



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Ordenação de Sequências Finitas por Reversões Usando Conjugações em Grupos de Permutações

José Luiz Correa de Moraes

Dissertação apresentada como requisito parcial  
para conclusão do Mestrado em Informática

Orientador  
Prof. Dr. Mauricio Ayala-Rincón

Brasília  
2012.

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Mestrado em Informática

Coordenador: Prof. Dr. Mylène Christine Queiroz de Farias

Banca examinadora composta por:

Prof. Dr. Mauricio Ayala-Rincón (Orientador) — CIC/UnB  
Prof. Dr. Norai Romeu Rocco — MAT/UnB  
Prof. Dr. Jorge Petrucio Viana — IME/UFF

### **CIP — Catalogação Internacional na Publicação**

Moraes, José Luiz Correa de.

Ordenação de Sequências Finitas por Reversões Usando Conjugações em Grupos de Permutações / José Luiz Correa de Moraes. Brasília : UnB, 2012..

70 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2012..

1. Ordenação de sequência finita por reversões, 2. Distância de reversão entre duas sequências finitas, 3. Reversão em sequência finita usando a conjugação e a operação de um grupo de permutações, 4. Biologia molecular computacional, 5. Rearranjos de genomas.

CDU 2010/0002854

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Definições e notações</b>	<b>7</b>
2.1	Grupos de permutações . . . . .	7
2.2	Operações básicas . . . . .	8
2.3	O problema da distância de reversão e o problema da distância de ordenação	11
2.4	Mais definições e notações . . . . .	13
<b>3</b>	<b>Contextualização</b>	<b>18</b>
3.1	Sequências crescentes . . . . .	19
3.2	Inversões . . . . .	19
3.3	Desconhecimento de padrão . . . . .	19
3.4	Ordenação por pilhas e desconhecimento de padrão . . . . .	20
3.5	Standard Young Tableau (SYT) . . . . .	21
3.6	Uma ordem sobre $S_n$ . . . . .	23
3.7	Recursos muito utilizados . . . . .	25
3.8	Dedução não trivial de um resultado em ordenação . . . . .	29
<b>4</b>	<b>Fundamentos de uma forma diferente de computar para o problema da distância de reversão</b>	<b>36</b>
4.1	Uma solução diferente . . . . .	36
4.2	A motivação para a formalização deste modelo e a equivalência entre este e outros formalismos . . . . .	40
4.2.1	Aplicando este modelo a um caso concreto: representando cromossomos . . . . .	40
4.2.2	Conceitos básicos em rearranjos de genomas por reversões . . . . .	41
4.2.3	Equivalência entre formalismos . . . . .	42
4.3	O Conjugador (para $\iota$ ) e duas simulações . . . . .	45
4.3.1	O conjugador (para $\iota$ ) detém as informações . . . . .	45
4.3.2	O conjugador (para $\iota$ ) e uma simulação independente . . . . .	46
4.3.3	Simulando em paralelo . . . . .	48
4.3.4	Obtendo um novo conjugador (para $\iota$ ) . . . . .	49
4.4	Aplicando este formalismo em algoritmos . . . . .	49
4.4.1	Construindo a permutação-reversão associada a dois elementos de um “ $n$ -ciclo”	50
<b>5</b>	<b>Contribuição, conclusão e trabalhos futuros</b>	<b>53</b>
	<b>Referências</b>	<b>57</b>

# Lista de Figuras

1.1	Componentes do nucleotídeo (ilustrações do aluno, com base em [15]). . . . .	2
1.2	A dupla hélice, conforme estabeleceram James D. Watson e Francis Crick, em 1953, a forma da construção do cromossomo, com duas cadeias ou bandas de nucleotídeos, complementares e invertidas e de orientações $5'-3'$ e $3'-5'$ (a ilustração (a) foi tirada da Internet e a (b) é do aluno, com base em [15]). . . . .	3
1.3	O nucleotídeo, uma unidade constituinte do cromossomo, e a banda $5'-3'$ , a do sentido canônico adotado em documentos e bancos e dados onde apenas uma das duas esteja representada (ilustrações do aluno, com base em [15]). . . . .	3
1.4	Um gene (uma cadeia de nucleotídeos, aqui representados pelas bases) em um cromossomo, os ternos de nucleotídeos com os correspondentes aminoácidos e a sua proteína (uma cadeia de aminoácidos) em uma simulação da codificação de uma proteína a partir de um gene fictício. . . . .	3
2.1	O cromossomo de oito blocos de genes numerados em ordem crescente, representado pela permutação 12345678, antes do evento que reverterá o bloco de entradas consecutivas 4, 5, 6 e 7, que resultará 12376548, a representação do seu homólogo, da Figura 2.2 (ilustração do aluno). . . . .	9
2.2	Como ficará o novo cromossomo 12376548 após o evento da Figura 2.1 (ilustração do aluno). . . . .	9
2.3	O evento da reversão do bloco das entradas consecutivas 4, 5, 6 e 7 do cromossomo 12345678, da Figura 2.1. A hipótese para o evento é que, em um encontro das ligações entre 3 e 4 e entre 7 e 8, por alguma razão, as mesmas são desfeitas e as entradas religadas de forma diferente, unindo 3 com 7 e 4 com 8. O resultado é o cromossomo 12376548, da Figura 2.2 (ilustrações do aluno). . . . .	10
2.4	O cromossomo original, Figura 2.1, e o resultado da reversão, Figura 2.2; em destaque o bloco das entradas consecutivas que foi revertido (ilustração do aluno). . . . .	10
2.6	Ordenação de 3417562 por três transposições. . . . .	10
2.5	Ordenações de $p = 3417562$ (sem sinal) por quatro e de $p$ (com uma de suas possíveis $2^7$ sinalizações) por cinco reversões, respectivamente. . . . .	11
2.7	Ordenação de 3417562 por dois intercâmbios de blocos. . . . .	11
2.8	A transformação da <i>Nicotiana tabacum</i> na <i>Lobelia fervens</i> , duas plantas homólogas, em cinco reversões e a distância de reversão, $d(\alpha, \beta) = 5$ , entre ambas. Este é o número mínimo de reversões (em destaque o bloco das entradas consecutivas que foi revertido em cada uma), necessário para transformar uma planta na outra (ilustração do aluno, com base em [23]). . . . .	12
2.9	Fotos da <i>Nicotiana tabacum</i> e <i>Lobelia fervens</i> (fotos tiradas da Internet). . . . .	13
2.10	A permutação 3412576 tem os descendentes 2 e 6 e os ascendentes 1, 3, 4 e 5. . . . .	13

2.11	A permutação 2415367. . . . .	14
2.12	A permutação 3451267 e os padrões 321 e 2134. . . . .	15
2.13	Um diagrama representando 3561247 com três sequências alternadas. . . . .	16
2.14	Ordenação da permutação (1 3 5 7 6 4 2), de Gollan, onde $n = 7$ é ímpar, com 6 reversões. . . . .	16
3.1	A permutação 3142 e o padrão 231. . . . .	21
3.2	Os passos do algoritmo aplicado à permutação 3142 e que não a ordena porque contém o padrão 231. . . . .	22
3.3	A forma de Ferrer da partição (5, 2, 1) com oito caixas. . . . .	22
3.4	O gancho $H_b$ , de comprimento seis, e todos os comprimentos da forma. . . . .	22
3.5	Um SYT com dez caixas. . . . .	23
3.6	Os comprimentos dos ganchos da forma $F = 2 \times 3$ . . . . .	23
3.7	Os cinco SYT da forma $F = 2 \times 3$ . . . . .	23
3.8	As ordens de Bruhat sobre $S_2$ e $S_3$ . Em $S_3$ , as reduções em 321 se referem às inversões (3, 2) e (3, 1); em 231, (2, 1) e (3, 1); em 312, (3, 1) e (3, 2); em 132, (3, 2); e, em 213, à inversão (2, 1). . . . .	24
3.9	As ordens fracas de Bruhat sobre $S_2$ e $S_3$ . Em $S_3$ , as reduções em 321 se referem às inversões (3, 2) e (2, 1); em 231, à inversão (3, 1); em 312, (3, 1); em 132, (3, 2); e, em 213, (2, 1). . . . .	25
3.10	Os dois SYT da forma em escada (2, 1). . . . .	25
3.11	Ordenação de 3412 com três reversões, sem corte de subsequência ordenada, e com duas, cortando a 34. . . . .	29
3.12	Grafo $\Gamma(4162573)$ , onde $c(\Gamma(4162573)) = 3$ . . . . .	29
3.13	Construção de $G(4312)$ , onde $4312 \in S_4$ . . . . .	30
3.14	O grafo da Figura 3.13 se decompõe em $c(G(4312)) = 3$ ciclos (ver, também, a Observação 3). . . . .	30
3.15	Construção de $G(1234)$ , onde $1234 \in S_4$ é a permutação identidade. . . . .	31
3.16	O grafo $G(1234)$ da Figura 3.15 se decompõe em $c(G(1234)) = 5$ ciclos (ver, também, a Observação 4). . . . .	31
3.17	Construção de $G(4213) \in S_4$ , que contém $c(G(4213)) = 1$ único ciclo (ver, também, a Observação 5). . . . .	31
3.18	Contagem, com base nos grafos $G$ (Observação 3), efetuada a partir do Teorema 6. . . . .	32
3.19	Contagem, com base nos ciclos de arestas pretas dos grafos $G$ (Observação 3), efetuada a partir do Corolário 1. . . . .	33
3.20	Como $z'$ é obtida de $z$ , de acordo com a Proposição 3. . . . .	34
4.1	As sucessivas conjugações de $\pi$ por $p_i^{-1}$ e as sucessivas multiplicações de $a$ por $p_i$ , dados um “n-ciclo” $\pi$ , $a$ (o seu conjugador (para $\iota$ )) e uma família $(p_i)$ , $i=1, \dots, m$ , de permutações. . . . .	47
4.2	Aplicação do Lema 7 aos elementos das linhas da Figura 4.1. . . . .	47
4.3	Simulação da ordenação de $\pi$ e da transformação de $a$ em $e$ . . . . .	49
4.4	Um passo na obtenção de uma sequência de reversões que, utilizando apenas conjugadores (para $\iota$ ), ordenam um “n-ciclo”. . . . .	49

# Lista de Tabelas

1.1	Cada terno de nucleotídeos de um gene produz um aminoácido e o gene produz uma cadeia desses aminoácidos, chamada proteína. Alguns aminoácidos são produzidos por mais de um terno; três ternos não produzem aminoácidos, TAA, TAG e TGA, apenas indicando o fim de um gene, e o metionina, ATG, indica o início de um gene. Na <i>transcrição</i> , uma fase da construção da proteína, onde o gene é copiado, a base T é substituída pela U, a uracil (tabela tirada de [15]). . . . .	4
1.2	Os vinte aminoácidos comumente encontrados nas proteínas (tabela tirada de [15]).	4
3.1	Os valores de $b(n, k)$ (número de permutações com $k$ inversões) para $n \leq 5$ . A linha $n$ começa com $b(n, 0)$ . . . . .	19
3.2	Os valores de $A(n, k)$ (número de Euler) para $n \leq 5$ . Cada diagonal NE-SO contém os valores para um $k$ fixo e a linha $n$ começa com $A(n, 1)$ . Nota-se que $A(n, k + 1) = A(n, n - k)$ ou, em outras palavras, os números de Euler são simétricos e tem-se que $\sum_{k=1}^n A(n, k) = n!$ . . . . .	26
3.3	Os valores de $c(n, k)$ (número de Stirling sem sinal de primeira espécie) para $n \leq 5$ . Uma diagonal NE-SO contém os valores para um $k$ fixo e a linha $n$ começa com $c(n, 0)$ . . . . .	27
3.4	Os valores de $s(n, k)$ (número de Stirling com sinal de primeira espécie) para $n \leq 5$ . A diagonal NE-SO contém os valores para um $k$ fixo e a linha $n$ começa com $s(n, 0)$ . . . . .	27
3.5	Os valores de $S(n, k)$ (número de Stirling de segunda espécie) para $n \leq 5$ . A diagonal NE-SO contém os valores de $S(n, k)$ para um $k$ fixo e a linha $n$ começa com $S(n, 0)$ . . . . .	27
3.6	Os valores de $G(n, k)$ (número de permutações com $k$ seqüências alternadas) para $n \leq 6$ . O primeiro valor da linha $n$ é $G(n, 1)$ . A diagonal NE-SO contém os valores de $G(n, k)$ para um $k$ fixo. . . . .	28
3.7	Os primeiros números de Hultman, onde os zeros resultam da paridade entre $c(G(p))$ e $c(G(i))$ (onde $p, i \in S_n$ e $i$ é a permutação identidade) e os uns (os últimos de cada linha) se referem às permutações identidades, conforme observações do Teorema 5. . . . .	32
4.1	Relação entre os elementos de $\pi$ e os do seu conjugador (para $\iota$ ) $a$ . . . . .	46
5.1	Alguns dados relacionados com as três operações básicas. . . . .	55

# Agradecimentos

Agradeço a todos os que compartilham comigo as suas existências, pelo muito que participam do sentido que tem a minha vida e apoio nas dificuldades; aos colegas do curso, pela presença e palavras amigas; ao professor Mauricio Ayala-Rincón, por sua disposição para discutir os trabalhos com os alunos, pelas palavras de apoio e por ter aceito ser o meu orientador, me permitindo aprender, com a sua experiência, a pesquisar, atividade esta que é e continuará sendo para mim uma das mais importantes; à professora Maria Emilia Machado Telles Walter, com quem aprendi os primeiros fundamentos em algoritmos e rearranjos de genomas, por ter me aceito no programa de mestrado; ao professor Norai Romeu Rocco, pela leitura do capítulo Fundamentos de uma forma diferente de computar para o problema da distância de reversão e sugestões que deram melhor legibilidade e clareza para a exposição; aos professores Jorge Petrúcio Viana e Norai Romeu Rocco, que aceitaram participar da banca examinadora; a todos os professores, por tudo que me ensinaram, nas pessoas dos professores Flávio Leonardo Cavalcanti de Moura e Mauricio Ayala-Rincón, da primeira disciplina que cursei no mestrado, Tópicos em Formalismos de Computação, pelos seus conhecimentos, didática, respeito e crença nos alunos, que confirmaram a minha expectativa de ser esta uma universidade competente e humana, o que foi mais um estímulo para prosseguir; aos funcionários da pós-graduação, pela ajuda e à universidade que, pela sua preocupação com o alto nível do ensino e por desempenhar de forma exemplar o seu papel de instrumento público de educação, democraticamente aberta a todos os que queiram aprender e pesquisar, interagindo com as outras universidades de dentro e de fora do país, como participante de um intercâmbio de culturas e de conhecimentos, contribui para formar e enobrecer pessoas e para o desenvolvimento geral.

Dedico este  
trabalho aos meus filhos  
Juliano, Lais e Isabel.

# Resumo

O problema da distância entre duas sequências finitas por reversões é estudado neste trabalho em uma abordagem formal e algébrica com base em um grupo de permutações onde a operação do grupo e a de conjugação são utilizadas para simular reversões.

A abordagem é geral por operar com sequências genéricas que podem representar estruturas utilizadas na especificação de uma grande variedade de problemas relevantes em computação e em matemática, entre as quais se incluem os genomas.

Uma estrutura de grupo foi estabelecida sobre um conjunto de famílias de sequências para possibilitar aplicar uma reversão em uma destas famílias mediante uma operação de conjugação, onde a família e uma certa permutação, atuando como um conjugador, participam como fatores ou termos.

Além disso, a mesma reversão pode ser simulada pela operação do grupo utilizando, como fatores ou termos, um conjugador especial da família e a mesma permutação que, acima, atuou como conjugador.

O trabalho propõe um método diferente de computar que conduz a uma maneira diferente de pensar no problema da distância de reversão o que, eventualmente, poderá contribuir na descoberta de respostas a questões nesta área.

Os programas computacionais que utilizam reversões podem ser adaptados ao método.

**Palavras-chave:** Ordenação de sequência finita por reversões, Distância de reversão entre duas sequências finitas, Reversão em sequência finita usando a conjugação e a operação de um grupo de permutações, Biologia molecular computacional, Rearranjos de genomas.

# Abstract

The problem of sorting finite sequences by reversals is studied in this work using a formal and algebraic approach based on a group of permutations where the operation of the group and the conjugation are used to implement reversals.

The approach is general in the sense that it treats generic sequences which can represent structures, used in the specification of a wide variety of relevant problems in computer science and mathematics, among them are included the genomes.

A permutation group structure was established on a set of families of sequences in order to apply a reversal on one of this family through the operation of conjugation, where the family and one certain permutation, acting as a conjugator, participate as factors or terms.

Moreover, the same reversal may be applied by means the group operation using, as factors or terms, a special conjugator of the family and the same permutation, above mentioned, that served as conjugator.

The work proposes a different method of computing resulting in a different way of thinking about the reversal distance problem which may possibly contribute to find answers in this area.

Computer programs that use reversals can be adapted for the method.

**Keywords:** Sorting of finite sequence by reversals, Reversal distance between finite sequences, Reversal in a finite sequence using the conjugation and the operation of a group of permutations, Computational molecular biology, Genome rearrangements.

# Capítulo 1

## Introdução

Os conhecimentos advindos da descoberta da estrutura do cromossomo no início da segunda metade do século passado levaram a grandes avanços na biologia e ao surgimento da biologia molecular computacional, uma nova ciência que de imediato se deparou com o desafio de processar os enormes volumes de dados gerados por sequenciadores.

Paralelamente ao desenvolvimento da bioinformática, houve a invenção e o aperfeiçoamento dos processadores e quantidades cada vez maiores de informações puderam ser analisadas em um espaço de tempo cada vez menor.

Apesar disso, permanecem intransponíveis as dificuldades para resolver certos problemas de grande complexidade para os quais os processadores mais velozes, em face da quantidade de dados e operações necessárias, também não produzem soluções eficientes em tempo.

A pesquisas com algoritmos passaram a ocupar o centro das atenções como talvez a única maneira de resolver tais problemas e emergiram as teorias voltadas para tentar reduzir as suas complexidades ou para aumentar a sua eficiência.

### Motivação

O material genético dos organismos é modificado a medida que evoluem [5]. Assim, as similaridades e as diferenças entre as sequências dos nucleotídeos dos cromossomos podem ser usadas para inferir a relação evolucionária entre espécies.

O genoma é o conjunto dos cromossomos de uma espécie; o genoma humano tem 23 pares com cerca de 27000 genes; um gene humano pode ter cerca de 10000 nucleotídeos e o genoma, centenas de milhões; o gene é uma cadeia de moléculas mais simples chamadas de nucleotídeos e cada gene tem por finalidade codificar uma proteína; a proteína também é uma cadeia de moléculas mais simples, os aminoácidos, e cada um é gerado a partir de uma codificação diferente de um terno de nucleotídeos, conforme o código de mapeamento genético das Tabelas 1.1 e 1.2; as proteínas são de diferentes tipos, têm diversas funções e a sua importância para a vida das espécies é tão grande que um importante cientista,

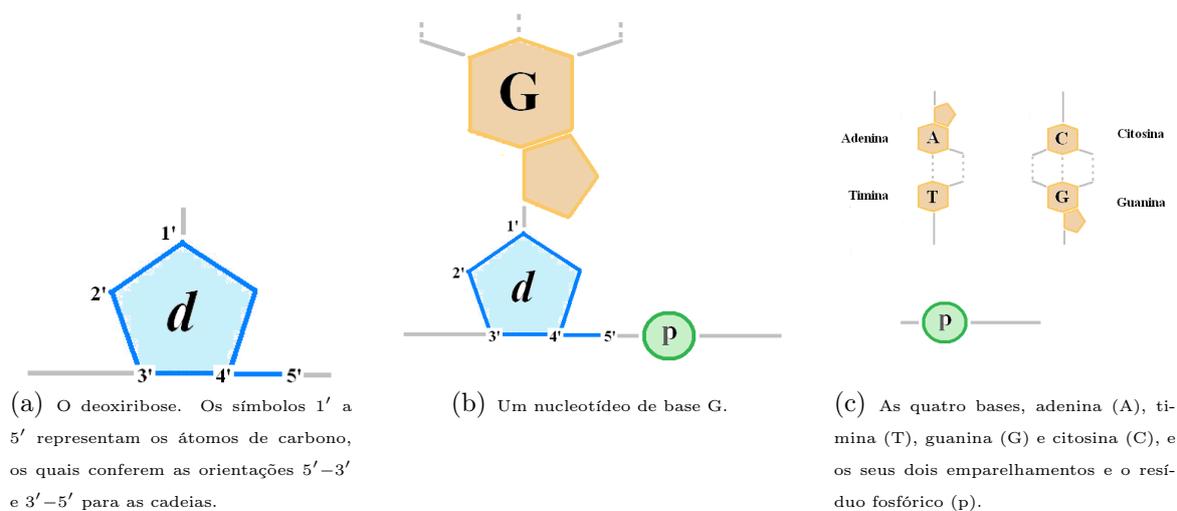


Figura 1.1: Componentes do nucleotídeo (ilustrações do aluno, com base em [15]).

Russel Doolittle, uma vez afirmou que “nós somos as nossas proteínas”.

O nucleotídeo é uma unidade, conforme a ilustração da Figura 1.3 (a), que constitui o gene e tem três componentes, como mostra a Figura 1.1: uma molécula de açúcar, denominado deoxiribose, um resíduo fosfórico e uma base. O deoxiribose contém cinco átomos de carbono, numerados de 1' a 5', o que confere uma orientação a cada uma das bandas do cromossomo, sendo 5'–3' a direção canônica, exemplificada na Figura 1.3 (b), encontrada em textos e bancos de dados. As bases são moléculas identificadas pelas letras *A*, *C*, *G* e *T*, que sempre se emparelham da seguinte maneira: *A* com *T* e *C* com *G*, sendo, por isso, *A* e *T*, ditas complementares e, da mesma forma, *C* e *G*. Conseqüentemente, as bandas da dupla hélice são, também, complementares.

A Figura 1.4 ilustra a codificação de uma proteína a partir do gene. A Figura 1.2 contém um esboço e uma visão molecular esquemática da dupla hélice, com as duas seqüências de nucleotídeos, que podem, também, ser representadas por seqüências de letras, cada uma simbolizando um destes nucleotídeos, como

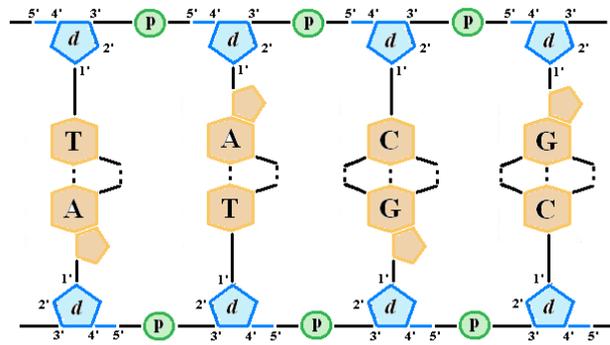
$$\begin{array}{l} 5' \dots \text{TACG} \dots 3' \\ 3' \dots \text{ATGC} \dots 5' \end{array}$$

Comparar as posições de genes ou de blocos de genes em dois diferentes organismos homólogos (de mesmos genes) e tentar determinar as operações de mutação (eventos de rearranjo) que transformam o genoma de um no do outro tem sido objeto de pesquisas.

Existem diversas proposições, usando diferentes técnicas e ferramentas, que buscam minimizar o trabalho computacional de calcular as operações de rearranjo que diferenciam dois organismos. Uma delas é o algoritmo, do tipo *branch-and-bound*, exato, “força bruta” de Kececioğlu e Sankoff, contido em [10], que obtém todas as seqüências de reversões que levam um genoma em outro.

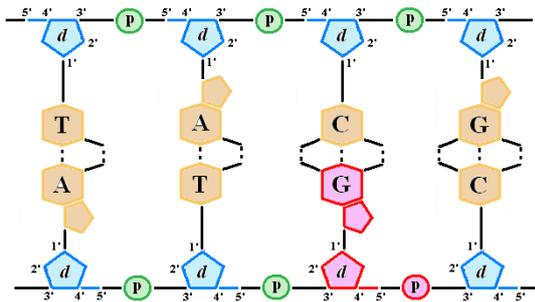


(a) Um esboço.

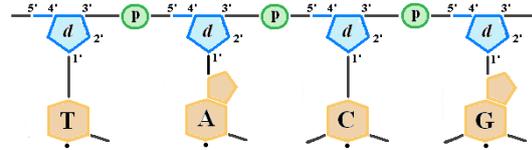


(b) Visão molecular esquemática.

Figura 1.2: A dupla hélice, conforme estabeleceram James D. Watson e Francis Crick, em 1953, a forma da construção do cromossomo, com duas cadeias ou bandas de nucleotídeos, complementares e invertidas e de orientações  $5'-3'$  e  $3'-5'$  (a ilustração (a) foi tirada da Internet e a (b) é do aluno, com base em [15]).



(a) O nucleotídeo.



(b) A banda de orientação  $5'-3'$ .

Figura 1.3: O nucleotídeo, uma unidade constituinte do cromossomo, e a banda  $5'-3'$ , a do sentido canônico adotado em documentos e bancos e dados onde apenas uma das duas esteja representada (ilustrações do aluno, com base em [15]).

