Homework 03: Value Iteration

1. Given the world defined by the following Transition Function `fmt(s, a)`, the Reward Function `fr(s, a, sf)` and `gamma = 0.9`:

$$
f_{M_T}(s,a) = \begin{matrix} & a_1 & a_2 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{matrix}\begin{bmatrix} s_2 & s_2 \\ s_1 & s_3 \\ s_3 & s_1 \\ s_1 & s_4 \end{bmatrix} \qquad f_R(s_f) = \begin{matrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{matrix}\begin{bmatrix} 2 \\ 1 \\ -1 \\ 10 \end{bmatrix}
$$

Figure 1: Image

    a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

- It was solved in 45 iterations, if we assume that **a change of less than 0.1 between iterations** means it converged:

```
1  ===========================
2  Iteration 1:
3  fr: [2, 1, -1, 10]
4  s1       s2       s3       s4
5  V(s) current
6  0.00     0.00     0.00     0.00
7  V(s) new
8  1.00     2.90     2.90     10.00
9  ===========================
10 Iteration 2:
11 fr: [2, 1, -1, 10]
12 s1       s2       s3       s4
13 V(s) current
14 1.00     2.90     2.90     10.00
15 V(s) new
16 3.61     5.25     5.25     19.00
17 ===========================
18 Iteration 3:
19 fr: [2, 1, -1, 10]
20 s1       s2       s3       s4
21 V(s) current
22 3.61     5.25     5.25     19.00
23 V(s) new
24 5.72     7.15     7.15     27.10
25 ===========================
26 Iteration 4:
27 fr: [2, 1, -1, 10]
28 s1       s2       s3       s4
29 V(s) current
30 5.72     7.15     7.15     27.10
31 V(s) new
32 7.44     8.69     8.69     34.39
33 ===========================
```

```
34  Iteration 5:
35  fr: [2, 1, -1, 10]
36  s1        s2        s3        s4
37  V(s) current
38  7.44      8.69      8.69      34.39
39  V(s) new
40  8.82      9.94      9.94      40.95
41  ==========================
42  Iteration 6:
43  fr: [2, 1, -1, 10]
44  s1        s2        s3        s4
45  V(s) current
46  8.82      9.94      9.94      40.95
47  V(s) new
48  9.95      10.95     10.95     46.86
49  ==========================
50  Iteration 7:
51  fr: [2, 1, -1, 10]
52  s1        s2        s3        s4
53  V(s) current
54  9.95      10.95     10.95     46.86
55  V(s) new
56  10.86     11.77     11.77     52.17
57  ==========================
58  Iteration 8:
59  fr: [2, 1, -1, 10]
60  s1        s2        s3        s4
61  V(s) current
62  10.86     11.77     11.77     52.17
63  V(s) new
64  11.59     12.43     12.43     56.95
65  ==========================
66  Iteration 9:
67  fr: [2, 1, -1, 10]
68  s1        s2        s3        s4
69  V(s) current
70  11.59     12.43     12.43     56.95
71  V(s) new
72  12.19     12.97     12.97     61.26
73  ==========================
74  Iteration 10:
75  fr: [2, 1, -1, 10]
76  s1        s2        s3        s4
77  V(s) current
78  12.19     12.97     12.97     61.26
79  V(s) new
80  12.68     13.41     13.41     65.13
81  ==========================
82  Iteration 11:
83  fr: [2, 1, -1, 10]
84  s1        s2        s3        s4
85  V(s) current
```

```
 86  12.68    13.41    13.41    65.13
 87  V(s) new
 88  13.07    13.76    13.76    68.62
 89  ==========================
 90  Iteration 12:
 91  fr: [2, 1, -1, 10]
 92  s1        s2        s3        s4
 93  V(s) current
 94  13.07    13.76    13.76    68.62
 95  V(s) new
 96  13.38    14.05    14.05    71.76
 97  ==========================
 98  Iteration 13:
 99  fr: [2, 1, -1, 10]
100  s1        s2        s3        s4
101  V(s) current
102  13.38    14.05    14.05    71.76
103  V(s) new
104  13.64    14.28    14.28    74.58
105  ==========================
106  Iteration 14:
107  fr: [2, 1, -1, 10]
