O ALGORITMO GENÉTICO BINÁRIO

III.1 ALGORITMOS GENÉTICOS: SELEÇÃO NATURAL COMPUTACIONAL

CAPÍTULO III : GA BINÁRIOS
CAPÍTULO IV : GA DE PARÂMETROS CONTÍNUOS

MENU:
MODELAR
- RECOMBINAÇÃO GENÉTICA E
- SELEÇÃO NATURAL.

BINÁRIOS

→ REPRESENTAM PARÂMETROS POR STRINGS BINÁRIAS

→ TRATAM STRINGS BINÁRIAS P/ MINIMIZAR O CUSTO

DE PARÂMETROS CONTÍNUOS

→ TRATAM OS PRÓPRIOS PARÂMETROS CONTÍNUOS P/ MINIMIZAR O CUSTO.

ANALOGIA ENTRE EVOLUÇÃO E UM GA BINÁRIO

GA BINÁRIO | **EVOLUÇÃO**

POPULAÇÃO

1010101010
1110001000
0011011100
0100001111
1101100011
0011101010
...

POPULAÇÃO INICIAL DE MEMBROS ALEATÓRIOS. CADA LINHA DE N.OS BINÁRIOS REPRESENTA CARACTERÍSTICAS DE UM DOS CACHORROS NA POPULAÇÃO.

FITNESS: ALGUM MAPEAMENTO BINÁRIO QUE CARACTERIZE CÃES LATIDORES.

DESEJAMOS REPRODUZIR O CÃO MAIS LATIDOR.

SELEÇÃO DE PROVÁVEIS REPRODUTORES

1111001101
1110001000
0001000011
0100001111

SELEÇÃO É FUNÇÃO CRESCENTE DO FITNESS.
POUCOS CÃES (4 NO CASO) SÃO SELECIONADOS P/ REPRODUÇÃO.

CRUZAMENTO

$\boxed{111}$0001000
000$\boxed{1000011}$

000 0001000
$\boxed{111}\boxed{1000011}$

DOIS CÃES SÃO SELECIONADOS ALEATORIAMENTE PARA CRIAR DOIS NOVOS FILHOTES.

PROLE

OS FILHOTES TÊM UMA ALTA PROBABILIDADE DE SEREM LATIDORES, PORQUE AMBOS OS PAIS TÊM GENES QUE OS TORNAM LATIDORES.

AS NOVAS SEQÜÊNCIAS BINÁRIAS DOS FILHOTES CONTÊM PORÇÕES DAS SEQÜÊNCIAS BINÁRIAS DE AMBOS OS PAIS.

NOVA POPULAÇÃO

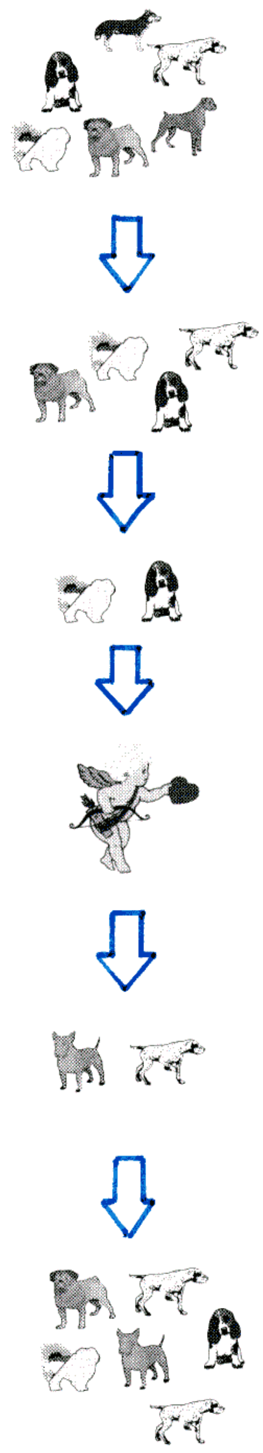
- 1110001000
- 0001000011
- 0000001000
- 1111000011

1111001101
0100001111

OS NOVOS FILHOTES SUBSTITUIRÃO OS DOIS CÃES DESCARTADOS (QUE NÃO ERAM LATIDORES)

UM N.º SUFICIENTE DE FILHOTES FOI GERADO PARA TRAZER A POPULAÇÃO DE VOLTA AO TAMANHO ORIGINAL.

ITERAÇÕES DESTE PROCESSO CONDUZIRÃO A UM CÃO COM UMA FORTE CARACTERÍSTICA DE LATIDOR.



III.2 COMPONENTES DE UM ALGORITMO GENÉTICO BINÁRIO

DEFINIÇÃO DE PARÂMETROS, FUNÇÃO DE CUSTO, ... = OUTROS ALGORITMOS DE OTIMIZAÇÃO

↓

REPRESENTAÇÃO PARAMÉTRICA

↓

CRIAÇÃO DA POPULAÇÃO

↓

AVALIAÇÃO DO CUSTO

↓

SELEÇÃO DOS REPRODUTORES

↓

REPRODUÇÃO

↓

MUTAÇÃO

↓

TESTE DE CONVERGÊNCIA = OUTROS ALGORITMOS DE OTIMIZAÇÃO

↓

FIM

(REPRESENTAÇÃO PARAMÉTRICA ... MUTAÇÃO: MUITO ≠)

A FUNÇÃO DE CUSTO É UMA SUPERFÍCIE COM PICOS E VALES NO ESPAÇO DE PARÂMETROS, COMO UM MAPA TOPOGRÁFICO.

VALE: ALGO BUSCA CUSTO MÍN.
PICO: ALGO BUSCA CUSTO MÁX.

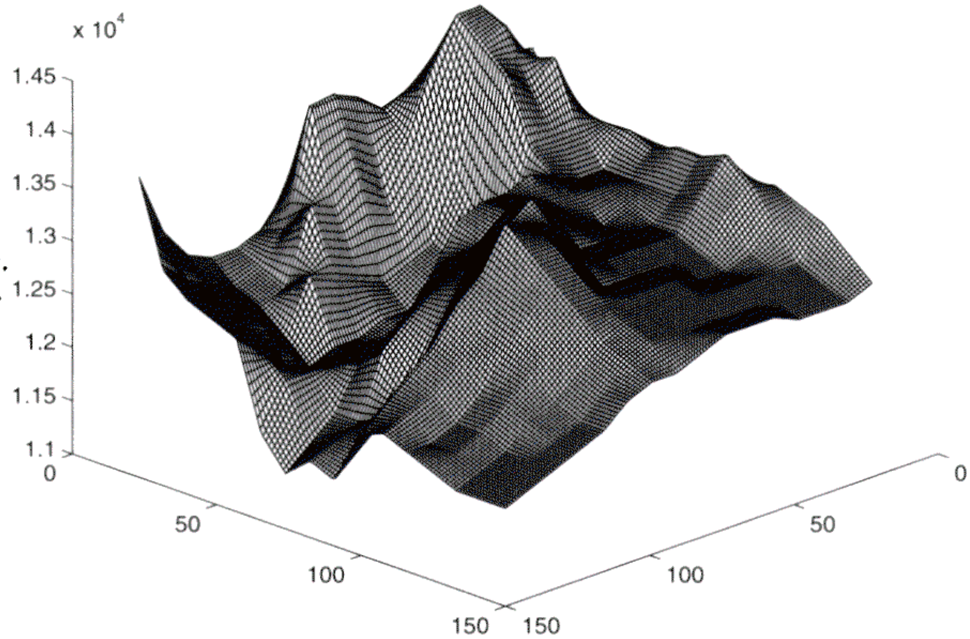


FIG. III.1: GRÁFICO TRI-DIM. DE UMA PORÇÃO DO ROCKY MOUNTAIN NATIONAL PARK.