$5' \dots \text{ATGGCAAGTCAAAGGATACAGAAATTTTAA} \dots 3'$   
 ATG GCA AGT CAA AGG ATA CAG AAA TTT TAA  
 Met Ala Ser Gln Arg Ile Gln Lis Fen FIM  
 MASQRIQKF

Figura 1.4: Um gene (uma cadeia de nucleotídeos, aqui representados pelas bases) em um cromossomo, os ternos de nucleotídeos com os correspondentes aminoácidos e a sua proteína (uma cadeia de aminoácidos) em uma simulação da codificação de uma proteína a partir de um gene fictício.

1ª posição	2ª posição				3ª posição
	G	A	C	U	
G	Gli	Glu	Ala	Val	G
	Gli	Glu	Ala	Val	A
	Gli	Asp	Ala	Val	C
	Gli	Asp	Ala	Val	U
A	Arg	Lis	Thr	Met	G
	Arg	Lis	Thr	Ile	A
	Ser	Asn	Thr	Ile	C
	Ser	Asn	Thr	Ile	U
C	Arg	Gln	Pro	Leu	G
	Arg	Gln	Pro	Leu	A
	Arg	His	Pro	Leu	C
	Arg	His	Pro	Leu	U
U	Tri	FIM	Ser	Leu	G
	FIM	FIM	Ser	Leu	A
	Cis	Tir	Ser	Fen	C
	Cis	Tir	Ser	Fen	U

Tabela 1.1: Cada terno de nucleotídeos de um gene produz um aminoácido e o gene produz uma cadeia desses aminoácidos, chamada proteína. Alguns aminoácidos são produzidos por mais de um terno; três ternos não produzem aminoácidos, **TAA**, **TAG** e **TGA**, apenas indicando o fim de um gene, e o metionina, **ATG**, indica o início de um gene. Na *transcrição*, uma fase da construção da proteína, onde o gene é copiado, a base T é substituída pela U, a uracil (tabela tirada de [15]).

	Código de uma letra	Código de três letras	Nome
1	A	Ala	Alanina
2	C	Cis	Cisteína
3	D	Asp	Aspartato
4	E	Glu	Glutamato
5	F	Fen	Fenilalanina
6	G	Gli	Glicina
7	H	His	Histidina
8	I	Ile	Isoleucina
9	K	Lis	Lisina
10	L	Leu	Leucina
11	M	Met	Metionina
12	N	Asn	Asparagina
13	P	Pro	Prolina
14	Q	Gln	Glutamina
15	R	Arg	Arginina
16	S	Ser	Serina
17	T	Thr	Treonina
18	V	Val	Valina
19	W	Tri	Triptofano
20	Y	Tir	Tirosina

Tabela 1.2: Os vinte aminoácidos comumente encontrados nas proteínas (tabela tirada de [15]).

Os cromossomos de um ser podem ser representados por permutações e essa possibilidade, aliada à de simular os eventos de rearranjos por operações algébricas, permitindo estudar por este meio a genética dos organismos vivos, motivou a escolha do tema.

A diversidade das ferramentas utilizadas e os conhecimentos acumulados pelo intensivo estudo das permutações ao longo do tempo aumenta as possibilidades de novas descobertas nos dois campos, da teoria e da prática.

## Objetivo

O objetivo do trabalho foi estudar as permutações com o fim de propor e formalizar um método para simular uma operação de reversão em sequências finitas por intermédio de uma operação de conjugação em um grupo de simetria e, por esta forma, estudar o problema da *distância de reversão* e, possivelmente, dar uma abordagem unificada para os problemas semelhantes de todas as operações.

## Organização do Trabalho

Os conteúdos mais importantes da dissertação estão distribuídos nos três capítulos adiante resumidos. O capítulo 2 reúne as definições e as notações importantes incluindo as três operações básicas para a transformação de uma sequência e a definição do problema que norteia o trabalho e está dividido em quatro seções, cada uma contendo: a seção 2.1, os conceitos básicos sobre grupos de permutações, que foram utilizados em todo o trabalho; a seção 2.2, as três operações básicas; a seção 2.3, a definição do problema e, a seção 2.4, outras definições. O capítulo 3 está dividido em seções, sendo que as primeiras constituem uma pequena amostra de algumas diferentes representações dadas às permutações, da sua adaptabilidade a diferentes recursos, algébricos, analíticos e gráficos, e de alguns resultados e, a última, contém a dedução de um resultado representativo, obtido em consequência dessa flexibilidade e a partir da operação de intercâmbio de blocos. Na seção 3.1, é dado um resultado combinatorial, obtido a partir do conceito de sequências crescentes; na seção 3.2, é dado um resultado combinatorial, utilizando o conceito de inversões; na seção 3.3, são dados resultados relacionados ao conceito de desconhecimento de padrão; na seção 3.4, é dados um resultado em ordenação, utilizando o conceito de pilhas junto com o de desconhecimento de padrão; na seção 3.5, é dado um resultado, utilizando o conceito dos Standard Young Tableau (SYT); na seção 3.6, é apresentada uma ordem parcial definida sobre  $S_n$ ; a seção 3.7 encerra a amostra, dando alguns conceitos muito utilizados no estudo das permutações e a seção 3.8 contém a dedução de um resultado em ordenação por intercâmbio de blocos, obtido com o uso de uma grande diversidade de recursos. No capítulo 4, está a parte principal do trabalho, onde é formalizado, detalhadamente, o método para considerar uma sequência finita qualquer como um “ciclo”  $\pi$  de um grupo simétrico e para simular uma reversão em  $\pi$  por meio de uma operação de conjugação do grupo; também é mostrada a sua aplicabilidade em algoritmos e são apresentadas duas maneiras equivalentes de se obter uma mesma família de reversões que transformam uma sequência em outra. Está dividido em quatro seções: na seção 4.1, é formalizado o modelo; na seção 4.2, como um exemplo de aplicação a casos concretos, são representados cromossomos; na seção 4.3, é definido um conjugador especial  $a$  de  $\pi$  e são feitas duas

simulações em paralelo, utilizando a operação do grupo e a de conjugação, para obter uma mesma família de reversões que ordenam  $\pi$  e, por último, na seção 4.4, são formalizadas as duas maneiras de construir a mesma permutação que simula uma reversão e onde fica demonstrada a aplicabilidade do método em algoritmos.

# Capítulo 2

## Definições e notações

### 2.1 Grupos de permutações

Nesta seção estão alguns conceitos básicos sobre grupos de permutações (presentes nos livros de álgebra dos cursos de graduação, dos quais, as citações [26] e [32] são apenas dois exemplos) que serão empregados daqui para frente.

Uma **permutação** de elementos de um conjunto  $E$  é uma bijeção de  $E$  em  $E$ . O conjunto das permutações do conjunto  $E = \{1, \dots, n\}$  (conjunto também denotado por  $[n]$ ), munido da operação de composição de funções, constitui o grupo chamado de *grupo das permutações* de  $E$  ou *grupo simétrico de grau  $n$* , denotado por  $S_n$ . A ordem de  $S_n$  é  $n!$ . Toda permutação de  $S_n$  será também chamada de  $n$ -permutação. O conjunto formado pelos elementos de um grupo é chamado de *suporte do grupo*. A inversa de uma permutação  $a$  é denotada por  $a^{-1}$ .

Sejam  $a, b \in S_n$ , então o **produto** de  $a$  por  $b$ , representado por  $ab$ , é definido como  $(ab)(x) = a(b(x))$ , para cada elemento  $x \in E$  e a **conjugação** de  $b$  por  $a$ , representada por  ${}^a b$ , é definida como  $({}^a b a^{-1})(x) = a(b(a^{-1}(x)))$ , para cada elemento  $x \in E$ . Além disso,  $aa^{-1} = a^{-1}a = e$ , onde  $e$  é a identidade, para todo  $a \in S_n$ .

Seja  $(a_i)_{1 \leq i \leq r}$  uma família de elementos distintos de  $[n]$  tem-se que  $r \leq n$ . A permutação  $a \in S_n$  definida por

$$\begin{aligned} a(a_i) &= a_{i+1}, & \text{para } & i = 1, \dots, r-1 \\ a(a_r) &= a_1 & \text{e} & \\ a(x) &= x & \text{para todo } & x \neq a_i \ (i = 1, \dots, r) \end{aligned}$$

é denominada **ciclo** determinado pelos elementos  $a_1, \dots, a_r$ , será indicada por  $(a_1 \dots a_r)$  e o número  $r$  denominado *comprimento* do ciclo; também, diz-se que a permutação é um  $r$ -ciclo e, se  $r = n$ , diz-se que  $(a_1 \dots a_r)$  é uma permutação *circular*. Todo ciclo de comprimento 2 é chamado de **transposição** e todo ciclo de comprimento 1 é o *elemento identidade*, que é denotado por  $e$ . Se  $a$  é o ciclo  $(a_1 a_2 \dots a_r)$ , o conjunto  $\{a_1, a_2, \dots, a_r\}$  é chamado de *suporte de  $a$* .

Dois ciclos  $a$  e  $b$  são disjuntos se os seus suportes forem disjuntos. Se  $a_1, \dots, a_m$ ,  $m \leq n$ , são ciclos dois a dois disjuntos e de comprimentos, respectivamente,  $t_1, \dots, t_m$ ,  $t_1 + \dots + t_m$

$\leq n$ , determinados pelas famílias  $(a_{ij}), 1 \leq i \leq m, 1 \leq j \leq t_i$ , então, uma permutação é definida como  $a_1 \dots a_m = (a_{11} \dots a_{1t_1}) \dots (a_{m1} \dots a_{mt_m})$  e a sua *inversa* como  $(a_{1t_1} \dots a_{11}) \dots (a_{mt_m} \dots a_{m1})$ . Toda permutação pode ser escrita como um produto de ciclos disjuntos. Essa representação é chamada de representação cíclica, não depende da ordem dos  $a_i, i = 1, \dots, m$ , e é única a menos dessa ordem.

**Observação 1** Se  $(a_1 a_2 \dots a_r)$  é um  $r$ -ciclo, então também o indicam as permutações  $(a_2 a_3 \dots a_r a_1), (a_3 \dots a_r a_1 a_2), \dots, (a_r a_1 \dots a_{r-1})$ .

**Observação 2** Se  $a$  é um produto de transposições disjuntas, então  $a = a^{-1}$ .

**Exemplo 1** Dados  $E = \{1, 2, 3, 4, 5, 6\}$ ,  $a = (1)(3\ 2\ 5)(4\ 6)$   $b = (2\ 6\ 1)$  e  $c = (2\ 6)(1\ 3)$ , então  $a^{-1} = (1)(5\ 2\ 3)(4\ 6)$ ;  $aa^{-1} = a^{-1}a = e$ ;  $ab = (1)(3\ 2\ 5)(4\ 6)(2\ 6\ 1) = (1\ 5\ 3\ 2\ 4\ 6)$ ;  $cc = e$  e  $a^2b = (1)(3\ 2\ 5)(4\ 6)(2\ 6\ 1)(1)(5\ 2\ 3)(4\ 6) = (1\ 5\ 4)(2)(3)(6) = (1\ 5\ 4)$ .

## 2.2 Operações básicas

Nesta seção são revistas algumas operações básicas, que representam eventos em rearranjos de genoma.

Em vista da estrutura de dupla hélice de um cromossomo (conforme estabeleceram James D. Watson e Francis Crick, em 1953), com duas cadeias, complementares e invertidas, de genes, na construção de uma sequência (ou permutação, etc.), admitindo-se que esta seja a maneira escolhida para representar o cromossomo, resultante de uma leitura efetuada sobre o mesmo e que represente ordenadamente os seus genes, estes podem provir ora de uma ora da outra.

As sequências representando cromossomos são estudadas em nível de genes ou de blocos de genes. Estará sendo admitido, sem prejuízo para os resultados obtidos, que cada um de seus elementos estará representando um bloco. Quando as orientações não são conhecidas, tem-se somente a ordem relativa entre eles. Caso contrário, pode-se, também, considerar a sua orientação no estudo do cromossomo.

As mutações ou eventos de rearranjos que podem ocorrer em um único cromossomo são as de interesse deste trabalho e, entre outros exemplos de eventos que afetam um único cromossomo, estão a *reversão*, a *transposição* e o *intercâmbio de blocos*.

Como se verá no Capítulo 4, os blocos de um cromossomo, admitindo-se que ele tenha  $n$  blocos, serão rotulados pelo conjunto  $[n]$  e ele passará a ser representado por uma sequência destes rótulos, dispostos na mesma ordem em que os blocos correspondentes ocorrem no cromossomo. Nesta seção será admitido que um cromossomo  $p$  será representado por uma sequência  $p_1p_2 \dots p_n$  e que, quando os rótulos da sequência estiverem ordenados, ou seja, quando a sequência for  $12 \dots n$ , dir-se-á que o cromossomo está ordenado (conforme a seção 2.3). Os blocos afetados pelos eventos aparecerão em destaque.

As Figuras 2.1, 2.2 e 2.3 têm apenas o objetivo de ilustrar o evento da reversão, que é a operação em estudo neste trabalho, de acordo com a representação, em sequência, que

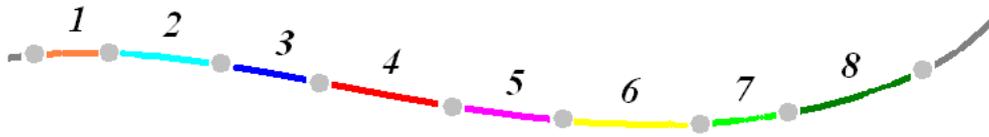


Figura 2.1: O cromossomo de oito blocos de genes numerados em ordem crescente, representado pela permutação 12345678, antes do evento que reverte o bloco de entradas consecutivas 4, 5, 6 e 7, que resultará 12376548, a representação do seu homólogo, da Figura 2.2 (ilustração do aluno).

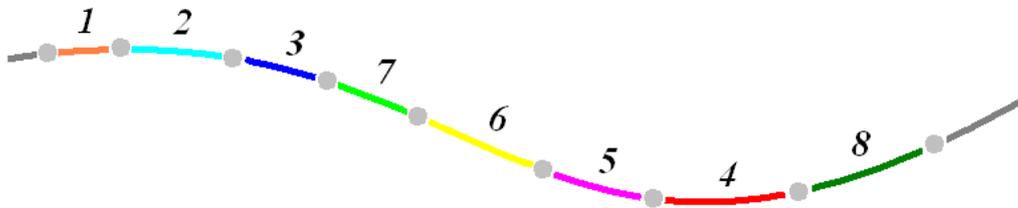


Figura 2.2: Como ficará o novo cromossomo 12376548 após o evento da Figura 2.1 (ilustração do aluno).

está sendo adotada; estão simplificadas tendo em vista que o evento ocorre na dupla hélice.

Dada uma sequência  $p = p_1 p_2 \dots p_n$ , uma *reversão* é uma operação que inverte uma subsequência de blocos consecutivos de  $p$ . Na reversão com sinal, caso em que os blocos têm orientações conhecidas, também são revertidos os sinais, simbolizados por  $\rightarrow$  ou  $\leftarrow$ , de cada um dos blocos revertidos.

Revertendo-se os blocos  $\underline{p_i \dots p_{i+a}}$ ,  $a \geq 0, i > 0, i + a \leq n$ , em  $p$ , obtém-se

$$p_1 \dots \underline{p_{i+a} \dots p_i} \dots p_n.$$

**Exemplo 2** Se em 3 4 1 7 5 6 2 forem revertidos os blocos 1756 resulta 3 4 6 5 7 1 2 e, se as suas orientações forem conhecidas, ou seja, se for um cromossomo com sinal, por exemplo,  $\overrightarrow{3} \overrightarrow{4} \overleftarrow{1} \overleftarrow{7} \overleftarrow{5} \overleftarrow{6} \overrightarrow{2}$ , resulta  $\overrightarrow{3} \overrightarrow{4} \overleftarrow{6} \overleftarrow{5} \overrightarrow{7} \overrightarrow{1} \overrightarrow{2}$ .

Uma *transposição* é uma operação que intercambia duas subsequências contíguas de blocos consecutivos sem mudar a ordem desses blocos em cada subsequência. Transpondo-

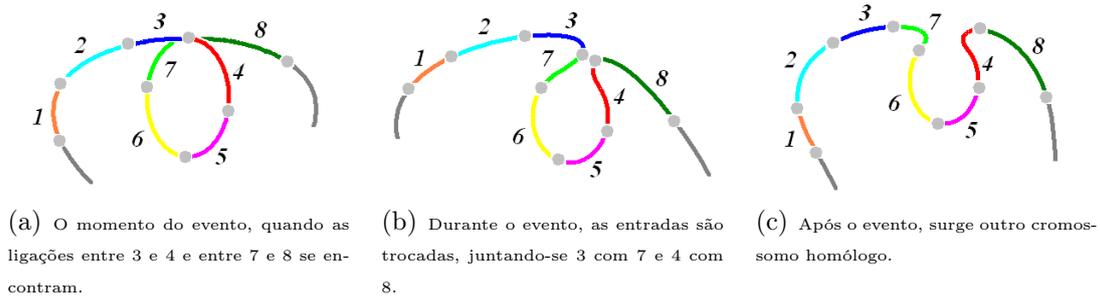


Figura 2.3: O evento da reversão do bloco das entradas consecutivas 4, 5, 6 e 7 do cromossomo 12345678, da Figura 2.1. A hipótese para o evento é que, em um encontro das ligações entre 3 e 4 e entre 7 e 8, por alguma razão, as mesmas são desfeitas e as entradas religadas de forma diferente, unindo 3 com 7 e 4 com 8. O resultado é o cromossomo 12376548, da Figura 2.2 (ilustrações do aluno).

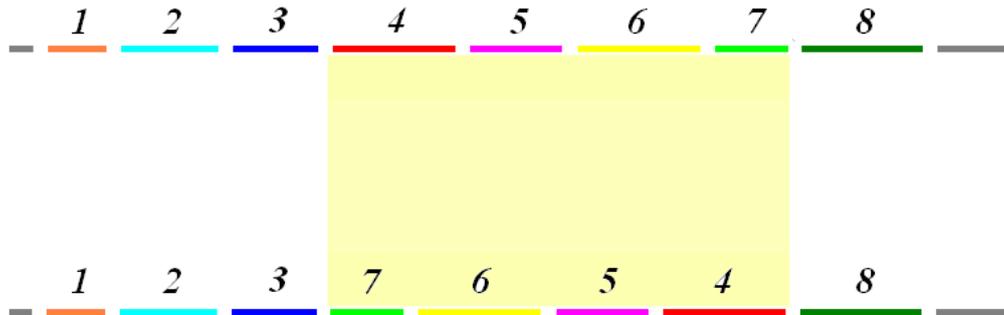


Figura 2.4: O cromossomo original, Figura 2.1, e o resultado da reversão, Figura 2.2; em destaque o bloco das entradas consecutivas que foi revertido (ilustração do aluno).

se os blocos  $\underline{p_i \dots p_{i+a}}$  e  $\underline{p_{i+a+1} \dots p_{i+a+1+b}}$ ,  $a, b \geq 0, i > 0, i+a+1+b \leq n$ , em  $p$ , obtém-se

$$p_1 \dots \underline{p_{i+a+1} \dots p_{i+a+1+b}} \underline{p_i \dots p_{i+a}} \dots p_n.$$

Nas transposições, não são consideradas as orientações dos blocos porque um bloco isoladamente não é revertido.

**Exemplo 3** Se em 3417562 forem transpostas as subsequências 175 e 62 resulta 3462175.

$$\begin{array}{cccccccc} \underline{3} & \underline{4} & \underline{1} & \underline{7} & \underline{5} & \underline{6} & \underline{2} & \\ \underline{2} & \underline{3} & \underline{4} & \underline{1} & 7 & 5 & 6 & \\ 1 & 2 & 3 & 4 & \underline{7} & \underline{5} & \underline{6} & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & \end{array}$$

Figura 2.6: Ordenação de 3417562 por três transposições.

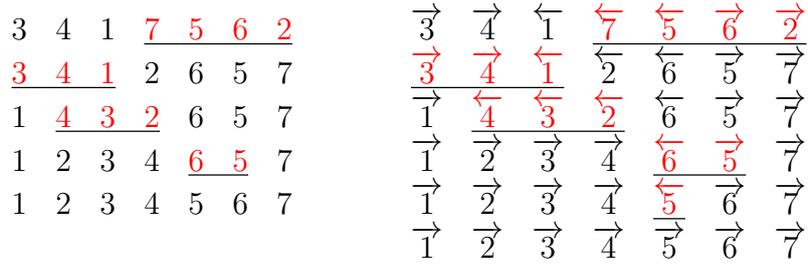


Figura 2.5: Ordenações de  $p = 3417562$  (sem sinal) por quatro e de  $p$  (com uma de suas possíveis  $2^7$  sinalizações) por cinco reversões, respectivamente.

Um *intercâmbio de blocos* é uma operação que intercambia duas subsequências de blocos consecutivos sem mudar a ordem dos blocos de cada subsequência. Intercambiando-se os blocos  $\underline{p_i \dots p_{i+a}}$  e  $\underline{p_j \dots p_{j+b}}$ ,  $a, b \geq 0, i > 0, i + a < j, j + b \leq n$ , em  $p$ , obtém-se

$$p_1 \dots \underline{p_j \dots p_{j+b}} \dots \underline{p_i \dots p_{i+a}} \dots p_n.$$

No intercâmbio de blocos, também não são consideradas as orientações dos blocos porque um bloco isoladamente não é revertido.

**Exemplo 4** Se em 3417562 forem intercambiadas as subsequências 34 e 562 resulta 5621734.

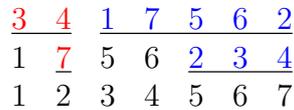


Figura 2.7: Ordenação de 3417562 por dois intercâmbios de blocos.

### 2.3 O problema da distância de reversão e o problema da distância de ordenação

Na natureza, um ser pode evoluir para outro pela mudança de lugar dentro do seu genoma, de um ou mais blocos de genes. Esses eventos mutacionais que são conhecidos pelos biólogos são também chamados de operações de rearranjo. Vários tipos são aceitos como eventos biológicos e mais de um pode ocorrer simultaneamente em um organismo em evolução.

De forma genérica, o problema de rearranjo de genomas agindo num único cromossomo pode ser enunciado como: dados dois cromossomos, obter o menor número de eventos de rearranjo que transformem um no outro.

A menor série desses eventos está relacionada com a hipótese da parcimônia, referida em [15], que assume que a natureza sempre encontra caminhos que requeiram o mínimo de mudanças. Assim, para se saber como um organismo pode ser transformado em outro,

deve-se encontrar uma série mínima de eventos que realize essa tarefa.

As Figura 2.8 ilustra o conceito de distância de reversão utilizando os cromossomos de duas plantas e a Figura 2.9 contém fotos de ambas onde aparecem algumas das suas diferentes características externas, como o formato e o tamanho das folhas e a coloração das flores, que refletem as diferenças entre os cromossomos.

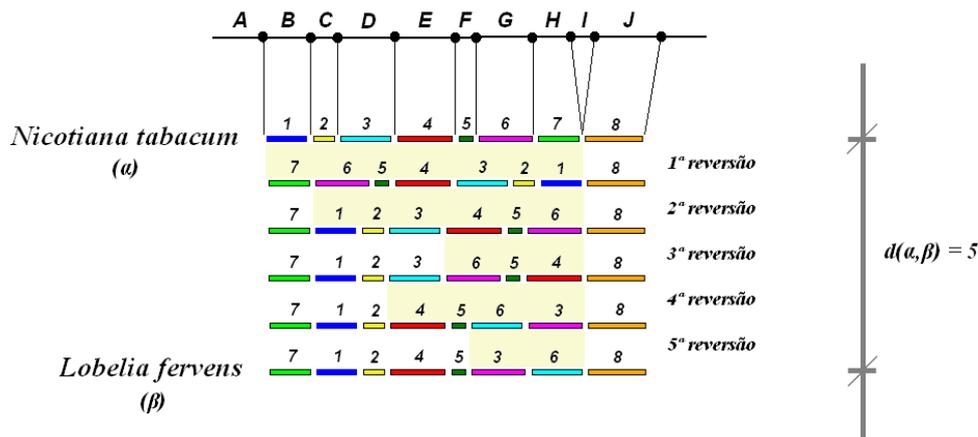


Figura 2.8: A transformação da *Nicotiana tabacum* na *Lobelia fervens*, duas plantas homólogas, em cinco reversões e a distância de reversão,  $d(\alpha, \beta) = 5$ , entre ambas. Este é o número mínimo de reversões (em destaque o bloco das entradas consecutivas que foi revertido em cada uma), necessário para transformar uma planta na outra (ilustração do aluno, com base em [23]).

Apenas o evento da reversão ocorrendo em um único cromossomo estará sendo considerado. Um cromossomo pode ser representado por uma sequência finita e, como este trabalho estuda a operação de reversão aplicada a essas sequências, os resultados também se aplicam aos cromossomos.

