

Pedro Lucas Damasceno

# Minimização de blocos consecutivos com Parallel Tempering e Lin-Kernighan-Helsgaun

Ouro Preto

2025

Pedro Lucas Damasceno

# Minimização de blocos consecutivos com Parallel Tempering e Lin-Kernighan-Helsgaun

Tese submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Ouro Preto para obtenção do título de Mestre em Ciência da Computação.

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Orientador: Prof. Dr. Marco Antonio Moreira de Carvalho

Ouro Preto

2025

*A simplicidade é o último grau de sofisticação.*  
(Leonardo da Vinci)

# Resumo

O problema de minimização de blocos consecutivos consiste em determinar uma permutação das colunas de uma matriz binária de modo a minimizar o número de blocos de 1's consecutivos em cada linha. Esse problema apresenta aplicações em diversas áreas, como organização de arquivos, programação de linhas de produção, escalonamento e compressão de dados. Por se tratar de um problema de otimização combinatória NP-difícil, métodos aproximados têm se mostrado mais adequados para sua resolução. Neste contexto, este estudo descreve o desenvolvimento de um algoritmo baseado na aplicação da metaheurística *Parallel Tempering* sobre soluções iniciais geradas pelo algoritmo Lin-Kernighan-Helsgaun, após a conversão do problema em um *Traveling Salesman Problem* utilizando a matriz de distâncias de Hamming. A abordagem demonstrou eficiência ao produzir resultados equivalentes ou superiores ao estado da arte em todas as instâncias analisadas, sem que o tempo de execução se tornasse proibitivo.

**Palavras-chave:** Metaheurística, Blocos Consecutivos, Parallel Tempering, Distâncias de Hamming, Lin-Kernighan-Helsgaun.

# Abstract

The problem of consecutive blocks minimization consists in determining a permutation of the columns of a binary matrix so as to minimize the number of consecutive 1's in each row. This problem has applications in several areas, such as file organization, production line scheduling, task sequencing, and data compression. Since it is an NP-hard combinatorial optimization problem, approximate methods have proven to be more suitable for obtaining high-quality solutions. In this context, this study presents the development of an algorithm based on the application of the Parallel Tempering metaheuristic over initial solutions generated by the Lin-Kernighan-Helsgaun algorithm, after transforming the problem into a Traveling Salesman Problem using the Hamming distance matrix. The proposed approach proved effective, producing results equivalent or superior to the current state of the art for all analyzed instances, without incurring prohibitive computational times.

**Keywords:** Metaheuristic, Consecutive Blocks, Parallel Tempering, Hamming Distances, Lin-Kernighan-Helsgaun.

## Lista de ilustrações

## Lista de tabelas

# Lista de abreviaturas e siglas

C1P	Propriedade dos 1's Consecutivos ( <i>Consecutive Ones Property</i> )
CBM	Problema de Minimização de Blocos Consecutivos ( <i>Consecutive Blocks Minimization</i> )
PT	Revenimento Paralelo ( <i>Parallel Tempering</i> )



# Lista de símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence

# Sumário

1	INTRODUÇÃO . . . . .	10
1.1	Justificativa . . . . .	11
1.2	Objetivos . . . . .	11
1.3	Organização do Trabalho . . . . .	12
2	REVISÃO BIBLIOGRÁFICA . . . . .	13
3	FUNDAMENTAÇÃO TEÓRICA . . . . .	14
4	DESENVOLVIMENTO . . . . .	15
5	EXPERIMENTOS COMPUTACIONAIS . . . . .	16
6	CONCLUSÃO . . . . .	17
	REFERÊNCIAS . . . . .	18

# 1 Introdução

A análise de matrizes binárias (0, 1) é fundamental para a abstração e modelagem de problemas de otimização discreta e combinatória. Um problema canônico neste contexto, formalizado por [Fulkerson e Gross \(1965\)](#), é a verificação da Propriedade dos 1's Consecutivos (*Consecutive Ones Property* - C1P). Esta propriedade é satisfeita se existir uma permutação das colunas da matriz que resulte em blocos contíguos de 1's em cada linha. A existência de tal permutação valida a consistência dos dados com um modelo estrutural perfeitamente linear, uma questão originalmente motivada pela análise da estrutura de genes. Para exemplificar os conceitos, considere a Matriz  $A$  (1.1) (C1P), e a Matriz  $B$  (1.2), para a qual nenhuma permutação de colunas é capaz de satisfazer a propriedade.

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (1.1) \quad B = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \quad (1.2)$$

Na prática, a maioria das matrizes que representam problemas reais não possui a C1P. Esta limitação motiva a formulação de um problema de otimização mais geral: o Problema de Minimização de Blocos Consecutivos (*Consecutive Blocks Minimization* - CBM). O objetivo do CBM é determinar a permutação de colunas que minimiza o número total de blocos de 1's na matriz, representando a melhor aproximação possível a uma estrutura linear ideal. A Matriz  $B$  (1.2) possui 5 blocos de 1's, entretanto, é possível permutar suas colunas de forma a diminuir essa quantidade. A Matriz  $C$  (1.3) apresenta uma permutação ótima das colunas da Matriz  $B$  (1.2), reduzindo o total de blocos para 4.

$$C = \begin{matrix} & 1 & 4 & 2 & 3 \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} & (1.3) \end{matrix}$$

O CBM é um problema de relevância multidisciplinar e possui aplicações em áreas diversas, como arqueologia (*Sequence Dating Problem*) ([Kendall, 1969](#)), genética computacional (*Physical Mapping Problem*) ([Alizadeh et al., 1995](#)), compressão de dados em larga escala ([Lemire; Kaser, 2011](#)), e várias outras. Por se tratar de um problema NP-difícil ([Kou, 1977](#)), métodos heurísticos têm se provado mais adequados à sua resolução.