PTO DE MÁXIMO GLOBAL: • LONG'S PEAK

PTOS DE MÁXIMOS LOCAIS:
- STORM PEAK
- MOUNT LADY WASHINGTON
- CHASM LAKE

↳ TÉCNICAS CONVENCIONAIS DE OTIMIZAÇÃO FALHAM EM ENCONTRAR O MÁX. GLOBAL. (A MENOS QUE INICIEM A BUSCA NA VIZINHANÇA IMEDIATA DO PICO)

→ TODOS OS MÉTODOS QUE REQUEIRAM UM GRADIENTE DA FUNÇÃO DE CUSTO IRÃO FALHAR COM DADOS DISCRETOS.

→ GA SE APLICAM.

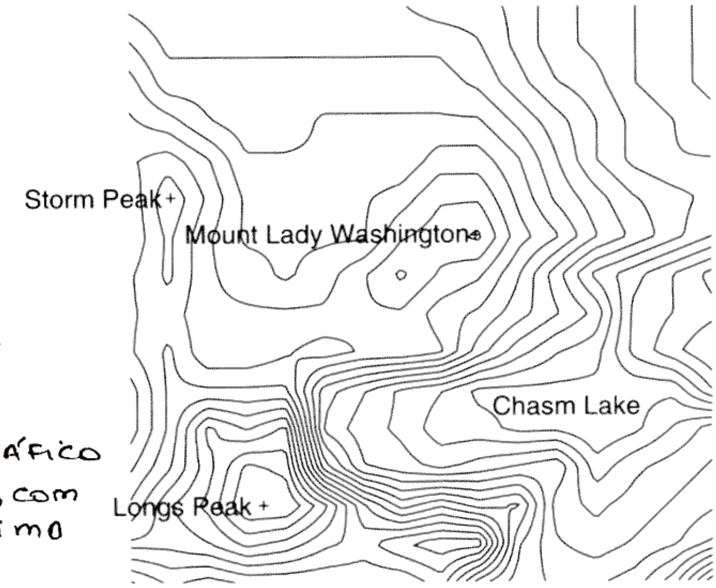


FIG. III.2: MAPA TOPOGRÁFICO DE 128 x 128 PONTOS, COM OS PONTOS DE MÁXIMO MARCADOS.

III. 2.1. SELEÇÃO DOS PARÂMETROS E DA FUNÇÃO DE CUSTO

EM UM ALGORITMO DE OTIMIZAÇÃO O OBJETIVO É MODIFICAR A SAÍDA, DE ACORDO COM ALGUM CRITÉRIO, ATRAVÉS DA ESCOLHA APROPRIADA DE VALORES PARA OS PARÂMETROS DE ENTRADA.

A FUNÇÃO DE CUSTO GERA UMA SAÍDA A PARTIR DE UM CONJUNTO DE PARÂMETROS DE ENTRADA (CROMOSSOMOS).

ASSIM, SÃO FUNDAMENTAIS, A ESCOLHA DA FUNÇÃO DE CUSTO APROPRIADA E DOS PARÂMETROS ADEQUADOS A SEREM UTILIZADOS.

→ OS VETORES DE PARÂMETROS (OU CROMOSSOMOS) QUE SERÃO UTILIZADOS SÃO VETORES N_{par}-DIMENSIONAIS, DA FORMA:

$$\text{CROMOSSOMO} = [p_1, p_2, \ldots, p_{N_{par}}] \qquad (III.1)$$

→ EM FUNÇÕES DE CUSTO TRIDIMENSIONAIS (MAPAS TOPOGRÁFICOS), A BUSCA PELO VALOR MÁXIMO REQUER 2 PARÂMETROS DE ENTRADA ($N_{par} = 2$), CROMOSSOMO $= [x, y]$ (III.2)

→ CADA CROMOSSOMO TEM O CUSTO DETERMINADO AVALIANDO A FUNÇÃO DE CUSTO f EM $p_1, p_2, \ldots, p_{N_{par}}$:

$$\text{CUSTO} = f(\text{CROMOSSOMO}) = f(p_1, p_2, \ldots, p_{N_{par}}) \qquad (III.3)$$

→ QUANDO BUSCAMOS O PICO, A FUNÇÃO DE CUSTO É ESCRITA COMO $f(x, y) = -\text{ELEVAÇÃO}$, (III.4) PARA QUE O PROBLEMA SEJA TRATADO COMO UM PROBLEMA DE MINIMIZAÇÃO.

QUANTO À ESCOLHA DE PARÂMETROS :

QUANDO CONHECEMOS A EXPRESSÃO ANALÍTICA DA FUNÇÃO DE CUSTO, OS PARÂMETROS SÃO AS VARIÁVEIS DA FUNÇÃO.

QUANDO NÃO CONHECEMOS, A ESCOLHA DOS PARÂMETROS É BASEADA EM ESTIMATIVAS EMPÍRICAS, OU EM TENTATIVAS.

MUITOS PROBLEMAS DE OTIMIZAÇÃO REQUEREM LIMITES OU RESTRIÇÕES ($>$, $<$, $\geqslant$, $\leqslant$).

VARIÁVEIS RESTRITAS PODEM SER TRANSFORMADAS EM VARIÁVEIS IRRESTRITAS, PERMITINDO QUE UM PROBLEMA DE OTIMIZAÇÃO RESTRITO POSSA SER TRATADO COMO IRRESTRITO.

A MUDANÇA DE VARIÁVEIS PODE SER DA FORMA :
SE x TEM LIMITES DEFINIDOS POR $0 \leqslant x \leqslant 10$, ENTÃO $x = 5 \operatorname{sen} y + 5$ É UMA TRANSFORMAÇÃO ENTRE A VARIÁVEL RESTRITA x E A VARIÁVEL IRRESTRITA y.

VARIAR y PARA QUALQUER VALOR É O MESMO QUE VARIAR x DENTRO DOS LIMITES DE x; $0 \leqslant x \leqslant 10$.

PARÂMETROS DEPENDENTES REPRESENTAM PROBLEMAS ESPECIAIS PARA ALGORITMOS DE OTIMIZAÇÃO.

NA LITERATURA DE GAS, INTERAÇÃO ENTRE PARÂMETROS É CHAMADA EPISTASE (TERMO BIOLÓGICO PARA INTERAÇÃO ENTRE GENES).