O foco deste trabalho está no evento da reversão (sem sinal ou com orientações não conhecidas), aplicado a sequências finitas, como foi dito, e tem-se o **problema da distância de reversão** em sequências finitas, que consiste em “dadas duas sequências finitas,  $\alpha$  e  $\beta$ , obter o menor número de reversões que transformem uma na outra”. A este número se dá o nome de **distância de reversão** entre  $\alpha$  e  $\beta$  que é denotado por  $d(\alpha, \beta)$  ou  $d_\beta(\alpha)$ . Ordenar uma sequência finita é transformá-la na sequência finita onde os blocos estão ordenados e que é chamada de sequência ordenada. Um caso particular é a **distância de ordenação**, que consiste em “dada uma sequência finita  $\alpha$ , obter o menor número de reversões que a transformem na sequência ordenada” e é denotada por  $d(\alpha)$ .



Figura 2.9: Fotos da *Nicotiana tabacum* e *Lobelia fervens* (fotos tiradas da Internet).

## 2.4 Mais definições e notações

Nesta seção estão outras definições e notações utilizadas no trabalho. A obra [2] será referenciada pelos números das páginas.

### Número de Euler.

**Definição 1** (p. 3) Sejam  $n$  um número inteiro positivo e  $p=p_1p_2 \dots p_n$  uma permutação de  $S_n$ . Diz-se que  $i$  é um **descendente** de  $p$  se  $p_i > p_{i+1}$ ,  $1 \leq i < n$ . Similarmente, diz-se que  $i$  é um **ascendente** de  $p$  se  $p_i < p_{i+1}$ . Se  $p$  tem  $k-1$  descentes, então é a união de  $k$  subsequências crescentes de entradas consecutivas, chamadas de **seqüências crescentes** de  $p$ .

**Exemplo 5** A permutação 3412576, da Figura 2.10, tem dois descentes e quatro ascentes.

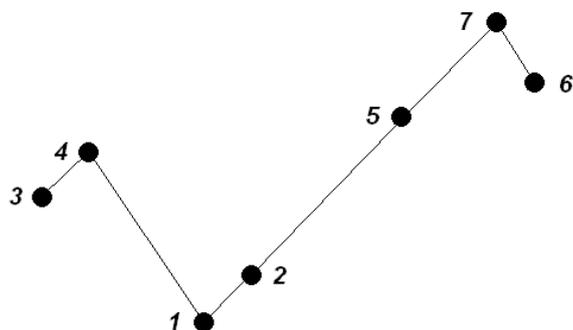


Figura 2.10: A permutação 3412576 tem os descentes 2 e 6 e os ascentes 1, 3, 4 e 5.

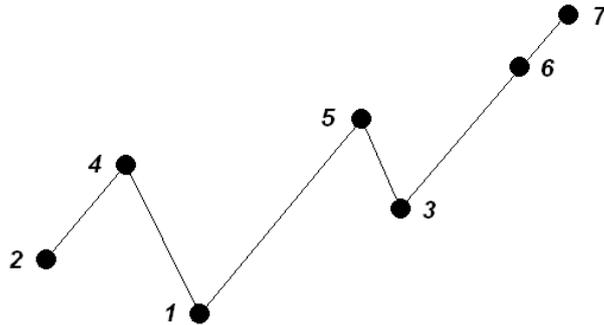


Figura 2.11: A permutação 2415367.

**Exemplo 6** A permutação 2415367, da Figura 2.11, tem três sequências crescentes que são 24, 15 e 367.

### Número de Stirling de segunda espécie

**Definição 2** (p. 11) Sejam  $n$  um número inteiro positivo,  $[n]$  e  $k \leq n$ . A distribuição dos elementos de  $[n]$  em  $k$  conjuntos de tal forma que cada elemento é colocado em exatamente um e nenhum é vazio será chamada de *partição de um conjunto*  $[n]$  em  $k$  blocos.

**Exemplo 7** Sendo  $n = 7$  e  $k = 3$ ,  $\{1, 2, 4\}$ ,  $\{3, 6\}$ ,  $\{5\}$  e  $\{7\}$  é uma partição de  $[7]$  em três blocos.

### Standard Young Tableau (SYT).

**Definição 3** (p. 47) Sejam  $n$ ,  $m \leq n$  e  $a_1 \geq \dots \geq a_m$  números inteiros positivos tais que  $a_1 + \dots + a_m = n$ . Então  $(a_1, \dots, a_m)$  é chamada de *partição de  $n$*  e os números  $a_i$  de partes da partição.

**Exemplo 8** O número 5 tem sete partições que são (5), (4, 1), (3, 2), (3, 1, 1), (2, 2, 1), (2, 1, 1, 1, 1) e (1, 1, 1, 1, 1).

### Inversões.

**Definição 4** (p. 43) Sejam  $n$  um número inteiro positivo e  $a = a_1 a_2 \dots a_n$  uma permutação de  $S_n$ . Diz-se que  $(a_i, a_j)$  é uma *inversão* de  $a$  se  $i < j$  e  $a_i > a_j$   $1 \leq i < j \leq n$ . O número de inversões de  $a$  será denotada por  $i(a)$ .

Tem-se que  $0 \leq i(a) \leq \binom{n}{2}$ , para todas as permutações e os dois extremos se referem a  $1 \dots n$  e  $n(n-1) \dots 1$ , respectivamente.

**Exemplo 9** A permutação

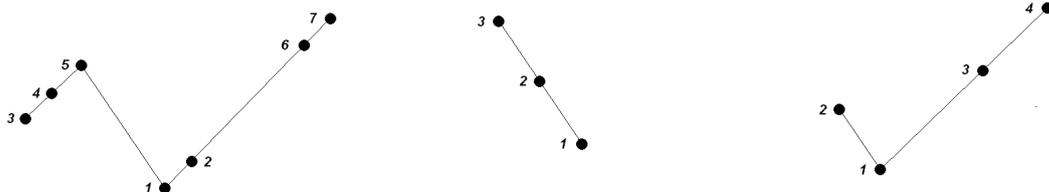
31524

tem quatro inversões, ou seja, (3, 1), (3, 2), (5, 2) e (5, 4).

**Desconhecimento de padrão.**

**Definição 5** (p. 129) Sejam  $n$  um número inteiro positivo,  $k < n$  e  $b = b_1 \dots b_k$  uma permutação de  $S_k$ . Diz-se que a permutação  $a = a_1 \dots a_n$  de  $S_n$  contém  $b$  como um padrão se existem  $k$  entradas  $a_{i_1}, a_{i_2}, \dots, a_{i_k}$  em  $a$  tais que  $i_1 < i_2 < \dots < i_k$  e  $a_{i_r} < a_{i_s}$  se, e somente se  $b_r < b_s$ ,  $1 \leq r < s \leq k$ . Caso contrário, diz-se que  $a$  **desconhece o padrão**  $b$ .

**Exemplo 10** A permutação 3451267, da Figura 2.12 (a), desconhece o padrão 321 (da Figura 2.12 (b)) porque não contém uma subsequência decrescente de 3 elementos. Ela contém o padrão 2134, Figura 2.12 (c).



(a) A permutação 3451267 contém o padrão 2134, como mostram as entradas 3167, 3267, 4167, 4267, 5167 e 5267, mas não o 321. (b) O padrão 321, contido em 3451267. (c) O padrão 2134, contido em 3451267.

Figura 2.12: A permutação 3451267 e os padrões 321 e 2134.

**Sequências alternadas.**

**Definição 6** (p. 24) Sejam  $n$  um número inteiro positivo e  $p = p_1 p_2 \dots p_n$  uma permutação de  $S_n$ . Diz-se que  $p$  muda de direção na posição  $i$ ,  $1 < i < n$  quando  $p_{i-1} < p_i > p_{i+1}$  ou  $p_{i-1} > p_i < p_{i+1}$ . Desses elementos  $p_i$  onde  $p$  muda de direção diz-se que são, respectivamente, um **pico** e um **vale**. Diz-se que  $p$  tem  $k$  sequências alternadas se existem  $k-1$  índices  $i$  onde  $p$  muda de direção. Denota-se por  $G(n, k)$  o número de permutações com  $k$  sequências alternadas.

**Exemplo 11** A permutação 3561247 da Figura 2.13, com duas mudanças de direção (onde  $i = 3$  é um pico e  $i = 4$  é um vale) tem três sequências alternadas (356, 61 e 1247).

**Ponto-de-quebra e subsequência ordenada.**

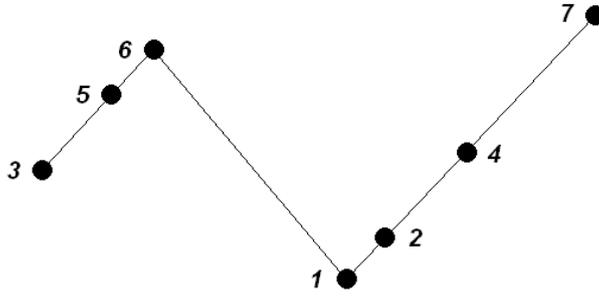


Figura 2.13: Um diagrama representando 3561247 com três seqüências alternadas.

**Definição 7** (em [11]) Sejam  $n$  um número inteiro positivo e  $a_1 \dots a_n$  uma permutação de  $S_n$  estendida para  $0 a_1 \dots a_n n+1$  com a adição de  $a_0 = 0$  e  $a_{n+1} = n+1$ . Um par de elementos consecutivos  $a_i$  e  $a_{i+1}$ , onde  $0 \leq i \leq n$  e  $|a_{i+1} - a_i| \neq 1$ , é chamado de *ponto-de-quebra* e  $q(a)$ , de número de pontos-de-quebra de  $a$ .

**Definição 8** Dada uma permutação  $a_1 \dots a_i \dots a_j \dots a_n$ , chama-se uma subsequência  $a_i \dots a_j$ , onde  $|a_i - a_{i-1}| \neq 1$  e  $|a_{j+1} - a_j| \neq 1$ ,  $i \leq k \leq j-1$ ,  $|a_{k+1} - a_k| = 1$ , de *subseqüência ordenada (maximal)* da permutação.

### Diâmetro de reversão do grupo simétrico e permutações de Gollan.

**Definição 9** Chamando-se de  $D(n) = \max_{\pi \in S_n} d(\pi)$  ao *diâmetro de reversão do grupo simétrico de grau  $n$* , onde  $d(\pi)$  é a distância de reversão entre a permutação  $\pi$  e a identidade, Gollan conjecturou (e Bafna e Pevzner provaram, em [23]) que  $D(n) = n-1$  e que apenas as permutações (em notação cíclica)  $\gamma_n = (1357 \dots n-1 n \dots 8642)$ , onde  $n$  é par, e  $\gamma_n = (1357 \dots n n-1 \dots 8642)$ , onde  $n$  é ímpar, e a suas inversas  $\gamma_n^{-1}$  requerem  $n-1$  reversões para serem ordenadas. Essas permutações são conhecidas como as *permutações de Gollan*.

<u>3</u>	<u>1</u>	5	2	7	4	6
1	<u>3</u>	<u>5</u>	<u>2</u>	7	4	6
1	2	<u>5</u>	<u>3</u>	7	4	6
1	2	3	<u>5</u>	<u>7</u>	<u>4</u>	6
1	2	3	4	<u>7</u>	<u>5</u>	<u>6</u>
1	2	3	4	<u>6</u>	<u>5</u>	7
1	2	3	4	5	6	7

Figura 2.14: Ordenação da permutação  $(1 3 5 7 6 4 2)$ , de Gollan, onde  $n = 7$  é ímpar, com 6 reversões.

## Diâmetro de reversão por prefixos do grupo simétrico.

O problema da ordenação por reversões e prefixos, também conhecido como *pancake flipping problem*, é definido da seguinte maneira: “dada uma permutação  $\pi_1 \dots \pi_n$  de  $S_n$ , encontrar uma menor série de reversões da forma  $\pi_1 \dots \pi_i$  que a ordenem,  $1 \leq i \leq n$ ”.

**Definição 10** O *diâmetro de reversão por prefixos do grupo simétrico de grau  $n$*  é definido como  $\max_{\pi \in S_n} d_{pref}(\pi)$  e denotado por  $d_{pref}(n)$ , onde  $d_{pref}(\pi)$  é a distância de reversão por prefixos entre  $\pi$  e a identidade. .

# Capítulo 3

## Contextualização

As permutações e os grupos de simetria são intensamente estudados e aplicados em diferentes contextos, entre os quais se encontra o dos rearranjos de genomas (um exemplo se encontra no artigo [23], onde Bafna e Pevzner as aplicam no estudo da operação de reversão) e antigos problemas relacionados com a distância de reversão são estudados por meio de novas ou diferentes ferramentas (no artigo [24], analisado neste capítulo, Bóna e Flynn utilizam uma diversidade de recursos, além de resultados obtidos por meio de outros, na dedução de uma fórmula importante em ordenação por intercâmbio de blocos).

Conjecturas milenares, como ilustra a brilhante solução de equações por meio de radicais construída por Galois, em fins do século XVIII, somente foram esclarecidas por meio dos grupos de simetria. Certos conceitos poderosos de fundo combinatorial, são estudados há muito tempo como é o caso dos *números de Euler*, empregados há 200 anos, tendo a maior parte das pesquisas direcionadas para conceitos analíticos. O estudo formal da combinatória começou pelo menos com Leibniz que, trabalhando em sua habilitação em Filosofia, publicou, em 1666, o “Dissertatio de arte combinatoria”, tendo esse ramo de estudo, no último meio século, um grande crescimento.

Os estudos e as conclusões acumulados em séculos de pesquisa estão espalhados pela literatura. Diversos conceitos aplicados no estudo das permutações, representativos em diversas áreas de pesquisa, foram extraídos de artigos e do livro [2] (neste capítulo esta obra será referenciada pelos números das páginas) onde o autor teve por objetivo reunir muitos resultados importantes, principalmente os de natureza combinatorial. Em seu prefácio fica destacado esse potencial ao afirmar que “as permutações têm uma estrutura combinatorial muito rica, devida em parte ao fato de que uma permutação de um conjunto finito pode ser representada de muitas maneiras equivalentes como *palavra* (sequência), função, coleção de ciclos disjuntos, etc., cada uma com uma coleção própria de operações, transformações, estruturas, etc., que podem ser aplicadas às permutações”.

Nesta contextualização, serão exibidos alguns desses recursos e resultados e finaliza com a análise de um artigo sobre ordenação de permutações, representativo pela importante fórmula que obtém e não trivialidade das deduções.

### 3.1 Sequências crescentes

As noções de ascente e descende (dadas na Definição 1) suscitam algumas questões combinatórias e enumerativas. Quantas permutações existem com um dado número de descentes? Quantas permutações existem com um dado conjunto de ascetes? Se duas permutações têm o mesmo conjunto ou o mesmo número de descentes, quais outras propriedades elas compartilham? Estas questões permitem estabelecer resultados combinatórios como, por exemplo, o seguinte lema.

**Lema 1** (p. 4) Sejam  $n$  um número inteiro positivo,  $k \leq n$ ,  $S = \{s_1, s_2, \dots, s_k\} \subseteq [n-1]$  e  $\alpha(S)$  o número de permutações cujo conjunto de descentes está contido em  $S$ . Então, tem-se

$$\alpha(S) = \binom{n}{s_1} \binom{n-s_1}{s_2-s_1} \binom{n-s_2}{s_3-s_2} \dots \binom{n-s_k}{n-s_k}.$$

### 3.2 Inversões

O estudo das inversões (dadas na Definição 4) começou em 1.901, em [33], e pode levar a interessantes resultados combinatórios como à função geradora que enumera as permutações de  $S_n$  com respeito ao seu número de inversões, que é dada pelo seguinte teorema.

**Teorema 1** (p. 43) Para todo inteiro positivo  $n > 2$ , tem-se

$$\sum_{a \in S_n} x^{i(a)} = I_n(x) = (1+x)(1+x+x^2) \dots (1+x+x^2+\dots+x^{n-1}).$$

O número de permutações com  $k$  inversões, denotado por  $b(n, k)$ , é o coeficiente de  $x^k$  em  $I_n(x)$ . Os primeiros são mostrados na Tabela 3.1 e dados, recursivamente, pela fórmula

$$b(n+1, k) = b(n+1, k-1) + b(n, k).$$

$n = 1$												
2												
3												
4												
5												
	1	4	9	15	20	22	20	15	9	4	1	

Tabela 3.1: Os valores de  $b(n, k)$  (número de permutações com  $k$  inversões) para  $n \leq 5$ . A linha  $n$  começa com  $b(n, 0)$ .

### 3.3 Desconhecimento de padrão

A noção de desconhecimento de padrão (dado na Definição 5), que também é aplicada a permutações, pode facilitar a obtenção de resultados interessantes.

**Exemplo 12** O número  $S_n(q)$  de permutações que desconhecem o padrão  $q$ , quando  $q=12$  (comprimento 2), é  $S_n(12)=S_n(21)=1$ . As permutações que desconhecem os padrões 12 e 21 são, respectivamente,  $n(n-1) \dots 321$  e  $123 \dots (n-1)n$ .

Quando o comprimento é 3,  $q$  admite 6 padrões diferentes, 123, 132, 213, 231, 312 e 321, mas existem algumas simetrias entre eles. Dada uma permutação  $p = p_1 p_2 \dots p_n$ , define-se um **reverso** de  $p$  como a permutação  $p^r = p_n p_{n-1} \dots p_1$  e o **complemento** de  $p$  como a permutação  $p^c$  cuja  $i$ -ésima entrada é  $n+1-p_i$ .

**Exemplo 13** A permutação  $p = 3451267$  tem reverso  $p^r = 7621543$  e complemento  $p^c = 1267345$ .

Assim, tem-se que se uma permutação desconhece 123, então, o seu reverso desconhece 321, logo,  $S_n(123)=S_n(321)$ . Da mesma forma, se uma permutação desconhece 132, então, o seu reverso desconhece 231, o seu complemento desconhece 312 e o reverso do seu complemento desconhece 213, assim,  $S_n(132)=S_n(231)=S_n(312)=S_n(213)$ .

**Exemplo 14** A permutação  $p = 3412$  desconhece o padrão 123 e o seu reverso  $p^r = 2143$  desconhece 321; também,  $p$  desconhece 132 e  $p^r$  desconhece 231;  $p^c$  desconhece 312 e  $p^{cr}$  desconhece 213.

Resta provar que  $S_n(123)=S_n(132)$  para se chegar ao resultado notável, devido a Simion e Schmidt, em [13], que afirma que todo padrão de comprimento 3 é desconhecido do mesmo número de permutações, e isso é dado pelo seguinte lema.

**Lema 2** (p. 130) Para todo inteiro positivo  $n$ ,  $S_n(123)=S_n(132)$ .

Um outro resultado enumerativo se segue.

**Teorema 2** (p. 132) Para todo inteiro positivo  $n$ ,  $S_n(132) = \binom{2n}{n}/(n+1)$  ( $\binom{2n}{n}$  indica a combinação de  $2n$  elementos tomados 2 a 2).

**Exemplo 15** Em  $S_4$ , das  $4! = 24$  permutações,  $\binom{8}{4}/(5) = 14$ , 1234, 2134, 2314, 2341, 3124, 3214, 3241, 3412, 3421, 4123, 4213, 4231, 4312 e 4321, desconhecem 132.

### 3.4 Ordenação por pilhas e desconhecimento de padrão

Um algoritmo que pode fazer (se verá uma condição em que não será possível) a ordenação de uma permutação por meio de uma pilha, do tipo *guloso*, é descrito a seguir, utilizando a permutação  $p = p_1 p_2 \dots p_n \in S_n$ .

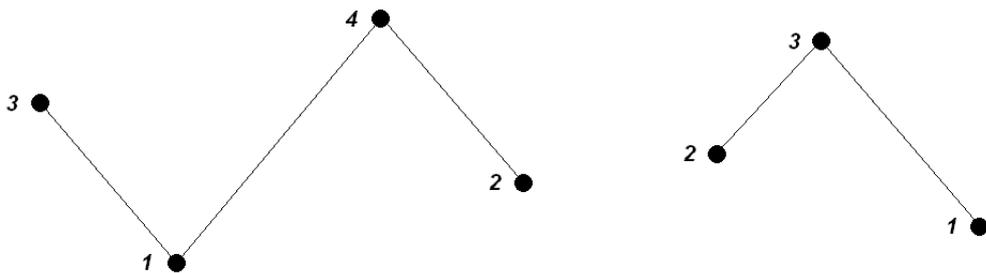
O algoritmo começa colocando  $p$  na entrada e  $p_1$  na pilha. Em seguida, verifica se  $p_2 < p_1$  e, neste caso, põe  $p_2$  em cima de  $p_1$  na pilha; caso contrário, primeiro põe  $p_1$  na saída e, em seguida,  $p_2$ , na pilha. Assim, continua até o passo  $i$ , quando o elemento  $p_i$  é comparado com o elemento  $p_{i-1}$  que está no topo da pilha. Se  $p_i < p_{i-1}$ ,  $p_i$  vai para o topo da pilha; caso contrário,  $p_{i-1}$  vai para a posição mais à direita da saída e  $p_i$  é comparado novamente com o elemento que ocupa o topo da pilha. O algoritmo termina quando as  $n$  entradas passarem pela pilha e estiverem na permutação  $s(p)$  da saída. Observa-se que os elementos da pilha se mantêm, durante a aplicação do algoritmo, em ordem decrescente.

Assim, se torna impossível passar pela pilha e em ordem crescente um conjunto  $\{p_i, p_j, p_k\}$  de entradas de  $p = p_i \dots p_j \dots p_k$ , onde  $p_k < p_i < p_j$ ,  $1 \leq i < j < k \leq n$ , que obedecem ao padrão, pois, antes de passar  $p_k$ , que é a menor,  $p_i$  ou  $p_j$  já terá passado para a saída para não violar a ordenação da pilha. Isso prova a seguinte proposição.

**Definição 11** (p. 288) Se a permutação  $s(p)$  de saída do algoritmo acima for a identidade  $123 \dots n$ , diz-se que  $p$  é ordenável por pilha.

**Proposição 1** (p. 289) Uma permutação é ordenável por pilha se, e somente se, desconhece o padrão 231 (dado na Definição 5).

**Exemplo 16** A Figura 3.2 mostra todos os passos do algoritmo aplicado a 3142, Figura 3.1 (a), que contém o padrão 231, Figura 3.1 (b), e onde se tem que  $s(3142) = 1324$ , o que prova que 3142 não é ordenável por pilha.



(a) A permutação 3142 contém o padrão 231, como mostram as entradas 342.

(b) O padrão 231, contido em 3142.

Figura 3.1: A permutação 3142 e o padrão 231.

### 3.5 Standard Young Tableau (SYT)

Os SYT existem há mais de um século tendo sido pela primeira vez definidos pelo Reverendo Young em uma série de artigos, iniciados pelo [37], no início do século passado. Os SYT e as permutações estão associados e as entradas de um SYT determinam uma permutação de  $[n]$ .

Uma **forma** de Ferrer é uma maneira de representar partições (dadas na Definição 3) por meio de um diagrama. A forma de Ferrer da partição  $(a_1, a_2, \dots, a_k)$  é um conjunto com  $n$  caixas postas em linhas paralelas de tal forma que na  $i$ -ésima linha se tenha  $a_i$  caixas e onde todas as linhas começam na mesma linha vertical.

Um SYT é uma forma de Ferrer de  $n$  caixas, como o da Figura 3.3, na qual cada uma contém um dos elementos de  $[n]$  de modo que cada caixa contém um número diferente e as linhas e colunas crescem para a direita e para baixo, respectivamente.

Saída	Pilha	Entrada
		3142
	3	142
	1	42
	3	
1	3	42
13		42
13	4	2
13	2	
	4	
132	4	
1324		

Figura 3.2: Os passos do algoritmo aplicado à permutação 3142 e que não a ordena porque contém o padrão 231.

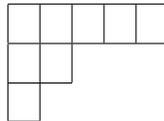


Figura 3.3: A forma de Ferrer da partição (5, 2, 1) com oito caixas.

Um diagrama como o da Figura 3.5 é também conhecido como um diagrama Young ou diagrama de Ferrer. Listando-se o número de caixas em cada linha obtém-se a partição  $\lambda=(5, 3, 2)$  relativa a 10, o número total de caixas do diagrama. O diagrama é dito como sendo de forma  $\lambda$ . O conteúdo do diagrama Young define uma ordem parcial sobre o conjunto de todas as partições, a qual é de fato uma estrutura reticulada conhecida como reticulado de Young. Listando-se o número de caixas em cada coluna obtém-se outra partição conhecida como a conjugada ou transposta de  $\lambda$ , que no exemplo corresponde a  $(3, 3, 2, 1, 1)$ . A pergunta natural é quantos SYT existem sobre uma dada forma de Ferrer ou, generalizando, sobre todas as formas consistindo de  $n$  caixas? Uma entrada de um SYT determina uma permutação de  $[n]$ .

**Definição 12** (p. 215) Sejam  $F$  uma forma de Ferrer e  $b$  uma caixa de  $F$ , então o *gancho* de  $b$  é o conjunto  $H_b$  formado por  $b$  e pelas caixas de  $F$  à direita (e mesma linha) e abaixo (e mesma coluna) de  $b$ . O tamanho de  $H_b$  é chamado de *comprimento do gancho* de  $b$  e é denotado por  $h_b$ .

