + 2 + 4 + 7 + 8 + 11 + 13 + 14 \\
620 &= 0^2 + 3^2 + 5^2 + 6^2 + 9^2 + 10^2 + 12^2 + 15^2 \\
&= 1^2 + 2^2 + 4^2 + 7^2 + 8^2 + 11^2 + 13^2 + 14^2 \\
7200 &= 0^3 + 3^3 + 5^3 + 6^3 + 9^3 + 10^3 + 12^3 + 15^3 \\
&= 1^3 + 2^3 + 4^3 + 7^3 + 8^3 + 11^3 + 13^3 + 14^3,
\end{aligned}$$

where we define $0^0 = 1$.

The first published proof of Prouhet's solution was given by D. H. Lehmer [4], who presented a more general construction of ESPs beyond those described by Prouhet's solution. This is achieved by considering products of polynomials whose coefficients are roots of unity. In particular, Lehmer defined

$$F(\theta) = \prod_{m=0}^{M} (1 + \omega e^{\mu_m \theta} + \omega^2 e^{2\mu_m \theta} + \cdots + \omega^{p-1} e^{(p-1)\mu_m \theta}), \tag{1}$$

where ω is a p-th root of unity and $\mu_0, \ldots, \mu_M$ are arbitrary positive integers. It is clear that $F(x)$ has a zero at $x = 0$ of order $M + 1$ so that its derivative vanishes up to order M, i.e., $F^{(m)}(0) = 0$ for $m = 0, 1, \ldots, M$. On the other hand, Lehmer expanded $F(x)$ to obtain

$$F(\theta) = \sum_{a_0, \ldots, a_M} \omega^{a_0 + \cdots + a_M} e^{(a_1\mu_0 + \cdots + a_M\mu_M)\theta}, \tag{2}$$

where the indices $a_0, \ldots, a_M$ take on all integers from 0 to $p - 1$. Since

$$F^{(m)}(0) = \sum_{a_0, \ldots, a_M} \omega^{a_0 + \cdots + a_M} (a_0\mu_0 + \cdots + a_M\mu_M)^m,$$

Lehmer proved using linear algebra that

$$\sum_{n \in S_0} n^m = \sum_{n \in S_1} n^m = \cdots = \sum_{n \in S_{p-1}} n^m$$

by assigning $n = a_0\mu_0 + \cdots + a_M\mu_M \in S_k$ whenever $a_0 + \cdots + a_M = k \bmod p$. This solves the PTE problem by setting $\mu_m = p^m$ for all $m = 0, 1, \ldots, M$. Other proofs of Prouhet's solution have been given by E. M. Wright [8] using multinomial expansion and J. B. Roberts [7] using difference operators (see also [9]).

Observe that in the aforementioned case where $\mu_m = p^m$ for all $m = 0, \ldots, M$, equating (1) with (2), together with the substitution $x = e^\theta$, yields the product generating function formula

$$\prod_{m=0}^{M}(1 + \omega x^{p^m} + \omega^2 x^{2p^m} + \cdots + \omega^{p-1}x^{(p-1)p^m}) = \sum_{n=0}^{p^{M+1}-1} \omega^{u_p(n)}x^n. \tag{3}$$

For $p = 2$, equation (3) reduces to the classical product generating function formula for the PTM sequence $u(n)$ (see [1, 2]):

$$\prod_{m=0}^{N}(1 - x^{2^m}) = \sum_{n=0}^{2^{N+1}-1} (-1)^{u(n)}x^n. \tag{4}$$