GAS, NO ENTANTO, DESEMPENHAM ADEQUADAMENTE QUANDO A EPISTASE FOR MÉDIA OU ALTA.

QUANTO À DEFINIÇÃO DA FUNÇÃO DE CUSTO:

FUNÇÕES DE CUSTO PODEM SER RELACIONADAS:

→ A UMA SAÍDA OBSERVÁVEL OU

→ A UMA COMBINAÇÃO DAS SAÍDAS OBSERVÁVEIS

QUANDO MAIS DE UMA SAÍDA PRECISA SER CONSIDERADA EM UMA FUNÇÃO DE CUSTO, AS SAÍDAS ENVOLVIDAS POSSUEM ORDENS DE GRANDEZA DIFERENTES E, NO ENTANTO, SÃO IGUALMENTE IMPORTANTES NA FORMAÇÃO DO CUSTO, É NECESSÁRIO NORMALIZAR S A

POR EXEMPLO: NO PROJETO DE UM AUTOMÓVEL, CONSIDEREMOS QUE O DINHEIRO ENVOLVIDO NA CONSTRUÇÃO DO CARRO (R$) E A TAXA DE CONSUMO DE COMBUSTÍVEL ($C = km/\ell$) SEJAM MEDIDAS DE PROJETO, COM PARÂMETROS DEFINIDOS POR POTÊNCIA DO MOTOR (p) E TORQUE DO MOTOR (t).

* TEREMOS UMA FUNÇÃO DE CUSTO DA FORMA

$$\underbrace{f(p,t)}_{\text{PARÂMETROS}} = \underbrace{R\$ + C}_{\text{SAÍDAS OBSERVÁVEIS A SEREM COMBINADAS}} \qquad (III.5)$$

* COMO AS SAÍDAS R$ E ℓ/km POSSUEM ORDENS DE GRANDEZA DIFERENTES (R$ 40K; $C = 12\ km/\ell$, P. EXPLO.), O VALOR DE $f(p,t)$ TENDERÁ AO VALOR MAIOR, O CUSTO MONETÁRIO DO VEÍCULO ⇒ NORMALIZAÇÃO.

$$f(p,t) = \frac{R\$ - R\$_{lo}}{R\$_{hi} - R\$_{lo}} + \frac{C - C_{lo}}{C_{hi} - C_{lo}} \qquad (III.6)$$

ONDE: hi INDICA VALOR MÁXIMO E lo INDICA VALOR MÍNIMO.

SE: $R\$ = 40K$; $R\$_{lo} = 30K$; $R\$_{hi} = 50K$ E

$C = 12\ km/\ell$; $C_{lo} = 10\ km/\ell$; $C_{hi} = 14\ km/\ell$,

ENTÃO: $f(p,t) = \dfrac{40K - 30K}{50K - 30K} + \dfrac{12 - 10}{14 - 10} = \dfrac{10K}{20K} + \dfrac{2}{4} = 1$

QUANDO MAIS DO QUE UMA SAÍDA PRECISA SER CONSIDERADA EM UMA FUNÇÃO DE CUSTO, AS SAÍDAS ENVOLVIDAS POSSUEM ORDENS DE GRANDEZA DIFERENTES E UMA PARCELA DO CUSTO É MAIS IMPORTANTE NA FORMAÇÃO DO CUSTO, AS PARCELAS PODEM SER PONDERADAS ADEQUADAMENTE $(0 < w < 1)$.

EM NOSSO EXEMPLO:

$$f(p,t) = w \left\{ \frac{R\$ - R\$_{lo}}{R\$_{hi} - R\$_{lo}} \right\} + (1-w) \left\{ \frac{C - C_{lo}}{C_{hi} - C_{lo}} \right\} \quad (\text{III}.7)$$

* ALGORITMOS GENÉTICOS SÃO SENSÍVEIS AOS INTERVALOS DE VARIAÇÃO DOS PARÂMETROS E DAS SAÍDAS (OU CUSTO).

* A EQUAÇÃO (III.7) PERMITE UM EXCELENTE CONTROLE SOBRE A SAÍDA $f(p,t)$ (OU CUSTO).

A ESCOLHA DE PONDERADORES (w) ADEQUADOS, NO ENTANTO, NÃO É SEMPRE IMEDIATA.

III.2.2 REPRESENTAÇÃO PARAMÉTRICA

CASO DISCRETO:

→ O GA BINÁRIO TRABALHA COM UM ESPAÇO DE PARÂMETROS FINITO (DISCRETO), MAS USUALMENTE MUITO GRANDE.

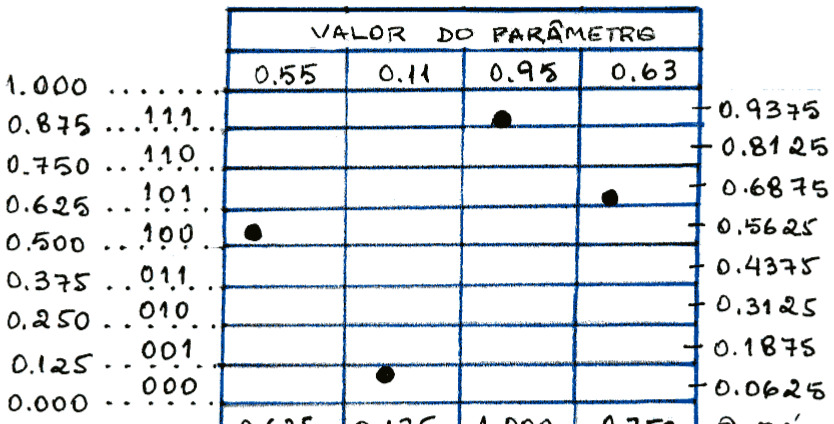
→ ESTA CARACTERÍSTICA TORNA OS GAS IDEAIS PARA OTIMIZAR UM CUSTO QUE É DEVIDO A PARÂMETROS QUE PODEM ASSUMIR UM Nº FINITO DE VALORES.

CASO CONTÍNUO:

→ SE UM PARÂMETRO É CONTÍNUO, PRECISA SER QUANTIZADO.

1. O INTERVALO CONTÍNUO É DIVIDIDO EM NÍVEIS DE QUANTIZAÇÃO IGUAIS.
2. QUALQUER VALOR QUE CAIA DENTRO DE UM DOS NÍVEIS É QUANTIZADO PARA O VALOR MÉDIO, ALTO, OU BAIXO DAQUELE NÍVEL.

- EM GERAL, ESCOLHER O VALOR MÉDIO É MAIS INDICADO, PORQUE, NESTE CASO, O MAIOR ERRO POSSÍVEL CORRESPONDERÁ À METADE DO NÍVEL.
- ESCOLHER O VALOR MÍNIMO OU O VALOR MÁXIMO DO NÍVEL PERMITIRÁ UM ERRO MÁXIMO IGUAL AO NÍVEL DE QUANTIZAÇÃO.



