



Exercício de Fixação de Conceitos 2

EA072 - Inteligência Artificial em Aplicações Industriais

Gustavo **CIOTTO PINTON** - RA 117136
Campinas, 16 de novembro de 2015

1 Síntese de controle PID

1. Antes de definirmos os operadores de mutação e *crossover*, é necessário diferenciarmos estes dois conceitos. Embora ambos ocorram em organismos biológicos e sejam responsáveis pela alteração do genótipo de um indivíduo, estes mecanismos ocorrem em diferentes etapas da vida do respectivo organismo. O termo *mutação* é utilizado para descrever o processo no qual um alelo de um gene é **aleatoriamente** substituído ou modificado por outro. Em termos matemáticos, sejam os índices r, \dots, u as posições que sofrerão mutação e que foram determinadas aleatoriamente. Cada posição possui uma probabilidade p_m de ser submetida a este processo. Ao final da mutação, os alelos referentes a estes índices serão modificados, no caso de uma codificação em ponto flutuante, segundo uma distribuição *uniforme* ou *não uniforme*. No caso da *não uniforme*, insere-se uma perturbação com distribuição *normal* com média nula na posição escolhida e desvio-padrão decrescente ao longo das gerações, o que garante consequentemente um refinamento da solução à medida que nos aproximamos de resultados ótimos.

O processo de *crossover*, por sua vez, trata-se de um mecanismo de recombinação genética de dois ou mais cromossomos. Para a codificação de algoritmos evolutivos em que os cromossomos possuem valores em ponto flutuante, destacam-se duas técnicas: o *crossover* aritmético e o uniforme. Para o primeiro, a partir dos cromossomos \mathbf{x} e \mathbf{y} , obtém-se uma combinação convexa dos valores dos genes de \mathbf{x} e \mathbf{y} , isto é, o cromossomo $a\mathbf{x} + (1 - a)\mathbf{y}$, em que $a \in [0, +1]$. Para o segundo, os genes dos cromossomos pais são escolhidos com igual probabilidade para formar um cromossomo filho. Por exemplo, se a probabilidade vale 0.5, então espera-se que o cromossomo resultante seja composto por metade dos genes de \mathbf{x} e metade de \mathbf{y} .

Outra etapa importante para os algoritmos evolutivos é a seleção dos indivíduos mais “adaptados” ao problema (que possuem maior função de *fitness*). Este processo deve garantir que os melhores indivíduos persistam, mas não deve ser extremamente radical a fim de garantir uma diversidade na população. Um exemplo de um operador de seleção é a técnica de *seleção por torneio*. Para selecionar N indivíduos, realizam-se N torneios com p participantes, escolhidos aleatoriamente. Quanto mais alto é p , maior é a pressão seletiva, isto é, para um indivíduo ruim ser escolhido ao menos em um torneio, é necessário que ele compita com $p - 1$ indivíduos piores que ele (para p grande, a probabilidade deste fato ocorrer é muito pequena). Cada torneio é vencido pelo indivíduo que apresenta maior *fitness*. No caso deste exercício, será realizado apenas um torneio, que é composto por 3 indivíduos.

2. As constantes rrier apresentadas na tabela 1 a seguir foram definidas no arquivo `prog_PID.m`, disponibilizado pelo professor.

Tabela 1: Constantes definidas em `prog_PID.m`

Atributos	Valor
Tamanho da População	100
Número máximo de Gerações	50
Taxa de Mutação	0.4
Taxa de Crossover	0.8

Observa-se taxas de mutação e *crossover* relativamente altas: para a primeira, um alelo tem probabilidade de 40% de ser mutado, isto é, ele possui aproximadamente a metade das chances de ser modificado. Em relação ao tamanho da população e o número de gerações, 100 e 50 são números suficientemente grandes para garantir, respectivamente, diversidade entre os indivíduos e uma boa qualidade no resultado visto a quantidade de recombinações entre os cromossomos que serão realizadas.

A população inicial é obtida pelo comando `pop = 5*rand(tam_pop, 3);`, em que `rand` é uma função do MATLAB que gera números aleatórios segundo uma distribuição uniforme e

t_{am_pop} vale 100. Será criada, portanto, uma matriz de 100 linhas, cada uma representando um indivíduo, e 3 colunas, uma para cada constante a ser determinada k_p , k_d e k_i .

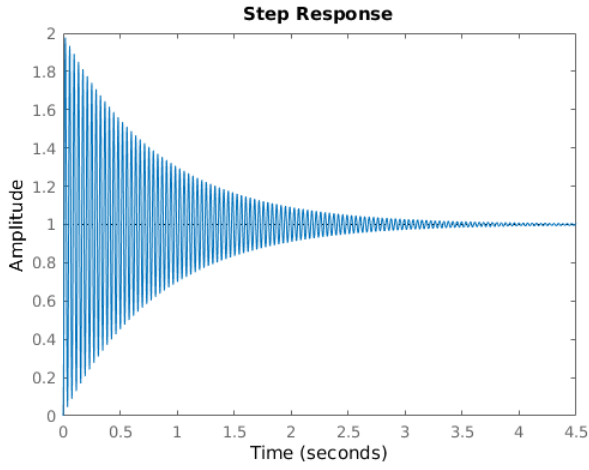
Finalmente, cada operador de *crossover* é escolhido com 50% de probabilidade. Espera-se portanto que o operador aritmético seja escolhido em metade das oportunidades e o uniforme, também.

3. Após adicionarmos as linhas referentes ao cálculo do *fitness*, as execuções do programa *prog_pid.m* resultam nos dados da tabela 2.

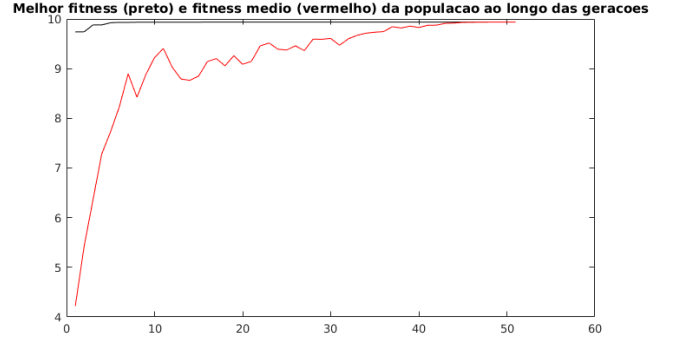
Tabela 2: Resultados das 5 execuções de *prog_PID.m*

Ex.	k_p	k_d	k_i	t_{subida} (s)	$t_{estabilizacao}$ (s)	<i>Overshoot</i> (%)	$\Delta\phi$ (°)	<i>Fitness</i>
1	4.9935	2.385e-05	4.46633	0.0064396	3.26286	97.7286	1.1935	9.9360
2	4.9960	2.296e-05	3.99303	0.0064393	3.14669	97.6506	1.2340	9.9360
3	4.9923	1.350e-05	3.95399	0.0064397	3.3217	97.7735	1.1687	9.9360
4	4.9940	2.240e-05	4.33452	0.0064395	3.24405	97.7234	1.1960	9.9360
5	4.9990	1.528e-05	4.4244	0.0064348	3.41617	97.8392	1.1346	9.9361

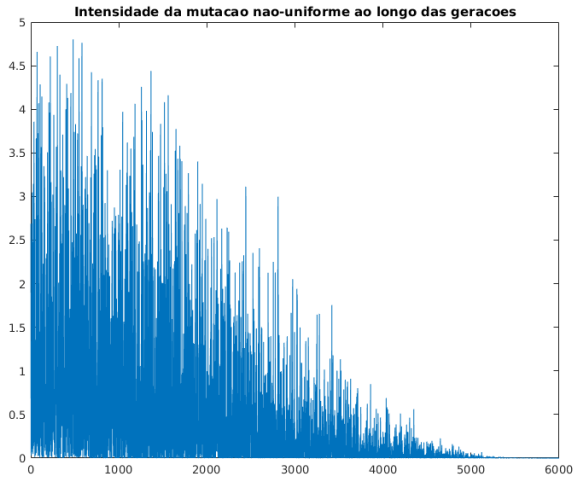
Observa-se que todos os valores k_d relativos à ação derivadora são praticamente nulos e que k_i e k_p encontram-se muito próximos a seus valores máximos, isto é, 5. Isto indica que a ação derivadora não exerce efeitos importantes sobre o tempo de resposta apresentado pelo sistema. Os valores de tempo dos primeiros máximos em todas as execuções concentram-se em torno do valor $t_{subida} = 0.00644$ s. Sabe-se também que este atributo é fortemente ligado à frequência de oscilação que a resposta ao degrau apresentará. Para sistemas de segunda ordem, por exemplo, a relação $w_0 t_m \approx 3$ evidencia o compromisso entre essas duas características. Portanto, para valores muito pequenos de t_{subida} , teremos frequências w_0 elevadas e vice-versa. É justamente esse comportamento que é observado neste caso: pelo fato de que a função de *fitness* só levar em conta o tempo de resposta, obtém-se respostas muito rápidas, porém que oscilam fortemente em torno de 1. A margem de fase ($\Delta\phi$) também possui uma relação com o *overshoot*: pequenas margens de fase produzem sistemas menos estáveis e que apresentam maior erros percentuais entre a entrada e a saída. No nosso caso, como não nos preocupamos com esse parâmetro, obtém-se pequenas margens de fase e, conseqüentemente, grandes erros antes da estabilização. A figura 1a, que é o resultado de uma das execuções acima, permite a visualização destes comportamentos. A figura 1c, por sua vez, mostra a quantidade de mutações realizadas pelo algoritmo evolutivo no decorrer no programa. Conforme explicado anteriormente, essa quantidade deve ser decrescente, visto que os melhores indivíduos devem ser mantidos nas gerações finais. As figuras 1b e 1d mostram, respectivamente, a evolução dos valores da função de *fitness* e dos parâmetros k_p , k_i e k_d . Para o primeiro, observa-se que o melhor *fitness* permanece sempre próximo de 10 e que o *fitness* médio aproxima-se gradualmente a este valor, indicando que a população esta se “especializando” e adaptando no problema. Para segundo, verifica-se que k_d é praticamente nulo em todas as gerações, k_p se estabiliza a partir da geração 10 e k_i varia fortemente até a geração 40. Estes fatos nos dizem que o parâmetro que mais influencia o tempo do primeiro máximo é a ação integradora.



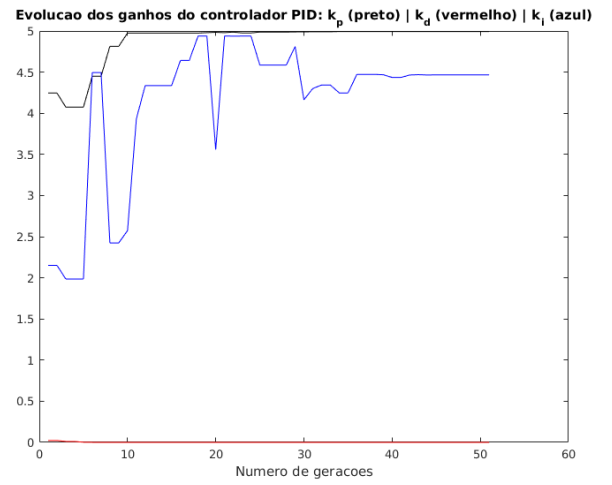
(a) Resposta ao degrau do sistema.



(b) Evolução dos valores da função *fitness*.



(c) Evolução da intensidade de mutações ao longo das gerações.



(d) Evolução dos parâmetros.