108  s1        s2        s3        s4
109  V(s) current
110  13.64    14.28    14.28    74.58
111  V(s) new
112  13.85    14.46    14.46    77.12
113  ==========================
114  Iteration 15:
115  fr: [2, 1, -1, 10]
116  s1        s2        s3        s4
117  V(s) current
118  13.85    14.46    14.46    77.12
119  V(s) new
120  14.02    14.62    14.62    79.41
121  ==========================
122  Iteration 16:
123  fr: [2, 1, -1, 10]
124  s1        s2        s3        s4
125  V(s) current
126  14.02    14.62    14.62    79.41
127  V(s) new
128  14.15    14.74    14.74    81.47
129  ==========================
130  Iteration 17:
131  fr: [2, 1, -1, 10]
132  s1        s2        s3        s4
133  V(s) current
134  14.15    14.74    14.74    81.47
135  V(s) new
136  14.27    14.84    14.84    83.32
137  ==========================
```

```
138  Iteration 18:
139  fr: [2, 1, -1, 10]
140  s1        s2        s3        s4
141  V(s) current
142  14.27    14.84    14.84    83.32
143  V(s) new
144  14.35    14.92    14.92    84.99
145  ==========================
146  Iteration 19:
147  fr: [2, 1, -1, 10]
148  s1        s2        s3        s4
149  V(s) current
150  14.35    14.92    14.92    84.99
151  V(s) new
152  14.43    14.98    14.98    86.49
153  ==========================
154  Iteration 20:
155  fr: [2, 1, -1, 10]
156  s1        s2        s3        s4
157  V(s) current
158  14.43    14.98    14.98    86.49
159  V(s) new
160  14.49    15.04    15.04    87.84
161  ==========================
162  Iteration 21:
163  fr: [2, 1, -1, 10]
164  s1        s2        s3        s4
165  V(s) current
166  14.49    15.04    15.04    87.84
167  V(s) new
168  14.53    15.08    15.08    89.06
169  ==========================
170  Iteration 22:
171  fr: [2, 1, -1, 10]
172  s1        s2        s3        s4
173  V(s) current
174  14.53    15.08    15.08    89.06
175  V(s) new
176  14.57    15.12    15.12    90.15
177  ==========================
178  Iteration 23:
179  fr: [2, 1, -1, 10]
180  s1        s2        s3        s4
181  V(s) current
182  14.57    15.12    15.12    90.15
183  V(s) new
184  14.60    15.14    15.14    91.14
185  ==========================
186  Iteration 24:
187  fr: [2, 1, -1, 10]
188  s1        s2        s3        s4
189  V(s) current
```

```
190  14.60    15.14    15.14    91.14
191  V(s) new
192  14.63    15.17    15.17    92.02
193  =========================
194  Iteration 25:
195  fr: [2, 1, -1, 10]
196  s1       s2       s3       s4
197  V(s) current
198  14.63    15.17    15.17    92.02
199  V(s) new
200  14.65    15.18    15.18    92.82
201  =========================
202  Iteration 26:
203  fr: [2, 1, -1, 10]
204  s1       s2       s3       s4
205  V(s) current
206  14.65    15.18    15.18    92.82
207  V(s) new
208  14.67    15.20    15.20    93.54
209  =========================
210  Iteration 27:
211  fr: [2, 1, -1, 10]
212  s1       s2       s3       s4
213  V(s) current
214  14.67    15.20    15.20    93.54
215  V(s) new
216  14.68    15.21    15.21    94.19
217  =========================
218  Iteration 28:
219  fr: [2, 1, -1, 10]
220  s1       s2       s3       s4
221  V(s) current
222  14.68    15.21    15.21    94.19
223  V(s) new
224  14.69    15.22    15.22    94.77
225  =========================
226  Iteration 29:
227  fr: [2, 1, -1, 10]
228  s1       s2       s3       s4
229  V(s) current
230  14.69    15.22    15.22    94.77
231  V(s) new
232  14.70    15.23    15.23    95.29
233  =========================
234  Iteration 30:
235  fr: [2, 1, -1, 10]