A Figura 3.4 mostra o gancho  $H_b$  e os comprimentos associados a cada caixa da forma.

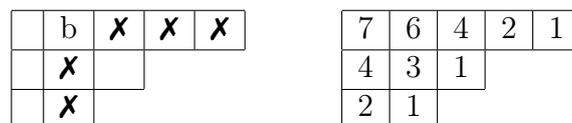


Figura 3.4: O gancho  $H_b$ , de comprimento seis, e todos os comprimentos da forma.

1	2	5	7	10
3	4	8		
6	9			

Figura 3.5: Um SYT com dez caixas.

Entre diversos resultados de teor probabilístico e enumerativo envolvendo os SYT, é dado, como exemplo, o que se segue.

**Teorema 3 (fórmula para os comprimentos dos ganchos)** (p. 215) Seja  $F$  uma forma de Ferrer com  $n$  caixas. Então o número de SYT com a forma  $F$  é igual a

$$\frac{n!}{\prod_b h_b},$$

calculado sobre todos os ganchos de  $F$ .

**Exemplo 17** Seja  $F$  um retângulo  $2 \times 3$ , então existem 5 SYT de forma  $F$  onde os comprimentos dos ganchos de  $F$  são, linha por linha, 4, 3 e 2 (os da primeira) e 3, 2 e 1 (os da segunda). Então, a fórmula para comprimento de gancho é verificada como

$$\frac{6!}{4 \cdot 3 \cdot 2 \cdot 3 \cdot 2 \cdot 1} = \frac{720}{144} = 5,$$

conforme as Figuras 3.6 e 3.7.

4	3	2
3	2	1

Figura 3.6: Os comprimentos dos ganchos da forma  $F = 2 \times 3$ .

1	2	3
4	5	6

1	2	4
3	5	6

1	2	5
3	4	6

1	3	4
2	5	6

1	3	5
2	4	6

Figura 3.7: Os cinco SYT da forma  $F = 2 \times 3$ .

### 3.6 Uma ordem sobre $S_n$

Há várias maneiras de definir uma ordem parcial sobre o conjunto das permutações de  $S_n$  para um  $n$  fixo e uma interessante e algumas de suas propriedades é dada pela seguinte definição.

**Definição 13** (p. 257) Seja  $P_n$  o conjunto parcialmente ordenado de todas as permutações no qual  $p < q$  se  $p$  pode ser obtido de  $q$  por uma série de operações onde cada uma intercambia as duas entradas de uma inversão (dada na Definição 4). Então,  $P_n$  é, também, chamada de *ordem de Bruhat* sobre  $S_n$ .

Uma operação que intercambia as duas entradas de uma inversão, por simplicidade, é chamada de *redução*. A ordem de Bruhat também é chamada de *ordem forte de Bruhat*.

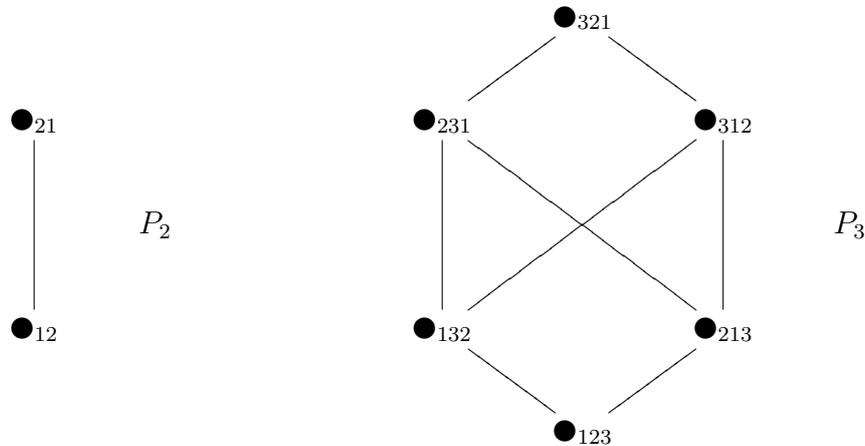


Figura 3.8: As ordens de Bruhat sobre  $S_2$  e  $S_3$ . Em  $S_3$ , as reduções em 321 se referem às inversões (3, 2) e (3, 1); em 231, (2, 1) e (3, 1); em 312, (3, 1) e (3, 2); em 132, (3, 2); e, em 213, à inversão (2, 1).

**Exemplo 18** Em um conjunto parcialmente ordenado se diz que  $y$  cobre  $x$  se  $x < y$ , mas não existe  $z$  tal que  $x < z < y$  ou, visualmente, quando  $y$  está “imediatamente acima” de  $x$ . Na Figura 3.8  $p = 231$  cobre  $q = 132$ , pois  $(p_1, p_3) = (2, 1)$ , sendo  $1 < 3$  e  $2 > 1$ , é uma inversão de 231 e 132 resulta de 231 após o intercâmbio das duas entradas de (2, 1).

**Proposição 2** (p. 258) Chamando-se de *comprimento de uma cadeia de um conjunto parcialmente ordenado* ao número dos seus elementos menos um e de *graduado* ao conjunto parcialmente ordenado cujas cadeias maximais têm o mesmo comprimento, tem-se que a ordem de Bruhat  $P_n$  é graduada.

Uma outra ordem parcial, baseada em inversões do tipo  $(y_i, y_{i+1})$ , onde  $y_i > y_{i+1}$ , é dada pela seguinte definição.

**Definição 14** (p. 260) Seja  $P'_n$  o conjunto parcialmente ordenado de todas as permutações, onde  $p < q$  se  $p$  pode ser obtido de  $q$  por uma série de reduções, considerando-se apenas as reduções formadas por entradas consecutivas. Então,  $P'_n$  é, também, chamada de *ordem fraca de Bruhat* sobre  $S_n$ .

Tem-se que a ordem fraca de Bruhat  $P'_n$  é, também, graduada, ou seja, as suas cadeias maximais tem o mesmo número de elementos, e que este número é  $\binom{n}{2}$ . Stanley provou em [35] e, um ano mais tarde, Greene e Edelman, em [7], que o número de cadeias maximais de  $P'_n$  é igual ao número de Standard Young Tableaux da forma em escada  $(n-1, n-2, \dots, 2, 1)$ , usando funções simétricas, que espera uma prova combinatorial. No caso da  $P'_3$ , da Figura 3.9, tem-se duas cadeias de comprimentos  $\binom{3}{2} = 3$  e dois SYT, de acordo com a Figura 3.10.

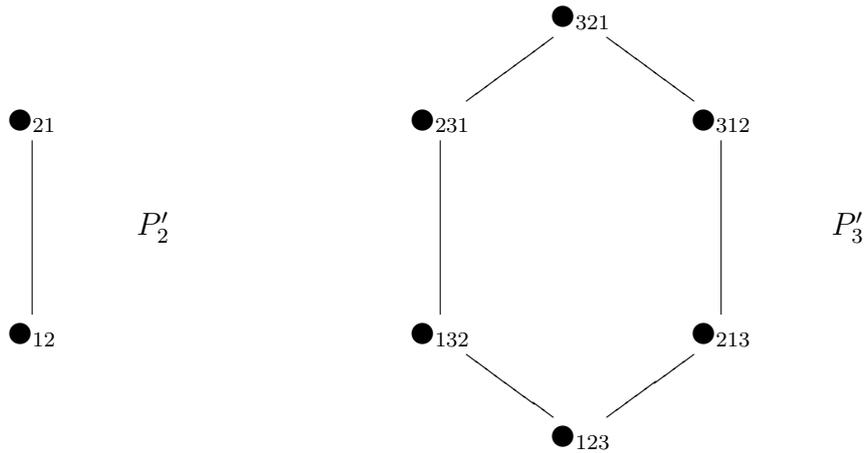


Figura 3.9: As ordens fracas de Bruhat sobre  $S_2$  e  $S_3$ . Em  $S_3$ , as reduções em 321 se referem às inversões (3, 2) e (2, 1); em 231, à inversão (3, 1); em 312, (3, 1); em 132, (3, 2); e, em 213, (2, 1).

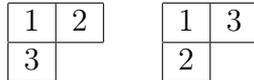


Figura 3.10: Os dois SYT da forma em escada (2, 1).

### 3.7 Recursos muito utilizados

Esta seção apresenta alguns recursos poderosos aplicados no estudo das permutações, que foram ferramentas importantes em muitas pesquisas e na demonstração dos resultados.

#### Número de Euler.

**Definição 15** (p. 4) Sejam  $n$  um número inteiro positivo e  $k < n$ . O *número de Euler*, denotado por  $A(n, k)$ , é definido como o número de permutações  $S_n$  com  $k-1$  descentes. Assim,  $A(n, k)$  também pode ser lido como “o número de permutações com  $k$  seqüências crescentes (dadas na Definição 1)”.

Durante sua longa história foi extensivamente estudado especialmente com respeito às suas propriedades relacionadas com a teoria dos números e os problemas combinatoriais, probabilísticos e estatísticos. A Tabela 3.2 contém os primeiros números de Euler.

**Exemplo 19** Existem quatro permutações em  $S_3$  com um desceste: 132, 213, 231 e 312. Assim,  $A(3, 2)=4$ . Similarmente,  $A(3, 3)=1$ , correspondendo à permutação 321, enquanto  $A(3, 1)=1$ , correspondendo à permutação 123.

#### Número de Stirling sem sinal de primeira espécie.

$n = 0$				1			
1				1		0	
2			1	1		0	
3			1	4		1	0
4		1	11	11		1	0
5	1	26	66	26		1	0

Tabela 3.2: Os valores de  $A(n, k)$  (número de Euler) para  $n \leq 5$ . Cada diagonal NE-SO contém os valores para um  $k$  fixo e a linha  $n$  começa com  $A(n, 1)$ . Nota-se que  $A(n, k+1) = A(n, n-k)$  ou, em outras palavras, os números de Euler são simétricos e tem-se que  $\sum_{k=1}^n A(n, k) = n!$ .

**Definição 16** (p. 85) Sejam  $n$  um número inteiro positivo e  $k \leq n$ . O *número de Stirling sem sinal de primeira espécie*, denotado por  $c(n, k)$ , é definido como o número de permutações de  $S_n$  com  $k$  ciclos (dados na seção 2.1).

Esse número é dado pela expressão: para todo inteiro positivo  $n > 1$ ,

$$c(n, n-2) = 2 \binom{n}{3} + \frac{1}{2} \binom{n}{2} \binom{n-2}{2},$$

e, recursivamente, pelo seguinte lema.

**Lema 3** (p. 86) Seja  $c(n, 0) = 0$ , se  $n \geq 1$ , e  $c(0, 0) = 1$ . Então,  $c(n, k) = c(n-1, k-1) + (n-1)c(n-1, k)$ .

Os números de Stirling sem sinal de primeira espécie  $c(n, 0), c(n, 1), \dots, c(n, n)$  também podem ser gerados como coeficientes do seguinte polinômio.

**Teorema 4** (p. 86) Para todo inteiro positivo  $n$ ,

$$x(x+1) \dots (x+n-1) = \sum_{k=0}^n c(n, k) x^k.$$

**Exemplo 20** Sendo  $n=4$  e  $k=2$ , tem-se que  $c(4, 2) = 11$ , correspondendo às permutações (em representação cíclica)  $(1\ 2\ 3)(4)$ ,  $(3\ 2\ 1)(4)$ ,  $(1\ 2\ 4)(3)$ ,  $(4\ 2\ 1)(3)$ ,  $(1\ 3\ 4)(2)$ ,  $(4\ 3\ 1)(2)$ ,  $(2\ 3\ 4)(1)$ ,  $(4\ 3\ 2)(1)$ ,  $(1\ 2)(3\ 4)$ ,  $(1\ 3)(2\ 4)$  e  $(1\ 4)(2\ 3)$ .

A Tabela 3.3 contém os primeiros números de Stirling sem sinal de primeira espécie.

### Números de Stirling com sinal de primeira espécie.

**Definição 17** (p. 90) O *número de Stirling com sinal de primeira espécie*, denotado por  $s=(n, k)$ , é definido como  $s(n, k) = (-1)^{n-k} c(n, k)$ .

A Tabela 3.4 contém os primeiros números de Stirling com sinal de primeira espécie.

### Número de Stirling de segunda espécie.

$n = 0$				1			
1			0	1			
2		0	1	1			
3		0	2	3	1		
4	0	6	11	6	1		
5	0	24	50	35	10	1	

Tabela 3.3: Os valores de  $c(n, k)$  (número de Stirling sem sinal de primeira espécie) para  $n \leq 5$ . Uma diagonal NE-SO contém os valores para um  $k$  fixo e a linha  $n$  começa com  $c(n, 0)$ .

$n = 0$				1			
1			0	1			
2		0	-1	1			
3		0	2	-3	1		
4	0	-6	11	-6	1		
5	0	24	-50	35	-10	1	

Tabela 3.4: Os valores de  $s(n, k)$  (número de Stirling com sinal de primeira espécie) para  $n \leq 5$ . A diagonal NE-SO contém os valores para um  $k$  fixo e a linha  $n$  começa com  $s(n, 0)$ .

**Definição 18** (p. 11) Sejam  $n$  um número inteiro positivo e  $k \leq n$ . O **número de Stirling de segunda espécie**, denotado por  $S(n, k)$ , é definido como o número de partições de  $[n]$  em  $k$  blocos (dados na Definição 2).

**Exemplo 21** Sendo  $n=4$  e  $k=3$ , o conjunto  $[4]$  tem seis partições de três blocos ( $\{1\}, \{2\}, \{3, 4\}$ ;  $\{1\}, \{3\}, \{2, 4\}$ ;  $\{1\}, \{4\}, \{2, 3\}$ ;  $\{2\}, \{3\}, \{1, 4\}$ ;  $\{2\}, \{4\}, \{1, 3\}$  e  $\{3\}, \{4\}, \{1, 2\}$ ). Assim,  $S(4, 3) = 6$ .

A Tabela 3.5 contém os primeiros números de Stirling de segunda espécie.

$n = 0$				1			
1			0	1			
2		0	1	1			
3		0	1	3	1		
4	0	1	7	6	1		
5	0	1	15	25	10	1	

Tabela 3.5: Os valores de  $S(n, k)$  (número de Stirling de segunda espécie) para  $n \leq 5$ . A diagonal NE-SO contém os valores de  $S(n, k)$  para um  $k$  fixo e a linha  $n$  começa com  $S(n, 0)$ .

### Sequências alternadas

**Definição 19** (p. 24) Sejam  $n$  um número inteiro positivo,  $k < n$ ,  $1 < i \leq n$  e  $p$  uma permutação de  $S_n$ . Diz-se que  $p$  tem  $k$  **sequências alternadas** (dadas na Definição 6)

se existem  $k-1$  índices  $i$  tais que  $p$  muda de direção nessas posições e o número de permutações com  $k$  seqüências alternadas é denotado por  $G(n, k)$ .

Na Tabela 3.6,  $G(n, k)$  é calculado recursivamente, por

$$G(n, k) = kG(n-1, k) + 2G(n-1, k-1) + (n-k)G(n-1, k-2),$$

onde  $G(1, 0) = 1$ ,  $G(1, k) = 0$  e  $k > 0$  (relação que foi primeiramente provada por André, em 1883). A origem desta linha de trabalho remonta o século XIX e, mais recentemente, D. E. Knuth, em [30], utilizou estes conceitos em ordenação e busca.

**Exemplo 22** A permutação 3561247, representada pelo diagrama da Figura 2.13 tem duas mudanças de direção (onde  $i = 3$  é um pico e  $i = 4$  é um vale) e três seqüências alternadas (356, 61 e 1247).

				2				
				2	4			
		2	12	58	32			
	2	28	236	300	122			
2	60							

Tabela 3.6: Os valores de  $G(n, k)$  (número de permutações com  $k$  seqüências alternadas) para  $n \leq 6$ . O primeiro valor da linha  $n$  é  $G(n, 1)$ . A diagonal NE-SO contém os valores de  $G(n, k)$  para um  $k$  fixo.

### Ponto-de-quebra e subsequência ordenada.

Sendo  $n$  um número inteiro positivo, ordenar uma permutação  $a_1 \dots a_n$  de  $S_n$  por reversões ( $1 \dots n$  representa a identidade) exige eliminar todos os pontos-de-quebra (dados na Definição 7) e, como cada reversão pode eliminar no máximo dois, tem-se que  $q(a)/2 \leq d(a)$  (distância dada na seção 2.3).

**Exemplo 23** A permutação 3412 tem quatro subsequências ordenadas (dadas na Definição 8) e tem-se que  $q(3412) = 3$ . Utilizando-se o símbolo  $\bullet$  para indicar um ponto-de-quebra, eles podem ser assinalados na permutação estendida, resultando  $0 \bullet 3 \ 4 \bullet 1 \ 2 \bullet 5$ .

Pode-se suspeitar que em uma menor série de reversões ordenando  $a$ , nenhuma delas corte uma subsequência ordenada. Isso, no entanto, é falso; por exemplo a permutação 3412 requer três reversões para ser ordenada sem cortes, no entanto, pode ser ordenada com dois, de acordo com a Figura 3.11.

O trabalho com permutações pode induzir a raciocínios não verdadeiros como ilustra o exemplo acima. Os algoritmos que atuam nos pontos-de-quebra podem não produzir resultados ótimos, pois normalmente são escritos para eliminar o maior número deles a cada reversão.

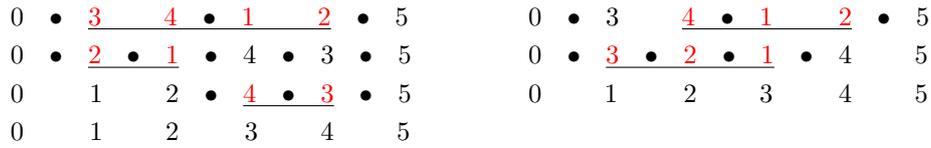


Figura 3.11: Ordenação de 3412 com três reversões, sem corte de subsequência ordenada, e com duas, cortando a 34.

### 3.8 Dedução não trivial de um resultado em ordenação

A artigo de Bóna e Flynn (em [24]), que é analisado a seguir, deduz um resultado notável em ordenação de permutações de  $S_n$ ,

$$b_n = \frac{n - \frac{1}{\lfloor (n+2)/2 \rfloor} - \sum_{i=2}^n \frac{1}{i}}{2},$$

de cunho combinatorial e probabilístico, para o “número médio de *intercâmbios de blocos* necessários para ordenar uma permutação”, que é o Teorema 9, no final desta seção. Inicia-se pela definição de dois tipos de grafos.

**Definição 20** O grafo  $\Gamma(\pi)$  de uma permutação  $\pi$  contém uma aresta  $(i, j)$  sempre que  $\pi(i) = j$  e tem uma decomposição única em ciclos disjuntos. O número de ciclos de  $\Gamma(\pi)$  é indicado por  $c(\Gamma(\pi))$ . Serão referenciados também por grafo  $\Gamma$  e por ciclos  $\Gamma$ .

Os ciclos de uma permutação  $\pi$  são equivalentes aos ciclos direcionados do grafo  $\Gamma(\pi)$ , assim, a permutação  $(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 1 & 6 & 2 & 5 & 7 & 3 \end{smallmatrix})$  tem três ciclos na representação (142)(367)(5) e em  $\Gamma(4162573)$ .

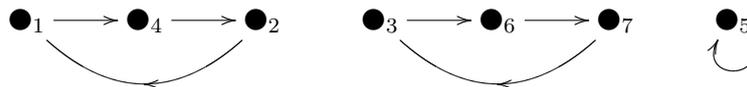


Figura 3.12: Grafo  $\Gamma(4162573)$ , onde  $c(\Gamma(4162573)) = 3$ .

**Definição 21** (grafo introduzido por Bafna e Pevzner, em [22], quando trabalhavam com transposições) O grafo  $G(p)$  da permutação  $p = p_1 p_2 \dots p_n$  é o grafo direcionado com vértices no conjunto  $\{0, 1, \dots, n\}$  e  $2(n+1)$  arestas, coloridas de preto ou pontilhadas como se segue. Supondo-se  $p_0 = 0$ :

1. Existe uma aresta preta de  $p_i$  para  $p_{i-1}$ ,  $0 \leq i \leq n$ , onde os índices são lidos módulo  $n+1$ , e
2. Existe uma aresta pontilhada de  $i$  para  $i+1$ ,  $0 \leq i \leq n$ , onde os índices são lidos módulo  $n+1$ .

Uma vez construído (conforme as Figuras 3.13, 3.15 e 3.17), o grafo  $G(p)$  pode ser decomposto em ciclos direcionados (subgrafos) da seguinte maneira (conforme as Figuras 3.14, e 3.16): toma-se um dos vértices  $p_i$  de  $G(p)$  e escolhe-se um tipo de aresta. Então,

alternando-se, a cada passo o tipo de aresta (em cada passo é percorrida uma aresta) percorre-se  $G(p)$  até retornar ao ponto de partida pela outra seta que não a de partida. Assim, é obtido um ciclo direcionado. Em seguida, exclui-se de  $G(p)$  as arestas que foram utilizadas neste ciclo e, se ainda sobraram arestas não utilizadas em  $G(p)$ , repete-se o procedimento para se obter outro ciclo e esse processo deve se repetir até que não mais restem arestas em  $G(p)$ . O grafo  $G(p)$  tem decomposição única em ciclos direcionados e o número desses ciclos é denotado por  $c(G(p))$ . Serão referenciados também por grafo  $G$  e por ciclos  $G$ .

**Observação 3 (Doignon e Labarre [6])**

- Sendo  $a = 4312 \in S_4$  e  $b = (01234)(02134) = (03)(1)(24) \in S_5$ , onde  $(01234)$  é o ciclo das arestas pontilhadas e  $(02134)$  o das pretas da Figura 3.13, tem-se que  $c(G(a)) = 3$  (da Figura 3.14) e que, portanto,  $c(G(a)) = c(G(b)) = 3$ .
  - A  $a$  está associado o produto  $b = (01234)(02134)$ .
- O ciclo  $(01234)$  é um 5-ciclo de  $S_5$  e é o ciclo das arestas pontilhadas de todas as  $4!$  permutações de  $S_4$ .
  - Em geral, o ciclo  $(01 \dots n)$  é um  $n+1$ -ciclo de  $S_{n+1}$  e é o ciclo das arestas pontilhadas de todas as  $n!$  permutações de  $S_n$ .
- Existem  $n!$   $n+1$ -ciclos em  $S_{n+1}$  e cada um é o ciclo das arestas pretas de uma única permutação de  $S_n$  e reciprocamente.
  - Existem, em  $S_{n+1}$ ,  $n!$  produtos da forma de  $(01 \dots n)(0c_n \dots c_1)$ , onde  $(0c_n \dots c_1)$  corresponde ao ciclo das arestas pretas de  $c = c_1 \dots c_n \in S_n$ .

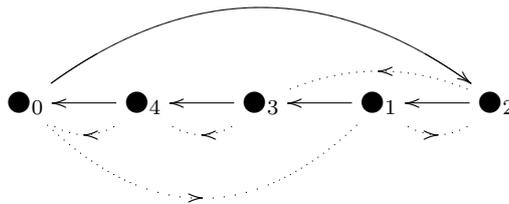


Figura 3.13: Construção de  $G(4312)$ , onde  $4312 \in S_4$ .

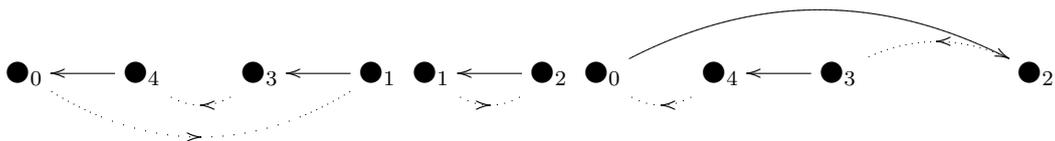


Figura 3.14: O grafo da Figura 3.13 se decompõe em  $c(G(4312)) = 3$  ciclos (ver, também, a Observação 3).

**Observação 4** Sendo  $(01234)$  o ciclo das arestas pontilhadas e  $(04321)$  o das arestas pretas da Figura 3.15 e  $c(G(a)) = 5$  (da Figura 3.16),  $c(G(a)) = c(\Gamma(b)) = 5$ , onde  $a = 1234$  e  $b = (01234)(04321) = (0)(1)(2)(3)(4) \in S_5$ . Em  $S_n$ , a permutação identidade é a única que se decompõe em  $n+1$  ciclos.

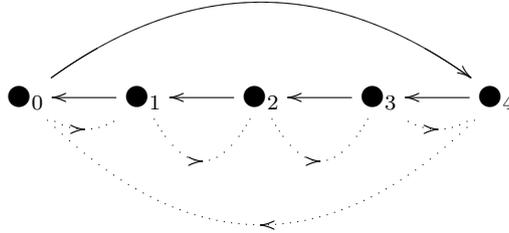


Figura 3.15: Construção de  $G(1234)$ , onde  $1234 \in S_4$  é a permutação identidade.