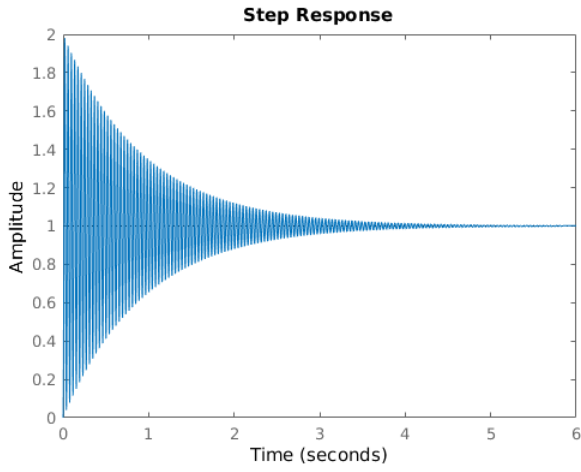
Figura 1: Resultado para a execução 1 do algoritmo evolutivo.

A modificação dos parâmetros da tabela 1 pode influenciar nos resultados acima. Se dobrarmos o número de indivíduos na população, isto é, 200, e aumentarmos as taxas de crossover e mutação para, respectivamente, 0.9 e 0.6, obtém-se os dados contidos na tabela 3.

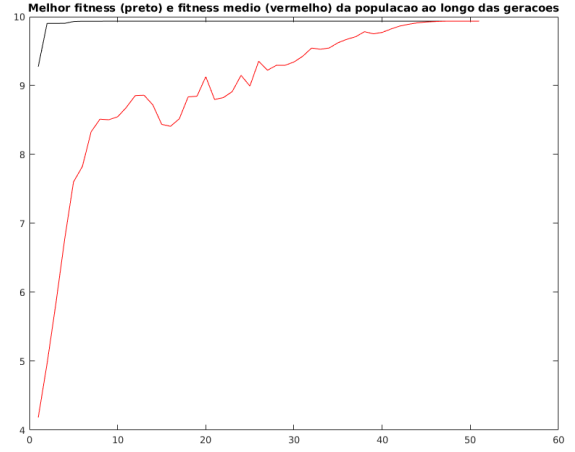
Tabela 3: Resultados da execução com os parâmetros modificados.

k_p	k_d	k_i	t_{subida} (s)	$t_{estabilizacao}$ (s)	$Overshoot$ (%)	$\Delta\phi$ (°)	$Fitness$
4.9975	7.80994e-06	4.6839	0.00643281	3.68641	97.9919	1.05405	9.93608

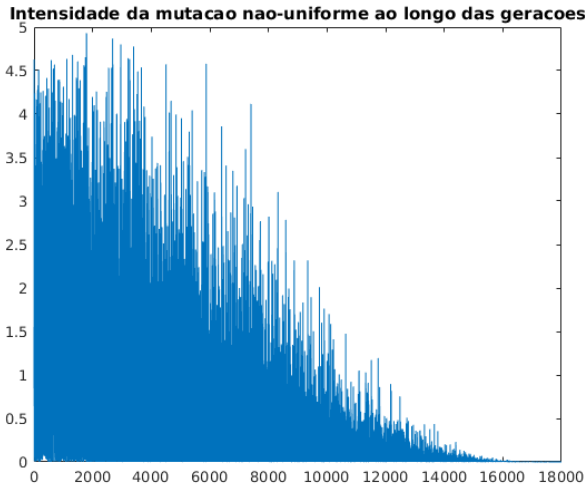
Observa-se que o tempo do primeiro máximo é ligeiramente inferior aos dos casos anteriores e, conseqüentemente, a função de *fitness* é superior. Os valores dos parâmetros continuam parecidos, sendo que k_d é inferior e k_i superior. De maneira geral, o aumento de recursos computacionais disponibilizados ao programa (aumento de população e quantidade de mutações) não causaram grandes diferenças nos resultados. As figuras 2b e 2d apresentam o mesmo comportamento descrito no parágrafo anterior e a 2c destaca que a quantidade de mutações realizadas foi muito superior aos casos anteriores: observa-se que a presença de quantidades superiores a 10000 quando a taxa de mutação é 0.6.



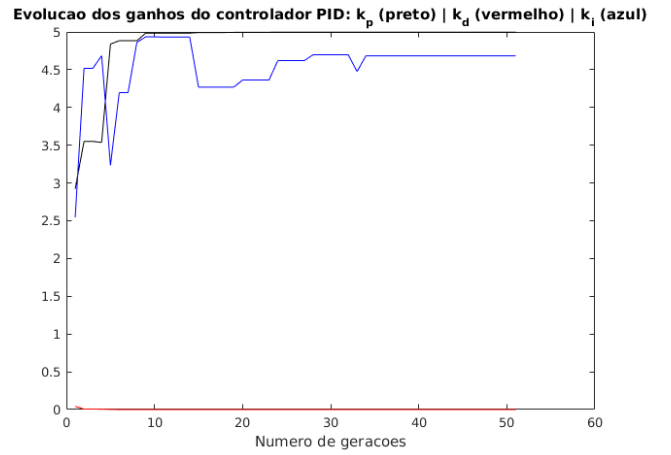
(a) Resposta ao degrau do sistema.



(b) Evolução dos valores da função *fitness*.



(c) Evolução da intensidade de mutações ao longo das gerações.



(d) Evolução dos parâmetros.

Figura 2: Resultado para a execução do algoritmo evolutivo com parâmetros modificados.

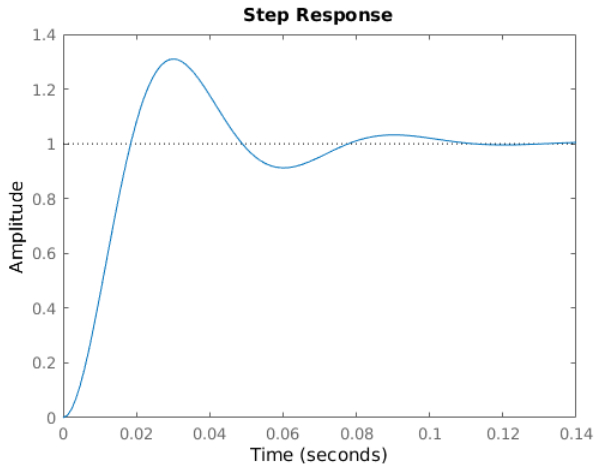
4. Os resultados das 5 execuções com a nova função de *fitness* estão presentes na tabela 4, logo abaixo. Dessa vez, considera-se também a margem de fase.

Tabela 4: Resultados das 5 execuções de `prog_PID.m`

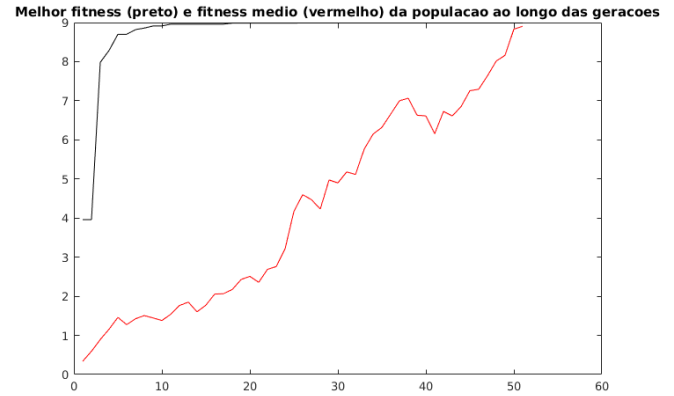
Ex.	k_p	k_d	k_i	t_{subida} (s)	$t_{estabilizacao}$ (s)	<i>Overshoot</i> (%)	$\Delta\phi$ (°)	<i>Fitness</i>
1	1.22881	0.718353	4.54691	0.705398	12.7319	56.4502	59.9346	5.53763
2	1.15518	0.57906	4.97227	0.605573	10.924	56.4079	59.9872	5.86308
3	0.812793	0.298811	4.76622	0.44429	8.01846	56.4355	59.9522	6.47451
4	1.67979	1.23205	4.94384	0.885861	15.9772	56.398	56.398	5.0356
5	2.36522	0.00523679	1.84369	0.0124718	0.100391	30.9773	59.986	8.98883

Ao contrário do item anterior, verifica-se neste caso uma grande variação nos melhores resultados nas diferentes execuções. As soluções 1 e 2 apresentam soluções muito parecidas em todos os aspectos. A solução 4 foi a pior a ser encontrada e possui o maior valor de k_d encontrado. Isso leva a crer que grandes valores deste parâmetro influenciam negativamente no resultado, principalmente em relação à margem de fase: a pior margem foi encontrada nesta execução. A solução 5 foi aquela que melhor encontrou um compromisso entre tempo do primeiro máximo e estabilidade e, portanto, possui o maior *fitness*. Remarcamos que os tempos de subida e estabilização e o *overshoot* desta resposta foram muito inferiores a aqueles encontrados nas outras

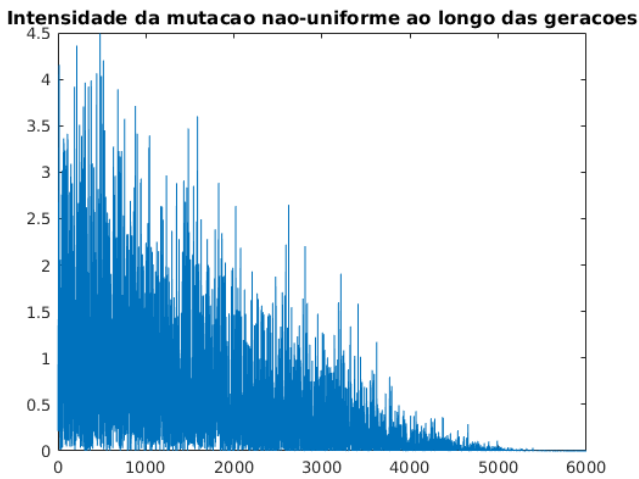
execuções, confirmando ainda mais a qualidade desta solução. Novamente, nos deparamos com o fato de que o ajuste de k_d é o mais sensível: grandes valores deste parâmetro não levam a bons resultados. Observa-se ainda que os outros dois parâmetros, k_p e k_i , também apresentam grandes variações em relação as outras execuções, sendo k_d muito superior e k_i inferior. As imagens a seguir contém os gráficos gerados para a melhor execução.



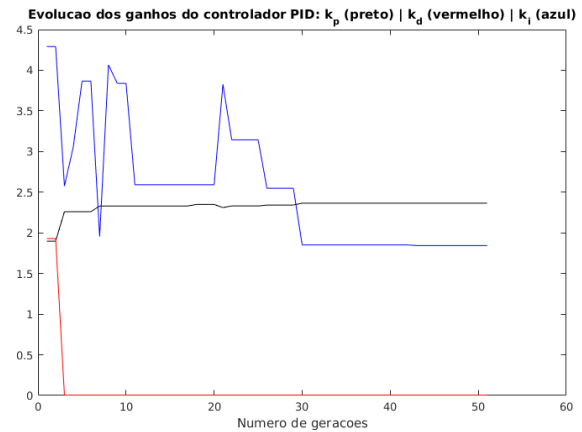
(a) Resposta ao degrau do sistema.



(b) Evolução dos valores da função *fitness*.



(c) Evolução da intensidade de mutações ao longo das gerações.



(d) Evolução dos parâmetros.

Figura 3: Resultado para a melhor execução do algoritmo evolutivo com *fitness* considerando a margem de fase.