Nesse contexto, este trabalho propõe um algoritmo híbrido para a resolução do CBM, combinando o Revenimento Paralelo (*Parallel Tempering* — PT) com o algoritmo

Lin-Kernighan-Helsgaun (LKH). A metodologia consiste em converter as instâncias do CBM em instâncias do Problema do Caixeiro Viajante (*Traveling Salesman Problem* — TSP), utilizando a matriz de distâncias de Hamming, e empregar o LKH para construir um conjunto de soluções iniciais de alta qualidade. Essas soluções são utilizadas como ponto de partida para o PT, que, por apresentar resultados expressivos em problemas de permutação análogos (Almeida; de Castro Lima; Carvalho, 2025), é utilizado como estratégia de diversificação para expandir o espaço de soluções explorado.

## 1.1 Justificativa

A relevância do CBM é evidenciada por sua vasta aplicabilidade em domínios práticos distintos, que incluem a seriação arqueológica, o mapeamento genético e a compressão de dados em larga escala (Kendall, 1969; Alizadeh et al., 1995; Lemire; Kaser, 2011).

Contudo, a natureza NP-difícil do problema (Kou, 1977) impõe uma complexidade computacional significativa. A obtenção de soluções ótimas por métodos exatos é, na maioria dos casos práticos, inviável, exigindo um tempo de execução que cresce exponencialmente com o tamanho da instância. Esta complexidade intrínseca é a principal motivação para o desenvolvimento e aprimoramento de métodos heurísticos e metaheurísticos, que buscam encontrar soluções de alta qualidade em tempo computacional razoável.

Embora diversas abordagens heurísticas tenham sido propostas para o CBM, a busca por métodos que avancem o estado da arte — ou seja, que encontrem soluções com um número de blocos ainda menor ou que o façam de forma mais eficiente — é um desafio contínuo. Este trabalho se justifica na exploração de um novo método que propõe a sinergia entre duas técnicas de vanguarda da otimização combinatória: uma das mais poderosas heurísticas para o problema do TSP simétrico e uma robusta metaheurística paralela de diversificação.

## 1.2 Objetivos

O principal objetivo deste trabalho é desenvolver um algoritmo híbrido, baseado na combinação dos métodos LKH e PT, capaz de avançar o estado da arte do CBM, identificando soluções de menor custo para as instâncias da literatura em tempo computacional viável. São objetivos específicos:

1. Elaborar uma rigorosa revisão da literatura relacionada ao CBM e problemas permutacionais análogos a fim de compreender abordagens anteriores e identificar lacunas;

2. Desenvolver um novo método baseado na combinação do algoritmo LKH e da metaheurística PT;
3. Realizar experimentos computacionais para configurar o método desenvolvido, analisar suas propriedades a partir das estatísticas de execução e comparar os resultados obtidos com o atual estado da arte.

## 1.3 Organização do Trabalho

Este trabalho está organizado como descrito a seguir. O Capítulo 2 apresenta a revisão da literatura relevante em relação ao CBM. Em seguida, no Capítulo 3, uma base teórica sólida é estabelecida, abrangendo a definição do problema, os objetivos específicos considerados neste trabalho e uma introdução às técnicas empregadas no seu desenvolvimento. O Capítulo 4 discute minuciosamente a implementação das técnicas incorporadas no algoritmo, enquanto o Capítulo 5 apresenta os resultados obtidos pelo método desenvolvido, bem como uma comparação com o estado da arte. Por fim, o Capítulo 6 encerra este trabalho, fornecendo uma síntese dos resultados alcançados e delineando as próximas etapas de pesquisa.

## 2 Revisão Bibliográfica

### 3 Fundamentação Teórica

## 4 Desenvolvimento



## 5 Experimentos Computacionais

## 6 Conclusão

# Referências

ALIZADEH, F. et al. Physical mapping of chromosomes: A combinatorial problem in molecular biology. *Algorithmica*, Springer, v. 13, n. 1, p. 52–76, 1995. Citado 2 vezes nas páginas 10 e 11.

ALMEIDA, A. L. B.; de Castro Lima, J.; CARVALHO, M. A. M. Revisiting the parallel tempering algorithm: High-performance computing and applications in operations research. *Computers & Operations Research*, v. 178, p. 107000, 2025. ISSN 0305-0548. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0305054825000280>>. Citado na página 11.

FULKERSON, D.; GROSS, O. Incidence matrices and interval graphs. *Pacific journal of mathematics*, Mathematical Sciences Publishers, v. 15, n. 3, p. 835–855, 1965. Citado na página 10.

KENDALL, D. Incidence matrices, interval graphs and seriation in archeology. *Pacific Journal of mathematics*, Mathematical Sciences Publishers, v. 28, n. 3, p. 565–570, 1969. Citado 2 vezes nas páginas 10 e 11.

KOU, L. T. Polynomial complete consecutive information retrieval problems. *SIAM Journal on Computing*, SIAM, v. 6, n. 1, p. 67–75, 1977. Citado 2 vezes nas páginas 10 e 11.

LEMIRE, D.; KASER, O. Reordering columns for smaller indexes. *Information Sciences*, Elsevier, v. 181, n. 12, p. 2550–2570, 2011. Citado 2 vezes nas páginas 10 e 11.