0.55	0.11	0.95	0.63	
0.625	0.125	1.000	0.750	Q. MÁX.
0.500	0.000	0.875	0.625	Q. MÍN.
0.5625	0.0625	0.9375	0.6875	Q. MÉDIO
100	000	111	101	CROMOSSOMO

- QUATRO VALORES DE PARS. CONTÍNUOS + OS NÍVEIS DE QUANTIZAÇÃO.
- O CROMOSSOMO CORRESPONDENTE INDICA O NÍVEL DE QUANTIZAÇÃO ONDE O VALOR DO PARÂMETRO CAIU.
- CADA CROMOSSOMO CORRESPONDE A UM NÍVEL Q.MÍN, MÁX OU MÉDIO.
- O PARÂMETRO É QUANTIZADO P/ O VALOR Q MÉDIO DO NÍVEL DE QUANTIZAÇÃO.

NO EXEMPLO CITADO NAS FIGURAS III.1 E III.2, O MAPA TOPOGRÁFICO DO PARQUE NACIONAL DAS MONTANHAS ROCHOSAS TEM 128 x 128 PONTOS DE ELEVAÇÃO (OU COTAS).

DE ACORDO COM A EQUAÇÃO (III.4), A FUNÇÃO DE CUSTO P/ O PROBLEMA (BUSCA DE MÁXIMO = LONG'S PEAK) É ESCRITA COMO

$$f(x,y) = -\text{ELEVAÇÃO}$$

↑ P/ TRANSFORMAR O PROBLEMA EM MINIMIZAÇÃO.

ONDE, PORTANTO, OS CROMOSSOMOS SERÃO FORMADOS POR DOIS GENES: $x \to$ LONGITUDE, $y \to$ LATITUDE } DO PONTO.

SE x E y SÃO CODIFICADOS EM DOIS GENES, CADA UM COM N_{GENE} = nº DE BITS DO GENE = 7 bits, ENTÃO HAVERÁ 2^7 POSSÍVEIS VALORES PARA x E y.

UM CROMOSSOMO ALEATÓRIO PODERÁ ASSUMIR A SEGUINTE REPRESENTAÇÃO BINÁRIA (DE $N_{BITS} = 14$ bits):

$$\text{CROMOSSOMO} = [\ \underbrace{1100011}_{x}\underbrace{0011001}_{y}\ ]$$

CORRESPONDENDO ÀS COORDENADAS [99, 25] DE UMA MATRIZ DE 128 x 128 PONTOS.

$1100011 \to 99$ $0011001 \to 25$

$2^6 + 2^5 + 2^1 + 2^0$ $2^4 + 2^3 + 2^0$

* NO CASO DAS MONTANHAS ROCHOSAS, OS VALORES POSSÍVEIS PARA LONGITUDE SÃO $105°37'30'' \geq x \geq 105°36'$ E PARA LATITUDE SÃO $40°16' \geq y \geq 40°15'$.

III.2.3 POPULAÇÃO INICIAL

O GA INICIA COM UM GRANDE Nº DE INDIVÍDUOS, OU CROMOSSOMOS, CHAMADO DE POPULAÇÃO INICIAL (OU IPOP).
QUANTO ↑ IPOP, MELHOR A AMOSTRAGEM DO ESPAÇO DE BUSCA P/ O GA.

IPOP TEM N_{IPOP} CROMOSSOMOS E É UMA MATRIZ DE $[N_{IPOP} \times N_{BITS}]$, CUJOS ELEMENTOS SÃO ZEROS E UNS GERADOS ALEATORIAMENTE A PARTIR DE

$$IPOP = ROUND \{ RANDOM (N_{IPOP}, N_{BITS}) \}$$

ONDE A FUNÇÃO

→ $RANDOM (N_{IPOP}, N_{BITS})$

GERA UMA MATRIZ DE $[N_{IPOP} \times N_{BITS}]$ NºS ALEATÓRIOS COM DISTRIBUIÇÃO UNIFORME ENTRE ZERO E UM.

E A FUNÇÃO

→ $ROUND \{\cdot\}$

ARREDONDA OS NºS GERADOS PARA O INTEIRO MAIS PRÓXIMO.

* CADA LINHA DA MATRIZ É UM CROMOSSOMO.
* CADA CROMOSSOMO CORRESPONDE A VALORES DISCRETOS DE LONGITUDE E LATITUDE.

→ A PARTIR DAS CURVAS DE NÍVEL MOSTRADAS NA FIGURA III.2, SÃO ATRIBUÍDOS OS RESPECTIVOS CUSTOS A CADA CROMOSSOMO (LEMBRANDO QUE $f(x,y) = -$ ELEVAÇÃO).

Chromosome	Cost
00000000000000	−13000
11111011010010	−11800
00010110000010	−13255
11000011001010	−12347
01111111101001	−12560
01000111010001	−12700
01010110000100	−13338
11101111001110	−11890
01111100111100	−12953
00100001011110	−12891
10001110111010	−12759
10111000111100	−12320
11011011101000	−11797
00100110011101	−13778
00010100011011	−13360
01110010101011	−12220
11000011001100	−12452
10011101110000	−12335
10100000000011	−12857
00001101010110	−13166
00010000110101	−13164
01101100110010	−12927
01101111000010	−13079
10001001011111	−12756

TABELA III.1: POPULAÇÃO INICIAL E CUSTO PARA NIPOP = 24 CROMOSSOMOS ALEATÓRIOS.