A figura 3a representa a resposta ao degrau do sistema com os parâmetros determinados na execução 5, cujas características já foram discutidas no parágrafo anterior. Na imagem 3b, que apresenta a evolução do *fitness*, observa-se que o melhor valor já é encontrado nas primeiras gerações. Entretanto, a quantia média ainda é muito inferior a tal valor, o que implica que a população é muito diversificada até a geração 50, isto é, poucos indivíduos se “adaptaram” tão bem quanto o melhor. A figura 3d revela que o parâmetro que mais variou entre as gerações foi k_p , assim como no caso do item anterior.

Os resultados obtidos com a função *fitness* considerando o tempo de subida e também a margem de fase foram muito superiores em muitos aspectos:

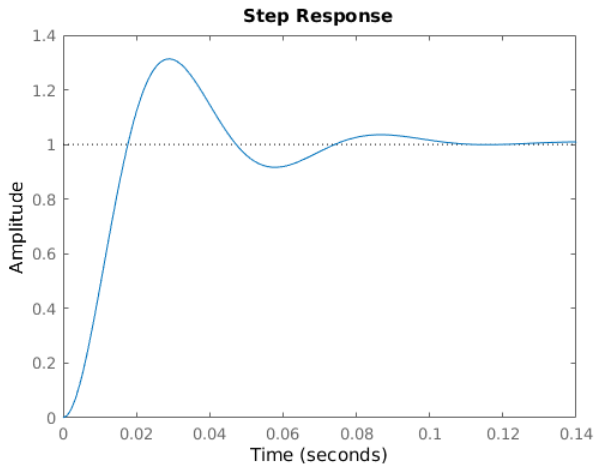
- os valores de *overshoot* no caso anterior eram, em média, de 97%, sendo que neste item, encontra-se 30%;

- a margem de fase no item anterior ($\approx 2^\circ$) era muito inferior ao caso atual, em que exige-se $\Delta\phi = 60^\circ$, comprometendo fortemente a estabilidade do sistema;
- o tempo de subida no caso anterior era muito inferior ao atual, porém sua frequência de oscilação em torno de 1 era inaceitável;
- no caso anterior, a resposta demorava muito a se estabilizar, o que pode representar um problema para algumas aplicações. No caso atual, encontra-se tempos de estabilização na ordem de centésimos de segundo;

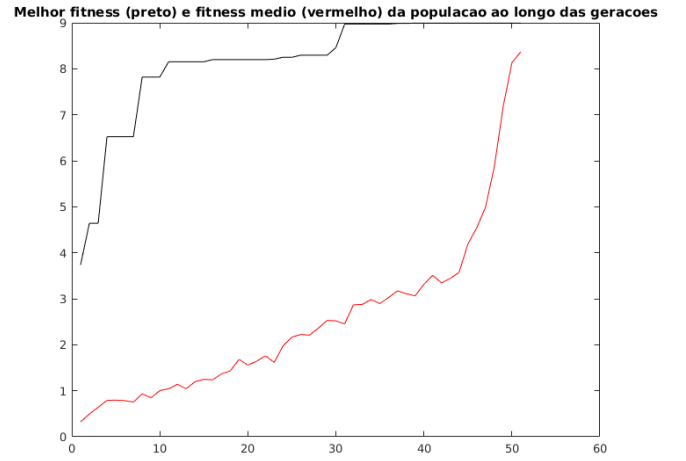
Enfim, efetuando as mesmas modificações realizadas anteriormente no tamanho da população e nas taxas de crossover e mutação, encontra-se os dados da tabela 5 e a imagens a seguir:

Tabela 5: Resultados da execução com os parâmetros modificados.

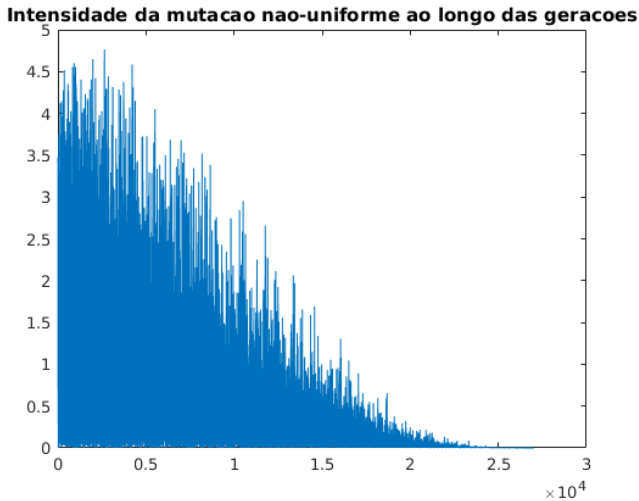
k_p	k_d	k_i	t_{subida} (s)	$t_{estabilizacao}$ (s)	<i>Overshoot</i> (%)	$\Delta\phi$ ($^\circ$)	<i>Fitness</i>
2.57579	0.005528	3.77055	0.011941	0.09798	31.3347	60.015	8.99308



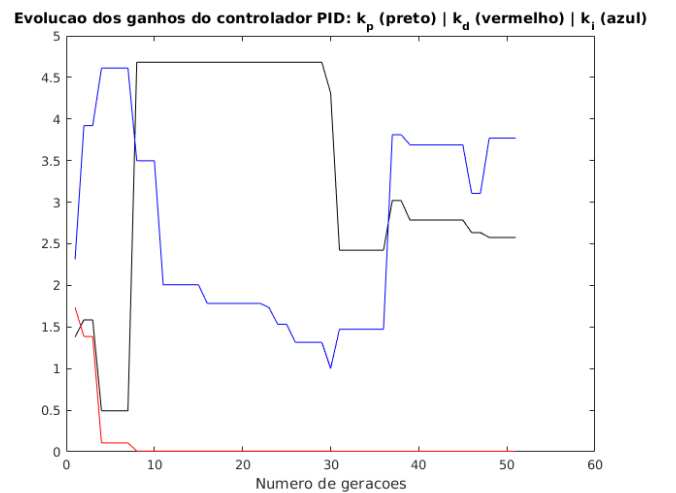
(a) Resposta ao degrau do sistema.



(b) Evolução dos valores da função *fitness*.



(c) Evolução da intensidade de mutações ao longo das gerações.



(d) Evolução dos parâmetros.

Figura 4: Resultado para a melhor execução do algoritmo evolutivo com *fitness* considerando a margem de fase e parâmetros modificados.

Neste caso, como os resultados de cada execução variam bastante, aumentar a população e as taxas de *crossover* e mutação contribuem para a obtenção de uma solução de melhor qualidade,

visto que a diversidade nos indivíduos é maior. Logo, a probabilidade que a solução ótima esteja presente em um dos indivíduos é consequentemente maior. Uma taxa de mutação elevada também garante que mais soluções sejam testadas, à medida que bons indivíduos podem dar origem a indivíduos melhores que eles mesmos através da mutação. Enfim, assim como no caso em que os parâmetros não tinham sido alterados, a figura 4b revela que a média dos valores de *fitness* é muito distante do maior, indicando que nas gerações mais jovens há muitos indivíduos ruins. A figura 4d mostra, por sua vez, que, para populações maiores, todos os parâmetros podem variar fortemente ao passar das gerações, ao contrário que é observado no caso em que tamanho da população é igual a 100. Este fato também é totalmente esperado, visto que, para populações maiores, a chance de um indivíduo encontrar um outro melhor que ele é maior.

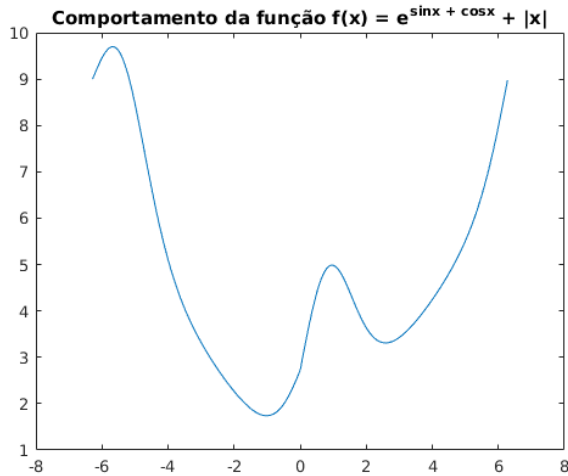
2 Aproximação de funções multidimensionais

2.1 Mapeamento $y = f(x)$

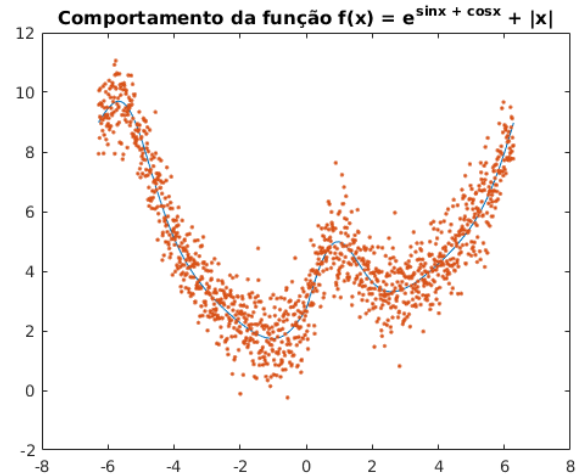
A função escolhida para o mapeamento $\mathbb{R}^1 \rightarrow \mathbb{R}^1$ está representada logo abaixo. Foi definido que o intervalo de x é $[-2\pi, 2\pi]$ e que este intervalo seria discretizado a cada 0.01, a fim de possuírmos uma quantidade grande de amostras, isto é, 1257.

$$y = \exp(\sin x + \cos x) + |x| \quad (1)$$

Sem a presença de ruídos, este mapeamento produz a imagem 5a. A imagem 8d, por sua vez, apresenta o resultado da soma da função com um ruído de distribuição normal de média nula e variância 0.64. É possível verificar que a adição do ruído prejudica fortemente a identificação do formato inicial do mapeamento.



(a) Mapeamento sem ruídos.



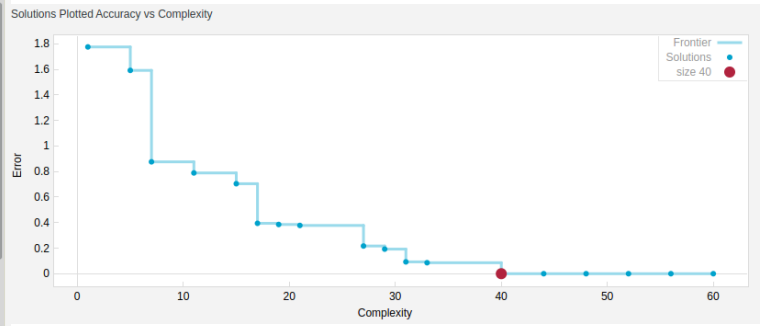
(b) Mapeamento com ruídos.

Figura 5: Mapeamento da função 1.

A simulação realizada no *Eureqa*, primeiramente com os dados sem ruídos e habilitando-se as funções-base $\exp()$, $\sin()$, $\cos()$, $\text{abs}()$ e as operações básicas, forneceu os resultados contidos nas figuras 6a e 6b. A primeira lista as soluções encontradas que mais se aproximaram dos dados amostrais. Destaca-se que a primeira solução, cujo peso é 40, é justamente aquela que utilizamos para produzir os dados. A segunda imagem ilustra o compromisso entre acurácia (erro) e simplicidade (complexidade da solução). Observa-se que inúmeras soluções apresentam erro praticamente nulo, porém existe apenas uma cuja complexidade é mínima para este caso. Essa função trata-se da solução de tamanho 40 que, como já foi dito, é igual à equação 1. A figura 6c apresenta um resumo da melhor solução encontrada. Destaca-se que o erro podem ser considerados como nulos, visto que o maior erro encontrado foi da ordem de 10^{-14} .