236  s1       s2       s3       s4
237  V(s) current
238  14.70    15.23    15.23    95.29
239  V(s) new
240  14.71    15.24    15.24    95.76
241  =========================
```

```
242  Iteration 31:
243  fr: [2, 1, -1, 10]
244  s1        s2        s3        s4
245  V(s) current
246  14.71    15.24    15.24    95.76
247  V(s) new
248  14.71    15.24    15.24    96.18
249  ==========================
250  Iteration 32:
251  fr: [2, 1, -1, 10]
252  s1        s2        s3        s4
253  V(s) current
254  14.71    15.24    15.24    96.18
255  V(s) new
256  14.72    15.25    15.25    96.57
257  ==========================
258  Iteration 33:
259  fr: [2, 1, -1, 10]
260  s1        s2        s3        s4
261  V(s) current
262  14.72    15.25    15.25    96.57
263  V(s) new
264  14.72    15.25    15.25    96.91
265  ==========================
266  Iteration 34:
267  fr: [2, 1, -1, 10]
268  s1        s2        s3        s4
269  V(s) current
270  14.72    15.25    15.25    96.91
271  V(s) new
272  14.72    15.25    15.25    97.22
273  ==========================
274  Iteration 35:
275  fr: [2, 1, -1, 10]
276  s1        s2        s3        s4
277  V(s) current
278  14.72    15.25    15.25    97.22
279  V(s) new
280  14.73    15.25    15.25    97.50
281  ==========================
282  Iteration 36:
283  fr: [2, 1, -1, 10]
284  s1        s2        s3        s4
285  V(s) current
286  14.73    15.25    15.25    97.50
287  V(s) new
288  14.73    15.26    15.26    97.75
289  ==========================
290  Iteration 37:
291  fr: [2, 1, -1, 10]
292  s1        s2        s3        s4
293  V(s) current
```

```
294  14.73    15.26    15.26    97.75
295  V(s) new
296  14.73    15.26    15.26    97.97
297  =========================
298  Iteration 38:
299  fr: [2, 1, -1, 10]
300  s1       s2       s3       s4
301  V(s) current
302  14.73    15.26    15.26    97.97
303  V(s) new
304  14.73    15.26    15.26    98.18
305  =========================
306  Iteration 39:
307  fr: [2, 1, -1, 10]
308  s1       s2       s3       s4
309  V(s) current
310  14.73    15.26    15.26    98.18
311  V(s) new
312  14.73    15.26    15.26    98.36
313  =========================
314  Iteration 40:
315  fr: [2, 1, -1, 10]
316  s1       s2       s3       s4
317  V(s) current
318  14.73    15.26    15.26    98.36
319  V(s) new
320  14.73    15.26    15.26    98.52
321  =========================
322  Iteration 41:
323  fr: [2, 1, -1, 10]
324  s1       s2       s3       s4
325  V(s) current
326  14.73    15.26    15.26    98.52
327  V(s) new
328  14.73    15.26    15.26    98.67
329  =========================
330  Iteration 42:
331  fr: [2, 1, -1, 10]
332  s1       s2       s3       s4
333  V(s) current
334  14.73    15.26    15.26    98.67
335  V(s) new
336  14.73    15.26    15.26    98.80
337  =========================
338  Iteration 43:
339  fr: [2, 1, -1, 10]
340  s1       s2       s3       s4
341  V(s) current
342  14.73    15.26    15.26    98.80
343  V(s) new
344  14.73    15.26    15.26    98.92
345  =========================
```

```
346  Iteration 44:
347  fr: [2, 1, -1, 10]
348  s1        s2        s3        s4
349  V(s) current
350  14.73    15.26    15.26    98.92
351  V(s) new
352  14.74    15.26    15.26    99.03
353  ==========================
354  Iteration 45:
355  fr: [2, 1, -1, 10]
356  s1        s2        s3        s4
357  V(s) current
358  14.74    15.26    15.26    99.03
359  V(s) new
360  14.74    15.26    15.26    99.13
361  Optimal Politic:
362  s1 = a1,          s2 = a1,          s3 = a2,          s4 = a2,
```

