ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

Phùng Thị Hương

SỐ XOẮN CỦA DÃY SỐ NGUYÊN

LUẬN VĂN THẠC SĨ TOÁN HỌC

Thái Nguyên, Năm 2014

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

Phùng Thị Hương

SỐ XOẮN CỦA DÃY SỐ NGUYÊN

Chuyên ngành: PHƯƠNG PHÁP TOÁN SƠ CẤP
Mã số: 60460113

LUẬN VĂN THẠC SĨ TOÁN HỌC

Người hướng dẫn khoa học:
GS. TSKH. HÀ HUY KHOÁI

Thái Nguyên, Năm 2014

Lời cảm ơn

Luận văn được thực hiện và hoàn thành tại trường Đại học Khoa học - Đại học Thái Nguyên dưới sự hướng dẫn khoa học của GS. TSKH. Hà Huy Khoái. Qua đây, tác giả xin được gửi lời cám ơn sâu sắc đến thầy giáo, người hướng dẫn khoa học của mình, GS. TSKH. Hà Huy Khoái, người đã đưa ra đề tài và tận tình hướng dẫn trong suốt quá trình nghiên cứu của tác giả.

Đồng thời tác giả cũng chân thành cám ơn các thầy cô giáo trong khoa Toán - Tin trường Đại học Khoa học - Đại học Thái Nguyên, đã tạo mọi điều kiện cho tác giả về tài liệu và thủ tục hành chính để tác giả hoàn thành bản luận văn này.

Tác giả cũng gửi lời cám ơn đến gia đình và các bạn trong lớp Cao học toán K6B, đã động viên và giúp đỡ tác giả trong quá trình học tập và làm luận văn.

Thái Nguyên, tháng 4 năm 2014

Tác giả

Phùng Thị Hương

Mục lục

Mở đầu

Giả thuyết toán về số xoắn (curling number) là một bài toán mở chưa có lời giải của lý thuyết số. Mặc dầu vậy, đã có nhiều nghiên cứu cho những kết quả ủng hộ tính đúng đắn của giả thuyết này.

Luận văn nhằm trình bày một số kết quả rất gần đây về số xoắn, cụ thể là kết quả trong bài báo sau đây:

B. Chaffin, J. P. Linderman, N. J. A. Sloane, Allan R. Wilks. *On Curling Number of Integer Sequences*, Journal of Integer Sequences, Vol. **16** (2013), 236; pp. 1-47.

Bố cục của luận văn như sau:

Chương 1: Giả thuyết số xoắn

Chương 2: Dãy các số 2 và 3

Chương 3: Số dãy nhị phân với số xoắn cho trước

Chương 4: Độ dài đuôi của dãy số 2 và 3

Chương 5: Dãy Gijswijt

Chương 1

Giả thuyết số xoắn

Cho một dãy khác rỗng hữu hạn S các số nguyên, biểu diễn nó bằng $S = XY^k$, trong đó X và Y là dãy các số nguyên và Y^k là lũy thừa với số mũ lớn nhất mà là đuôi (hậu tố) của S: số k này gọi là *số xoắn* của S, ký hiệu là $\mathrm{cn}(S)$. X có thể là dãy rỗng ε; có thể có một số cách chọn dãy Y khác nhau, mặc dù dãy Y ngắn nhất để đạt được k là duy nhất (như ta sẽ thấy trong §3.1, là dãy nguyên thủy).

Ví dụ, nếu $S = 0122122122$ thì ta có thể viết $S = XY^2$, trong đó $X = 01221221$ và $Y = 2$, hay ta có thể viết $S = XY^3$, trong đó $X = 0$ và $Y = 122$. Cách biểu diễn thứ hai sẽ được chọn bởi vì nó có $k = 3$, và vì $k = 4$ không thể xảy ra, do đó số xoắn của dãy S là 3.