In this paper, we present a new proof of Prouhet's solution by generalizing (3) to polynomials whose coefficients sum to zero while preserving the form of (4). This is achieved by observing that the key ingredient in the proof of (3) relies on the property that all p-th roots of unity sum to zero, namely,

$$\omega^0 + \omega^1 + \cdots + \omega^{p-1} = 0,$$

where ω is a primitive p-th root of unity. To this end, let $A = (a_0, a_1, \ldots, a_{p-1})$ be a vector consisting of p arbitrary complex values that sum to zero, i.e.,

$$a_0 + a_1 + \cdots + a_{p-1} = 0.$$

For any positive integer N, we define $F_N(x; A)$ to be the polynomial of degree $p^N - 1$ whose coefficients belong to A, and repeat according to $u_p(n)$, i.e.,

$$F_N(x; A) = \sum_{n=0}^{p^N-1} a_{u_p(n)}x^n. \tag{5}$$

In Theorem 1 we prove that there exists a polynomial $P_N(x)$ such that

$$F_N(x; A) = P_N(x)\prod_{m=0}^{N-1}(1 - x^{p^m}). \tag{6}$$

For example, if $p = 3$ so that $a_0 + a_1 + a_2 = 0$, then (6) becomes

$$a_0 + a_1x + a_2x^2 = (a_0 + (a_0 + a_1)x)(1 - x)$$

and

$$\begin{aligned} &a_0 + a_1x + a_2x^2 + a_1x^3 + a_2x^4 + a_0x^5 + a_2x^6 + a_0x^7 + a_1x^8 \\ &\quad = (a_0 + (a_0 + a_1)x + (a_0 + a_1)x^3 + a_1x^4)(1 - x)(1 - x^3) \end{aligned}$$

for $N = 1$ and $N = 2$, respectively. In the case where $p = 2$, $a_0 = 1$, and $a_1 = -1$, then $P_N(x) = 1$ for all N and therefore (6) reduces to (4).

Equation (6) is useful in that it allows us to establish that the polynomial $F_N(x, A)$ has a zero of order N at $x = 1$. Prouhet's solution now follows easily by setting $N = M + 1$ and differentiating $F_N(x; A)$ up to order m as demonstrated in Theorem 2.

2. Proof of the Prouhet-Tarry-Escott Problem

Let $p \geq 2$ be a fixed integer. We begin with a lemma that describes a recurrence for $F_N(x; A)$ and whose proof follows from the fact that

$$u_p(n + kp^m) = (u_p(n) + k)_p \quad (0 \leq n < p^m, 0 \leq k < p), \tag{7}$$

where we define $(n)_p = n \bmod p$. Moreover, let A_k denote the k-th left cyclic shift of the elements of A, i.e.,

$$A_k = (a_{(k)_p}, a_{(k+1)_p}, \ldots, a_{(p-1+k)_p}).$$

Lemma 1. *For any integer $N > 1$, we have*

$$F_N(x; A) = F_{N-1}(x; A_0) + x^{p^{N-1}} F_{N-1}(x; A_1) + \cdots + x^{(p-1)p^{N-1}} F_{N-1}(x; A_{p-1}). \tag{8}$$

Proof. We first decompose $F_N(x; A)$ as follows:

$$\begin{aligned} F_N(x; A) &= \sum_{n=0}^{p^N-1} a_{u_p(n)} x^n \\ &= \sum_{n=0}^{p^{N-1}-1} a_{u_p(n)} x^n + \sum_{n=p^{N-1}}^{2p^{N-1}-1} a_{u_p(n)} x^n + \ldots \\ &\quad + \sum_{n=(p-1)p^{N-1}}^{p^N-1} a_{u_p(n)} x^n \\ &= \sum_{n=0}^{p^{N-1}-1} a_{u_p(n)} x^n + x^{p^{N-1}} \sum_{n=0}^{p^{N-1}-1} a_{u_p(n+p^{N-1})} x^n + \ldots \\ &\quad + x^{(p-1)p^{N-1}} \sum_{n=0}^{p^{N-1}-1} a_{u_p(n+(p-1)p^{N-1})} x^n. \end{aligned}$$

It follows from (7) that

$$F_N(x;A) = \sum_{n=0}^{p^{N-1}-1} a_{u_p(n)}x^n + x^{p^{N-1}} \sum_{n=0}^{p^{N-1}-1} a_{(u_p(n)+1)_p}x^n + \dots + x^{(p-1)p^{N-1}} \sum_{n=0}^{p^{N-1}-1} a_{(u_p(n)+p-1)_p}x^n.$$