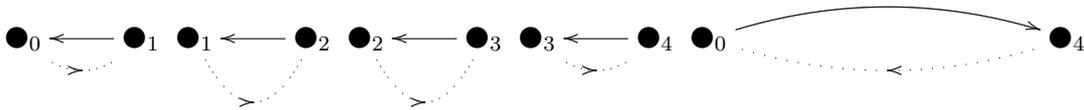


Figura 3.16: O grafo  $G(1234)$  da Figura 3.15 se decompõe em  $c(G(1234)) = 5$  ciclos (ver, também, a Observação 4).

**Observação 5** Sendo  $(01234)$  o ciclo das arestas pontilhadas e  $(03124)$  o das arestas pretas da Figura 3.17, onde  $c(G(a)) = 1$ ,  $c(G(a)) = c(\Gamma(b)) = 1$ , onde  $a = 1234$  e  $b = (01234)(03124) = (02143) \in S_5$ .

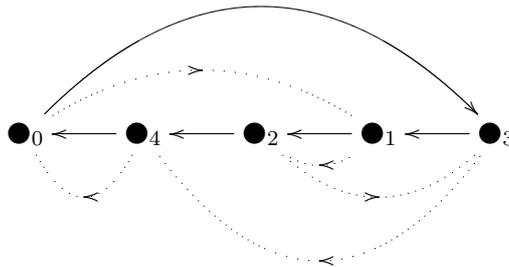


Figura 3.17: Construção de  $G(4213) \in S_4$ , que contém  $c(G(4213)) = 1$  único ciclo (ver, também, a Observação 5).

**Definição 22** (Hultman, em [29]) O número de Hultman  $\mathcal{S}_H(n, k)$  é o número de permutações  $p$  satisfazendo  $c(G(p)) = k$ , ou seja

$$\mathcal{S}_H(n, k) = |\{p \in S_n \mid c(G(p)) = k\}|.$$

Os primeiros números de Hultman [6] estão na Tabela 3.7. O Teorema 5 dá um resultado de Christie, o primeiro significativo sobre o tópico de ordenação por intercâmbios de blocos.

	$\mathcal{S}_H(n, 1)$	$\mathcal{S}_H(n, 2)$	$\mathcal{S}_H(n, 3)$	$\mathcal{S}_H(n, 4)$	$\mathcal{S}_H(n, 5)$	$\mathcal{S}_H(n, 6)$
$n=1$	0	1				
2	1	0	1			
3	0	5	0	1		
4	8	0	15	0	1	
5	0	84	0	35	0	1

Tabela 3.7: Os primeiros números de Hultman, onde os zeros resultam da paridade entre  $c(G(p))$  e  $c(G(i))$  (onde  $p, i \in S_n$  e  $i$  é a permutação identidade) e os uns (os últimos de cada linha) se referem às permutações identidades, conforme observações do Teorema 5.

**Teorema 5 (Christie, em [4]).** O número mínimo de intercâmbios de blocos necessários para ordenar uma permutação  $p$  é

$$\frac{n+1 - c(G(p))}{2}.$$

- $n+1$  (o número de ciclos da permutação identidade, que é a única que necessita de zero intercâmbios de blocos para ser ordenada) e  $c(G(p))$  têm a mesma paridade ( $\forall p \in S_n$ ). A diferença, um número par, é o número de ciclos que faltam para  $p$  ficar ordenada e
- um intercâmbio de blocos aplicado sobre  $p$  altera  $c(G(p))$  em dois.

**Observação 6** O Teorema 5 se aplica em  $S_n$ .

Uma consequência do Teorema 5 é que achar o número médio de intercâmbios de blocos necessários para ordenar uma permutação equivale a achar o número médio dos  $c(G(p))$  de todas.

**Teorema 6 (Doignon e Labarre, em [6]).** O número de Hultman  $\mathcal{S}_H(n, k)$  é igual ao número de maneiras de se obter o ciclo  $(1\ 2\ \dots\ n\ (n+1)) \in S_{n+1}$  como um produto  $qr$  de permutações, onde  $q \in S_{n+1}$  é qualquer ciclo de comprimento  $n+1$  e a permutação  $r \in S_{n+1}$  tem exatamente  $k$  ciclos, ou seja,  $c(\Gamma(r)) = k$ .

O Teorema 6 permite efetuar a contagem da Figura 3.18 e um seu resultado importante para este contexto é a seguinte bijeção (ver a Observação 3):

$$\begin{aligned} S_n &\rightarrow \{(1\ 2\ \dots\ n\ (n+1))q \mid q \in S_{n+1} \text{ é um } n+1\text{-ciclo}\} \\ a &\mapsto b = (1\ 2\ \dots\ n\ (n+1))q, \\ &\text{onde } c(G(a)) = c(\Gamma(b)). \end{aligned}$$

$$\mathcal{S}_H(n, k) = |\{qr \mid (1\ 2\ \dots\ n\ (n+1)) = qr, q \text{ é um } n+1\text{-ciclo e } c(\Gamma(r)) = k\}|$$

Figura 3.18: Contagem, com base nos grafos  $G$  (Observação 3), efetuada a partir do Teorema 6.

**Corolário 1** O número de Hultman  $\mathcal{S}_H(n, k)$  é igual ao número de  $n+1$ -ciclos  $q$  tais que o produto  $(1\ 2\ \dots\ n\ (n+1))q$  é uma permutação com exatamente  $k$  ciclos, isto é,  $c(\Gamma((1\ 2\ \dots\ n\ (n+1))q)) = k$ .

**Prova** Se  $(1\ 2\ \dots\ n\ (n+1))q = w$ , onde  $w$  tem  $k$  ciclos e  $q$  é um  $n+1$ -ciclo, tem-se

$$\begin{aligned} (1\ 2\ \dots\ n\ (n+1))q &= w \\ (1\ 2\ \dots\ n\ (n+1))qq^{-1} &= wq^{-1} \\ (1\ 2\ \dots\ n\ (n+1)) &= wq^{-1}. \end{aligned}$$

Sendo  $wq^{-1} = (q^{-1}q)wq^{-1} = q^{-1}(qwq^{-1}) = q^{-1}\bar{w}$ , vem que  $(1\ 2\ \dots\ n\ (n+1)) = q^{-1}\bar{w}$ . Agora, aplicando-se o Teorema 6, considerando que  $q^{-1}$  é um ciclo de comprimento  $n+1$  e que  $c(\Gamma(\bar{w})) = k$ , pois  $\bar{w}$  é um conjugado de  $w$ , segue o resultado do corolário.  $\square$

O Corolário 1 permite efetuar a contagem da Figura 3.19 e uma sua consequência é: achar (em  $S_n$ ) a média dos números  $c(G(p))$  sobre todas as permutações é equivalente a achar (em  $S_{n+1}$ ) a média dos números  $c(\Gamma((1\ 2\ \dots\ n\ (n+1))q))$ , onde  $q$  é um  $n+1$ -ciclo. Assim, será calculada a média do número de ciclos  $\Gamma$  do subconjunto  $\{(1\ 2\ \dots\ n\ (n+1))q \mid q \text{ é um } n+1\text{-ciclo}\}$ , que corresponde à média do total de ciclos  $G$  de  $S_n$ .

No próximo passo, considerando-se  $\{(1\ 2\ \dots\ n)z \mid z \text{ é um } n\text{-ciclo de } S_n\}$  como o con-

$$\mathcal{S}_H(n, k) = |\{q \mid q \text{ é um } n+1\text{-ciclo e } c(\Gamma((1\ 2\ \dots\ n\ (n+1))q)) = k\}|$$

Figura 3.19: Contagem, com base nos ciclos de arestas pretas dos grafos  $G$  (Observação 3), efetuada a partir do Corolário 1.

junto dos  $n-1!$  produtos da forma de  $b$ , da Observação 3, correspondentes a todas as permutações de  $S_{n-1}$ , será construído o conjunto  $\{(1\ 2\ \dots\ n\ (n+1))z' \mid z' \text{ é um } n+1\text{-ciclo de } S_{n+1}\}$ , correspondente às de  $S_n$ , pela inserção do elemento  $n+1$  em todas as posições de todos ciclos das arestas pretas  $z$  e a Proposição 3 mostra como varia o número dos ciclos  $\Gamma$  de cada grafo antes e depois da inserção de  $n+1$ .

Inserção da entrada  $n+1$  em um  $n$ -ciclo  $z \in S_n$ , obtendo a permutação  $z' \in S_{n+1}$ , de tal forma que  $n+1$  esteja entre duas entradas específicas,  $a$  e  $b$  (da Figura 3.20):

$$z'(i) = \begin{cases} z(i) & \text{se } i \notin \{a, n+1\} \\ n+1 & \text{se } i = a \\ b & \text{se } i = n+1. \end{cases}$$

Efeitos em  $c(\Gamma(s))$  da inserção de  $n+1$  em  $z$ :

**Proposição 3** Sejam  $a$ ,  $b$  e  $z$  definidas como acima e sejam  $s = (1\ 2\ \dots\ n)z$  e  $s' = (1\ 2\ \dots\ n\ (n+1))z'$ . Então, tem-se:

$$c(\Gamma(s')) = \begin{cases} c(\Gamma(s)) - 1 & \text{se } 2 \leq a \text{ e } a-1, z(1) \notin \text{mesmo ciclo de } s \\ c(\Gamma(s)) + 1 & \text{se } 2 \leq a \text{ e } a-1, z(1) \in \text{mesmo ciclo de } s \\ c(\Gamma(s)) + 1 & \text{se } a=1. \end{cases}$$

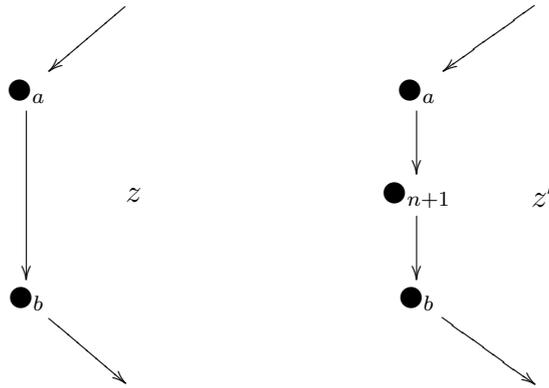


Figura 3.20: Como  $z'$  é obtida de  $z$ , de acordo com a Proposição 3.

$$\begin{array}{lll}
 z = (\dots a b \dots) & s = (1\ 2 \dots n)z & s = (1\ 2 \dots n)(\dots a b \dots) \\
 z' = (\dots a\ n+1\ b \dots) & s' = (1\ 2 \dots n\ n+1)z' & s' = (1\ 2 \dots n\ n+1)(\dots a\ n+1\ b \dots)
 \end{array}$$

**Definição 23** Seja  $n$  um número inteiro positivo, então  $TC_n = \{(x, y) \mid x, y \text{ são permutações que consistem de um } n\text{-ciclo cada}\}$ .

**Observação 7** Existem  $(n-1)!$   $n$ -ciclos em  $S_n$  e  $(n-1)!^2$  pares ordenados de  $n$ -ciclos em  $TC_n$ .

- A inserção de um elemento  $n+1$  em uma posição de  $z$  pode aumentar ou diminuir em um o número de ciclos,  $c(\Gamma(s))$ , do produto  $s = (1\ 2 \dots n)z$ .
- Saber quantas vezes esse número aumenta (ou diminui) equivale a verificar quantas vezes  $z(1)$  e  $a-1$  estão (ou não) no mesmo ciclo de  $s$ .

A seguinte questão decorre da Proposição 3 e da Observação 7.

**Questão** Sejam  $i$  e  $j$  dois elementos fixados do conjunto  $[n] = \{1, 2, \dots, n\}$ . Seleccionando-se um elemento qualquer  $(x, y)$  de  $TC_n$ , qual é a probabilidade de que o produto  $xy$  contenha  $i$  e  $j$  no mesmo ciclo?

Resultado recente de Stanley (prova não elementar que usou funções simétricas, funções geradoras exponenciais, integrais e uma fórmula de Boccara [1].)

**Teorema 7 (Stanley, em [36]).** Sejam

- $i$  e  $j$  dois elementos distintos do conjunto  $[n]$ , onde  $n > 1$ ,
- $(x, y)$  um elemento selecionado aleatoriamente de  $TC_n$  e
- $p(n)$  a probabilidade de que  $i$  e  $j$  estejam no mesmo ciclo de  $xy$ ,

então:

$$p(n) = \begin{cases} \frac{1}{2} & \text{se } n \text{ ímpar} \\ \frac{1}{2} - \frac{2}{(n-1)(n+2)} & \text{se } n \text{ par.} \end{cases}$$

Seja  $a_n$  o número médio de ciclos  $\Gamma$  de todas as permutações do conjunto  $\{(1\ 2\ \dots\ n)y \mid y \in TC_n\}$  ( $a_n$  é também, pelo Corolário 1, o número médio  $\frac{\sum_{i=1}^n (i S_H(n,i))}{(n-1)!}$  dos ciclos  $G$  de todas as permutações de  $S_{n-1}$ ), então, tem-se o Lema 4, que mostra como cresce  $a_n$ .

**Lema 4** Seja  $n > 1$ , então os números  $a_n$  crescem como segue:

1. se  $n = 2m + 2$ , então  $a_n = a_{n-1} + \frac{1}{n-1}$  e
2. se  $n = 2m + 1$ , então  $a_n = a_{n-1} + \frac{1}{n-1} - \frac{1}{m(m+1)}$ .

**Observação 8** A média  $a_n$  pode ser calculada em  $S_n$  (com base nos  $(n-1)!$   $n$ -ciclos e em ciclos  $\Gamma$ ) e utilizada em  $S_{n-1}$  (como o total de ciclos  $G$  de todas as permutações,  $(n-1)!$ ).

O Teorema 8 transforma os dois resultados do Lema 4 nas fórmulas não recursivas

$$a_n = a_1 + \sum_{i=1}^{n-1} \frac{1}{i} - \frac{n-2}{n} \quad (n \text{ pares}) \quad \text{e} \quad a_n = a_1 + \sum_{i=1}^{n-1} \frac{1}{i} - \frac{n-1}{n+1} \quad (n \text{ ímpares})$$

e, sendo  $a_1 = 1$ , as unifica.

**Teorema 8** Para todo inteiro positivo  $n$ , tem-se:

$$a_n = \frac{1}{\lfloor \frac{n+1}{2} \rfloor} + \sum_{i=1}^{n-1} \frac{1}{i}.$$

**Observação 9** O resultado do Teorema 8,  $a_n$ , dá a média do número de ciclos  $G$  em  $S_{n-1}$ .

O Teorema 9 ajusta o resultado do Teorema 8 (ver a Observação 9) à fórmula de Christie,  $\frac{n+1-c(G(p))}{2}$  (ver a Observação 6), obtendo

$$a_{n+1} = \frac{1}{\lfloor \frac{n+2}{2} \rfloor} + \sum_{i=1}^n \frac{1}{i} = \frac{1}{\lfloor \frac{n+2}{2} \rfloor} + \sum_{i=2}^n \frac{1}{i} + 1,$$

que, depois de substituído, é o Teorema 9, o resultado principal do artigo.

**Teorema 9** O número médio de intercâmbios de blocos necessários para ordenar uma permutação é:

$$b_n = \frac{n - \frac{1}{\lfloor \frac{n+2}{2} \rfloor} - \sum_{i=2}^n \frac{1}{i}}{2}.$$

Não há um resultado similar para as outras operações incluindo a de reversão.

# Capítulo 4

## Fundamentos de uma forma diferente de computar para o problema da distância de reversão

### 4.1 Uma solução diferente

Nesta seção será formalizado um modelo que permitirá considerar uma sequência finita como um “ciclo”, à semelhança dos ciclos das permutações da seção 2.1. Inicia pela criação do grupo de permutações  $\Phi_n$ , com suporte formado por classes de famílias de sequências, onde é definida uma leitura sequencial em um “ciclo”. O modelo se mostrará adequado para formalizar o problema da distância de reversão entre sequências finitas com o uso da operação de conjugação de um grupo de simetria.

#### Formalizando o grupo de permutações $\Phi_n$ .

Sejam  $n$  um inteiro positivo fixado,  $G$  um conjunto com  $n$  elementos e  $\Gamma = \{1, \dots, n\}$  o conjunto dos  $n$  primeiros números inteiros positivos.

Rotulem os elementos de  $G$  pelos elementos de  $\Gamma$  mediante uma bijeção  $\phi : \Gamma \rightarrow G$ ; uma vez fixada esta bijeção e abusando da notação o elemento  $\phi(i) \in G$  será indicado simplesmente por  $i$ . Assim, a partir deste ponto, e a menos de menção em contrário,  $G$  fica representado por  $\Gamma$ .

A seguir, é definida uma *sequência finita de elementos de  $\Gamma$*  (tendo em vista que todas as sequências referidas neste trabalho serão finitas, uma sequência finita será chamada apenas de *sequência*).

**Definição 24** *Sejam  $n$  um número inteiro positivo fixado,  $m \leq n$  e  $E = [m]$ , o conjunto dos  $m$  primeiros números inteiros positivos. A cada função injetora  $\pi : E \rightarrow \Gamma$ , está associada uma ***m-sequência*** de elementos de  $\Gamma$ ,  $\hat{\pi} = [\pi_1, \dots, \pi_m]$ , constituída pelas imagens de  $\pi$ , ordenadas segundo a ordem de  $E$ . A função  $\pi$  é chamada ***função de escolha*** da sequência  $\hat{\pi}$ . O número  $m$  é o ***comprimento*** e o conjunto  $\{\pi_1, \dots, \pi_m\}$  é o ***suporte*** da sequência  $\hat{\pi}$ . Uma *m-sequência* poderá também ser referida apenas por ***sequência***.*

Duas seqüências  $\hat{\pi}$  e  $\hat{\sigma}$  são ditas **disjuntas** se têm suportes disjuntos.

Reciprocamente, a toda  $m$ -seqüência  $\Sigma=[\Gamma_1, \dots, \Gamma_m]$  de elementos distintos de  $\Gamma$  corresponde uma função  $\sigma : E \rightarrow \Gamma$ ,  $i \mapsto \Gamma_i$  tal que  $\hat{\sigma}=\Sigma$ , ou seja,  $\sigma$  é a função de escolha de  $\Sigma$ . Assim, uma  $m$ -seqüência de elementos de  $\Gamma$  e a sua função de escolha são associadas.

A seguir, serão consideradas certas famílias de seqüências de elementos de  $\Gamma$  adequadas a este contexto.

**Definição 25** *Seja  $\mathcal{F}=\{\hat{\pi}_1, \dots, \hat{\pi}_t\}$  uma família de seqüências de elementos de  $\Gamma$ . A família  $\mathcal{F}$  é dita **disjunta** se as seqüências que a compõem são duas a duas disjuntas. Para simplificar a notação, a família  $\mathcal{F}$  será denotada apenas por  $\hat{\pi}_1 \dots \hat{\pi}_t$ .*

**Definição 26** *O conjunto  $\Gamma_n$  é definido como o conjunto de todas as famílias disjuntas de  $t$  seqüências,  $1 \leq t \leq n$ , tais que  $m_1 + \dots + m_t = n$ , onde  $m_i$  é o comprimento da seqüência  $\hat{\pi}_i$  (eventualmente,  $m_i=1$ , para alguns valores de  $i$ ,  $1 \leq i \leq t$ ). Um elemento de  $\Gamma_n$  será chamado apenas de **família**.*

**Observação 10** Admitindo-se que as seqüências de comprimento 1 que compõem uma dada família  $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_t\} \in \Gamma_n$  sejam  $\widehat{\pi_{s+1}}, \dots, \widehat{\pi_t}$ , por simplicidade,  $\hat{\Pi}$  será representada por  $\hat{\pi}_1 \dots \hat{\pi}_s$ , omitindo-se as seqüências de comprimento 1. A ordem em que as seqüências aparecem na expressão  $\hat{\pi}_1 \dots \hat{\pi}_s$ , sendo esta um conjunto, não é relevante e, como se verá, atende aos propósitos deste trabalho.

**Exemplo 24** Sendo  $\Gamma = \{1, 2, 3, 4, 5, 6\}$  e  $m = 3$ , pela aplicação da Definição 24, pode ser tomada a seqüência  $\hat{\pi}_1 = [\pi_{11}, \pi_{12}, \pi_{13}] = [1, 3, 4]$ , associada à função  $\pi_1 = \{(1, 1), (2, 3), (3, 4)\}$ . Da mesma forma, se  $m=2$ ,  $\hat{\pi}_2 = [\pi_{21}, \pi_{22}] = [2, 5]$ , associada à função  $\pi_2 = \{(1, 2), (2, 5)\}$ . Pode-se tomar as seqüências  $\hat{\pi}_1$  e  $\hat{\pi}_2$  para formar a família  $\hat{\Pi} = \hat{\pi}_1 \hat{\pi}_2 = [1, 3, 4][2, 5]$ . Esta é a família  $[1, 3, 4][2, 5][6]$ , onde foi omitida a seqüência  $[6]$ , de comprimento 1. O conjunto  $\{1, 3, 4\}$  dos elementos de  $\hat{\pi}_1$  é disjunto do conjunto  $\{2, 5\}$  dos elementos de  $\hat{\pi}_2$ . Exemplos de destaque neste trabalho são as famílias  $\hat{E}=[1][2][3][4][5][6]$ , que é chamada de **família identidade**, e  $\hat{I}=[1, 2, 3, 4, 5, 6]$ , que é chamada de **família ordenada**.

A seguir, será definida uma função que faz a leitura seqüencial de uma seqüência através da leitura seqüencial da imagem da função a ela associada.

**Definição 27** *A **leitura seqüencial de uma seqüência**  $\hat{\pi}=[\pi_1, \dots, \pi_m]$ ,  $m \leq n$ , associada da função  $\pi$ , é a função  $\bar{\pi}$ , definida como*

$$\begin{aligned} \bar{\pi}(\pi_i) &= \pi_{i+1} & \text{para} & \quad i=1, \dots, m-1, \\ \bar{\pi}(\pi_m) &= \pi_1 \text{ e} \\ \bar{\pi}(j) &= j & \text{para} & \quad j \in [n] - \{\pi_1, \dots, \pi_m\}, \end{aligned}$$

ou seja, é o  $m$ -ciclo  $\bar{\pi}=(\pi_1 \dots \pi_m) \in S_n$ .

Em geral, pode ser feita a leitura seqüencial de uma família qualquer de seqüências, conforme a definição a seguir.

**Definição 28** Seja  $\widehat{\Pi}=\widehat{\pi}_1 \dots \widehat{\pi}_t$  uma família de seqüências. A *leitura sequencial de uma família de seqüências*  $\widehat{\Pi}$  é, por definição, a permutação  $\overline{\Pi}=\overline{\pi}_1 \dots \overline{\pi}_t$ , isto é,  $\overline{\Pi}$  é o produto das leituras sequenciais das seqüências componentes de  $\widehat{\Pi}$ .

**Observação 11** Uma seqüência  $\widehat{\pi}=[\pi_1, \dots, \pi_m]$  com a leitura sequencial da Definição 27 também será chamada de “*m-ciclo*” ou, genericamente, apenas de “*ciclo*” e o seu comprimento é o comprimento da seqüência. Se  $m=n$ , diz-se que o “*m-ciclo*” ou, genericamente, “*ciclo*”, é “*circular*”.

**Observação 12** Pela Definição 25 e subsequente Observação 10 a permutação  $\overline{\Pi}$  independe da ordem da leitura sequencial dos ciclos componentes; os eventuais ciclos de comprimento 1 que aparecem em  $\widehat{\Pi}$  não interferem na decomposição cíclica de  $\overline{\Pi}$ .

**Observação 13** “Ciclos” diferentes podem ter leituras sequenciais iguais e, consequentemente, famílias diferentes também.

**Exemplo 25** Sendo  $\Gamma=\{1, 2, 3, 4, 5, 6\}$ , então, os “3-ciclos”  $[1, 3, 4]$ ,  $[3, 4, 1]$  e  $[4, 1, 3]$  são distintos e têm como leitura sequencial o 3-ciclo  $(1\ 3\ 4)$ . Analogamente, as famílias  $[1, 3, 4][2, 5][6]$ ,  $[3, 4, 1][2, 5][6]$ ,  $[4, 1, 3][2, 5][6]$ ,  $[1, 3, 4][5, 2][6]$ ,  $[3, 4, 1][5, 2][6]$  e  $[4, 1, 3][5, 2][6]$  são distintas e têm como leitura sequencial a permutação  $(1\ 3\ 4)(2\ 5)$ , mas é fácil caracterizá-las.

As famílias de mesma leitura sequencial podem ser agrupadas particionando  $\Gamma_n$  em classes disjuntas o que define uma relação de equivalência sobre  $\Gamma_n$ . Simbolizando esta relação por  $\sim$ , o conjunto constituído pelas classes de equivalência de  $\Gamma_n$  módulo  $\sim$ ,  $\Gamma_n/\sim$ , passa a ser chamado de  $\Phi_n$  e será chamada de  $\widetilde{\Lambda}$  a classe a que pertence a família  $\widehat{\Lambda}$ . Dada uma classe  $\widetilde{\Lambda}$ , o número de suas famílias é igual ao produto dos comprimentos dos ciclos da leitura sequencial que a ela corresponde.