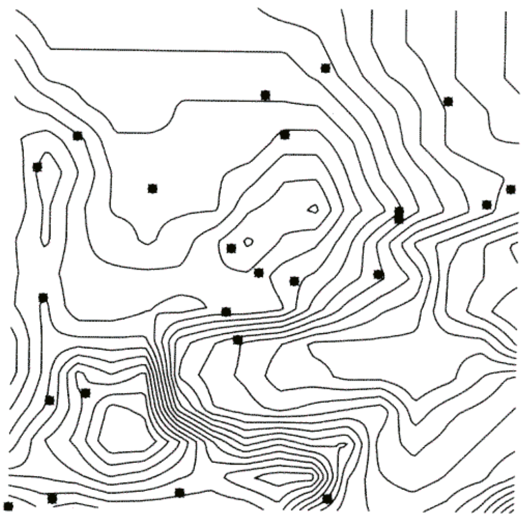
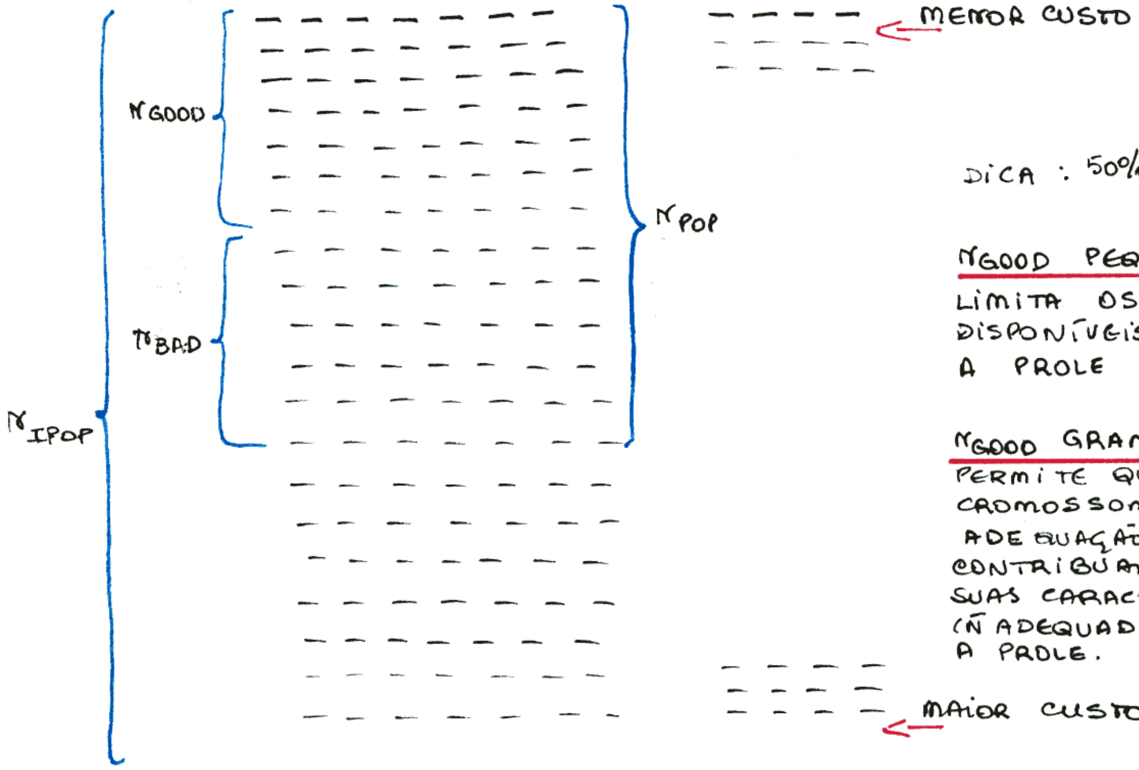


FIGURA III.3: MAPA TOPOLÓGICO COM A LOCALIZAÇÃO DOS CROMOSSOMOS.

III.2.4 SELEÇÃO NATURAL

A POPULAÇÃO INICIAL (IPOP) É, EM GERAL, MUITO GRANDE PARA SER INTEIRAMENTE UTILIZADA AO LONGO DOS PASSOS ITERATIVOS DO ALGORITMO GENÉTICO.

UMA PORÇÃO DE CROMOSSOMOS DE CUSTO ELEVADO SÃO, ENTÃO, DESCARTADOS POR MEIO DE SELEÇÃO NATURAL (OU: SOBREVIVÊNCIA DO MAIS ADAPTADO – FITTEST).



DICA: 50% NGOOD

NGOOD PEQUENO
LIMITA OS GENES DISPONÍVEIS PARA A PROLE

NGOOD GRANDE
PERMITE QUE CROMOSSOMOS DE ADEQUAÇÃO INFERIOR CONTRIBUAM COM SUAS CARACTERÍSTICAS (Ñ ADEQUADAS) PARA A PROLE.

→ APENAS OS $N_{POP} \leq N_{IPOP}$ MELHORES MEMBROS DA POPULAÇÃO SÃO MANTIDOS PARA SEREM OPERADOS POR CADA ITERAÇÃO DO GA. OS DEMAIS SÃO DESCARTADOS.

* SELEÇÃO NATURAL OCORRE A CADA GERAÇÃO DO ALGORITMO (OU ITERAÇ.).

* DOS NPOP CROMOSSOMOS EM UMA GERAÇÃO, APENAS OS NGOOD SUPERIORES IRÃO SOBREVIVER P/ SEREM COMBINADOS (CRUZADOS).

* OS NBAD CROMOSSOMOS INFERIORES SÃO DESCARTADOS PARA "ABRIREM ESPAÇO" PARA OS CROMOSSOMOS DA FUTURA GERAÇÃO.

→ NO EXEMPLO CONSIDERADO, $N_{IPOP} = 24$.

→ NO ENTANTO, APENAS $N_{POP} = 12$ SERÃO MANTIDOS PARA A POPULAÇÃO, A CADA ITERAÇÃO DO GA.

→ A CADA ITERAÇÃO $N_{GOOD} = 6$ CROMOSSOMOS SÃO SELECIONADOS PARA REPRODUÇÃO. E $N_{BAD} = 6$ CROMOSSOMOS SÃO DESCARTADOS.

		Chromosome	Cost	
N_{ipop}	N_{good}	00100110011101	−13778	N_{pop}
		00010100011011	−13360	
		01010110000100	−13338	
		00010110000010	−13255	
		00001101010110	−13166	
		00010000110101	−13164	
	N_{bad}	01101111000010	−13079	
		00000000000000	−13000	
		01111100111100	−12953	
		01101100110010	−12927	
		00100001011110	−12891	
		10100000000011	−12857	
		10001110111010	−12759	
		10001001011111	−12756	
		01000111010001	−12700	
		01111111101001	−12560	
		11000011001100	−12452	
		11000011001010	−12347	
		10011101110000	−12335	
		10111000111100	−12320	
		01110010101011	−12220	
		11101111001110	−11890	
		11111011010010	−11800	
		11011011101000	−11797	

TABELA III.2: SELEÇÃO NATURAL SOBRE NIPOP

THRESHOLDING:

ABORDAGEM PARA SELEÇÃO NATURAL EM QUE TODOS OS CROMOSSOMOS QUE TÊM UM CUSTO INFERIOR DO QUE ALGUM LIMIAR IRÃO SOBREVIVER (EQUIVALE A UM CORTE NA POPULAÇÃO).

1ª SELEÇÃO: APENAS POUCOS CROMOSSOMOS SOBREVIVERÃO.

2ª SELEÇÃO: UM Nº MAIOR DE CROMOSSOMOS SOBREVIVERÃO (DESDE QUE O LIMIAR NÃO SEJA ALTERADO).

⋮

III.2.5 SELEÇÃO DE PARES

DOIS CROMOSSOMOS SÃO SELECIONADOS PARA REPRODUÇÃO DENTRE OS n_{GOOD} SOBREVIVENTES.

O PROCESSO "SELEÇÃO DE PARES" OCORRE ATÉ QUE n_{BAD} CROMOSSOMOS FILHOS TENHAM SIDO GERADOS (PARA QUE SUBSTITUAM OS n_{BAD} DESCARTADOS).

HÁ DIVERSAS HEURÍSTICAS P/ A SELEÇÃO DE PARES:

1. SELEÇÃO DE PARES "DE CIMA P/ BAIXO": (FROM TOP TO BOTTOM)

- OS PARES SÃO FORMADOS A PARTIR DO TOPO DA LISTA, ATÉ QUE OS n_{GOOD} CROMOSSOMOS SEJAM SELECIONADOS.

PARES FORMADOS: $\begin{cases} \text{CROMOSSOMO}_{2i-1} \text{ COM CROMOSSOMO}_{2i}\ ; \\ i = 1, 2, \ldots \end{cases}$

(CROMOSSOMO_1 & CROMOSSOMO_2 ;
CROMOSSOMO_3 & CROMOSSOMO_4 ;
...)