Which gives out a result of:

	s1	s2	s3	s4
V(s)	14.74	15.26	15.26	99.13
f_pi(s)	a1	a1	a2	a2

Below is a convergence plot for all V(s) values for each of the states s = {s1, s2, s3, s4}, where x axis is the **amount of iterations** and y axis is the **V(s) result for that iteration**:
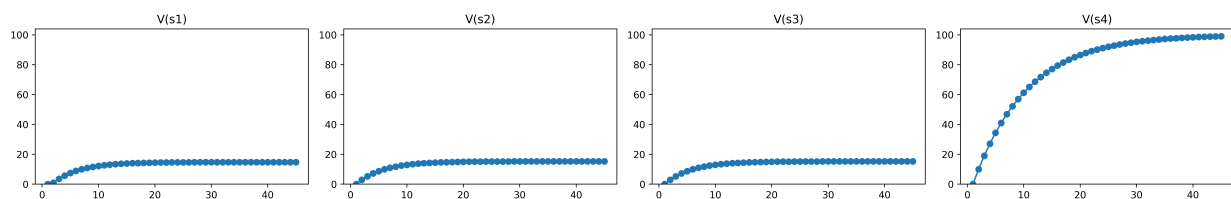
Figure 2: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s, a).

- It was solved in 45 iterations as well, if we assume that **a change of less than 0.1 between iterations** means convergence:

```
1         ==========================
2  Iteration 1:
3  fr: [2, 1, -1, 10]
4  s1        s2        s3        s4
5  Q(s,a) current
6  0.00     0.00     0.00     0.00
```

```
 7  0.00      0.00      0.00      0.00
 8  Q(s,a) new
 9  1.00      2.90      -1.00     2.90
10  3.61      -1.00     5.25      12.61
11  =========================
12  Iteration 2:
13  fr: [2, 1, -1, 10]
14  s1        s2        s3        s4
15  Q(s,a) current
16  1.00      2.90      -1.00     2.90
17  3.61      -1.00     5.25      12.61
18  Q(s,a) new
19  3.61      5.25      3.72      5.25
20  5.72      3.72      7.15      21.35
21  =========================
22  Iteration 3:
23  fr: [2, 1, -1, 10]
24  s1        s2        s3        s4
25  Q(s,a) current
26  3.61      5.25      3.72      5.25
27  5.72      3.72      7.15      21.35
28  Q(s,a) new
29  5.72      7.15      5.44      7.15
30  7.44      5.44      8.69      29.21
31  =========================
32  Iteration 4:
33  fr: [2, 1, -1, 10]
34  s1        s2        s3        s4
35  Q(s,a) current
36  5.72      7.15      5.44      7.15
37  7.44      5.44      8.69      29.21
38  Q(s,a) new
39  7.44      8.69      6.82      8.69
40  8.82      6.82      9.94      36.29
41  =========================
42  Iteration 5:
43  fr: [2, 1, -1, 10]
44  s1        s2        s3        s4
45  Q(s,a) current
46  7.44      8.69      6.82      8.69
47  8.82      6.82      9.94      36.29
48  Q(s,a) new
49  8.82      9.94      7.95      9.94
50  9.95      7.95      10.95     42.66
51  =========================
52  Iteration 6:
53  fr: [2, 1, -1, 10]
54  s1        s2        s3        s4
55  Q(s,a) current
56  8.82      9.94      7.95      9.94
57  9.95      7.95      10.95     42.66
58  Q(s,a) new
```

```
59  9.95      10.95    8.86      10.95
60  10.86     8.86     11.77     48.40
61  =========================
62  Iteration 7:
63  fr: [2, 1, -1, 10]
64  s1       s2       s3       s4
65  Q(s,a) current
66  9.95      10.95    8.86      10.95
67  10.86     8.86     11.77     48.40
68  Q(s,a) new
69  10.86     11.77    9.59      11.77
70  11.59     9.59     12.43     53.56
71  =========================
72  Iteration 8:
73  fr: [2, 1, -1, 10]
74  s1       s2       s3       s4
75  Q(s,a) current
76  10.86     11.77    9.59      11.77
77  11.59     9.59     12.43     53.56
78  Q(s,a) new
79  11.59     12.43    10.19     12.43
80  12.19     10.19    12.97     58.20
81  =========================
82  Iteration 9:
83  fr: [2, 1, -1, 10]
84  s1       s2       s3       s4
85  Q(s,a) current
86  11.59     12.43    10.19     12.43
87  12.19     10.19    12.97     58.20
88  Q(s,a) new