Size	Fit	Solution
40	0.000	$y = \text{abs}(x) + \exp(\sin(x) + \cos(x))$
44	0.000	$y = \text{abs}(x) + \exp(\sin(x) + \cos(2.27e-15 + x))$
48	0.000	$y = \text{abs}(4.01e-15 + x) + \exp(\sin(x) + \cos(3.04e-15 + x))$
52	0.000	$y = \text{abs}(3.82e-15 + x) + \exp(\sin(2.27e-15 + x) + \cos(3.82e-15 + x))$
56	0.000	$y = \text{abs}(3.82e-15 + x) + \exp(\sin(2.7e-15 + x) + \cos(3.82e-15 + x))$
60	0.000	$y = 1 \exp(\sin(4.01e-15 + x) + \cos(1.96e-15 + x)) + \text{abs}(3.82e-15 + x)$
33	0.048	$y = 1.57 + 1.27 \sin(x) + 1.27 \cos(x) + 1.13 \sin(x) \cos(x) + \text{abs}(x)$
31	0.092	$y = 1.56 + 1.25 \cos(x) + 1.25 \sin(x) + \sin(x) \cos(x) + \text{abs}(x)$
29	0.108	$y = 1.55 + 1.17 \sin(x) + \sin(x) \cos(x) + \cos(x) + \text{abs}(x)$
27	0.122	$y = 1.5 + \sin(x) \cos(x) + \sin(x) + \cos(x) + \text{abs}(x)$
21	0.213	$y = 1.54 + 1.17 \sin(x) + 1.17 \cos(x) + \text{abs}(x)$
19	0.217	$y = 1.5 + 1.17 \cos(x) + \sin(x) + \text{abs}(x)$
17	0.222	$y = 1.46 + \sin(x) + \cos(x) + \text{abs}(x)$
15	0.397	$y = 1.91 + 0.186x^2 - 0.00732x^3$
11	0.444	$y = 7 + 0.183x^2 - 0.141x$

(a) Soluções encontradas pelo software.



(b) Compromisso acurácia x simplicidade das soluções.

Solution Details (calculated on validation data)	
Solution	$y = \text{abs}(x) + \exp(\sin(x) + \cos(x))$
R ² Goodness of Fit	1
Correlation Coefficient	1
Maximum Error	2.30926e-14
Mean Squared Error	5.44583e-29
Mean Absolute Error	5.78929e-15
Coefficients	0
Complexity	40

(c) Resumo da melhor solução encontrada.

Figura 6: Resultados para mapeamento da função 1 sem ruídos.

A simulação executada com os dados com ruído produziu o mapeamento representado pela equação 2. A figura 7 compara o mapeamento produzido nesta fase com aquele utilizado para gerar os dados. Observa-se que ambos os mapeamento estão muito próximos, porém há regiões em que as diferenças são notáveis. Os resultados gerados pelo *software* são, portanto, bastante satisfatórios.

$$y = 1.61175 + 1.830094 * \sin(0.792661 + x) + 0.55764 * \sin(1.99325 * x) + |x| \quad (2)$$

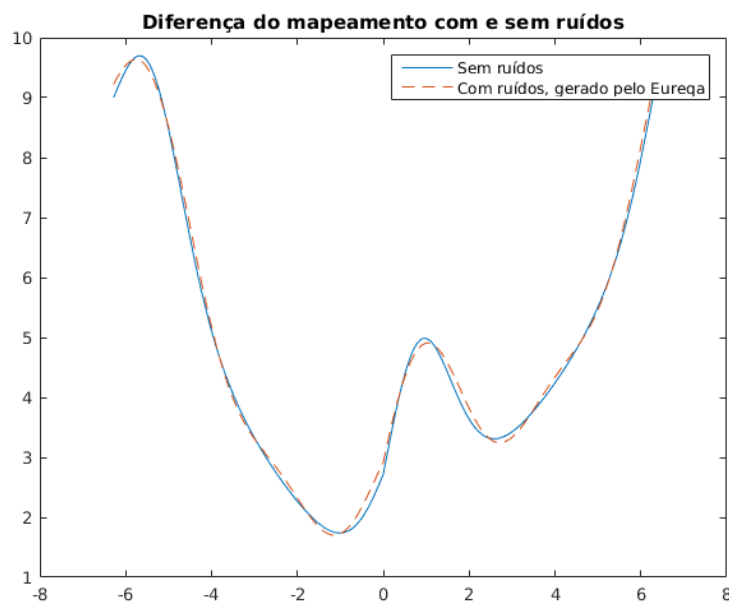
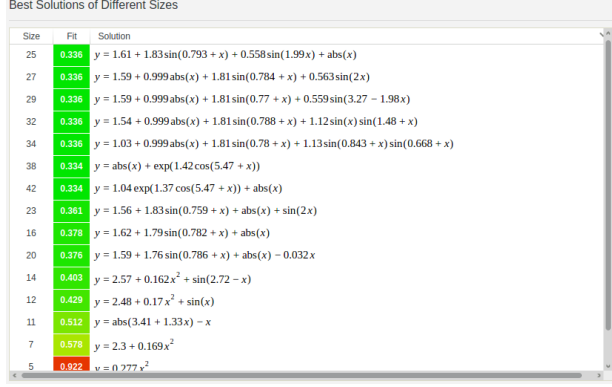
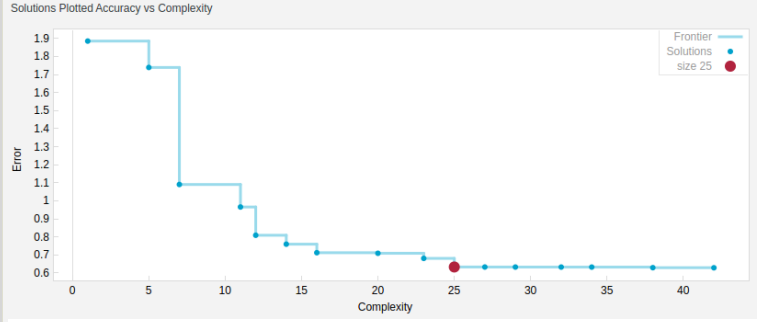


Figura 7: Comparação entre mapeamento com e sem ruídos.

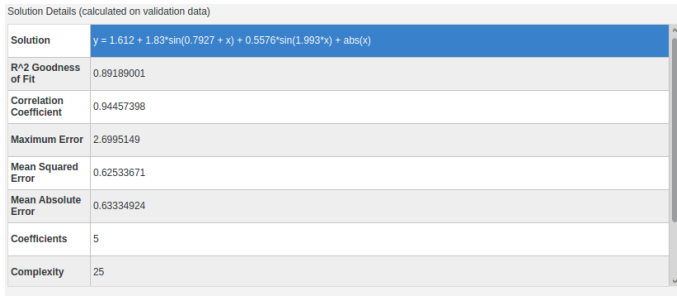
As figuras 8a, 8b, 8c e 8d foram geradas pelo *Eureqa* e possuem informações sobre a solução encontrada. A primeira é a lista de soluções encontradas pelo *software*, a segunda mostra o compromisso acurácia x complexidade (neste caso, a solução apresentada na equação 2 possui melhor compromisso, visto que seu erro é próximo de 0 e sua complexidade, mínima), a terceira contém algumas informações sobre a melhor solução (o erro quadrático médio neste caso é considerável, contrariamente ao caso sem ruído, e vale 0.625) e a última revela quais dados foram usados para treinamento e validação, assim como o mapeamento encontrado.



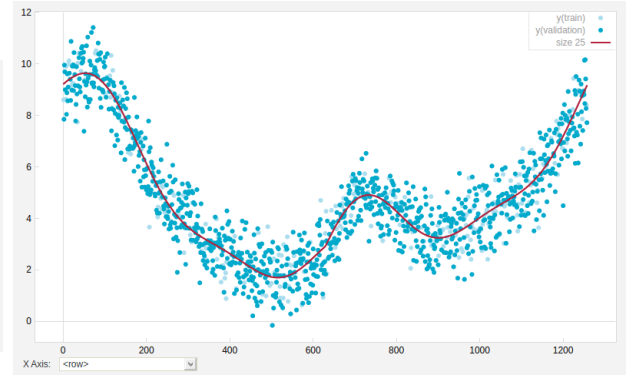
(a) Soluções encontradas pelo software.



(b) Compromisso acurácia x simplicidade das soluções.



(c) Resumo da melhor solução encontrada.



(d) Resumo do mapeamento encontrado e amostras usadas para treinamento e validação.

Figura 8: Resultados para mapeamento da função 1 com ruídos.

2.2 Mapeamento $y = f(x_1, x_2, x_3)$

A função escolhida para este caso está representada pela equação 3, sendo que $x_1 \in [-3, 1]$, $x_2 \in [1, 5]$ e $x_3 \in [-4, 4]$. As duas primeiras variáveis foram amostradas a cada 0.01 e a segunda, 0.02. Temos, portanto, 401 amostras no total.

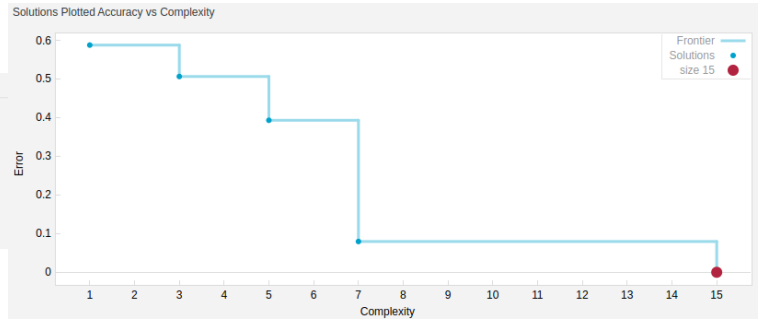
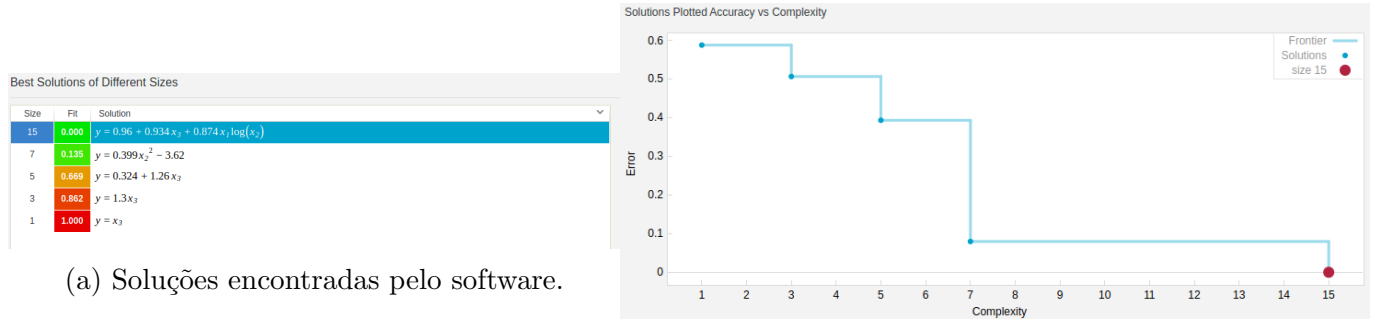
$$y = x_1 * \log_{\pi} \left(\frac{x_2}{3} \right) + \sqrt{2} * x_3 = x_1 * \frac{\log_{10} \left(\frac{x_2}{3} \right)}{\log_{10}(\pi)} + \sqrt{2} * x_3 \quad (3)$$

O mapeamento produzido pela simulação utilizando os dados sem ruídos foi

$$y = 0.959713 + 0.93436 * x_3 + 0.873569 * x_1 * \log_{10}(x_2) \quad (4)$$

O mapeamento produzido não corresponde ao utilizado para a geração dos dados, porém os erros associados ao primeiro são praticamente nulos, isto é, o maior erro é da ordem de 10^{-15} , conforme figura 9c. Isto indica que, para os intervalos considerados para as variáveis, alguns termos podem ser substituídos por outros sem perdas de precisão. Nota-se também, pela figura 9a, que menos soluções foram encontradas em relação ao item anterior. A imagem 9b mostra o compromisso entre

erro e complexidade da melhor solução e é possível concluir, a partir desta imagem, que realmente a solução de complexidade 15 supera as demais.