Hence,

$$F_N(x;A) = F_{N-1}(x;A_0) + x^{p^{N-1}}F_{N-1}(x;A_1) + \cdots + x^{(p-1)p^{N-1}}F_{N-1}(x;A_{p-1})$$

as desired. □

For example, let $p = 3$ and $A = (a_0, a_1, a_2)$. Then

$$\begin{aligned} F_1(x;A) &= a_0 + a_1x + a_2x^2 \\ F_2(x;A) &= a_0 + a_1x + a_2x^2 + a_1x^3 + a_2x^4 + a_0x^5 + a_2x^6 + a_0x^7 + a_1x^8 \\ &= F_1(x;A_0) + x^3F_1(x;A_1) + x^6F_1(x;A_2). \end{aligned}$$

Next, define a recursive sequence of vectors C_N consisting of unknown constants as follows:

$$C_1 = (c_0, \dots, c_{p-2}),$$

and for $N > 1$,

$$C_N = C_{N-1}(0)\#C_{N-1}(1)\#\dots\#C_{N-1}(p-2) \tag{9}$$

where # denotes concatenation of vectors and

$$C_{N-1}(k) = (c_{j+kp^{N-1}} : c_j \in C_{N-1})$$

for $k = 0, 1, \dots, p-2$. For example, if $p = 3$, then

$$\begin{aligned} C_1 &= (c_0, c_1) \\ C_2 &= C_1(0)\#C_1(1) = (c_0, c_1, c_3, c_4) \\ C_3 &= C_2(0)\#C_2(1) = (c_0, c_1, c_3, c_4, c_9, c_{10}, c_{12}, c_{13}). \end{aligned}$$

Note that if $p = 2$, then $C_N = (c_0)$ for all $N \geq 1$.

Moreover, define a sequence of polynomials $P_N(x; C_N)$ recursively as follows:

$$P_1(x;C_1) = c_0 + c_1x + \cdots + c_{p-2}x^{p-2},$$

and for $N > 1$,

$$P_N(x;C_N) = P_{N-1}(x;C_{N-1}(0)) + x^{p^{N-1}}P_{N-1}(x;C_{N-1}(1)) + \dots + x^{(p-2)p^{N-1}}P_{N-1}(x;C_{N-1}(p-2)). \tag{10}$$

We are now ready to prove that $F_N(x;A)$ has the following factorization.

Theorem 1. *Let N be a positive integer. There exists a polynomial $P_N(x;C_N)$ such that*

$$F_N(x;A) = P_N(x;C_N)\prod_{m=0}^{N-1}(1-x^{p^m}). \tag{11}$$

Proof. We prove (11) by induction. First, define $Q_N(x) = \prod_{m=0}^{N-1}(1-x^{p^m})$ by $Q_1(x) = (1-x)$ and for $N > 1$,

$$Q_N(x) = Q_{N-1}(x)(1-x^{p^{N-1}}). \tag{12}$$

To establish the base case $N = 1$, we expand $F_1(x;A) = P_1(x;C_1)Q_1(x)$ to obtain

$$a_0 + a_1x + \cdots + a_{p-1}x^{p-1} = c_0 + (c_1 - c_0)x + \cdots + (c_{p-2} - c_{p-1})x^{p-2} - c_{p-2}x^{p-1}.$$

Then equating coefficients yields the system of equations

$$\begin{aligned} c_0 &= a_0 \\ c_1 - c_0 &= a_1 \\ &\vdots \\ c_{p-2} - c_{p-1} &= a_{p-2} \\ -c_{p-2} &= a_{p-1}. \end{aligned}$$

Since $a_0 + a_1 + \cdots + a_{p-1} = 0$, this system is consistent with the solution $c_m = \sum_{k=0}^{m} a_k$ for $m = 0, 1, \ldots, p-2$ where $c_{p-2} = a_0 + \cdots + a_{p-2} = -a_{p-1}$. Thus, $P_1(x;C_1)$ is given by

$$P_1(x;C_1) = \sum_{m=0}^{p-2}\left(\sum_{k=0}^{m} a_k\right)x^m.$$

Note that if $p = 2$, then $P_1(x;C_1) = a_0$.