- CONSTITUI A ABORDAGEM MAIS SIMPLES.
- NÃO MODELA A NATUREZA.

2. SELEÇÃO DE PARES ALEATÓRIA:

- UTILIZA UM GERADOR DE N^{OS} ALEATÓRIOS COM DISTRIBUIÇÃO DE PROBABILIDADE UNIFORME P/ SELECIONAR OS CROMOSSOMOS.
- OS CROMOSSOMOS SÃO ORDENADOS EM FUNÇÃO DO CUSTO, DE 1 A n_{GOOD} E DOIS N^{OS} ALEATÓRIOS SÃO GERADOS PARA DETERMINAR OS DOIS 1OS CROMOSSOMOS A CONSTITUIREM UM PAR.

CROMOSSOMOS PAIS = ROUNDUP{NGOOD × RANDOM}

ARREDONDA {·} P/ O PRÓX. INTEIRO SUPERIOR

RANDOM	×	n_{GOOD}		
0.1535	×	6	= 0.921	⇒ 1
0.6781	×	6	= 4.0686	⇒ 5
0.0872	×	6	= 0.5232	⇒ 1
0.1936	×	6	= 1.1616	⇒ 2
0.7021	×	6	= 4.2126	⇒ 5
0.3933	×	6	= 2.3598	⇒ 3

3. SELEÇÃO DE PARES ALEATÓRIA PONDERADA:

ATRIBUI PROBABILIDADES AOS CROMOSSOMOS SELECIONADOS PARA REPRODUÇÃO, DE ACORDO COM AS FUNÇÕES DE CUSTO DOS CROMOSSOMOS.

CROMOSSOMO ↓ F DE CUSTO ⟺ ↑ PROBABILIDADE DE REPROD.

UM Nº ALEATÓRIO DETERMINA QUAL CROMOSSOMO SERÁ SELECIONADO.

PONDERAÇÃO É REFERIDA COMO "ROULETTE WHEEL WEIGHTING".

3.1 PONDERAÇÃO POR POSIÇÃO DO CROMOSSOMO EM NGOOD:

A PROBABILIDADE É DETERMINADA A PARTIR DA POSIÇÃO DO CROMOSSOMO POR:

$$P_n = \frac{N_{GOOD} - n + 1}{\sum_{n=1}^{N_{GOOD}} n} = \frac{6 - n + 1}{1+2+3+4+5+6} = \frac{7-n}{21}$$

A TABELA III.3 APRESENTA OS RESULTADOS P/ NOSSO EXPLO:

n	Chromosome	P_n	$\sum_{i=1}^{n} P_i$
1	00100110011101	0.2857	0.2857
2	00010100011011	0.2381	0.5238
3	01010110000100	0.1905	0.7143
4	00010110000010	0.1429	0.8572
5	00001101010110	0.0952	0.9524
6	00010000110101	0.0476	1.0000

AS PROBABILIDADES CUMULATIVAS MOSTRADAS NA COL. 4 SÃO USADAS NA SELEÇÃO DO CROMOSSOMO.

UM Nº ALEATÓRIO É GERADO (ENTRE 0 E 1) E, A COMEÇAR DE CIMA, SÃO SELECIONADOS OS CROMOSSOMOS CUJA PROB. CUMULATIVA SEJA > DO QUE O Nº ALEATÓRIO.

0.1535 → 0.1535 < 0.2857 → CROM. 1
0.6781 → 0.6781 < 0.7143 → CROM. 3
0.0872 → 0.0872 < 0.2857 → CROM. 1
0.1936 → 0.1936 < 0.2857 → CROM. 1
0.7021 → 0.7021 < 0.7143 → CROM. 3
0.3933 → 0.3933 < 0.5238 → CROM. 2

← PODE FICAR ASSIM OU ALEATORIAMENTE ESCOLHER OUTRO CROMOSSOMO.

- MAIS SEMELHANTE A PROCESSOS NATURAIS.
- ABORDAGEM SIMPLES.

3.2 PONDERAÇÃO POR CUSTO:

A PROBABILIDADE É CALCULADA A PARTIR DO CUSTO DO CROMOSSOMO.

UM CUSTO NORMALIZADO É CALCULADO PARA CADA CROMOSSOMO, SUBTRAINDO O MENOR CUSTO DOS CROMOSSOMOS DESCARTADOS ($CUSTO_{NGOOD+1}$) DO CUSTO DE TODOS OS CROMOSSOMOS CANDIDATOS À REPRODUÇÃO:

$$C_n = CUSTO_n - CUSTO_{NGOOD+1}$$

A OPERAÇÃO RESULTARÁ EM C_n NEGATIVOS.

A TABELA III.4 LISTA OS CUSTOS NORMALIZADOS ASSUMINDO QUE $CUSTO_{NGOOD+1} = -13079$.

P_n É CALCULADO POR: $P_n = \left| \dfrac{C_n}{\sum_{p=1}^{NGOOD} C_n} \right|$

n	Chromosome	$C_n = c_n - c_{N_{good}+1}$	$\mathbf{P}_n$	$\sum_{i=1}^{n} \mathbf{P}_i$
1	00100110011101	$-13778 + 13079 = -699$	0.4401	0.4401
2	00010100011011	$-13360 + 13079 = -281$	0.1772	0.6174
3	01010110000100	$-13338 + 13079 = -259$	0.1632	0.7805
4	00010110000010	$-13255 + 13079 = -176$	0.1109	0.8915
5	00001101010110	$-13166 + 13079 = -87$	0.0547	0.9461
6	00010000110101	$-13164 + 13079 = -85$	0.0539	1.0000

USANDO OS MESMOS 6 N^OS ALEATÓRIOS E A COLUNA 5, TEREMOS:

0.1535 < 0.4401 → CROMOSSOMO 1
0.6781 < 0.7805 → " 3
0.0872 < 0.4401 → " 1
0.1936 < 0.4401 → " 1
0.7021 < 0.7805 → " 3
0.3933 < 0.4401 → " 1

- TENDE A PONDERAR MAIS OS CROMOSSOMOS DE MENOR CUSTO QUANDO HÁ UMA DISTÂNCIA MAIOR ENTRE O MAIOR E O MENOR CUSTO, EM NGOOD.
- TENDE A PONDERAR IGUALMENTE QUANDO OS CUSTOS DOS CROMOSSOMOS DE NGOOD ≃ OS MESMOS.

3.3 TOURNAMENT :

SE ASSEMELHA A PROCESSOS NATURAIS.

CONSISTE EM PEGAR ALEATORIAMENTE UM PEQUENO SUBCONJUNTO DE CROMOSSOMOS SELECIONADOS PARA REPRODUÇÃO, E ESCOLHER O CROMOSSOMO COM O MENOR CUSTO PARA SER O REPRODUTOR.

ESTA SELEÇÃO SE REPETE A CADA CROMOSSOMO REPRO- QUE SE QUEIRA BUSCAR.