**Definição 29** O conjunto  $\Phi_n$  é definido como o conjunto das classes de equivalência de  $\Gamma_n$  módulo  $\sim$ .

**Exemplo 26** A família  $\widehat{\Lambda}=[1, 3, 4][2, 5][6]$ , cujos “ciclos” têm, respectivamente, comprimentos 3, 2 e 1, pertence a classe  $\widetilde{\Lambda}$  formada pelas famílias cuja leitura sequencial é  $(1\ 3\ 4)(2\ 5)$  e que tem 6 elementos (as seis famílias do Exemplo 25).

À cada classe de equivalência  $\widetilde{\Lambda}$  está associada uma leitura sequencial  $\overline{\Lambda}$  e, consequentemente, existe uma bijeção  $\varphi : \Phi_n \rightarrow S_n$ ,  $\widetilde{\Lambda} \mapsto \overline{\Lambda}$ , onde  $S_n$  indica o grupo simétrico de grau  $n$ . Também, pode ser definida uma função  $\gamma : \Gamma_n \rightarrow \Phi_n$ ,  $\widehat{\Lambda} \mapsto \widetilde{\Lambda}$ , que associa uma família à classe de equivalência relativa à sua leitura sequencial. A função  $\gamma$  é sobrejetora pois a imagem inversa de qualquer elemento  $\widetilde{\Lambda} \in \Phi_n$ , chamado de  $\gamma^{-1}(\widetilde{\Lambda})$ , tem pelo menos o elemento  $\widehat{\Lambda}$ . A bijeção  $\varphi$  define uma multiplicação em  $\Phi_n$  da maneira que se segue.

**Definição 30** Sejam  $\widetilde{A}, \widetilde{B} \in \Phi_n$ . Então, a operação  $\Phi_n \times \Phi_n \rightarrow \Phi_n$ , chamada de *multiplicação*, é definida por

$$\widetilde{A} \cdot \widetilde{B} := \varphi^{-1}(\varphi(\widetilde{A}) \varphi(\widetilde{B})) = \varphi^{-1}(\overline{A} \overline{B}),$$

para todos  $\widetilde{A}, \widetilde{B} \in \Phi_n$ .

**Observação 14** Multiplicar duas classes,  $\tilde{A}$  e  $\tilde{B}$ , em  $\Phi_n$ , equivale a multiplicar as suas leituras sequenciais,  $\overline{A}$  e  $\overline{B}$ , em  $S_n$ , e, depois, voltar para  $\Phi_n$ . Por simplicidade, o produto  $\tilde{A} \cdot \tilde{B}$  será indicado por  $\tilde{A}\tilde{B}$ .

**Observação 15** A multiplicação de duas classes é feita a partir de um representante de cada uma delas, utilizando a função  $\gamma$ , independentemente, portanto, da escolha desses representantes e está bem definida por ser efetuada em  $S_n$ .

**Exemplo 27** Sendo  $\Gamma = \{1, 2, 3, 4, 5, 6\}$ ,  $\hat{A} = [1, 3, 2, 5]$  e  $\hat{B} = [2, 6, 4]$ , então  $\tilde{A}\tilde{B} = \tilde{\Delta}$ , onde  $\tilde{\Delta} = [1, 3, 2, 6, 4, 5]$  e  $\tilde{A}^{-1}$  é a classe da família  $[1, 5, 2, 3][4][6]$ .

A proposição seguinte é uma consequência imediata desta definição.

**Proposição 4** Com a multiplicação acima definida,  $\Phi_n$  é um grupo e  $\varphi$  é um isomorfismo de  $\Phi_n$  em  $S_n$ , onde o **elemento neutro** de  $\Phi_n$  é a classe de equivalência unitária  $\tilde{E}$  constituída pela **família identidade** e a **classe de equivalência inversa** de uma classe  $\tilde{A}$  é a classe  $\tilde{A}^{-1} = \varphi^{-1}(\overline{A}^{-1})$ .

*Demonstração.* A função  $\varphi$ , pela definição de multiplicação, acima, é um homomorfismo, pois  $\varphi(\tilde{A}\tilde{B}) = \varphi(\tilde{A})\varphi(\tilde{B})$  e, sendo uma bijeção, é, também, um isomorfismo. Assim,  $\Phi_n$  é um grupo. Aplicando a definição aos elementos  $\tilde{E}$ ,  $\tilde{A}$  e  $\tilde{A}^{-1}$ , tem-se que  $\tilde{A}\tilde{E} = \varphi^{-1}(\overline{A} \overline{E}) = \varphi^{-1}(\overline{A}) = \tilde{A} = \varphi^{-1}(\overline{A}) = \varphi^{-1}(\overline{E} \overline{A}) = \tilde{E}\tilde{A}$  e que  $\tilde{A}\tilde{A}^{-1} = \varphi^{-1}(\overline{A} \overline{A}^{-1}) = \varphi^{-1}(\overline{E}) = \tilde{E} = \varphi^{-1}(\overline{E}) = \varphi^{-1}(\overline{A}^{-1} \overline{A}) = \tilde{A}^{-1}\tilde{A}$ , o que completa a demonstração da proposição.  $\square$

**Definição 31** Sejam  $\tilde{\Lambda} \in \Phi_n$  e  $a \in S_n$ . Então fica definida uma operação  $S_n \times \Phi_n \rightarrow \Phi_n$ , chamada de **multiplicação à esquerda de uma permutação por uma classe de famílias**, que é indicada por

$$a\tilde{\Lambda} = \varphi^{-1}(a\varphi(\tilde{\Lambda})) = \varphi^{-1}(a\overline{\Lambda}).$$

Analogamente, pode ser definida uma **multiplicação à direita de uma permutação por uma classe de famílias**:

$$\Phi_n \times S_n \rightarrow \Phi_n, (\tilde{\Lambda}, a) \mapsto \tilde{\Lambda}a = \varphi^{-1}(\varphi(\tilde{\Lambda})a) = \varphi^{-1}(\overline{\Lambda}a).$$

Fica definida, também, uma **conjugação (à esquerda) de uma permutação sobre uma classe de famílias**,  $S_n \times \Phi_n \rightarrow \Phi_n, (a, \tilde{\Lambda}) \mapsto {}^a\tilde{\Lambda} = \varphi^{-1}(a\varphi(\tilde{\Lambda})a^{-1}) = \varphi^{-1}(a\overline{\Lambda}a^{-1}) = \varphi^{-1}({}^a\overline{\Lambda})$ . Uma conjugação à direita pode ser definida de maneira análoga. É fácil ver que as multiplicações acima definidas, à esquerda e à direita, comutam com o isomorfismo  $\varphi$ , isto é:  $\varphi(a\tilde{\Lambda}) = a\varphi(\tilde{\Lambda})$  e  $\varphi(\tilde{\Lambda}a) = \varphi(\tilde{\Lambda})a$ . Com isto, obtém-se que  ${}^a\tilde{\Lambda} = \varphi^{-1}(a\varphi(\tilde{\Lambda})a^{-1}) = \varphi^{-1}(\varphi(a\tilde{\Lambda})a^{-1}) = \varphi^{-1}(\varphi(a\overline{\Lambda}a^{-1})) = a\tilde{\Lambda}a^{-1}$ .

**Observação 16** A classe  $\tilde{A}$  é o resultado de  $\tilde{B}$  pela conjugação da permutação  $a$  quando  ${}^a\tilde{B} = \tilde{A}$ . A permutação  $a$  será referida por **o conjugador que leva  $\tilde{B}$  em  $\tilde{A}$**  ou **o conjugador de  $\tilde{B}$  para  $\tilde{A}$** . Quando  $\tilde{A}$  for igual a  $\tilde{I}$ , a classe da família ordenada  $\tilde{I}$ ,  $a$  será referida por **conjugador de  $\tilde{B}$  (para  $\tilde{I}$ )**, para distinguir este conjugador de uma permutação-reversão, que será definida adiante (Definição 33) e que também é um conjugador.

**Observação 17** Multiplicar e conjugar uma classe  $\tilde{\Lambda}$  por  $a$ , pela Definição 31, equivale, respectivamente, a multiplicar e a conjugar a leitura sequencial  $\bar{\Lambda}$  da classe  $\tilde{\Lambda}$  por  $a$  em  $S_n$  e depois voltar para  $\Phi_n$ , de forma análoga a da Observação 14.

**Exemplo 28** Dados  $\Gamma = \{1, 2, 3, 4, 5, 6\}$ ,  $\hat{\Sigma} = [6, 3, 1, 4, 2, 5]$ ,  $\hat{I} = [1, 2, 3, 4, 5, 6] \in \Gamma_6$  e a permutação  $a = (6\ 1\ 3\ 2\ 5)(4) \in S_6$ , então  ${}^a\tilde{\Sigma}$  é a classe da família  $[1, 2, 3, 4, 5, 6] = \hat{I}$  e  $a$  é o conjugador de  $\tilde{\Sigma}$  (para  $\hat{I}$ ).

**Exemplo 29** Dados  $\Gamma = \{1, 2, 3, 4, 5, 6\}$ , então  $\hat{\Lambda} = [1, 3, 2, 5, 4, 6] \in \Gamma_6$  e  $a = (3\ 1\ 2) \in S_6$ ,  $\tilde{\Lambda}a$ ,  $a\tilde{\Lambda}$  e  ${}^a\tilde{\Lambda}$  são, respectivamente, as classes das famílias  $[1, 5, 4, 6][2][3]$ ,  $[1][2, 5, 4, 6][3]$  e  $[1, 3, 5, 4, 6, 2]$ .

**Observação 18** A conjugação simula uma reversão em uma sequência genérica (que poderá estar representando um cromossomo, por exemplo), onde a ordem dos seus elementos é importante. Para os propósitos deste trabalho, tem relevância a escolha dos representantes das classes de  $\Phi_n$ , resultantes da conjugação (à esquerda) de uma permutação sobre uma classe de famílias e o procedimento de escolha desses representantes será detalhado na subseção 4.2.3, Convenção 1.

O grupo  $\Phi_n$ , como se verá, é um modelo adequado para representar e transformar por reversões as entidades representáveis por meio de sequências.

#### Mudança de notação.

Daqui para frente e para simplificar a notação, serão omitidos os símbolos  $\hat{\quad}$  e  $\tilde{\quad}$  e adotadas letras gregas minúsculas para representar: uma sequência  $\hat{\pi}$ , que será referida simplesmente por sequência (ou por “ciclo” ou, agregando o seu comprimento,  $m$ , ao seu nome, por “ $m$ -ciclo”)  $\pi$ ; uma família de sequências  $\hat{\Pi}$ , por família  $\pi$  e uma classe de famílias de sequências  $\tilde{\Pi}$ , por classe  $\pi$ . Se houver necessidade de mais clareza, será feita uma observação nesse sentido.

## 4.2 A motivação para a formalização deste modelo e a equivalência entre este e outros formalismos

Nesta seção os cromossomos serão representados no modelo; serão dados alguns conceitos básicos sobre rearranjos de genomas e as representações de um cromossomo e das respectivas operações de reversão em diferentes formalismos e será mostrada a equivalência deste com os outros diferentes formalismos apresentados na simulação de uma reversão biológica.

### 4.2.1 Aplicando este modelo a um caso concreto: representando cromossomos

A escolha de sequências nesta formalização foi motivada pelo fato de que os blocos de genes que aparecem (ordenadamente) num cromossomo ficam adequadamente representados por uma sequência finita e sem repetições de elementos de um dado conjunto. Com base no que foi visto na seção 4.1, se  $G$  for um conjunto com  $n$  blocos de genes

e  $\Gamma = \{1, \dots, n\}$  o conjunto que rotula  $G$  pela bijeção  $\phi : \Gamma \rightarrow G$ , então, o “ $m$ -ciclo”  $[\pi_1, \dots, \pi_m]$  de elementos de  $\Gamma$ , chamado de  $\pi$ , representa um cromossomo com blocos de genes em  $G$ .

**Exemplo 30** Seja  $G = \{a, b, c, d, e\}$  o conjunto dos cinco blocos de genes que constituem os organismos homólogos repolho (*Brassica campestris*) e nabo (*Brassica oleracea*). De acordo com a convenção acima estabelecida,  $G$  fica representado pelo conjunto  $\Gamma = \{1, 2, 3, 4, 5\}$ . Admitindo que esses blocos apareçam no cromossomo do repolho na ordem  $a, b, c, d$  e  $e$ , o cromossomo do repolho,  $a b c d e$ , fica representado pelo “5-ciclo”  $\iota = [1, 2, 3, 4, 5]$ , ou seja, o cromossomo do repolho está associado à função  $\iota : E \rightarrow \Gamma$ , dada por  $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}$ , que faz corresponder a cada elemento de  $E = \{1, 2, 3, 4, 5\}$  o elemento de  $\Gamma$  que representa o  $i$ -ésimo bloco de genes do repolho. Uma vez estabelecida esta ordem, com base no cromossomo do repolho, obtém-se que o “5-ciclo”  $\pi = [1, 5, 4, 3, 2]$ , associado à função  $\pi : E \rightarrow \Gamma$ , dada por  $\{(1, 1), (2, 5), (3, 4), (4, 3), (5, 2)\}$ , representa o cromossomo do nabo,  $a e d c b$ .

O exemplo seguinte mostra como utilizar apenas parte do conjunto de blocos de genes para constituir cromossomos (sequências ou “ciclos”).

**Exemplo 31** Mantendo o mesmo conjunto de blocos de genes do exemplo 30, se  $E = \{1, 2, 3\}$ , a função  $\pi : E \rightarrow \Gamma$ , dada por  $\{(1, 5), (2, 3), (3, 2)\}$ , define o “3-ciclo”  $\pi = [5, 3, 2]$ , que representa o cromossomo  $e c b$ .

#### 4.2.2 Conceitos básicos em rearranjos de genomas por reversões

Os conceitos básicos sobre rearranjos de genomas apresentados nesta subseção são encontrados em [15] e [20]. Como foi visto na seção 2.2, existem dois tipos de reversões agindo nos cromossomos dos organismos, sem e com sinais. O foco deste trabalho está na primeira, sendo o seu objetivo saber quais e o número mínimo de reversões que levam um cromossomo em um outro. Isso é sempre possível, pelo lema que se segue, cuja demonstração decorre de uma proposição, que tem uma prova em [31], onde um cromossomo, representado pela palavra  $\pi_1 \dots \pi_n$ , denotada por  $\pi_{str}$ , pode ser ordenado, isto é, transformando na palavra  $1 \dots n$ , mediante a aplicação de uma certa sequência  $\rho = \rho_1 \dots \rho_r$  de reversões, indicada como  $\pi_{str}\rho = 1 \dots n$ .

**Lema 5** *Dados dois cromossomos  $\alpha$  e  $\beta$  existe uma sequência de reversões,  $\rho_1, \dots, \rho_m$ , tais que  $\alpha\rho_1 \dots \rho_m = \beta$ .*

*Demonstração.* Pela proposição acima citada, existem duas sequências de reversões,  $\rho_\alpha$  e  $\rho_\beta$ , tais que  $\alpha\rho_\alpha = \beta\rho_\beta$ . Assim, admitindo-se que uma sequência de reversões  $\rho = \rho_1 \dots \rho_r$  tenha  $\rho^{-1} = \rho_r^{-1} \dots \rho_1^{-1}$  como inversa, tem-se que  $\alpha\rho_\alpha\rho_\beta^{-1} = \beta\rho_\beta\rho_\beta^{-1} = \beta$ . Daí,  $\rho_\alpha\rho_\beta^{-1}$  é a sequência de reversões procurada.  $\square$

**Definição 32** A distância de reversão entre os cromossomos  $\alpha$  e  $\beta$  é definida como o menor número de reversões que levam  $\alpha$  em  $\beta$  e é denotada por  $d(\alpha, \beta)$ ,  $d_\beta(\alpha)$  ou, simplesmente,  $d(\alpha)$ , quando  $\beta$  é conhecido.

### 4.2.3 Equivalência entre formalismos

A seguir, a mesma reversão será simulada em dois formalismos e, em seguida, pela operação de conjugação de um grupo de permutações.

#### Reversão em outros formalismos.

Em [31], uma reversão é uma permutação  $\rho \in S_n$  da forma  $(i\ j)(i+1\ j-1) \dots (i + \lfloor (j-i)/2 \rfloor + 1\ i + \lceil (j-i)/2 \rceil + 1)$ ,  $1 \leq i < j \leq n$ , é aplicada a palavras e permuta elementos nas imagens do intervalo  $[i, j]$ , de uma palavra  $\pi_1 \dots \pi_{i-1} \pi_i \pi_{i+1} \dots \pi_{j-1} \pi_j \pi_{j+1} \dots \pi_n$ , denotada por  $\pi_{str}$ , onde  $\pi_i \in \{1, \dots, n\}$ ,  $1 \leq i < j \leq n$  e  $\pi_i \neq \pi_j, \forall i \neq j$ , isto é,  $\pi_{str}\rho(i) = \pi_{str}(j), \pi_{str}\rho(j) = \pi_{str}(i)$ , etc., e, portanto, dada a

$$\pi_{str} = \pi_1 \dots \pi_{i-1} \underline{\pi_i \pi_{i+1} \dots \pi_{j-1} \pi_j \pi_{j+1} \dots \pi_n}, \quad (4.1)$$

a reversão  $\rho$  reverte a subpalavra do intervalo  $[i\ j]$  de  $\pi_{str}$  e resulta

$$\pi_{str}\rho = \pi_1 \dots \pi_{i-1} \underline{\pi_j \pi_{j-1} \dots \pi_{i+1} \pi_i \pi_{j+1} \dots \pi_n},$$

que é a reversão sublinhada. Em [15], página 220, onde uma reversão é aplicada a listas, dado um cromossomo, representado pela lista

$$\pi = (\pi_1, \dots, \pi_{i-1}, \underline{\pi_i, \pi_{i+1}, \dots, \pi_{j-1}, \pi_j}, \pi_{j+1}, \dots, \pi_n), \quad (4.2)$$

sendo  $i$  e  $j$  dois índices,  $1 \leq i < j \leq n$ , a reversão do trecho contíguo que vai de  $\pi_i$  até  $\pi_j$ , indicada por  $[i, j]$ , é dada por  $\pi[i, j](k) = \pi_{i+j-k}$ , se  $i \leq k \leq j$ , e  $\pi[i, j](k) = \pi_k$ , nos demais casos. Aplicando o reverso  $[i, j]$  a  $\pi$  resulta

$$\pi[i, j] = (\pi_1, \dots, \pi_{i-1}, \underline{\pi_j, \pi_{j-1}, \dots, \pi_{i+1}, \pi_i}, \pi_{j+1}, \dots, \pi_n),$$

que é a reversão sublinhada.

#### Reversão pela operação de conjugação de um grupo de permutações.

Neste modelo uma reversão é realizada em um “ciclo”, pela aplicação de uma operação de conjugação de um grupo de permutações (com base no modelo da seção 4.1).

O conjugador, que é uma permutação de  $S_n$ , e o elemento conjugado, um “ $n$ -ciclo” de  $\Phi_n$ , participam como termos ou fatores da operação de conjugação de um grupo (num abuso de linguagem, pois uma conjugação é o resultado de duas multiplicações consecutivas) e o “ $n$ -ciclo” resultante é o produto que incorpora uma reversão.

Sendo a classe  $\pi$  um “ $n$ -ciclo”, o conjugador,  $p$ , que realiza a reversão do trecho de posições  $i$  a  $j$  de  $\pi$ , é dado pela seguinte definição.

**Definição 33** Uma permutação  $p$  que realiza a reversão da porção contígua de posições  $i$  a  $j$ ,  $1 \leq i < j \leq n$ , do “ $n$ -ciclo”

$$\pi = [\pi_1, \dots, \pi_{i-1}, \underline{\pi_i, \pi_{i+1}, \dots, \pi_{j-1}, \pi_j}, \pi_{j+1}, \dots, \pi_n], \quad (4.3)$$

mediante a operação de conjugação de um grupo de permutações,  ${}^p\pi$ , é definida como:

$$p = \prod_{k=0}^{\lfloor (j-i)/2 \rfloor} (\pi_{i+k} \pi_{j-k}).$$

**Observação 19** A permutação  $p$  é um produto de transposições algébricas disjuntas e, portanto,  $p=p^{-1}$  (de acordo com a Observação 2).

**Observação 20** Neste modelo, um cromossomo, que é representado por um “ $n$ -ciclo”, e as permutações construídas conforme a Definição 33 são elementos de grupos de permutações isomorfos e utilizados como termos ou fatores da operação de multiplicação dos grupos na realização de reversões no problema da distância. Essas são particularidade deste formalismo e refletem a diferença de pontos de vista entre este e os demais na abordagem do problema.

A permutação  $p$  da Definição 33 proporciona o resultado esperado, pois a conjugação de  $\pi$  por  $p$  resulta:

$${}^p\pi = p\pi p^{-1} = [\pi_1, \dots, \pi_{i-1}, \underline{\pi_j, \pi_{j-1}, \dots, \pi_{i+1}}, \pi_i, \pi_{j+1}, \dots, \pi_n],$$

que é a reversão sublinhada. Sendo um  ${}^p\pi$  um “ $n$ -ciclo”, conforme a Observação 20 e a Definição 33, espera-se obter  ${}^p\pi(\pi_1)=\pi_2, \dots, {}^p\pi(\pi_{i-1})=\pi_j, {}^p\pi(\pi_j)=\pi_{j-1}, \dots, {}^p\pi(\pi_{i+1})=\pi_i, {}^p\pi(\pi_i)=\pi_{j+1}, \dots, {}^p\pi(\pi_n)=\pi_1$ . De fato, tomando-se  $\pi_k, k=1, \dots, n$ , em  $\pi$ , a sua imagem  ${}^p\pi(\pi_k)$  em  ${}^p\pi$ , considerando-se a Observação 2, coincide com o resultado esperado, pois:

$$\begin{aligned} {}^p\pi(\pi_k) &= p(\pi(p(\pi_k))) &= p(\pi(\pi_k)) &= p(\pi_{k+1}) &= \pi_{k+1}, & \text{se } 1 \leq k \leq i-2; \\ {}^p\pi(\pi_{i-1}) &= p(\pi(p(\pi_{i-1}))) &= p(\pi(\pi_{i-1})) &= p(\pi_i) &= \pi_j, & \text{se } k=i-1; \\ {}^p\pi(\pi_i) &= p(\pi(p(\pi_i))) &= p(\pi(\pi_j)) &= p(\pi_{j+1}) &= \pi_{j+1}, & \text{se } k=i; \\ {}^p\pi(\pi_k) &= p(\pi(p(\pi_k))) &= p(\pi(\pi_{i+j-k})) &= p(\pi_{i+j-k+1}) &= \pi_{k-1}, & \text{se } i+1 \leq k \leq j-1; \\ {}^p\pi(\pi_j) &= p(\pi(p(\pi_j))) &= p(\pi(\pi_i)) &= p(\pi_{i+1}) &= \pi_{j-1}, & \text{se } k=j; \\ {}^p\pi(\pi_k) &= p(\pi(p(\pi_k))) &= p(\pi(\pi_k)) &= p(\pi_{k+1}) &= \pi_{k+1}, & \text{se } j+1 \leq k \leq n-1 \text{ e} \\ {}^p\pi(\pi_n) &= p(\pi(p(\pi_n))) &= p(\pi(\pi_n)) &= p(\pi_1) &= \pi_1, & \text{se } k=n. \end{aligned}$$

Uma permutação obtida conforme a Definição 33, que é um produto de transposições algébricas disjuntas e um conjugador com a propriedade de reproduzir uma reversão em um “ $n$ -ciclo”, será referida daqui para frente como uma **permutação-reversão** desse “ $n$ -ciclo”.

**Observação 21** Uma reversão é aplicada a um representante de uma classe de elementos “circulares”, conforme a Observação 11.

**Exemplo 32** No formalismo da Equação 4.1, dadas as permutações  $\pi=(2 \ 4 \ 3)(5 \ 7)$  e  $\rho=(1 \ 6)(2 \ 5)(3 \ 4)$ , de  $S_7$ , tem-se que  $\pi_{str}=\underline{1 \ 3 \ 4 \ 2 \ 7 \ 6} \ 5$  e que  $\pi_{str}\rho=6 \ 7 \ 2 \ 4 \ 3 \ 1 \ 5$ . A subpalavra do intervalo  $[1 \ 6]$  de  $\pi_{str}$  é revertida em  $\pi_{str}\rho$ .

**Exemplo 33** No formalismo da Equação 4.2, a reversão  $[1, 3]$  aplicada nas posições de 1 a 3 do cromossomo representado por  $\pi=(\underline{5}, \underline{1}, 3, 2, 4)$ , produz  $\pi[1, 3]=(3, 1, 5, 2, 4)$ , pois  $\pi[1, 3](1)=\pi_{3+1-1} = \pi_3 = 3$ ,  $\pi[1, 3](2)=\pi_{3+1-2} = \pi_2 = 1$ , etc..