Giả thuyết sau được phát biểu bởi nhóm nghiên cứu của van de Bult :

Giả thuyết 1.1. (Giả thuyết số xoắn). *Xuất phát từ dãy số nguyên ban đầu tùy ý S, thác triển nó bằng cách ghép thêm liên tiếp số xoắn của dãy ở mỗi thời điểm, cuối cùng dãy sẽ đạt tới 1.*

Nói cách khác, nếu $S_0 = S$ là một dãy khác rỗng hữu hạn các số nguyên, và ta định nghĩa S_{m+1} là chuỗi ghép

$$S_{m+1} := S_m \, \mathrm{cn}(S_m) \quad \text{với} \quad m \geqslant 0, \tag{1.1}$$

thì khi đó giả thuyết phát biểu rằng với một số giá trị $t \geqslant 0$ ta sẽ có $\mathrm{cn}(S_t) = 1$. Giá trị nhỏ nhất của t này được gọi là *độ dài đuôi* của S_0, ký hiệu là $\tau(S_0)$ (và ta đặt $\tau(S_0) = \infty$ nếu giả thuyết không đúng).

Ví dụ, giả sử xuất phát với $S_0 = 2323$. Bằng cách chọn $X = \varepsilon, Y = 23$, có $S_0 = Y^2$, nên $\mathrm{cn}(S_0) = 2$, và do đó ta có $S_1 = 23232$. Bằng cách chọn $X = 2, Y = 32$ ta có $\mathrm{cn}(S_1) = 2, S_2 = 232322$. Lấy $X = 2323, Y = 2$ ta có $\mathrm{cn}(S_2) = 2, S_3 = 2323222$. Một lần nữa lấy $X = 2323, Y = 2$ ta có $\mathrm{cn}(S_3) = 3, S_4 = 23232223$. Rủi thay, ta không thể viết $S_4 = XY^k$ với $k > 1$, do đó $\mathrm{cn}(S_4) = 1, S_5 = 232322231$, và ta đã đạt tới 1, như dự đoán trong giả thuyết. Với ví dụ này, $\tau(S_0) = 4$. (nếu ta tiếp tục từ dãy này, nó sẽ hợp thành dãy Gijswijt, như trình bày trong §5.)

Một số chứng minh của nhóm van de Bult có thể được rút gọn lại và kết quả mạnh hơn nếu giả thuyết được cho là đúng. Tất cả bằng chứng cho thấy rằng giả thuyết đó đúng, nhưng cho đến nay nó chống lại mọi nỗ lực chứng minh nó.

Ta sẽ trình bày về một số nghiên cứu sâu rộng cho trường hợp dãy ban đầu chứa các số 2 và 3 (mặc dù trong trường hợp đặc biệt này giả thuyết vẫn là bài toán mở).

Trong mục 2 ta nghiên cứu bao xa để một dãy bắt đầu bằng n số 2 và 3 có thể được khai triển trước khi tiến đến 1. Gọi chiều dài cực đại này là $\Omega(n)$. Tức là, $\Omega(n)$ là giá trị cực đại của độ dài đuôi $\tau(S_0)$ của tất cả các dãy S_0 các số 2 và 3 có chiều dài n. Ta xác định $\Omega(n)$ cho tất cả $n \leqslant 48$, và giả sử cho tất cả $n \leqslant 80$ (Bảng 2.1 và Hình 2.1). Dữ liệu cho thấy một số tính chất mà các dãy khởi đầu đẹp đặc biệt sẽ có các tính chất này (Tính chất P2, P3, P4 trong §2.2). Mặc dù ta vẫn chưa tìm ra cấu trúc đại số cho dãy khởi đầu đẹp, Mục 2.3 miêu tả phương pháp mà đôi khi thành công trong việc xây dựng dãy xuất phát với chiều dài lớn hơn. Trong §2.4 miêu tả thuật toán cho phép ta mở rộng nghiên cứu với chiều dài 80. Cũng không nên ngạc nhiên nếu giả thuyết trong trường hợp đặc biệt này hóa ra là một hệ quả được biết đến của mô hình tất yếu của dãy nhị phân dài - được miêu tả ngắn gọn trong §2.5

Mục 3 được dành cho câu hỏi tổ hợp: số $c(n, k)$ của dãy nhị phân với chiều dài n và số xoắn k là số gì? Đây dường như là một bài toán vô cùng khó, và chỉ thành công trong việc liên hệ $c(n, k)$ với hai đại lượng bổ trợ: số các dãy nguyên thủy $p(n, k)$, và số các dãy nguyên thủy và mạnh. Kết quả chính của mục này là công thức cho $c(n, k)$ trong Định lý 3.5 và 3.17. Nhờ đó có thể tính được số xoắn của tất cả dãy nhị phân có chiều dài $n \leqslant 104$.. Số các dãy nhị phân có số xoắn bằng 1, $c(n, 1)$, là vấn đề vô cùng thú vị và được trình bày trong §3.2. Hiệu $d(n, k) := 2c(n-1, k) - c(n, k)$ cho thấy cấu trúc của bảng $c(n, k)$ rõ ràng hơn chính các số $c(n, k)$, và là chủ đề của §3.6.