Solution Details (calculated on validation data)	
Solution	$y = 0.9597 + 0.9344 \cdot x_2 + 0.8736 \cdot x_1 \cdot \log(x_2)$
R^2 Goodness of Fit	1
Correlation Coefficient	1
Maximum Error	6.21725e-15
Mean Squared Error	6.87771e-30
Mean Absolute Error	2.01243e-15
Coefficients	3
Complexity	15

(c) Resumo da melhor solução encontrada.

Figura 9: Resultados para mapeamento da função 3 sem ruídos.

Enfim, a execução para os dados com ruídos gera o mapeamento

$$y = 0.39025 * x_2^2 - 3.555333 \quad (5)$$

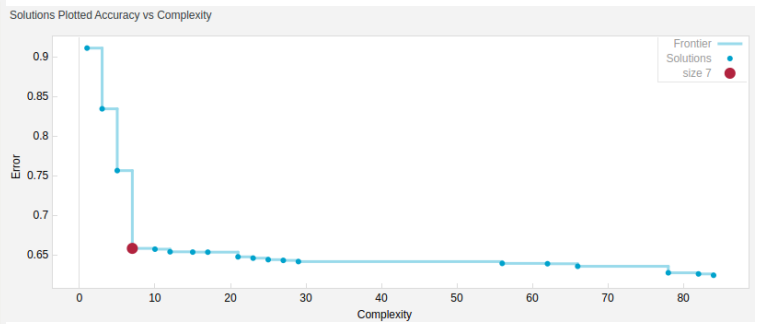
Observa-se que y não depende de x_1 e x_3 neste caso. As figuras a seguir foram geradas pelo *software*. Destacam-se os seguintes fatos:

- A figura 10b mostra o compromisso entre erro e complexidade para as melhores soluções. Nota-se que, apesar de a solução não apresentar o menor erro, ela possui uma maior simplicidade comparada às demais. Sendo assim, ela é escolhida como a melhor solução atendendo aos critérios de acurácia e complexidade simultaneamente.
- Conforme figura 10c, o erro quadrático médio é considerável e vale, aproximadamente, 0.70.
- A figura 10d representa o mapeamento e as amostras utilizadas. Observa-se que este mapeamento aproxima-se relativamente bem aos dados.

Best Solutions of Different Sizes

Size	Fit	Solution
7	0.723	$y = 0.39x_1^2 - 3.55$
10	0.721	$y = 0.875 + x_3 + 1.1 \sin(x_1)$
12	0.718	$y = 0.962 + 0.962x_3 + 1.22 \sin(x_1)$
15	0.717	$y = 2.2x_2 \log(x_2) - 0.822 - 2.16x_2$
17	0.717	$y = 0.0772x_2^3 - 2.09 - 0.0797x_1x_2x_3$
21	0.711	$y = 0.856 + x_3 + 0.191 \cos(13.7x_2 + 2.4x_3) + \sin(x_1)$
23	0.709	$y = 0.856 + x_3 + 1.1 \sin(x_1) + 0.225 \cos(13.7x_2 + 2.37x_3)$
25	0.707	$y = 0.911 + 0.961x_3 + 1.24 \sin(x_1) + 0.221 \cos(13.7x_2 + 2.37x_3)$
27	0.705	$y = 0.835 + x_3 + 0.262 \cos(x_1 + 2.98x_3 + 0.915x_2x_3^2) + \sin(x_1)$
29	0.704	$y = 0.835 + x_3 + 1.11 \sin(x_1) + 0.26 \cos(x_1 + 2.98x_3 + 0.915x_2x_3^2)$
5	0.830	$y = 0.499 + 1.2x_3$
3	0.916	$y = 0.429 + x_3$
1	1.000	$y = x_3$
56	0.702	$y = x_3 + 1.1 \cos(\sin(x_3 \sin(x_1) - 83.6x_1x_3)) + \sin(x_1)$
62	0.701	$y = 0.338 + x_3 + 1.12 \sin(x_1) + \cos(\sin(86.1x_3)) \cos(\sin(-94.1x_2))$

(a) Soluções encontradas pelo software.

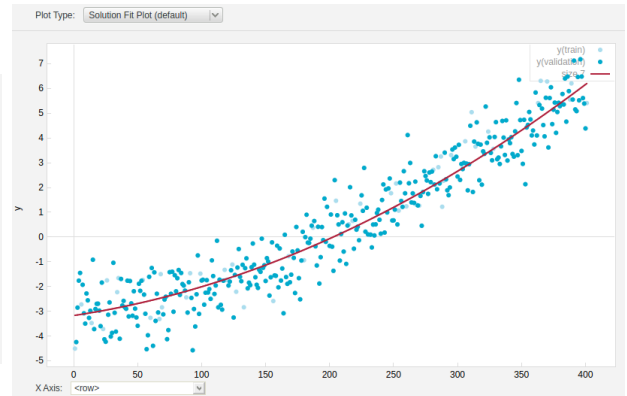


(b) Compromisso acurácia x simplicidade das soluções.

Solution Details (calculated on validation data)

Solution	$y = 0.3903x_1^2 - 3.555$
R ² Goodness of Fit	0.91685337
Correlation Coefficient	0.95755118
Maximum Error	2.6168161
Mean Squared Error	0.70094645
Mean Absolute Error	0.65845331
Coefficients	2
Complexity	7

(c) Resumo da melhor solução encontrada.



(d) Resumo do mapeamento encontrado e amostras usadas para treinamento e validação.

Figura 10: Resultados para mapeamento da função 3 com ruídos.

3 Controle Nebuloso e Robótica Evolutiva

3.0 Definição dos Consequentes

O primeiro passo no controle do robô é a definição dos consequentes das regras que serão levadas em conta. Considerando que o robô dispõe de 3 sensores (d_1 , d_2 e d_3) e que as funções de pertinência associadas às medidas realizadas por eles estão representadas na seção 3 do enunciado, sugestões de consequentes estão representadas na tabela 6, logo abaixo. As regras são da forma **SE** (d_1 **É** x) **E** (d_2 **É** y) **E** (d_3 **É** z) **ENTÃO** ($\Delta\theta$ **É** w).

Tabela 6: Regras que controlarão o robô.

d_1	d_2	d_3	$\Delta\theta$
P	MP	P	MN
P	MP	M	MN
P	MP	G	MN
P	PP	P	Z
P	PP	M	MN
P	PP	G	MN
P	PG	P	Z
P	PG	M	MN
P	PG	G	MN
P	MG	P	Z
P	MG	M	PN
P	MG	G	MN

d_1	d_2	d_3	$\Delta\theta$
M	MP	P	PP
M	MP	M	MP
M	MP	G	MN
M	PP	P	PP
M	PP	M	Z
M	PP	G	PN
M	PG	P	PP
M	PG	M	Z
M	PG	G	PN
M	MG	P	PP
M	MG	M	Z
M	MG	G	PN

d_1	d_2	d_3	$\Delta\theta$
G	MP	P	MP
G	MP	M	MP
G	MP	G	MP
G	PP	P	MP
G	PP	M	MP
G	PP	G	MP
G	PG	P	MP
G	PG	M	PP
G	PG	G	Z
G	MG	P	MP
G	MG	M	PP
G	MG	G	Z

3.1 Controlando o Robô

Uma vez definidas as regras que deverão ser seguidas, o próximo passo é então simular o comportamento do robô através do *software* MATLAB. Para isso, 5 funções foram escritas. A primeira, chamada `trap_pertinencia`, recebe 6 parâmetros, sendo eles a coordenada x , as informações a , b , c e d referentes os trapézios que compõem as funções de pertinência e o valor a ser considerado caso x esteja fora do intervalo $[a, d]$, e retorna o respectivo valor da função de pertinência. O programa 1 abaixo possui a implementação da respectiva função.

Programa 1: `trap_pertinencia.m` - Avalia pertinência de um regra em função de x .

```
1 %% FUNCAO trap_pertinencia
2 % Funcao responsavel por calcular a pertinencia em uma dada coordenada x.
3 % Parametros: x eh a coordenada para qual calcula-se a pertinencia.
4 %             a, b, c e d sao os parametros referentes ao trapezio.
5 %             out_of_range representa o resultado que deve ser atribuido
6 %             caso x esteja fora do intervalo [a, d].
7 % Autor: Gustavo CIOTTO PINTON
8 function pert = trap_pertinencia( x, a, b, c, d , out_of_range)
9
10 if (x < a) || (x > d)
11     pert = out_of_range;
12
13 elseif (x < b)
14     if (a ~= b)
15         pert = (x - a)/(b -a);
16     else
17         pert = 1;
18     end
19
20 elseif (x < c)
21     pert = 1;
22
23 else
24     if (c ~= d)
25         pert = (d - x)/(d - c);
26     else
27         pert = 1;
28     end
29 end
30 end
```

A segunda função, chamada de `get_D1_D3_Rule` e representada pelo programa 2, calcula a pertinência de cada um dos possíveis estados referentes aos sensores D_1 e D_3 . Essa função recebe como parâmetro a distância medida pelo sensor e retorna um vetor com três valores contendo a pertinência para os estados **P**, **M** e **G**.

Programa 2: Avalia pertinência da regras de D_1 e D_3 em função das distâncias medidas.

```
1 %% FUNCAO get_D1_D3_Rule
2 % Calcula pertinencia para todos os possiveis estados de D1 e D3.
3 % Parametros: sensor_capture eh a distancia medida pelo sensor.
4 % Retorno : Vetor com os valores das pertinencias atribuidas a cada
5 % regra.
6 % Autor: Gustavo CIOTTO PINTON
7 function S = get_D1_D3_Rule( sensor_capture )
8
9 % Calcula pertinencia para P
10 relevance_P = trap_pertinencia (sensor_capture, 0, 0, 1, 2, (sensor_capture < 0));
11
12 % Calcula pertinencia para M
13 relevance_M = trap_pertinencia (sensor_capture, 1, 2, 3, 4, 0);
```

```

14
15 % Calcula pertinencia para G
16 relevance_G = trap_pertinencia (sensor_capture, 3, 4, 5, 5, (sensor_capture > 5));
17
18 S = [relevance_P relevance_M relevance_G];
19
20 end

```

A função `get_D2_Rule` realiza o mesmo que o descrito anteriormente, mas agora com o sensor D_2 . A implementação desta função pode ser encontrada logo abaixo.