Next, assume there exists a polynomial $P_{N-1}(x;C_{N-1})$ that solves

$$F_{N-1}(x;A) = P_{N-1}(x;C_{N-1})Q_{N-1}(x).$$

To prove that there exists a solution $P_N(x;C_N)$ for

$$F_N(x;A) = P_N(x;C_N)Q_N(x), \tag{13}$$

we expand (13) using recurrences (8), (10), and (12):

$$\sum_{k=0}^{p-1} x^{kp^{N-1}}F_{N-1}(x;A_k) = \left(\sum_{k=0}^{p-2} x^{kp^{N-1}}P_{N-1}(x;C_{N-1}(k))\right)Q_{N-1}(x)(1-x^{p^{N-1}}).$$

We then equate coefficients corresponding to the terms $x^{kp^{N-1}}$. This yields the system of equations

$$\begin{aligned}
F_{N-1}(x;A_0) &= P_{N-1}(x;C_{N-1}(0))Q_{N-1}(x)\\
F_{N-1}(x;A_1) &= (P_{N-1}(x;C_{N-1}(1)) - P_{N-1}(x;C_{N-1}(0)))Q_{N-1}(x)\\
&\vdots\\
F_{N-1}(x;A_{p-2}) &= (P_{N-1}(x;C_{N-1}(p-2)) - P_{N-1}(x;C_{N-1}(p-3)))Q_{N-1}(x)\\
F_{N-1}(x;A_{p-1}) &= -P_{N-1}(x;C_{N-1}(p-2))Q_{N-1}(x).
\end{aligned}$$

Now, each equation above corresponding to $F_{N-1}(x;A_k)$ for $k = 1,\dots,p-2$ can be replaced by one obtained by summing all equations up to k, namely

$$F_{N-1}(x;B_k) = P_{N-1}(x;C_{N-1}(k))Q_{N-1}(x)$$

where $B_k = A_0+\cdots+A_k$ is defined by vector summation. This yields the equivalent system of equations

$$\begin{aligned}
F_{N-1}(x;B_0) &= P_{N-1}(x;C_{N-1}(0))Q_{N-1}(x)\\
F_{N-1}(x;B_1) &= P_{N-1}(x;C_{N-1}(1))Q_{N-1}(x)\\
&\vdots\\
F_{N-1}(x;B_{p-2}) &= P_{N-1}(x;C_{N-1}(p-2))Q_{N-1}(x)\\
F_{N-1}(x;A_{p-1}) &= -P_{N-1}(x;C_{N-1}(p-2))Q_{N-1}(x).
\end{aligned}$$

From our inductive assumption, each of the equations above corresponding to $F_{N-1}(x;B_k)$ has a solution in $C_{N-1}(k)$. Moreover, the last equation corresponding to $F_{N-1}(x;A_{p-1})$ is equivalent to the equation corresponding to $F_{N-1}(x;B_{p-2})$, since $B_{p-2} = A_0 + \cdots + A_{p-2} = -A_{p-1}$. This proves that (13) has a solution in C_N because of (9). □

We now present our proof of the solution to the Prouhet-Tarry-Escott problem.

Theorem 2 ([6, 4]). *Let M be a positive integer, $L = p^{M+1}$, and $\{S_0, S_1,\dots,S_{p-1}\}$ a partition of $\{0,1,\dots,L-1\}$ defined by the condition*

$$n \in S_{u_p(n)}$$

for $0 \le n \le L-1$. Then $S_0, S_1,\dots,S_{p-1}$ have equal sums of like powers of degree M, i.e.,

$$\sum_{n\in S_0} n^m = \sum_{n\in S_1} n^m = \cdots = \sum_{n\in S_{p-1}} n^m$$

for all $m = 0,1,\dots,M$.