Trong Mục 4, nghiên cứu số $t(n, i)$ cho dãy có chiều dài n và độ dài đuôi i, trong đó $0 \leqslant i \leqslant \Omega(n)$. Bằng nghiên cứu trực tiếp, đã xác định $t(n, i)$ cho $n \leqslant 48$, mặc dù không dùng bất kỳ phép truy toán (ngoại trừ cho $t(n, 0)$, tương tự như cho $c(n, 1)$) . Số hạng trong mỗi hàng của bảng $t(n, i)$ xảy ra theo từng cụm, ít nhất là với $n \leqslant 48$. Trong §4.1 và §4.2 nghiên cứu thống kê bảng $t(n, i)$, mặc dù vẫn còn rất xa để tìm được mô hình giải thích các cụm. Mục 4.3, 4.4, 4.5 thảo luận một số câu hỏi tổ hợp liên quan đến độ dài đuôi. Nếu dãy xuất phát S_0 đủ dài, có vẻ đáng tin cậy rằng tiền tố của S_0 với 2 hay 3 không làm giảm độ dài đuôi. Nếu một trong các số này làm giảm độ dài đuôi, ta gọi S_0 là *suy giảm* (rotten), nếu cả hai tiền tố 2 và 3 làm giảm độ dài đuôi, chúng ta gọi nó là *suy giảm gấp đôi*. Dãy suy giảm chắc chắn tồn tại, nhưng cho tới độ dài 34 thì không tồn tại dãy suy giảm gấp đôi, và giả định hơn nữa là không tồn tại dãy với chiều dài bất kỳ. Nếu giả thuyết này là đúng, nó có thể giải thích một hiện tượng xác định mà quan sát thấy trong §2.2, và nó kéo theo rằng $\Omega(n+1) \geqslant \Omega(n)$ với mọi n, điều mà chưa biết ở thời điểm hiện tại.

Trong Mục 5 miêu tả ngắn gọn dãy Gijswijt, là dãy xuất phát cho nghiên cứu này.

Ghi chú. Bởi vì dãy xuất phát S có thể là dãy các số nguyên bất kỳ, dùng từ "dãy" thích hợp hơn từ "từ" trong một số bảng chữ cái. Tuy nhiên, ta sử dụng một số thuật ngữ (như "tiền tố", "hậu tố") từ lý thuyết ngôn ngữ chính thức.

Dãy được ký hiệu bằng chữ Latin viết hoa. S^k nghĩa là $SS \cdots S$, trong đó S được lặp k lần. Độ dài của S được ký hiệu bằng $|S|$. ε ký hiệu dãy rỗng.

Tập các dãy được ký hiệu bằng kiểu chữ viết tay (ví dụ $\mathcal{C}(n, k)$) và lực lượng ký hiệu bằng chữ Latin thường tương ứng (ví dụ $c(n, k)$). Chữ Hy Lạp và các chữ Latin thường khác được dùng để ký hiệu số. Ký tự # ký hiệu lực lượng của một tập.

Số xoắn của S ký hiệu bằng $\mathrm{cn}(S)$. Với dãy xuất phát $S_0 := s_1 s_2 \cdots s_n$ có chiều dài n, trong đó s_i là số nguyên bất kỳ, định nghĩa S_{m+1} là phép ghép $S_m \mathrm{cn}(S_m) = s_1 \cdots s_{n+m+1}$ với $m \geqslant 0$. Nếu $\mathrm{cn}(S_t) = 1$ với $t \geqslant 0$, thì ta gọi giá trị t nhỏ nhất là *độ dài đuôi* của S_0, ký hiệu là $\tau(S_0)$, và dãy tương ứng $S^{(e)} := S_t = s_1 \cdots s_{n+t}$ là *mở rộng* của S_0. Nếu không tồn tại t như vậy, thì đặt $\tau(S_0) = \infty$, $S^{(e)} = S_\infty$ (và giả thuyết số xoắn sẽ bị sai).