PARTINDO DE 6 SELEÇÕES POR MEIO DESTE PROCESSO, EM QUE SEJAM SELECIONADOS ALEATORIAMENTE 2 CROMOSSOMOS DE NGOOD, SERÃO ESCOLHIDOS OS 6 CROMOSSOMOS QUE SERÃO OS REPRODUTORES, CONFORME TABELA III.5.

OS CROMOSSOMOS SE FORMARÃO OS PARES REPRODUTORES SERÃO : 1 & 5 ; 3 & 4 ; 1 & 4 .

Tournament	6 × random(2, 1)	Parent
1	2, 1	1
2	5, 5	5
3	6, 3	3
4	4, 5	4
5	1, 1	1
6	4, 5	4

III.2.6 CRUZAMENTO

→ É A CRIAÇÃO DE UM OU MAIS CROMOSSOMOS "FILHOS" A PARTIR DOS CROMOSSOMOS "PAIS" (REPRODUTORES) SELECIONADOS NO PROCESSO DE SELEÇÃO DE PARES.

→ É UMA OPERAÇÃO QUE UTILIZA OS BITS QUE, EM SEQÜÊNCIA, CONSTITUEM OS CROMOSSOMOS.

→ DOIS CROMOSSOMOS REPRODUTORES PRODUZEM DOIS FILHOS.

→ UM PONTO DE CRUZAMENTO É SELECIONADO NA STRING DE BITS DOS CROMOSSOMOS PAIS.

→ OS CROMOSSOMOS FILHOS CONTERÃO PORÇÕES DOS CÓDIGOS BINÁRIOS DE AMBOS OS PAIS.

→ Nº FILHOS = N_{BAD} P/ QUE A POPULAÇÃO VOLTE A N_{POP}.

• A TABELA III.6 ILUSTRA O PROCESSO.

Chromosome	Family	Binary String	Cost
1	*parent*$_1$	00100110011 101	−13778
3	*parent*$_2$	01010110000 100	−13338
7	*offspring*$_1$	00100110011 100	−13372
8	*offspring*$_2$	01010110000101	−13563
1	*parent*$_3$	001001100 11101	−13778
1	*parent*$_4$	001001100 11101	−13778
9	*offspring*$_3$	001001100 11101	−13778
10	*offspring*$_4$	00100110011101	−13778
3	*parent*$_5$	0 1010110000100	−13338
1	*parent*$_6$	00100110011101	−13778
11	*offspring*$_5$	0 0100110011101	−13778
12	*offspring*$_6$	01010110000100	−13338

III.2.7 MUTAÇÃO

→ MUTAÇÕES ALEATÓRIAS ALTERAM UM PEQUENO Nº DE BITS NA LISTA DE CROMOSSOMOS.

→ CONSTITUEM A 2ª FORMA DE EXPLORAÇÃO DE UMA SUPERFÍCIE DE CUSTO POR UM GA.

→ PODEM PRODUZIR CARACTERÍSTICAS NA POPULAÇÃO ORIGINAL QUE NÃO ESTAVAM PRESENTES.

→ MUTAÇÃO DE UM ÚNICO PONTO (PUNTUAL) MUDA UM "1" EM UM "0" E VICE-VERSA.

→ PONTOS DE MUTAÇÃO SÃO ALEATORIAMENTE SELECIONADOS ENTRE OS $N_{POP} \times N_{BITS}$ BITS TOTAIS NA MATRIZ QUE CONSTITUI A POPULAÇÃO.

! { ↑ O Nº DE MUTAÇÕES ↑ A LIBERDADE DO ALGORITMO NA BUSCA FORA DA REGIÃO ATUAL NO ESPAÇO DE PARÂMS.

! { DIFICULTA A CONVERGÊNCIA DO GA.

TIPICAMENTE : • MUTAR DE 1 A 5% DOS BITS / ITERAÇÃO.
• MELHORES SOLUÇÕES SÃO PRESERVADAS.

- DESEJAMOS MUTAR 5% DA POPULAÇÃO ($\mu = 0.05$) EXCETO O MELHOR CROMOSSOMO.
- UM GERADOR DE NºS ALEATÓRIOS GERA 7 PARES DE INTEIROS QUE CORRESPONDEM ÀS LINHAS E COLUNAS DOS BITS QUE SERÃO MUTADOS.
- O 1º PAR ALEATÓRIO É (4,11). ASSIM, O BIT NA LINHA 4, COLUNA 11 SERÁ MUTADO: 00010110000010 ⇒ 00010110001010
- MUTAÇÕES OCORRERÃO MAIS 6 VEZES NAS POSIÇÕES ALEATÓRIAS (9,3), (2,2), (2,1), (5,14), (8,10) E (5,8).

→ MUITAS MUTAÇÕES ↑ CUSTO DOS CROMOSSOMOS.

→ PERMITE, NO ENTANTO, EXPLORAR A SUPERFÍCIE DE CUSTO DE FORMA MAIS ABRANGENTE.

III.2.8 A PRÓXIMA GERAÇÃO

APÓS AS MUTAÇÕES, O CUSTO ASSOCIADO É CALCULADO PARA A NOVA POPULAÇÃO.

O PROCESSO É ITERATIVO.

A PRÓXIMA GERAÇÃO É MOSTRADA NA TABELA III.7

OS BITS EM ITÁLICO SÃO OS QUE FORAM MUTADOS.

Chromosome	Cost
00100110011101	−13778
*1*1010100011011	−11956
01010110000100	−13338
0001011000*1*010	−13553
0000110001011*1*	−13289
00010000110101	−13164
00100110000100	−13372
010101100*0*1101	−13632
00*0*00110011101	−13036
00100110011101	−13778
00100110011101	−13778
01010110000100	−13338

TAB. III.7: PRÓXIMA GERAÇÃO

O PRÓXIMO PASSO É ORDENAR OS CROMOSSOMOS E SELECIONAR OS REPRODUTORES (POSSÍVEIS), CONFORME MOSTRA A TABELA III.8

Chromosome	Cost
00100110011101	−13778
00100110011101	−13778
00100110011101	−13778
01010110001101	−13632
00010110001010	−13552
00100110000100	−13372
01010110000100	−13338
01010110000100	−13338
00001100010111	−13289
00010000110101	−13164
00000110011101	−13036
11010100011011	−11956

TAB. III.8: PRÓXIMA GERAÇÃO (2ª) ORDENADA

O CUSTO MÉDIO DA 1ª GERAÇÃO ERA -12738.
O CUSTO MÉDIO DA 2ª GERAÇÃO É -13334.

A FIGURA III.4 MOSTRA A LOCALIZAÇÃO DOS 12 CROMOSSOMOS.