Programa 3: Avalia pertinência da regras de D_2 em função das distâncias medidas.

```

1 %% FUNCAO get_D2_Rule
2 % Calcula pertinencia para todos os possiveis estados de D2.
3 % Parametros: sensor_capture eh a distancia medida pelo sensor.
4 % Retorno : Vetor com os valores das pertinencias atribuidas a cada
5 % regra.
6 % Autor: Gustavo CIOTTO PINTON
7 function S = get_D2_Rule( sensor_capture )
8
9 % Testar para MP
10 relevance_MP = trap_pertinencia (sensor_capture, 0, 0, 1, 2, (sensor_capture < 0));
11
12 % Testar para PP
13 relevance_PP = trap_pertinencia (sensor_capture, 1, 2, 3, 4, 0);
14
15 % Testar para PG
16 relevance_PG = trap_pertinencia (sensor_capture, 3, 4, 5, 6, 0);
17
18 % Testar para MG
19 relevance_MG = trap_pertinencia (sensor_capture, 5, 6, 7, 7, (sensor_capture > 7));
20
21 S = [relevance_MP relevance_PP relevance_PG relevance_MG];
22
23 end

```

A quarta função, `get_Angle_Rule`, implementa as regras descritas na tabela 6. Por ser extensa, a sua implementação foi colocada na seção **Anexos** ao fim deste documento no programa `??`. Em poucas palavras, esta função testa se cada uma das regras está ativa e, caso uma determinada regra esteja, atribui a ela o menor valor dentre as pertinências dos estados que a compõem. Por exemplo, se D_1 é **P** com 50%, D_2 é **MG** com 100% e D_3 é **M** com 33%, então $\Delta\theta$ é **PN** com 33%.

A quinta função, `get_Angle`, é responsável pelo processo de *defuzzificação* e está representada pelo programa 4. Ela utiliza o método de centro de massa para determinar $\Delta\theta$ que melhor convém dadas as regras ativas.

Programa 4: Avalia pertinência da regras de D_2 em função das distâncias medidas.

```

1 %% FUNCAO get_Angle
2 % Funcao responsavel por calcular o desvio no angulo do robo com base na
3 % regras que foram ativadas pelos sensores.
4 % Parametros: active_rules possui as regras que foram ativadas e suas
5 % respectivas pertinencias.
6 % Retorno: o desvio que devera ser feito na direcao do robo
7 % Autor: Gustavo CIOTTO PINTON
8 function angle = get_Angle (active_rules)
9
10 % Discretizacao do intervalo dos possiveis angulos.
11 available_angles = -20:0.5:20;
12

```

```

13 % c_values eh uma matriz, em que cada componente representa o resultado
14 % representa o minimo entre a funcao de pertinencia calculada em um ponto e
15 % o valor da pertinencia de uma regra ativa. Como ha 5 possiveis funcoes
16 % de pertinencia (MN, PN, Z, PP e MP), a matriz possui 5 linhas.
17 c_values = zeros (5, length(available_angles));
18
19 % parameters eh uma matriz que contem os parametros que descrevem os
20 % trapezios relativos a cada uma das possiveis funcoes de pertinencia de
21 % do desvio angular.
22 parameters = [ -20 -20 -15 -10 0 ;
23                -15 -10 -5  0 0 ;
24                -5  0  0  5 0 ;
25                0  5  10  15 0 ;
26                10  15  20  20 0 ];
27
28 % Itera todos os possiveis angulos.
29 for k = 1: length(available_angles)
30
31     % Para cada angulo e cada funcao de pertinencia, calcula-se o
32     % respectivo valor de pertinencia.
33     for n = 1:5
34
35         % O valor de pertinencia eh o minimo entre o resultado da n-esima funcao
36         % de pertinencia e a pertinencia da regra que corresponde a aquela
37         % mesma funcao.
38         c_values(n, k) = min (active_rules(n), trap_pertinencia(available_angles(k),
39             ...
40             parameters(n ,1), parameters(n ,2), parameters(n ,3), parameters(n ,4),
41             ...
42             parameters(n ,5)));
43     end
44 end
45
46 % Recupera os maximos dos valores de pertinencia entre todas as funcoes de
47 % pertinencia segundo o metodo do MAXIMO dos MINIMOS.
48 max_values = max(c_values);
49
50 % Calcula o centro de massa.
51 angle = sum(available_angles.*max_values)/sum(max_values);
52
53 % Converte o resultado em radianos.
54 angle = angle*pi/180;
55 end

```

Enfim, a sexta função, `robot_control`, realiza o controle do robô conforme regras especificadas acima. Sua implementação está representada pelo programa 5.

Programa 5: Automatiza controle do robô.

```

1 %% FUNCAO robot_control
2 % Funcao responsavel por controlar o robo segundo as regras formalizadas
3 % na secao '3.0 Definicao dos Consequentes'.
4 % Parametros: matrix eh o labirinto onde o robo sera colocado.
5 %             xStart e yStart sao as coordenadas iniciais do robo.
6 %             rDirection eh a direcao inicial do robo.
7 %             speed eh a velocidade do robo.
8 % Autor: Gustavo CIOTTO PINTON
9 function robot_control(matrix, xStart, yStart, rDirection, speed)
10
11 labyrinth_matrix = matrix;
12 robot_direction = rDirection;

```

```

13 x_robot = xStart;
14 y_robot = yStart;
15
16 m_length = length (labyrinth_matrix);
17
18 steps = 0;
19
20 % Movimenta o robo ate achar um ponto de destino, isto e, 'F'
21 while (labyrinth_matrix (round(m_length + 1 - y_robot), round(x_robot)) ~= 'F')
22
23     labyrinth_matrix (round(m_length + 1 - y_robot), round(x_robot)) = 'r';
24
25     % Procura o o obstaculo mais proximo na direcao d1.
26     % Itera ate encontrar tal obstaculo.
27     x_d1 = x_robot + cos(robot_direction + pi/4 );
28     y_d1 = y_robot + sin(robot_direction + pi/4 );
29     while (round(x_d1) >= 1) && (round(x_d1) <= length(labyrinth_matrix (1,:)) ) ...
30         && (round(m_length + 1 - y_d1) >= 1) ...
31         && (round(m_length + 1 - y_d1) <= m_length) ...
32         && (labyrinth_matrix (round(m_length + 1 - y_d1), round(x_d1)) ~= '#')
33
34         x_d1 = x_d1 + cos(robot_direction + pi/4);
35         y_d1 = y_d1 + sin(robot_direction + pi/4);
36
37     end
38     % Calcula distancia ao obstaculo mais proximo encontrado.
39     distance_d1 = sqrt((x_d1 - x_robot)^2 + (y_d1 - y_robot)^2);
40
41     % Procura o o obstaculo mais proximo na direcao d2.
42     % Itera ate encontrar tal obstaculo.
43     x_d2 = x_robot + cos(robot_direction);
44     y_d2 = y_robot + sin(robot_direction);
45     while (round(x_d2) >= 1) && (round(x_d2) <= length(labyrinth_matrix (1,:)) ) ...
46         && (round(m_length + 1 - y_d2) >= 1) ...
47         && (round(m_length + 1 - y_d2) <= m_length) ...
48         && (labyrinth_matrix (round(m_length + 1 - y_d2), round(x_d2)) ~= '#')
49
50         x_d2 = x_d2 + cos(robot_direction);
51         y_d2 = y_d2 + sin(robot_direction);
52
53     end
54
55     % Calcula distancia ao obstaculo mais proximo encontrado.
56     distance_d2 = sqrt((x_d2 - x_robot)^2 + (y_d2 - y_robot)^2);
57
58     % Procura o o obstaculo mais proximo na direcao d3.
59     % Itera ate encontrar tal obstaculo.
60     x_d3 = x_robot + cos(robot_direction - pi/4);
61     y_d3 = y_robot + sin(robot_direction - pi/4);
62     while (round(x_d3) >= 1) && (round(x_d3) <= length(labyrinth_matrix (1,:)) ) ...
63         && (round(m_length + 1 - y_d3) >= 1) ...
64         && (round(m_length + 1 - y_d3) <= m_length) ...
65         && (labyrinth_matrix (round(m_length + 1 - y_d3), round(x_d3)) ~= '#')
66
67         x_d3 = x_d3 + cos(robot_direction - pi/4);
68         y_d3 = y_d3 + sin(robot_direction - pi/4);
69
70     end
71
72     % Calcula distancia ao obstaculo mais proximo encontrado.
73     distance_d3 = sqrt((x_d3 - x_robot)^2 + (y_d3 - y_robot)^2);
74
75     % Verifica valores das funcoes de pertinencia para cada sensor.

```



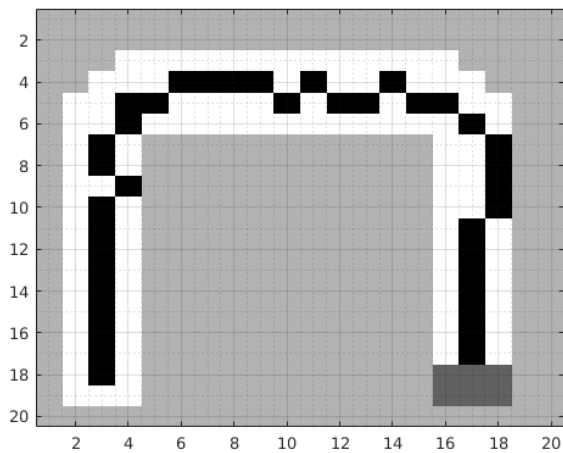
```

76 d1_rules = get_D1_D3_Rule(distance_d1);
77 d2_rules = get_D2_Rule(distance_d2);
78 d3_rules = get_D1_D3_Rule(distance_d3);
79
80 % Verifica quais regras foram ativadas.
81 active_rules = get_Angle_Rule(d1_rules, d2_rules, d3_rules);
82
83 % Processo de defuzzyficacao. A partir das regras ativas, calcula-se o
84 % desvio a ser feito na direcao do robo.
85 d_angle = get_Angle (active_rules);
86
87 robot_direction = robot_direction + d_angle;
88
89 % Calcula nova posicao.
90 x_robot = x_robot + speed*cos(robot_direction);
91 y_robot = y_robot + speed*sin(robot_direction);
92 end
93
94 % Imprime matriz do labirinto.
95 figure
96 imagesc(labyrinth_matrix);
97 colormap(flipud(gray));
98 grid on;
99 grid minor;
100 end

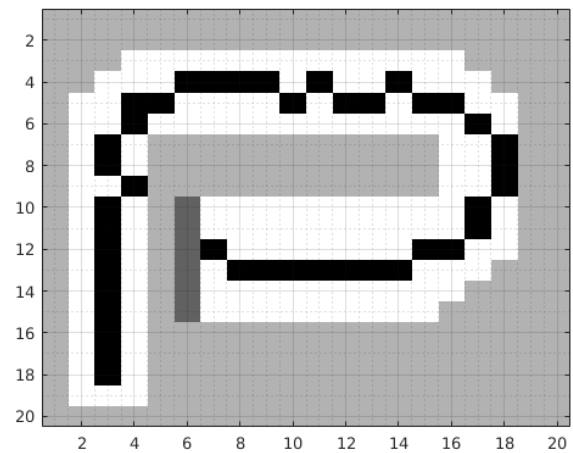
```

3.2 Resultados

A fim de testar o controlador implementado na seção anterior, dois mapas foram criados. Os resultados das execuções estão representados na figuras 11a e 11b, em que a trajetória está representada em preto, a parede em cinza claro, os pontos de destino em cinza escuro e as posições livres, em branco. Para ambos, a direção inicial do robô é $\frac{\pi}{2}$, as coordenadas iniciais são (3,3) e a velocidade é $v = 1$ *pixel/iteração*. Verifica-se que o controle é bem-sucedido, sendo que o robô é capaz de realizar as curvas presentes nos mapas. É importante comentar que os cantos *suavizados* possuem uma grande importância para o bom resultado, uma vez que o controle nas áreas de “ponta” não é tão eficiente visto que, quando o controlador verifica a proximidade a uma parede, ele pode não saber para que lado virar.



(a) Controle do robô no mapa 1.



(b) Controle do robô no mapa 2.

Figura 11: Resultados para o controle nebuloso do robô para dois mapas distintos.