**Exemplo 34** No formalismo da Equação 4.3, dada a classe  $\pi$ , representada pelo “5-ciclo”  $[5, 1, 3, 2, 4]$ , a permutação-reversão, dada pela Definição 33, que simula a reversão da porção contígua de posições 1 a 3 de  $\pi$  é  $p=(\pi_1 \pi_3)=(5 3)$ , pois a conjugação de  $\pi$  por  $p$ ,  ${}^p\pi$ , produz o novo “5-ciclo”  $(5 3) [5, 1, 3, 2, 4] (3 5)=[3, 1, 5, 2, 4]$ .

Sejam, por exemplo,  $[1, 3, 2]$ ,  $[3, 2, 1]$  e  $[2, 1, 3]$  as sequências de leitura sequencial  $(1 3 2)$ . As três sequências, enquanto “3-ciclos”, têm o mesmo comportamento do 3-ciclo  $(1 3 2)$ , de acordo com a Observação 1. Porém, dependendo da natureza dos elementos contituíntes de  $G$  (que pode ser um conjunto de blocos de genes, conforme os Exemplos 30 e 31), podem ser diferentes, o que é confirmado pelo fato de as suas funções associadas, que são, respectivamente,  $\{(1, 1), (2, 3), (3, 2)\}$ ,  $\{(1, 3), (2, 2), (3, 1)\}$ , e  $\{(1, 2), (2, 1), (3, 3)\}$ , serem diferentes.

A abordagem em estudo é genérica no sentido de poder ser aplicada à entidades que sejam representáveis por meio de sequências finitas e como uma reversão (conjugação) é aplicada a um representante de uma classe, com o objetivo de reverter um trecho de elementos contíguos desse representante, mantendo os demais fixos em suas respectivas posições, fica estabelecida a seguinte convenção, que é dada em forma de exemplo, para a escolha do representante da classe resultante.

**Convenção 1** Sejam  $\Gamma=\{1, 2, 3, 4, 5, 6\}$  e o “6-ciclo”  $[1, 3, 2, 5, 4, 6]$ , representante de uma classe  $\pi$ . Se for necessário obter a reversão  $[1, 3, 2, 5, 4, 6]$ , será utilizado como conjugador, de acordo com a Definição 33, a permutação-reversão  $a=(1 5)(3 2)$  e, como o primeiro elemento da  $\pi$ , que é 1, coincide com o primeiro elemento da porção a ser revertida, fica estabelecido que o representante da classe que é o resultado da conjugação de  $\pi$  por  $a$  deve ser a família que inicia pelo último elemento desta porção, que é 5. Assim, o resultado desta conjugação terá como representante o novo “6-ciclo”:

$${}^a\pi = a\pi a^{-1} = (1 5)(3 2) [1, 3, 2, 5, 4, 6] (2 3)(5 1) = [5, 2, 3, 1, 4, 6],$$

que traz a reversão esperada (a família  $[5, 2, 3, 1, 4, 6]$  é o conveniente representante da classe produto). Se for necessário obter a reversão  $[1, 3, 2, 5, 4, 6]$ , será utilizado como conjugador a permutação-reversão  $a=(2 4)$  e, como o primeiro elemento da  $\pi$ , que é 1, não coincide com o primeiro elemento da porção a ser revertida, o representante da classe que é o resultado da conjugação de  $\pi$  por  $a$  deve iniciar pelo primeiro elemento de  $\pi$ , que é 1, e, assim, o resultado desta conjugação terá como representante o novo “6-ciclo”:

$${}^a\pi = a\pi a^{-1} = (2 4) [1, 3, 2, 5, 4, 6] (4 2) = [1, 3, 4, 5, 2, 6],$$

que tem a reversão esperada (a família  $[1, 3, 4, 5, 2, 6]$  é o conveniente representante da classe produto).

#### Redefinição do problema.

O problema da distância de reversão entre sequências finitas é redefinido, para se adequar a este método, da seguinte maneira: “dados dois “ $n$ -ciclos”,  $\alpha$  e  $\beta$ , obter o menor

número de permutações-reversões que transformem um no outro”. A este número se dá o nome de **distância de reversão** entre  $\alpha$  e  $\beta$  que é denotado por  $d(\alpha, \beta)$  ou  $d_\beta(\alpha)$ . Um caso particular é a **distância de ordenação**, que consiste em “dado um “*n-ciclo*”  $\alpha$ , obter o menor número de permutações-reversões que o transformem na sequência ordenada” e é denotada por  $d(\alpha)$ .

Daqui para frente, será tratado apenas deste problema.

**Definição 34** Dados os “*n-ciclos*”  $\alpha$ ,  $\beta$  e  $\iota$ , transformar  $\alpha$  em  $\beta$  por reversões é achar uma sequência de permutações-reversões  $a_i, i=1, \dots, r$ , tais que:

$$a_r(\dots(a_2(a_1\pi))) = a_r \dots a_2 a_1 \alpha a_1^{-1} a_2^{-1} \dots a_r^{-1} = \beta$$

e ordenar  $\alpha$  por reversões é achar uma sequência de permutações-reversões  $b_i, i=1, \dots, s$ , tais que:

$$b_s(\dots(b_2(b_1\pi))) = b_s \dots b_2 b_1 \alpha b_1^{-1} b_2^{-1} \dots b_s^{-1} = \iota.$$

### 4.3 O Conjugador (para $\iota$ ) e duas simulações

Nesta seção será definido o conjugador (para  $\iota$ ) de um “*n-ciclo*” (representante de uma classe) e detalhado o seu papel; mostrado como as informações necessárias para a montagem de uma permutação-reversão se acham nos conjugadores (para  $\iota$ ); serão feitas duas simulações, paralelas e independentes, de uma ordenação de um “*n-ciclo*” e da transformação do seu *conjugador* (para  $\iota$ ) na permutação identidade utilizando as mesmas permutações-reversões e será mostrado como obter um conjugador (para  $\iota$ ) a partir de outro sem utilizar o “*n-ciclo*”. O “*n-ciclo*”  $[\pi_1, \dots, \pi_n]$  será referido, por abuso de notação, como um elemento de  $\Gamma_n$ , pois representa uma classe de  $\Phi_n$ .

#### 4.3.1 O conjugador (para $\iota$ ) detém as informações

Dados um “*n-ciclo*”  $\pi=[\pi_1, \dots, \pi_n] \in \Gamma_n$  e uma permutação  $a \in S_n$ , é denotado por  $p_\pi(a_x)$  o elemento  $a_y$  que é a posição do elemento  $a_x$  em  $\pi$  ( $\pi_{a_y} = a_x$ ).

**Exemplo 35** Dados  $\pi=[5, 1, 2, 3, 4] \in \Gamma_5$  e a permutação  $a=(5 \ 3 \ 2 \ 1 \ 4) \in S_5$ , então  $p_\pi(a_1)=p_\pi(5)$  é a posição do elemento  $a_1=5$  em  $\pi$  que, neste caso, é  $a_4=1$  ( $\pi_{a_4} = a_1, \pi_1 = 5$ ).

Dado um “*n-ciclo*”  $\pi \in \Gamma_n$ , a permutação  $a \in S_n$ , onde  $a_1$  é a posição do elemento  $a_n$  e  $a_{i+1}$  é a posição do elemento  $a_i, 1 \leq i \leq n-1$ , em  $\pi$ , será chamada de **representação posicional de  $\pi$** . O lema seguinte afirma que a representação posicional de um “*n-ciclo*” é o seu conjugador (para  $\iota$ ).

**Lema 6** *Sejam  $\pi \in \Gamma_n$  e  $a \in S_n$ , onde  $a$  é a representação posicional de  $\pi$ , então  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ .*

*Demonstração.* Sendo  $a$  a representação posicional de  $\pi$ , conjugando-se  $\pi$  por  $a$  ( ${}^a\pi = a\pi a^{-1}$ ) tem-se que obter  $\iota$  e, de fato, calculando-se  $a\pi a^{-1}$  em  $i$ , decorre que  $(a\pi a^{-1})(i) = a(\pi(a^{-1}(i))) = a(\pi(\pi_i)) = a(\pi_{i+1}) = i+1, i=1, \dots, n$ , assim,  $a\pi a^{-1} = \iota$  e, portanto,  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ . Reciprocamente, sendo  $a$  o conjugador (para  $\iota$ ) de  $\pi$ , tem-se

que  ${}^a\pi = a\pi a^{-1} = \iota$  do que decorre que  $a = \iota a \pi^{-1}$ . Tomando-se um elemento qualquer  $\pi_i$  de  $\pi$ , segue que  $a(\pi_i) = (\iota a \pi^{-1})(\pi_i) = \iota(a(\pi^{-1}(\pi_i))) = \iota(a(\pi_{i-1})) = \iota(i-1) = i$ ,  $i=1, \dots, n$ , o que mostra que  $a$  é a representação posicional de  $\pi$ . Isto completa a prova.  $\square$

**Exemplo 36** Para mostrar a relação entre  $\pi$  e  $a$ , o seu conjugador (para  $\iota$ ), considere o “6-ciclo” da Figura 4.3  $\pi = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6] = [3, 6, 4, 5, 2, 1]$  e  $a = (a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6) = (3 \ 1 \ 6 \ 2 \ 5 \ 4)$ . De fato,  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ , pois, a conjugação de  $\pi$  por  $a$  resulta  $\iota: {}^a\pi = a\pi a^{-1} = (3 \ 1 \ 6 \ 2 \ 5 \ 4) [3, 6, 4, 5, 2, 1] (4 \ 5 \ 2 \ 6 \ 1 \ 3) = [1, 2, 3, 4, 5, 6] = \iota$  e tem-se que  $a_2=1$  é a posição do elemento  $a_1=3$  em  $\pi$  ( $a_2 = p_\pi(a_1)$ );  $a_3=6$  é a posição do elemento  $a_2=1$  em  $\pi$  ( $a_3 = p_\pi(a_2)$ ); etc., como mostra a Tabela 4.1.

**Observação 22** Para facilitar a leitura, para primeiro elemento da permutação  $a$  estará sendo utilizado o primeiro elemento de  $\pi$  (pois, sendo  $a$  um ciclo, pode iniciar por qualquer um dos seus elementos).

$a_i$	$a_{i+1(\text{mod } 6)}$	$\pi_{a_{i+1(\text{mod } 6)}} = a_i$	
$a_1 = 3$	$a_2 = 1$	$\pi_{a_2} = a_1$	$\pi_1 = 3$
$a_3 = 6$	$a_4 = 2$	$\pi_{a_4} = a_3$	$\pi_2 = 6$
$a_6 = 4$	$a_1 = 3$	$\pi_{a_1} = a_6$	$\pi_3 = 4$
$a_5 = 5$	$a_6 = 4$	$\pi_{a_6} = a_5$	$\pi_4 = 5$
$a_4 = 2$	$a_5 = 5$	$\pi_{a_5} = a_4$	$\pi_5 = 2$
$a_2 = 1$	$a_3 = 6$	$\pi_{a_3} = a_2$	$\pi_6 = 1$

Tabela 4.1: Relação entre os elementos de  $\pi$  e os do seu conjugador (para  $\iota$ )  $a$ .

### 4.3.2 O conjugador (para $\iota$ ) e uma simulação independente

O Lema 7, embora tendo uma demonstração imediata, garante o paralelismo segundo o qual, dados um “ $n$ -ciclo”  $\pi$  e o seu conjugador (para  $\iota$ )  $a$ , uma mesma família de permutações, mediante multiplicações, transforma  $a$  na permutação identidade  $e$ , mediante conjugações, ordena  $\pi$  (o transformando em  $\iota$ ).

**Lema 7** Dados  $\pi \in \Gamma_n$ ,  $a \in S_n$ , onde  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ , e  $p \in S_n$ , uma permutação qualquer, então  $ap$  é o conjugador (para  $\iota$ ) de  $p^{-1}\pi p$  ( ${}^{p^{-1}}\pi$ ).

*Demonstração.* Sabendo-se que  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ , segue que  $a\pi a^{-1} = \iota$ . Então, tomando-se  $p^{-1}\pi p$ , que é o resultado da conjugação de  $\pi$  por  $p^{-1}$ , e  $ap$ , o resultado da conjugação de  $p^{-1}\pi p$  por  $ap$  é igual a  $\iota$ . De fato:  ${}^{ap}(p^{-1}\pi p) = (ap)(p^{-1}\pi p)(p^{-1}a^{-1}) = a\pi a^{-1} = \iota$ .  $\square$

**Observação 23** Se a permutação  $p$  do Lema 7 for um produto de transposições disjuntas, a conjugação de  $\pi$  e a multiplicação de  $a$  passam a ser feitas pela mesma  $p$  e não, respectivamente, por  $p^{-1}$  e  $p$  (conforme a Observação 19).

$$\begin{array}{ccc}
p_1^{-1}\pi p_1 & & ap_1 \\
\hline
p_2^{-1}(p_1^{-1}\pi p_1)p_2 & & (ap_1)p_2 \\
\hline
\dots & & \dots \\
\hline
p_m^{-1}(\dots(p_2^{-1}(p_1^{-1}\pi p_1)p_2)\dots)p_m & & (\dots((ap_1)p_2)\dots)p_m
\end{array}$$

Figura 4.1: As sucessivas conjugações de  $\pi$  por  $p_i^{-1}$  e as sucessivas multiplicações de  $a$  por  $p_i$ , dados um “n-ciclo”  $\pi$ ,  $a$  (o seu conjugador (para  $\iota$ )) e uma família  $(p_i)$ ,  $i=1, \dots, m$ , de permutações.

$$\begin{array}{ccc}
a^{p_1}(p_1^{-1}\pi p_1) = ap_1p_1^{-1}\pi p_1p_1^{-1}a^{-1} = \iota & & \\
\hline
a^{p_1p_2}(p_2^{-1}p_1^{-1}\pi p_1p_2) = ap_1p_2p_2^{-1}p_1^{-1}\pi p_1p_2p_2^{-1}p_1^{-1}a^{-1} = \iota & & \\
\hline
\dots & & \\
\hline
a^{p_1p_2\dots p_m}(p_m^{-1}\dots p_2^{-1}p_1^{-1}\pi p_1p_2\dots p_m) = ap_1p_2\dots p_m p_m^{-1}\dots p_2^{-1}p_1^{-1}\pi p_1p_2\dots p_m p_m^{-1}\dots p_2^{-1}p_1^{-1}a^{-1} = \iota
\end{array}$$

Figura 4.2: Aplicação do Lema 7 aos elementos das linhas da Figura 4.1.

**Corolário 2** Dados  $\pi \in \Gamma_n$ ,  $a \in S_n$ , onde  $a$  é o conjugador (para  $\iota$ ) de  $\pi$ , e  $p \in S_n$ , uma permutação-reversão qualquer de  $\pi$ , então  $ap$  é o conjugador (para  $\iota$ ) de  $p\pi p$ .

*Demonstração.* De fato, considerando-se a Observação 23, tem-se que  $p = p^{-1}$  e segue que  $a^{p(p\pi p)} = ap(p\pi p)p^{-1}a^{-1} = a\pi a^{-1} = \iota$ , o que prova o lema.  $\square$

**Exemplo 37** Seja  $\pi=[5, 3, 1, 2, 4]$  e  $a=(5\ 1\ 3\ 2\ 4)$  o seu conjugador (para  $\iota$ ), então  ${}^a\pi = a\pi a^{-1} = (5\ 1\ 3\ 2\ 4)[5, 3, 1, 2, 4](4\ 2\ 3\ 1\ 5) = [1, 2, 3, 4, 5] = \iota$ . Tomando-se uma permutação  $p$ , por exemplo  $p=(4\ 5)(3\ 2)$ , segue que  $ap = (5\ 1\ 3\ 2\ 4)(4\ 5)(3\ 2) = (1\ 3\ 4)(2)(5)$  e que  ${}^{p^{-1}}\pi = p^{-1}\pi p = p\pi p = (4\ 5)(3\ 2)[5, 3, 1, 2, 4](4\ 5)(3\ 2) = [1, 3, 5, 4, 2]$ . Conjugando-se  ${}^{p^{-1}}\pi$  por  $ap$ , segue o resultado esperado, que é  $(1\ 3\ 4)[1, 3, 5, 4, 2](4\ 3\ 1) = [1, 2, 3, 4, 5] = \iota$ .

Adiante, estão os quatro casos possíveis, dependendo de a família  $(p_i)$ ,  $i=1, \dots, m$ , ser uma família qualquer ou constituída apenas de permutações-reversões e de ser ou não uma decomposição de  $a$ , baseados nas Figuras 4.1 e 4.2.

- Se a família  $(p_i)$ ,  $i=1, \dots, m$ , da Figura 4.1, for uma família qualquer, ter-se-á que:
  - se  $(p_i)$ ,  $i=1, \dots, m$ , for tal que  $p_m^{-1}\dots p_2^{-1}p_1^{-1} = a$ , ou seja, se  $p_m^{-1}\dots p_2^{-1}p_1^{-1}$  for uma decomposição de  $a$ , tem-se  $p_1p_2\dots p_m = a^{-1}$  e

$$\begin{array}{ccc}
p_m^{-1}\dots p_2^{-1}p_1^{-1}\pi p_1p_2\dots p_m & & ap_1p_2\dots p_m \\
a\pi a^{-1} & & aa^{-1} \\
\iota & & e
\end{array}$$

e, neste caso, a mesma família, por sucessivas conjugações, ordenará  $\pi$  e, por sucessivas multiplicações, transformará  $a$  em  $e$  e

- caso contrário,  $\pi$  será transformado em algum “n-ciclo” diferente de  $\iota$  e  $a$  em alguma permutação diferente de  $e$ .
- Se a família  $(p_i), i=1, \dots, m$ , da Figura 4.1, for constituída apenas de permutações-reversões, em cada passo de ambas as simulações da figura estará sendo empregada a mesma  $p_i$ , de acordo com a Observação 23, e ter-se-á que:
  - se  $(p_i), i=1, \dots, m$ , for tal que  $p_m \dots p_2 p_1 = a$ , ou seja, se  $p_m \dots p_2 p_1$  for uma decomposição de  $a$ , tem-se  $p_1 p_2 \dots p_m = a^{-1}$  e

$$\begin{array}{ccc} p_m \dots p_2 p_1 \pi p_1 p_2 \dots p_m & & a p_1 p_2 \dots p_m \\ a \pi a^{-1} & & a a^{-1} \\ \iota & & e \end{array}$$

e, neste caso, a mesma família, por sucessivas conjugações, ordenará  $\pi$  por reversões e, por sucessivas multiplicações, transformará  $a$  em  $e$  e

- caso contrário,  $\pi$  será transformado em algum “n-ciclo”  $\alpha$ , diferente de  $\iota$ , por reversões e  $a$  em alguma permutação diferente de  $e$ .

Os dois últimos casos são de interesse para o trabalho, pois, no primeiro deles, um “ciclo”  $\pi$  é ordenado e, no segundo, transformado em outro por reversões. Nestes casos, pode-se obter de duas maneiras uma mesma família de reversões que executa ambas as tarefas. Concluindo, encontrar uma família de reversões que ordena um “ciclo” equivale a encontrar uma família que transforma o seu conjugador (para  $\iota$ ) em  $e$ .

### 4.3.3 Simulando em paralelo

Através de exemplos, serão mostradas a ordenação de um “6-ciclo”  $\pi$  e a transformação do seu conjugador (para  $\iota$ ),  $a$ , em  $e$ , pelas mesmas permutações-reversões  $p_i, i=1, \dots, 4$ , respectivamente, por sucessivas conjugações e multiplicações.

**Exemplo 38** Sejam  $\pi=[3, 6, 4, 5, 2, 1]$  um “6-ciclo” e  $a=(3\ 1\ 6\ 2\ 5\ 4)$  o seu conjugador (para  $\iota$ ). A Figura 4.3 mostra a sequência de quatro passos (os trechos revertidos em cada um dos passos estão em destaque) de uma ordenação de  $\pi=[3, 6, 4, 5, 2, 1]$  (colunas da esquerda) e de uma transformação de  $a=(3\ 1\ 6\ 2\ 5\ 4)$  na permutação identidade (colunas da direita), mediante as quatro permutações-reversões  $p_1=(1\ 3)(2\ 6)(4\ 5)$ ,  $p_2=(3\ 5)(4\ 6)$ ,  $p_3=(5\ 6)$  e  $p_4=(4\ 5)$ .

As quatro permutações-reversões, sucessivamente, conjugando “6-ciclos”, a partir do “6-ciclo”  $\pi$ , o ordenam e, multiplicando 6-permutações, a partir de  $a$ , a transformam em  $e$ . É fácil verificar, por intermédio do Lema 6, que as cinco 6-permutações que vão de  $a$  a  $e$ :  $a_1=(3\ 1\ 6\ 2\ 5\ 4)$ ,  $a_2=(1)(2)(4)(3\ 6\ 5)$ ,  $a_3=(1)(2)(3)(4\ 5\ 6)$ ,  $a_4=(1)(2)(3)(4\ 5)(6)$  e  $a_5=(1)(2)(3)(4)(5)(6)=e$  são, respectivamente, os conjugadores (para  $\iota$ ) dos “6-ciclos” que vão de  $\pi$  a  $\iota$ :  $\pi_1=[3, 6, 4, 5, 2, 1]$ ,  $\pi_2=[1, 2, 5, 4, 6, 3]$ ,  $\pi_3=[1, 2, 3, 6, 4, 5]$ ,  $\pi_4=[1, 2, 3, 5, 4, 6]$  e  $\pi_5=[1, 2, 3, 4, 5, 6]=\iota$ .

$\pi$	$[3, 6, 4, 5, 2, 1]$	$(3\ 1\ 6\ 2\ 5\ 4)$	$a$
$\pi_2 = p^1 \pi_1$	$(1\ 3)(2\ 6)(4\ 5)[3, 6, 4, 5, 2, 1]$	$(3\ 1\ 6\ 2\ 5\ 4)\ (1\ 3)(2\ 6)(4\ 5)$	$a_1 p_1 = a_2$
$\pi_3 = p^2 \pi_2$	$(3\ 5)(4\ 6)[1, 2, 5, 4, 6, 3]$	$(1)(2)(4)(3\ 6\ 5)\ (3\ 5)(4\ 6)$	$a_2 p_2 = a_3$
$\pi_4 = p^3 \pi_3$	$(5\ 6)[1, 2, 3, 6, 4, 5]$	$(1)(2)(3)(4\ 5\ 6)\ (5\ 6)$	$a_3 p_3 = a_4$
$\iota = \pi_5 = p^4 \pi_4$	$(4\ 5)[1, 2, 3, 5, 4, 6]$	$(1)(2)(3)(4\ 5)(6)\ (4\ 5)$	$a_4 p_4 = a_5 = e$
$\iota$	$[1, 2, 3, 4, 5, 6]$	$(1)(2)(3)(4)(5)(6)$	$e$

Figura 4.3: Simulação da ordenação de  $\pi$  e da transformação de  $a$  em  $e$ .

**Observação 24** As famílias  $\pi_1=[3, 6, 4, 5, 2, 1]$ ,  $\pi_2=[1, 2, 5, 4, 6, 3]$ ,  $\pi_3=[1, 2, 3, 6, 4, 5]$ ,  $\pi_4=[1, 2, 3, 5, 4, 6]$  e  $\pi_5=[1, 2, 3, 4, 5, 6]=\iota$  são as representantes das suas respectivas classes que devem ser aplicadas, de acordo com a Convenção 1. Por outro lado, diferentemente das famílias, as permutações  $a_1=(3\ 1\ 6\ 2\ 5\ 4)$ ,  $a_2=(1)(2)(4)(3\ 6\ 5)$ ,  $a_3=(1)(2)(3)(4\ 5\ 6)$ ,  $a_4=(1)(2)(3)(4\ 5)(6)$  e  $a_5=(1)(2)(3)(4)(5)(6)=e$ , podem ser substituídas, conforme a Observação 1.

**Observação 25** Descobrir uma sequência de permutações-reversões que ordenem um “ $n$ -ciclo” por intermédio dos conjugadores (para  $\iota$ ) é, computacionalmente, mais eficiente porque, para dar um passo na ordenação, aplica apenas uma operação de multiplicação.

#### 4.3.4 Obtendo um novo conjugador (para $\iota$ )

O corolário do Lema 7, Corolário 2, formaliza um método para obter, dados um “ $n$ -ciclo”  $\pi$ ,  $a$ , o seu conjugador (para  $\iota$ ) e uma permutação-reversão  $p$ , o conjugador (para  $\iota$ ) de  ${}^p\pi$ , por meio de apenas uma multiplicação algébrica,  $ap$ , utilizando, como termos,  $a$  e  $p$  (conforme resume a Figura 4.4). O “ $n$ -ciclo”  ${}^p\pi$  veio do “ $n$ -ciclo”  $\pi$  pela reversão sofrida

$$\begin{array}{ll} \pi & a \text{ é o conjugador (para } \iota \text{) de } \pi \quad (a\pi a^{-1} = \iota) \\ {}^p\pi & ap \text{ é o conjugador (para } \iota \text{) de } {}^p\pi \quad (ap({}^p\pi) = \iota) \end{array}$$

Figura 4.4: Um passo na obtenção de uma sequência de reversões que, utilizando apenas conjugadores (para  $\iota$ ), ordenam um “ $n$ -ciclo”.

com a conjugação pela permutação-reversão  $p$  (na Figura 4.3, os conjugadores (para  $\iota$ )  $a_{i-1}p_{i-1}=a_i$ ,  $i=2, 3, 4, 5$ , são exemplos). Depois de obter uma permutação-reversão  $p$ , a partir de um conjugador (para  $\iota$ )  $a$ , calcula-se o produto  $ap$ . Na pesquisa de uma sequência de reversões de uma ordenação, se  $ap = e$ , chegou-se à última reversão da sequência e se  $ap \neq e$ , o próprio  $ap$  será o próximo conjugador (para  $\iota$ ) a ser utilizado.