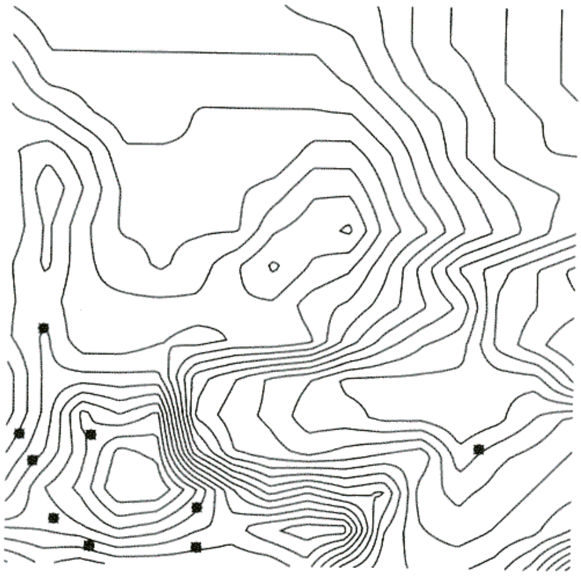


FIG. III.4 : LOCALIZAÇÃO DOS CROMOSSOMOS DA 2ª GERAÇÃO.

→ APENAS OS 6 MELHORES CROMOSSOMOS MOSTRADOS NA FIG. III.4 SÃO MANTIDOS P/ REPRODUÇÃO.

→ APÓS CRUZAMENTO, MUTAÇÃO E ORDENAÇÃO, A 3ª GERAÇÃO É MOSTRADA NA TABELA III.9.
A MÉDIA É −13403.

→ A FIGURA III.5 MOSTRA A LOCALIZAÇÃO DOS 12 MEMBROS PERTENCENTES À 3ª GERAÇÃO.

Chromosome	Cost
00100110011101	−13778
00100110011101	−13778
00100110011101	−13778
00100110011101	−13778
01010110001101	−13632
01010110001101	−13632
00010110001100	−13584
00010110001010	−13553
01010110001010	−13539
00100111010100	−12921
01101111011101	−12602
01100110011101	−12255

TAB. III.9: 3ª GERAÇÃO

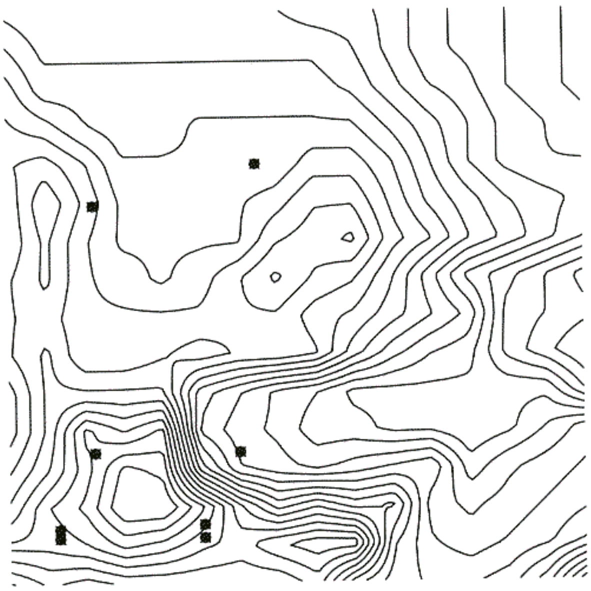


FIGURA III.5: LOCALIZAÇÃO DOS CROMOSSOMOS DA 3ª GERAÇÃO.

A 4ª GERAÇÃO, MOSTRADA NA FIGURA III.6, TEM CUSTO MÉDIO DE -13676.

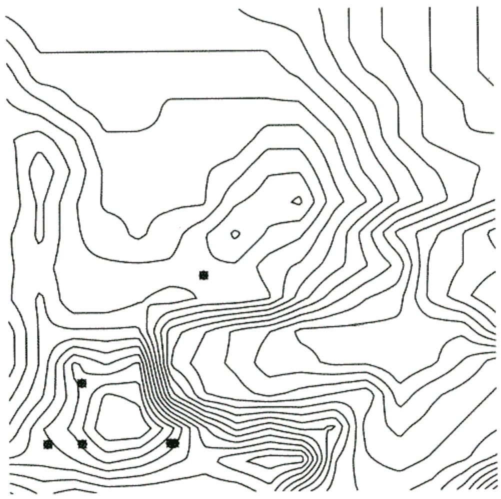


FIG. III.6: LOCALIZAÇÃO DOS CROMOSSOMOS DA 4ª GERAÇÃO.

III.2.9 CONVERGÊNCIA DO ALGORITMO

O Nº DE GERAÇÕES EVOLUTIVAS PODE SER DETERMINADO

→ POR TERMOS ENCONTRADO UMA SOLUÇÃO ACEITÁVEL OU

→ POR TERMOS EXCEDIDO UM DET. Nº DE ITERAÇÕES.

- EM NOSSO EXEMPLO, APÓS 9 ITERAÇÕES O MÍNIMO GLOBAL FOI ENCONTRADO, SENDO = -14199 ($\neq$ 14255 ERRO DE QUANTIZAÇÃO)

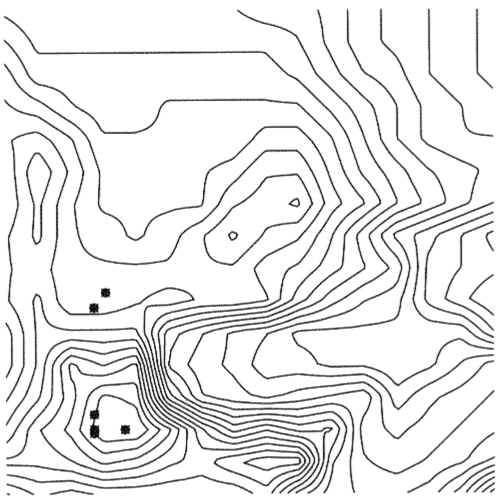


FIG. III.7: POPULAÇÃO FINAL

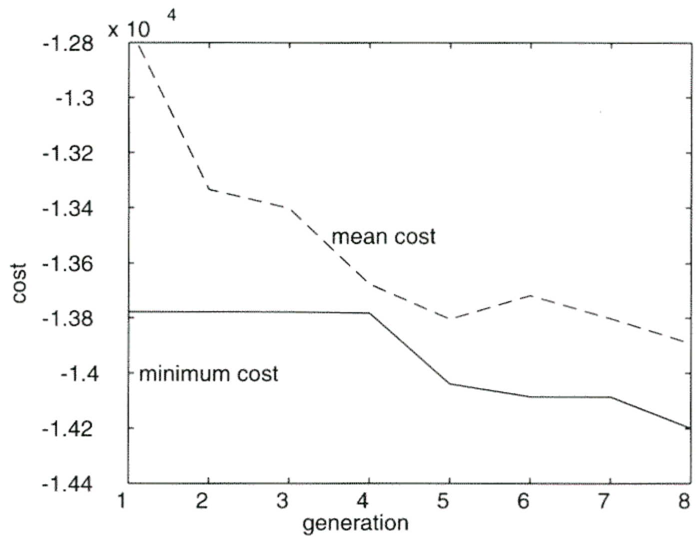


FIG. III.8: CONVERGÊNCIA DO ALGORITMO EM TERMOS DOS CUSTOS MÍNIMOS E MÉDIOS, A CADA GERAÇÃO.