O *script* utilizado nos testes pode ser encontrado no programa 6.

```

1 clear all;
2 close all;
3
4 % Constroi mapa 1.
5 labyrinth_matrix = zeros(20,20);
6 labyrinth_matrix (1:2, :) = '#';
7 labyrinth_matrix (20, :) = '#';
8 labyrinth_matrix (:, 1) = '#';
9 labyrinth_matrix (:, 19:20) = '#';
10 labyrinth_matrix (7:20, 5:15) = '#';
11 labyrinth_matrix (2:4, 2:4) = fliplr(triu(ones(3))) * '#';
12 labyrinth_matrix (2:5, 16:19) = triu(ones(4)) * '#';
13 labyrinth_matrix (18:19, 16:18) = 'F';
14
15 robot_control(labyrinth_matrix, 3,3, pi/2, 1);
16
17 % Constroi mapa 2.
18 labyrinth_matrix = zeros(20,20);
19 labyrinth_matrix (1:2, :) = '#';
20 labyrinth_matrix (20, :) = '#';
21 labyrinth_matrix (:, 1) = '#';
22 labyrinth_matrix (:, 19:20) = '#';
23 labyrinth_matrix (7:20, 5:15) = '#';
24 labyrinth_matrix (16:20, 5:20) = '#';
25 labyrinth_matrix (2:4, 2:4) = fliplr(triu(ones(3))) * '#';
26 labyrinth_matrix (2:5, 16:19) = triu(ones(4)) * '#';
27 labyrinth_matrix (12:15, 16:19) = flipud(triu(ones(4))) * '#';
28 labyrinth_matrix (10:15, 6:15) = 0;
29 labyrinth_matrix (10:15, 6) = 'F';
30
31 robot_control(labyrinth_matrix, 3,3, pi/2, 1);

```

3.3 Controle via Rede Neural MLP

A adaptação do controle do robô via redes neurais exigiu algumas modificações nos programas fornecidos anteriormente nesta seção e também no algoritmo de evolução fornecido pelo professor, que foi utilizado na seção 1. Foram utilizados, neste caso, $N = 10$ para a camada intermediária, 100 gerações e 100 indivíduos na população.

Para o algoritmo evolutivo, foram alteradas a função de *fitness*, a estrutura do cromossomo e as taxas de mutação e *crossover*, conforme explicado a seguir. A função completa pode ser encontrada no programa ?? na seção **Anexos**.

- A função de *fitness* utilizada foi $f_{fitness} = \frac{1}{n_{colisao}+1}$. Logo, soluções que obtiverem muitas colisões serão piores, isto é, terão um *fitness* menor. A linha adicionada ao código foi, portanto:

```
fitness(i,1) = 1/(colision_i + 1);
```

- Anteriormente, tínhamos apenas 3 parâmetros a evoluir, sendo eles k_d , k_p e k_i . Neste caso, o número de parâmetros depende do número de neurônios na camada intermediária. Sejam N neurônios nesta camada, cada um deles terá 4 pesos, sendo um deles para a entrada constante e o restante para as distâncias dos sensores. Há ainda mais $N + 1$ pesos a serem ajustáveis, correspondentes à soma dos resultados de cada neurônio. A estrutura utilizada foi, portanto:

```

% Numero de entradas: d1, d2, d3 e termo constante
E = 3 + 1;
pop = randn(tam_pop, E * n_neurons + n_neurons + 1);

```

Utilizou-se também uma distribuição normal de média 0 e variância 1 para os pesos iniciais da rede neural.

- A fim de maximizar a diversidade na rede, utilizou-se taxas de mutação $p_m = 0.8$ e *crossover* $p_c = 0.9$.

O controle do robô, realizado pela `robot_control_mlp`, também foi alterado. As linhas referentes ao controle nebuloso (linhas 75 a 86 do programa 5) foram retiradas e substituídas pelo trecho de código abaixo.

Programa 7: Trecho de código do controle do robô feito via rede MLP.

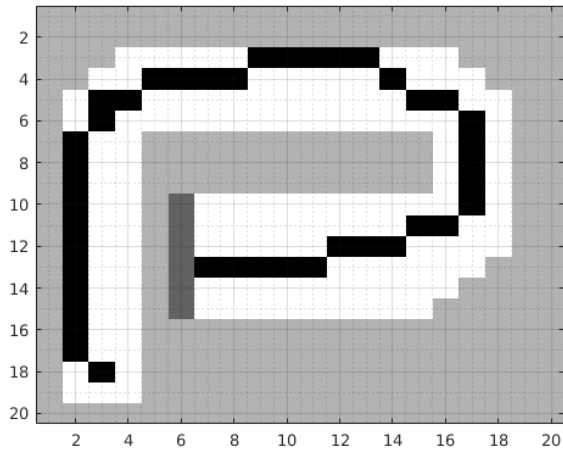
```

1  % Constroi vetor das entradas de cada neurônio
2  mlp_input = [distance_d1 distance_d2 distance_d3 1];
3
4  % middle_layer_output contera as saidas de cada neurônio da camada
5  % intermediária
6  middle_layer_output = zeros (1, n_neurons);
7
8  % Calcula a saída de cada neurônio. mlp_weights contem todos os pesos
9  % sinápticos da rede
10 for n = 1 : n_neurons
11
12     middle_layer_output (1, n) = ...
13         tanh(mlp_input * mlp_weights (1 + (n - 1)* 4 : 4 + (n - 1)* 4)');
14
15 end
16
17 % a entrada da última camada é a saída da camada intermediária e uma
18 % termo constante.
19 last_layer_input = [middle_layer_output 1];
20
21 % Calcula o desvio na direção do robô. Seleciona últimos n_neurons + 1 pesos
22 % sinápticos.
23 d_angle = ...
24     last_layer_input * mlp_weights (1 + 4 * n_neurons : 1 + 4 * n_neurons + n_neurons)';
25
26 robot_direction = robot_direction + d_angle;
27
28 teste (round(m_length + 1 - y_robot), round(x_robot)) = 'r';
29
30 % Calcula nova posição.
31 x_robot = x_robot + speed*cos(robot_direction);
32 y_robot = y_robot + speed*sin(robot_direction);
33
34 % Verifica se a nova posição está dentro do mapa.
35 if (round(x_robot) > length(labyrinth_matrix (1,:))) ...
36     || (round(x_robot) < 1) ...
37     || (round(m_length + 1 - y_robot) < 1) ...
38     || (round(m_length + 1 - y_robot) > m_length)
39
40     % Penaliza movimentos que levam o robô fora do mapa.
41     colisions = colisions + 1000;
42
43 % Detecta colisão.
44 elseif (labyrinth_matrix (round(m_length + 1 - y_robot), round(x_robot)) == '#')
45
46     % Aumenta contador de colisões.
47     colisions = colisions + 1;
48 end

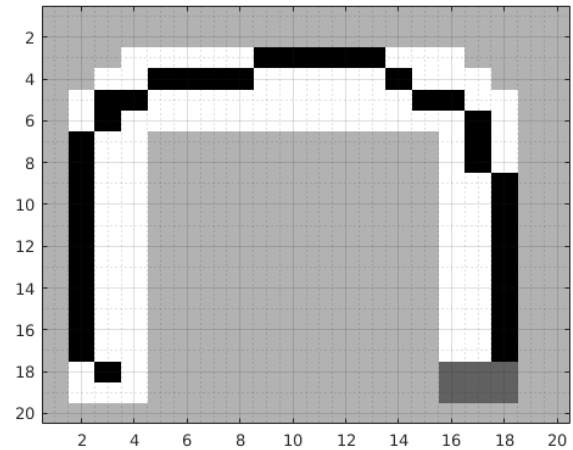
```

O mapa da figura 11b foi utilizado no treinamento da rede neural, visto que apresenta situações mais desafiadoras que o mapa da figura 11a. Sendo assim, acredita-se que se o programa for bem-

sucedido em calcular os pesos sinápticos para o mapa mais complexo, então, os mesmos pesos serão adequados para mapas mais simples. As figuras 12a e 12b contém resultados para um dos vetores de pesos sinápticos encontrado pelo algoritmo evolutivo em que essa afirmação é comprovada.



(a) Controle do robô via MLP no mapa 2.



(b) Controle do robô via MLP no mapa 1

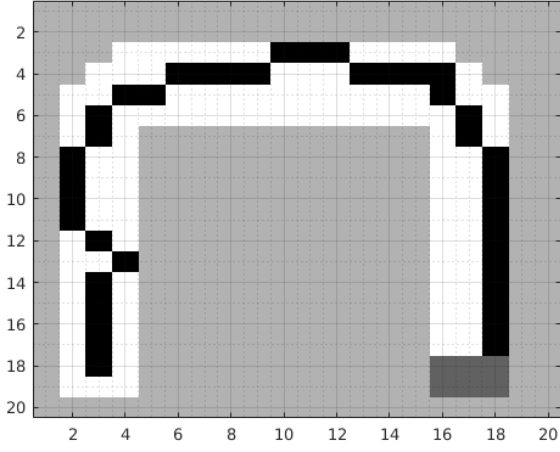
Figura 12: Resultados para o controle do robô via MLP para dois mapas distintos.

A figura 13 representa a evolução do *fitness* ao decorrer da execução do algoritmo evolutivo. Verifica-se que, apesar da melhor solução ser encontrada nas primeiras gerações, o *fitness* médio é muito pequeno em todas as gerações. Esse fato pode ser explicado pelos valores das taxas de mutação e *crossover*, responsáveis por manter uma grande diversidade na população. Dessa forma, sempre serão adicionados à população indivíduos ruins, isto é, que produzem baixos *fitness*.

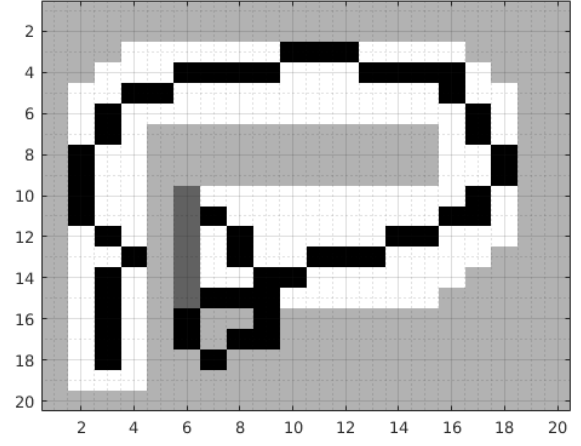


Figura 13: Evolução do *fitness* médio e melhor.

As figuras 14a e 14b foram obtidas pelo treinamento da rede neural no mapa da figura 11a.



(a) Controle do robô via MLP no mapa 1.



(b) Controle do robô via MLP no mapa 2.

Figura 14: Resultados para o controle do robô via MLP para dois mapas distintos.

Verifica-se que a solução encontrada para o primeiro mapa, que é mais simples, controla corretamente o robô. Entretanto, se o mesmo controle fosse utilizado para o mapa 2, então o robô colidiria com a parede, conforme mostrado na figura 14b. Isso é explicado pelo fato de que o mapa exige um controle mais complexo que o primeiro.