Proof. Define $s_k(m) = \sum_{n \in S_k} n^m$ and $A = (a_0, a_1, \ldots, a_{p-1})$ to be a vector consisting of p arbitrary complex values that sum to zero, i.e., $a_0 + a_1 + \cdots + a_{p-1} = 0$. Set $N = M + 1$ and define $F_N(x; A)$ as in (5). Next, substitute $x = e^\theta$ into $F_N(x; A)$ and compute the m-th derivative of $G_N(\theta) := F_N(e^\theta; A)$ at $\theta = 0$. Then, on the one hand, we have from the standard rules of differentiation that

$$\begin{aligned} G_N^{(m)}(0) &= \sum_{n=0}^{p^N-1} n^m a_{u_p(n)} \\ &= \sum_{n \in S_0} n^m a_0 + \cdots + \sum_{n \in S_{p-1}} n^m a_{p-1} \\ &= a_0 s_0(m) + \cdots + a_{p-1} s_{p-1}(m). \end{aligned}$$

On the other hand, we have from (11) that $G_N(\theta)$ has a zero of order N at $\theta = 0$. It follows that

$$G_N^{(m)}(0) = 0$$

for $m = 0, 1, \ldots, N-1$. Thus,

$$a_0 s_0(m) + \cdots + a_{p-1} s_{p-1}(m) = 0. \tag{14}$$

Now, recall that the values a_0, a_1, $\ldots$, a_{p-1} can be chosen arbitrarily as long as they sum to zero. Therefore, we choose them as follows: for any two distinct nonnegative integers j and k satisfying $0 \le j, k \le p-1$, set $a_j = 1$, $a_k = -1$, and $a_l = 0$ for all $l \ne j, k$. Then (14) reduces to

$$s_j(m) - s_k(m) = 0,$$

or equivalently, $s_j(m) = s_k(m)$. But since this holds for all distinct j and k, we have that

$$s_0(m) = s_1(m) = \cdots = s_{p-1}(m)$$

for $m = 0, 1, \ldots, M$ as desired. □

We conclude by explaining our motivation for studying the polynomials $F_N(x; A)$. In [5], the author and G. E. Coxson showed that these polynomials arise in radar as ambiguity functions of pulse trains generated by complementary codes that repeat according to the Prouhet-Thue-Morse sequence. Prouhet's solution is used to demonstrate that these pulse trains, called complementary PTM pulse trains, are tolerant of Doppler shifts due to a moving target by establishing that their Taylor series coefficients vanish up to order $N - 1$.

Acknowledgment The author wishes to thank the referee for a careful reading of this paper.

References

[1] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in *Sequences and Their Applications*, Proc. SETA'98 (Eds. C. Ding, T. Helleseth, and H. Niederreiter), Springer-Verlag, 1999, 1–16.

[2] P. Borwein and C. Ingalls, The Prouhet-Tarry-Escott problem revisited, *Enseign. Math.* **40** (1994), 3–27.

[3] H. L. Dorwart and O. E. Brown, The Tarry-Escott problem, *Amer. Math. Monthly* **44** (1937), 613–626.

[4] D. H. Lehmer, The Tarry-Escott problem, *Scripta Math.* **13** (1947), 37–41.

[5] H. D. Nguyen and G. E. Coxson, Doppler tolerance, complementary code sets, and the generalized Thue-Morse sequence, arXiv:1406.2076 [cs.IT] (2014).

[6] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *C. R. Math. Acad. Sci. Paris* **33** (1851), 225.

[7] J. B. Roberts, A new proof of a theorem of Lehmer, *Canad. J. Math.* **10** (1958), 191–194.

[8] E. M. Wright, Equal sums of like powers, *Proc. Edinb. Math. Soc. (2)* **8** (1949), 138–142.

[9] E. M. Wright, Prouhet's 1851 solution of the Tarry-Escott problem of 1910, *Amer. Math. Monthly* **102** (1959), 199–210.