## 4.4 Aplicando este formalismo em algoritmos

Ambas as simulações apresentadas na seção 4.3 podem ser feitas algoritmicamente, como se verá na sequência.

Um “ $n$ -ciclo” qualquer admite um total de  $C_{n,2}=n!/2!(n-2)!$  (combinação de  $n$  elementos, tomados 2 a 2) reversões, pois a cada dois elementos diferentes do seu suporte equivale uma das suas possíveis reversões (dois elementos do suporte equivalem à reversão onde eles são o primeiro e o último elementos).

Sabendo-se que os suportes de um “ $n$ -ciclo” e do seu conjugador (para  $\iota$ ) são iguais e com base na subseção 4.3.1 e Lemas 6 e 7 e o seu Corolário 2, pode-se construir qualquer uma das permutações-reversões de um “ $n$ -ciclo” olhando-se apenas o próprio “ $n$ -ciclo” ou apenas para o seu conjugador (para  $\iota$ ).

Isso significa poder, ao aplicar o método em um algoritmo qualquer que opere com reversões, utilizar apenas “ $n$ -ciclos” ou apenas conjugadores (para  $\iota$ ) para encontrar as permutações-reversões procuradas.

#### 4.4.1 Construindo a permutação-reversão associada a dois elementos de um “ $n$ -ciclo”

Dados um “ $n$ -ciclo”  $\pi$ , o seu conjugador (para  $\iota$ ),  $a$ , e dois elementos quaisquer,  $r$  e  $s$ , do suporte de  $\pi$ , será construída a permutação-reversão  $p$  de  $\pi$ , associada a esses elementos, primeiramente, atuando apenas sobre  $\pi$  e, em seguida, apenas sobre  $a$ .

**Olhando para o “ $n$ -ciclo”  $\pi$ :**

Para obter as transposições de  $p$  a partir de  $\pi$ , é suficiente aplicar a Definição 33 à porção contígua de posições  $i$  a  $j$ ,  $1 \leq i < j \leq n$ , de  $\pi$ , onde: se  $\pi^{-1}(r) < \pi^{-1}(s)$ , então  $i = \pi^{-1}(r)$  ( $r = \pi_i$ ) e  $j = \pi^{-1}(s)$  ( $s = \pi_j$ ), e, se  $\pi^{-1}(s) < \pi^{-1}(r)$ , então  $i = \pi^{-1}(s)$  ( $s = \pi_i$ ) e  $j = \pi^{-1}(r)$  ( $r = \pi_j$ ) (aqui, o “ $n$ -ciclo” e a sua função de escolha estão sendo representados pelo mesmo símbolo  $\pi$ , conforme observação feita no final da seção 4.1, ou seja,  $\pi = [\pi_1, \dots, \pi_i, \dots, \pi_n]$ ;  $\pi(i) = \pi_i$ ;  $\pi^{-1}(\pi_i) = i$  e, se, por exemplo,  $r = \pi_i$ , então,  $i = \pi^{-1}(r)$ ,  $1 \leq i \leq n$ ). A reversão associada a  $r$  e  $s$  é realizada pela conjugação de  $\pi$  por  $p$ .

**Olhando para  $a$ , o conjugador (para  $\iota$ ), do “ $n$ -ciclo”  $\pi$ :**

Por outro lado, para encontrar as transposições de  $p$  a partir de  $a$ , é suficiente seguir os seguintes três passos, que serão adiante detalhados:

- (1) tomar, de início,  $(a_u a_v)$  como primeira transposição de  $p$ , onde  $a_u = r$  e  $a_v = s$ ;
- (2) tendo obtido uma, verificar se existe ou não outra transposição em  $p$  e, em caso afirmativo, encontrar esta transposição e
- (3) daí para frente, repetir o passo (2), utilizando a última transposição encontrada até que não mais existam transposições em  $p$ .

Antes de explicar cada um dos passos, será definida a distância, em  $\pi$ , entre dois elementos distintos de  $a$ .

Sendo  $a$  a permutação que é a representação posicional de  $\pi$  (Lema 6),  $a_w$ , um elemento de um subciclo qualquer de  $a$  (lembrar que  $a$  tem os mesmos elementos de  $\pi$  e que pode ter de 1 a  $n$  subciclos) e  $a'_w$ , o elemento que o sucede no subciclo, ou seja,  $a'_w = a(a_w)$ , tem-se que  $a'_w$  é a posição de  $a_w$  em  $\pi$ , ou seja,  $\pi_{a'_w} = a_w$ .

Voltando-se, agora, aos elementos  $a_u$  e  $a_v$ , supondo-se que estejam em dois subciclos (eventualmente, podem estar no mesmo) de  $a$ , então, tem-se que  $a(a_u)$  e  $a(a_v)$  são as suas respectivas posições em  $\pi$ , ou seja,  $a_u = \pi_{a(a_u)}$  e  $a_v = \pi_{a(a_v)}$ , e, assim,  $d_\pi(a_u, a_v)$  é dada pela seguinte definição.

**Definição 35** Dados um “ $n$ -ciclo”  $\pi$ ,  $a$ , o seu conjugador (para  $\iota$ ) e os elementos  $a_u$  e  $a_v$ , fica definida por  $|a(a_v) - a(a_u)|$ , denotada por  $d_\pi(a_u, a_v)$ , a distância, em  $\pi$ , entre  $a_u$  e  $a_v$ .

Passa-se a explicar os passos (1), (2) e (3), partindo-se de  $(a_u a_v)$  (passo (1)). Para verificar a existência ou não de outra transposição em  $p$  (passo (2)) calcula-se  $d_\pi(a_u, a_v)$ . O resultado vai indicar se há ou não outros elementos entre eles em  $\pi$  e, se houver, quantos são.

A diferença  $a(a_v) - a(a_u)$  pode ser um número inteiro positivo ou negativo, dependendo, respectivamente, de  $a_u$  aparecer antes ( $u < v$ ) ou depois ( $v < u$ ) de  $a_v$  em  $\pi$ . A inversão da ordem dos seus elementos não altera uma transposição, então, tendo-se que  $(a_u a_v) = (a_v a_u)$ , será admitido que a diferença é um número inteiro positivo.

Como, então, é considerada cada uma das possíveis distâncias  $d_\pi(a_u, a_v)$  ( $1 \leq d_\pi(a_u, a_v) \leq n-1$ )?

Há três casos a considerar:

- (a) se  $d_\pi(a_u, a_v) = 1$ , então  $a_u$  e  $a_v$  são contíguos em  $\pi$  e a permutação-reversão  $p$  associada a  $r$  e  $s$  é a transposição algébrica  $(a_u a_v)$ ;
- (b) se  $d_\pi(a_u, a_v) = 2$ , então existe um único elemento entre  $a_u$  e  $a_v$  em  $\pi$  e, como ele não será movido pela reversão, a permutação-reversão  $p$  associada a  $r$  e  $s$  também é a transposição algébrica  $(a_u a_v)$ ;
- (c) se  $d_\pi(a_u, a_v) = q \geq 3$ , então, existem, em  $\pi$ ,  $q-1 \geq 2$  elementos entre  $a_u$  e  $a_v$ , o que significa que há pelo menos mais uma transposição em  $p$  a ser pesquisada.

No caso (c), é guardada a transposição algébrica  $(a_u a_v)$  e são procurados, em  $a$ , os dois elementos que formam a seguinte. Como, então, determinar esses elementos quando  $d_\pi(a_u, a_v) \geq 3$ ?

Sabendo-se que  $a_u$  e  $a_v$  ocupam em  $\pi$ , respectivamente, as posições  $a(a_u)$  e  $a(a_v)$ , ( $a_u = \pi_{a(a_u)}$  e  $a_v = \pi_{a(a_v)}$ , respectivamente) vem que os dois elementos procurados estão em  $\pi$ , nas posições  $a(a_u) + 1 \pmod{n}$  (sucede  $a_u$  em  $\pi$ ) e  $a(a_v) - 1 \pmod{n}$  (antecede  $a_v$  em  $\pi$ ) (sabe-se que estas posições existem pelo item (c)), sendo, então, os elementos  $\pi_{a(a_u)+1 \pmod{n}}$  e  $\pi_{a(a_v)-1 \pmod{n}}$ , respectivamente.

Para encontra-los em  $a$ , sendo  $a$  a representação posicional de  $\pi$ , basta localizar em  $a$  os elementos  $a_{x'} = a(a_u) + 1 \pmod{n}$  e  $a_{y'} = a(a_v) - 1 \pmod{n}$ , que são as posições dos elementos que constituirão a transposição seguinte de  $p$ , e tomar os dois que estão imediatamente antes, ou seja,  $a^{-1}(a_{x'})$  e  $a^{-1}(a_{y'})$ . Supondo-se que  $a_{x'}$  e  $a_{y'}$  estejam em dois subciclos diferentes de  $a$ , de comprimentos, respectivamente,  $c_u$  e  $c_v$ , tomar o elemento

que está imediatamente antes de um deles, por exemplo, de  $a_{x'}$ , significa voltar um elemento no seu subciclo, ou seja, tomar o elemento que ocupa a posição anterior no seu subciclo, que é calculada subtraindo-se 1 da posição de  $a_{x'}$  módulo  $c_u$ . Supondo-se que os elementos resultantes,  $a^{-1}(a_{x'})$  e  $a^{-1}(a_{y'})$ , sejam, respectivamente,  $a_x$  e  $a_y$ , então  $(a_x a_y)$  será a transposição procurada.

Ao ser encontrada uma transposição de  $p$  deve ser aplicado o passo (3) e esse procedimento deve parar quando a distância em  $\pi$  entre os elementos da última transposição encontrada for menor do que três.

A permutação-reversão obtida desta forma, ou seja, a partir de  $a$ , será o produto das transposições encontradas e coincide com a contruída pela Definição 33, a partir de  $\pi$ .

# Capítulo 5

## Contribuição, conclusão e trabalhos futuros

### Contribuição

Tendo como motivação os cromossomos que podem ser representados por sequências, os problemas das distâncias de reversão e de ordenação em rearranjos de genomas, as interessantes questões em aberto e muitos dos conhecimentos algébricos e combinatórios que não estão sendo utilizados, buscou-se um modelo para expressar qualquer conjunto de objetos que sejam representáveis por sequências, considerar sequências como “ciclos” em um grupo de simetria e simular uma reversão por uma operação de conjugação.

Chegou-se a um método para simular a operação, onde o “ciclo” e uma certa permutação participam como termos ou fatores, que proporciona duas maneiras para descobrir uma família de reversões que transforma um “ciclo” em outro, simular reversões nesses “ciclos”, ordenar um “ciclo” e encontrar a distância de reversão entre dois “ciclos”. Assim, os problemas das distâncias de reversão e ordenação puderam ser estudados em sequências em geral por operações algébricas (no caso dos cromossomos, reversões biológicas).

As contribuições foram formalizadas em definições e demonstradas em lemas: o Lema 6, prova a igualdade entre o conjugador (para  $\iota$ ) e a representação posicional de um “ciclo”; o Lema 7, garante o paralelismo das duas formas de encontrar, algebricamente, uma mesma sequência de reversões que transformam um “ciclo” em outro e o corolário do Lema 7, Corolário 2, mostra como obter, mediante um único produto, um conjugador (para  $\iota$ ) a partir do anterior, estando em um passo de uma simulação onde se busca uma sequência de reversões que transformam um “ciclo” em outro.

Existem muitos trabalhos em rearranjos de genomas utilizando grupos de permutações e aplicando todas as operações. Entre eles está o artigo de Meidanis e Dias, em [28], pioneiro em relacionar os grupos de permutações com o estudo dos rearranjos de genomas, introduzindo uma forma algébrica de simular, em permutações representando genomas, eventos como os de reversões, transposições e intercâmbios de blocos, com e sem sinais. A técnica algébrica aqui desenvolvida foi, posteriormente, utilizada por outros pesquisadores, como, por exemplo, Huang e Lu, no artigo [9]. Outro exemplo de utilização de

grupos de permutações, desta vez com a operação de intercâmbio de blocos, é o artigo de Mira e Meidanis, em [14].

Resultados importantes têm sido alcançados pela combinação de diferentes recursos e, assim, uma eventual contribuição do método talvez seja a forma diferente das demais de computar.

Em vista de que existem maneiras mais eficientes para simular reversões, a aplicação do método em um algoritmo poderá não reduzir o seu tempo computacional, mas, por possibilitar novas combinações de recursos, talvez facilite obter outros tipos de resultados como, por exemplo, um, para reversões, similar ao do artigo que foi apresentado na seção 3.8, que é um problema em aberto.

Trata-se apenas de uma maneira diferente de simular uma reversão, que usa efetivamente, para essa finalidade, a operação de um grupo de permutações, manipulando os elementos permutados e não os seus índices. Sendo uma maneira nova e que ainda não foi experimentada de se pensar nas questões relacionadas com o problema da distância de reversão, pode, embora ainda haja muito por fazer, contribuir para a descoberta de respostas.

Nenhuma proposição similar ao método que foi apresentado ou aos lemas demonstrados foi encontrada nos textos pesquisados.

## Conclusão

Neste trabalho foram utilizados resultados conhecidos em álgebra para simular uma reversão em uma sequência finita utilizando a operação de conjugação de um grupo de simetria e para, dadas duas sequências, formalizar dois métodos para descobrir uma família de reversões que transforma uma na outra: um, pela aplicação direta da operação do grupo, e outro, pela aplicação da operação de conjugação.

## Trabalhos futuros

As sequências e as mudanças nelas produzidas pelas diferentes operações podem ser estudadas por meio das permutações que já foram representadas de diversas formas e intensivamente estudadas em todas áreas e disciplinas da matemática (recentemente, aumentou o interesse, de muitos pesquisadores da área combinatorial, em estudar as permutações) produzindo resultados importantes, que ainda não estão sendo utilizados, além de diversas questões que permanecem em aberto. Adiante, são mencionadas algumas destas questões, relacionadas com a operação de reversão, tiradas de [17]:

- quando as permutações contêm duplicações, a ordenação por reversões é transformada no problema da ordenação de palavras por reversões e uma questão em aberto e relacionada com o problema é:

- definir as complexidades relacionadas ao problema da ordenação de palavras por reversões;
- com base no conceito de subsequência ordenada (dado na Definição 8), Kececioglu e Sankoff, em [11], fizeram a seguinte conjectura:
  - para toda permutação existe uma menor série de reversões que não corta subsequências ordenadas de três ou mais elementos;
- com base no conceito de pontos-de-quebra (dado na Definição 7), Kececioglu e Sankoff, em [11], também, conjecturaram que:
  - para toda permutação existe uma menor série de reversões que nunca aumenta o número de pontos-de-quebra;
- Bafna e Pevzner, em [23], provaram a conjectura de Gollan (dada na Definição 9) estudaram o problema da distância esperada de reversão,  $E(d)$ , entre duas permutações selecionadas aleatoriamente e demonstraram que está muito perto do diâmetro de reversão, sendo  $E(d) \geq (1 - \frac{4.5}{\log n})n$ . Um problema por resolver, excluídos os casos triviais de Gollan, é:
  - melhorar o limite inferior e dar um limite superior não trivial para a distância de reversão;
- Gates e Papadimitriou, em [8], e Gyori e Turan, em [12], trabalharam no problema da ordenação por reversões e prefixos (dado na Definição 10) e os primeiros mostraram que  $\frac{17}{16} \leq D_{pref}(n) \leq \frac{5}{3}n + \frac{5}{3}$ . São questões em aberto, relacionadas com este problema:
  - melhorar os limites inferior e superior da ordenação por reversões e prefixos e
  - encontrar o diâmetro de reversão por prefixos do grupo simétrico.

Em seguida, são comentados alguns parâmetros conhecidos, relativos às três operações básicas da Tabela 5.1, .

Operação	Aproximação	Média (0.2)	Pior caso	Complexidade
<i>Reversão sem sinal</i>	1.375 [25] (1.1)	?	$n-1$ (1.3)	NP-Difícil [3]
<i>Reversão com sinal</i>	exato	?	$2n-1$	$\mathcal{P}(O(n)/O(n \log n))$ (2.4)
<i>Transposição</i>	1.375 [27] (3.1)	?	$n-1$	NP-Completo [18] (3.4)
<i>Intercâmbio de blocos</i>	exato	(4.2)	$n-1$	$\mathcal{P}(O(n^2))$ [4]

Tabela 5.1: Alguns dados relacionados com as três operações básicas.

- (0.2) O número médio de operações, entre as  $n!$  permutações de  $S_n$ , necessárias para ordenar cada uma delas.
- (1.1) (3.1) É importante melhorar a aproximação dos algoritmos.
- (1.3) Somente as permutações de Gollan (dadas na Definição 9) se enquadram no pior caso. O seu estudo e o de novas permutações interessantes pode contribuir para a solução das questões (1.2), (2.2) e (3.2).
- (2.4) Existem algoritmos de complexidade  $O(n)$  para a distância, provado por Bader e outros, em [19], e Swenson, Rajan, Lin e Moret, em [21], demonstraram a complexidade  $O(n \log n)$  para a construção das reversões.
- (3.4) O resultado foi, antes, apresentado por Pinch, em [34], tendo, posteriormente, a sua demonstração, correção feita por Lima e Ayala-Rincón, em [16].
- (4.2) O número médio  $\frac{n - \frac{1}{\lfloor \frac{n+2}{2} \rfloor} - \sum_{i=2}^n \frac{1}{i}}{2}$  do Teorema 9.

Os resultados conhecidos em grupos de simetria e em outras áreas, como a combinatória, são poderosos e continuarão sendo utilizados juntos no intuito de resolver as questões mais complexas. Alguns trabalhos futuros podem ser:

- combinar o método com outros conhecimentos para estudar e chegar (talvez facilite) a resultados para questões em aberto como a da média de operações de reversão necessárias para ordenar uma permutação;
- tentar dar o mesmo tipo de tratamento, por um único tipo de formalismo, aos diferentes tipos de operações (em alguns casos, mais de uma ocorrem em um evento), que hoje são estudadas de formas bastante diversas, o que pode, também, facilitar a descoberta de resultados como a identificação de complexidades, ainda não conhecidas, no problema da distância de transposição;
- propor outros algoritmos.

# Referências

- [1] G. Boccara. Nombres de Répresentations d'une Permutation comme Produit de deux Cycles de Longueur Données. *Discrete Mathematics*, (29), páginas 105-134, 1980. 34
- [2] M. Bóna. *Combinatorics of Permutations*. Chapman and Hall/CRC, 2004. 13, 18
- [3] A. Caprara. Sorting by Reversals is Difficult. *ACM Press, EUA*, páginas 75-83, 1997. 55
- [4] D. A. Christie. Sorting Permutations by Block-interchanges. *Information Processing Letters*, 60, páginas 165-169, 1996. 32, 55
- [5] C. Darwin. *A Origem das Espécies*. Editora Itatiaia, Brasil, 2002. 1
- [6] J.-P. Doignon e A. Labarre. On Hultman Numbers. *Journal of Integers Sequences*, 10, 2007. 30, 31, 32
- [7] P. H. Edelman e C. Greene. Balanced Tableaux. *Advances in Mathematics*, 63:1, 42-99, 1987. 24
- [8] W. H. Gates e C. H. Papadimitriou. Bounds for Sorting by Prefix Reversals. *Discrete Mathematics*, 27, páginas 194-207, 1979. 55
- [9] Y.-L. Huang e C. L. Lu. Sorting by Reversals, Generalized Transpositions and Translocations Using Permutation Groups. *Journal of Computational Biology*, volume 17, número 5, páginas 685-705, 2010. 53
- [10] J. Kececioglu e D. Sankoff. Exact and Approximation Algorithm for Sorting by Reversals, with Application to Genome Rearrangement. *Algorithmica*, 13:180-210, 1995. 2
- [11] J. Kececioglu e D. Sankoff. Exact and Approximation Algorithms for the Inversion Distance Between two Chromosomes. *Proc. of 4th Ann. Symp. on Combinatorial Pattern Matching, Lecture Notes in Computer Science*, 684, páginas 87-105, 1993. 16, 55
- [12] E. Giori e E. Turan. Stack on Pancakes. *Studia Scientiarum Mathematicarum Hungarica*, 13, páginas 133-137, 1979. 55
- [13] R. Simion e F. W. Schmidt. Restricted Permutations. *European Journal of Combinatorics*, 6, páginas 383-406, 1985. 20

- [14] C. V. G. Mira e J. Meidanis. Sorting by Block-Interchanges and Signed Reversals. *Fourth International Conference on Information Technology, IEEE Computer Society, EUA*, páginas 670-676, 2007. 54
- [15] J. C. Setubal e J. Meidanis. *Introduction to Computational Molecular Biology*. PWS, 1997. v, vii, 2, 3, 4, 11, 41, 42
- [16] T. A. Lima e M. Ayala-Rincón. Complexity of Cayley Distance and other General Metrics on Permutation Groups. *UnB/ICE/Departamentos de Matemática e Ciência da Computação, Brasília, DF, Brazil*, 2012. 56
- [17] P. A. Pevzner e M. S. Waterman. Open Combinatorial Problems in Computational Molecular Biology. 1995. 54
- [18] C. Buchheim e outros. On the Subgroup Distance Problem. *Discrete Mathematics*, 309(4), páginas 962-968, 2009. 55
- [19] D. A. Bader e outros. A Linear-time Algorithm for Computing Inversion Distance Between Signed Permutations with an Experimental Study. *7th International Workshop on Algorithms and Data Structures*, páginas 2365-2376, 2001. 56
- [20] G. Fertin e outros. *Combinatorics of Genome Rearrangements*. The MIT Press, Londres, 2004. 41
- [21] K. M. Swenson e outros. Sorting Signed Permutations by Inversions in  $O(n \log n)$  Time. *Laboratory for Computational Biology and Bioinformatics, EPFL, École Polytechnique Fédérale de Lausanne, Suíça*, 17(3), páginas 489-501, março 2010. 56
- [22] V. Bafna e P. A. Pevzner. Sorting by Transpositions. *SIAM, Journal of Discrete Mathematics*, 11(2), páginas 224-240, maio 1998. 29
- [23] V. Bafna e P. A. Pevzner. Genome Rearrangements and Sorting by Reversals. *SIAM, Journal on Computing*, páginas 148-157, 1993. v, 12, 16, 18, 55
- [24] M. Bóna e R. Flynn. The Average Number of Block Interchanges Needed to Sort a Permutation and a Recent Result of Stanley. *Information Processing Letters*, páginas 927-931, 2009. 18, 29
- [25] P. Berman e S. Hannenhalli. Fast Sorting by Reversals. *Springer, Lecture Notes in Computer Science - Proceedings of the 7th Annual Symposium of the Combinatorial Pattern Matching (CPM'96), Berlim*, 1075, páginas 168-185, 2001. 55
- [26] G. Birkhoff e S. MacLane. *Algebra*. Macmillan, Londres, 1971. 7
- [27] I. Elias e T. Hartman. A 1.375-approximation Algorithm for Sorting by Transpositions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4), 2005. 55
- [28] J. Meidanis e Z. Dias. An Alternative Algebraic Formalism for Genome Rearrangements. *Editores D. Sankoff e J. H. Nadeau. Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and Evolution of Gene Families*. Kluwer Academic Press, Amsterdam, páginas 213-223, 2000. 53

- [29] A. Hultman. Toric Permutations. *Master's Thesis, Department of Mathematics, KTH, Suécia*, 1999. 31
- [30] D. E. Knuth. *The Art of Computer Programming*, volume 3. Addison-Wesley, 1973. 28
- [31] T. A. Lima. Representação Combinatória e Algébrica das Permutações na Análise do Problema de Rearranjo de Genomas por Reversões. *UnB/ICE/Departamento de Matemática, Dissertação de Mestrado*, páginas 12-15, março 2010. 41, 42
- [32] L. H. J. Monteiro. *Elementos de Álgebra*. LTC, Rio de Janeiro, 1969. 7
- [33] E. Netto. *Lehrbuch der Combinatoric*. Chelsea, New York, NY, 1901. 19
- [34] R. G. E. Pinch. The Distance of a Permutation from a Subgroup of  $S_n$ . *Editores G Brightwell, I. Leader, A. Scott e A. Thomason, Combinatorics and Probability. Cambridge University Press*, páginas 493-480, 2007. 56
- [35] R. Stanley. On the Number of Reduced Decompositions of Elements of Coxeter Groups. *European Journal of Combinatorics*, 5:359–372, 1984. 24
- [36] R. Stanley. Two Enumerative Results in Cycles of Permutations. *pré impressão*, 2009. 34
- [37] A. Young. On Quantitative Substitutional Analysis II. *Mathematics Society, Londres*, número 1, páginas 361-397, 1902. 21