4 Sistema de recomendação empregando k -NN

4.1 Abordagens *User-based* e *Item-based*

De acordo com o *paper State-of-the-art Recommender Systems*, sistemas de recomendação que utilizam a abordagem *user-based* exploram as similaridades entre usuários que possuam características em comum para prever uma classificação que um usuário faria potencialmente para um item. Sendo assim, define-se o conjunto \mathbb{U} formado pelos usuários, o conjunto \mathbb{I} formado pelos itens e o conjunto \mathbb{R} que possui todas as classificações dadas. Uma classificação r_{ui} é dada pelo usuário $u \in \mathbb{U}$ para o item $i \in \mathbb{I}$ e cada usuário u possui um conjunto de itens avaliados $\mathbb{S}_u \subseteq \mathbb{I}$. A predição da classificação que um usuário $a \in \mathbb{U}$ faria para o item $i \notin \mathbb{S}_a$ e que considere K elementos do conjunto \mathbb{T}_a , que contém K outros usuários cujas preferências sejam as mais parecidas com as de a , é então:

$$p_{ai} = \frac{\sum_{\{u \in \mathbb{T}_a | i \in \mathbb{S}_u\}} (sim(a, u) * r_{ui})}{\sum_{\{u \in \mathbb{T}_a | i \in \mathbb{S}_u\}} |sim(a, u)|} \quad (6)$$

ou

$$p_{ai} = \bar{r}_a + \frac{\sum_{\{u \in \mathbb{T}_a | i \in \mathbb{S}_u\}} (sim(a, u) * (r_{ui} - \bar{r}_u))}{\sum_{\{u \in \mathbb{T}_a | i \in \mathbb{S}_u\}} |sim(a, u)|} \quad (7)$$

em que $sim(a, u)$ mede o grau de similaridade entre a e u e \bar{r}_v é a média das classificações de um usuário v . A segunda expressão leva em consideração as diferenças de intensidade entre os usuários (por exemplo, em um sistema de 0 a 5, para u , *muito bom* pode ser 3.5, mas, para a , isso é 5). Essas equações nos dizem, portanto, que a predição realizada se aproximará da classificação feita pelo usuário u cujas características são as mais parecidas em relação a aquelas de a .

A abordagem *item-based*, por sua vez, utiliza os *ratings* de itens j já classificados pelo usuário para prever a possível classificação que o respectivo usuário daria a um item i , dado o grau de semelhança entre este item não avaliado e aqueles avaliados. Define-se também a vizinhança \mathbb{T}_i ,

contendo K itens próximos de i em termos de características. Assim como no caso anterior, essa predição p_{ai} pode ser realizada de duas maneiras:

$$p_{ai} = \frac{\sum_{\{j \in \mathbb{S}_a \cap \mathbb{T}_i\}} (sim(i, j) * r_{aj})}{\sum_{\{j \in \mathbb{S}_a \cap \mathbb{T}_i\}} |sim(i, j)|} \quad (8)$$

ou

$$p_{ai} = \bar{r}_i + \frac{\sum_{\{j \in \mathbb{S}_a \cap \mathbb{T}_i\}} (sim(i, j) * (r_{aj} - \bar{r}_j))}{\sum_{\{j \in \mathbb{S}_a \cap \mathbb{T}_i\}} |sim(i, j)|} \quad (9)$$

em que $sim(i, j)$ mede o grau de similaridade entre os itens i e j e \bar{r}_k é a média das classificações feitas para o item k . Como na abordagem anterior, considera-se que itens muito parecidos terão *ratings* também muito parecidos.

4.2 Exemplo de Aplicação

Utilizando o programa `exemplo.m` como referência e o *toolbox* fornecido, foi possível implementar a função `movieLens_impute_data`, representada pelo programa 8 logo abaixo. A função define um usuário, cuja identificação está contida na variável `user_id`, para qual serão calculadas as respectivas recomendações. O programa possui também outro parâmetro, chamado `k_neighbors`, que define a quantidade de vizinhos que serão utilizados no cálculo. Ao final, são impressas as recomendações encontradas e os *ratings* dos vizinhos mais próximos do usuário.

Programa 8: Recomenda filmes para um data usuário segundo o método kNN.

```

1  %% Script movieLens_impute_data
2  % Script responsavel por predizer ratings que um dado usuario poderia
3  % fazer baseado no seu grau de semelhanca com k vizinhos proximos.
4
5  clear all;
6
7  empty_element = -1;
8  load ms;
9  data = build_Data (empty_element);
10
11 k_neighbors = 3;
12 user_id = 898;
13
14 %% Verifica k_neighnors usuarios com maior similaridade
15 [A B] = sort(ms(user_id, :), 'descend');
16 col_max_user = B(1:k_neighbors);
17
18 %% Recupera indices dos elementos vazios
19 k_col = find (data(user_id, :) == empty_element);
20 k_indexes = [user_id*ones(length(k_col), 1) k_col'];
21
22 %% Aplica algoritmo k-NN
23 est = knnImputeRo( data, k_neighbors, k_indexes, empty_element, ms ); % estimativas
24 not_NaN_indexes = find(~isnan(est)); % Recupera somente recomendacoes bem sucedidas
25
26 %% Imprime recomendacoes e valores dos ratings feitos pelos vizinhos
27 display('=====');
28 display(sprintf('Estimativas para usuario %d', user_id));
29 display('=====');
30
31 for k = 1:length(not_NaN_indexes)
32
33     item_id = k_indexes(not_NaN_indexes(k),2);
34     display(sprintf('Recomendacao para filme %d : %.2f', item_id , est(
        not_NaN_indexes(k))));

```

```

35
36 end
37
38 for n=1:k_neighbors
39
40     display('=====');
41     display(sprintf('Ratings do usuario %d de maior similaridade (%.3g)',
42                     col_max_user(n), ...
43                     ms(user_id, col_max_user(n))));
44     display('=====');
45
46     for k = 1:length(not_NaN_indexes)
47
48         item_id = k_indexes(not_NaN_indexes(k),2);
49
50         if data(col_max_user(n), item_id) ~= empty_element
51             display(sprintf('Rating para filme %d : %.2f', item_id, data(col_max_user
52                                 (n), item_id)));
53         end
54     end
55 end

```

Os resultados da execução do programa acima estão representados nas tabelas abaixo. Foi escolhido aleatoriamente o usuário 898 e foram utilizados os 3 vizinhos mais próximos. No bloco superior da coluna da esquerda, estão representadas as recomendações calculadas e que foram bem-sucedidas, isto é, as recomendações que a função `knnImputeRo.m` não retornou como *NaN*. Observa-se que os filmes para os quais uma recomendação foi calculada devem possuir ao menos uma *rating* não nula em algum dos dos vizinhos. Caso contrário, não haveria dados suficientes para o cálculo das respectivas predições. Nos outros 3 blocos estão contidos as *ratings* realizadas pelos vizinhos, em ordem de semelhança com o usuário. Por exemplo, o vizinho 408 possui uma semelhança de 64.8% com o usuário e assim por diante.

É possível alguns fatos:

- O item 294 foi classificado por 408 e 205, sendo que o primeiro possui uma semelhança de 64.8% com o usuário 898 e o segundo, apenas 20.6%. 408 deu a pontuação máxima, enquanto 205 deu apenas uma nota 3.0. A nota recomendada para 898 foi 4.49, que é muito mais próxima de 5 do que de 3. Esse resultado está correta, visto que 408 é muito mais semelhante a 898 do que 205 e, portanto, espera-se que eles dêem notas parecidas.
- O item 1025 só foi avaliado por 205, que não possui uma grande semelhança com 898. A nota recomendada estará, portanto, ao redor do *rating* feito por 205, mas possuirá uma pequena margem que levará em conta essa falta de *total* semelhança.
- O item 333 foi avaliado pelos usuários 384 e 205. Ambos deram nota 4. A nota recomendada para 898 deve ser, portanto, muito próxima de 4, já que dois outros vizinhos deram a mesma nota. Observa-se justamente este fato: a nota recomendada para 898 é 3.99.

```

=====
Estimativas para usuario 898
=====
Recomendacao para filme 268 : 2.64
Recomendacao para filme 269 : 3.64
Recomendacao para filme 289 : 4.50
Recomendacao para filme 294 : 4.49
Recomendacao para filme 304 : 3.64
Recomendacao para filme 322 : 3.64
Recomendacao para filme 326 : 4.64
Recomendacao para filme 329 : 2.36
Recomendacao para filme 333 : 3.99
Recomendacao para filme 355 : 3.36
Recomendacao para filme 678 : 1.64
Recomendacao para filme 875 : 2.64
Recomendacao para filme 878 : 3.36
Recomendacao para filme 879 : 3.36
Recomendacao para filme 984 : 1.64
Recomendacao para filme 989 : 3.36
Recomendacao para filme 1025 : 1.64

```

```

=====
Ratings do usuario 408
de maior similaridade (0.648)
=====
Rating para filme 294 : 5.00

```

```

=====
Ratings do usuario 384
de maior similaridade (0.212)
=====
Rating para filme 289 : 5.00
Rating para filme 329 : 3.00
Rating para filme 333 : 4.00
Rating para filme 355 : 4.00
Rating para filme 878 : 4.00
Rating para filme 879 : 4.00
Rating para filme 989 : 4.00

```

```

=====
Ratings do usuario 205
de maior similaridade (0.206)
=====
Rating para filme 268 : 2.00
Rating para filme 269 : 3.00
Rating para filme 289 : 4.00
Rating para filme 294 : 3.00
Rating para filme 304 : 3.00
Rating para filme 322 : 3.00
Rating para filme 326 : 4.00
Rating para filme 333 : 4.00
Rating para filme 678 : 1.00
Rating para filme 875 : 2.00
Rating para filme 984 : 1.00
Rating para filme 1025 : 1.00

```

5 Sistema de recomendação – Paradigma alternativo

5.1 Funcionamento do SCOAL

Conforme explorado na seção anterior, o algoritmo k -NN primeiramente divide a matriz de dados em *co-clusters* e, em seguida, realiza uma predição baseada em um desses *co-clusters* gerados. Tal algoritmo, portanto, não leva em consideração os atributos dos filmes e dos usuários, isto é, a predição é baseada exclusivamente nos *ratings* dos usuários em uma vizinhança. O algoritmo *SCOAL*, por sua vez, utiliza ambas as informações de vizinhança e atributos para predizer uma nota. A ideia do *SCOAL* é dividir toda a matriz em grupos, ou *clusters*, de maneira que cada um possa ser bem caracterizado por um único modelo preditivo. A similaridade é dada, portanto, pela similaridade dos modelos preditivos e não somente dos valores dos *ratings*, como acontece no k -NN. Além disso, o *SCOAL* realiza o agrupamento, ou *co-clustering*, simultaneamente com a obtenção dos respectivos modelos de classificação, a fim de melhorar a designação dos dados aos *clusters* e a precisão dos modelos.

5.2 A base de dados *MovieLens*

Os dados utilizados nesta seção foram disponibilizados pelo grupo *GroupLens* no mês de Abril de 1998. São 100.000 *ratings*, compreendidas no intervalo [1, 5], dadas por 943 usuários em 1682 filmes e coletadas durante 7 meses, indo de 19 de setembro de 1997 até 22 de abril de 1998. É importante dizer que esses dados foram filtrados, de forma que usuários com menos de 20 *ratings* ou que não possuísem informações demográficas completas foram eliminados da base. Essas informações demográficas são compostas por idade, sexo, ocupação e endereço *zip*. Filmes, além de informações sobre título, data de lançamento e *url* do IMDb, podem possuir diversos gêneros (ação, comédia, aventura etc).

Esses dados foram disponibilizados em alguns arquivos, cujos os principais são:

- `u.data`: arquivo contendo as classificações dos usuários. Cada linha, além de possuir a respectiva *rating*, associa um *id* de usuário a um *id* de filme. Adicionalmente, há um outro atributo que descreve a data que a respectiva classificação foi dada.
- `u.item`: possui as informações apresentadas acima sobre os filmes.
- `u.user`: possui as informações apresentadas acima sobre os usuários.

5.3 Execução do *toolbox* fornecido