89  12.19     12.97    10.68     12.97
90  12.68     10.68    13.41     62.38
91  =========================
92  Iteration 10:
93  fr: [2, 1, -1, 10]
94  s1       s2       s3       s4
95  Q(s,a) current
96  12.19     12.97    10.68     12.97
97  12.68     10.68    13.41     62.38
98  Q(s,a) new
99  12.68     13.41    11.07     13.41
100 13.07     11.07    13.76     66.14
101 =========================
102 Iteration 11:
103 fr: [2, 1, -1, 10]
104 s1       s2       s3       s4
105 Q(s,a) current
106 12.68     13.41    11.07     13.41
107 13.07     11.07    13.76     66.14
108 Q(s,a) new
109 13.07     13.76    11.38     13.76
110 13.38     11.38    14.05     69.53
```

```
111  ==========================
112  Iteration 12:
113  fr: [2, 1, -1, 10]
114  s1        s2        s3        s4
115  Q(s,a) current
116  13.07    13.76    11.38    13.76
117  13.38    11.38    14.05    69.53
118  Q(s,a) new
119  13.38    14.05    11.64    14.05
120  13.64    11.64    14.28    72.58
121  ==========================
122  Iteration 13:
123  fr: [2, 1, -1, 10]
124  s1        s2        s3        s4
125  Q(s,a) current
126  13.38    14.05    11.64    14.05
127  13.64    11.64    14.28    72.58
128  Q(s,a) new
129  13.64    14.28    11.85    14.28
130  13.85    11.85    14.46    75.32
131  ==========================
132  Iteration 14:
133  fr: [2, 1, -1, 10]
134  s1        s2        s3        s4
135  Q(s,a) current
136  13.64    14.28    11.85    14.28
137  13.85    11.85    14.46    75.32
138  Q(s,a) new
139  13.85    14.46    12.02    14.46
140  14.02    12.02    14.62    77.79
141  ==========================
142  Iteration 15:
143  fr: [2, 1, -1, 10]
144  s1        s2        s3        s4
145  Q(s,a) current
146  13.85    14.46    12.02    14.46
147  14.02    12.02    14.62    77.79
148  Q(s,a) new
149  14.02    14.62    12.15    14.62
150  14.15    12.15    14.74    80.01
151  ==========================
152  Iteration 16:
153  fr: [2, 1, -1, 10]
154  s1        s2        s3        s4
155  Q(s,a) current
156  14.02    14.62    12.15    14.62
157  14.15    12.15    14.74    80.01
158  Q(s,a) new
159  14.15    14.74    12.27    14.74
160  14.27    12.27    14.84    82.01
161  ==========================
162  Iteration 17:
```

```
163  fr: [2, 1, -1, 10]
164  s1       s2       s3       s4
165  Q(s,a) current
166  14.15    14.74    12.27    14.74
167  14.27    12.27    14.84    82.01
168  Q(s,a) new
169  14.27    14.84    12.35    14.84
170  14.35    12.35    14.92    83.81
171  ==========================
172  Iteration 18:
173  fr: [2, 1, -1, 10]
174  s1       s2       s3       s4
175  Q(s,a) current
176  14.27    14.84    12.35    14.84
177  14.35    12.35    14.92    83.81
178  Q(s,a) new
179  14.35    14.92    12.43    14.92
180  14.43    12.43    14.98    85.43
181  ==========================
182  Iteration 19:
183  fr: [2, 1, -1, 10]
184  s1       s2       s3       s4
185  Q(s,a) current
186  14.35    14.92    12.43    14.92
187  14.43    12.43    14.98    85.43
188  Q(s,a) new
189  14.43    14.98    12.49    14.98
190  14.49    12.49    15.04    86.88
191  ==========================
192  Iteration 20:
193  fr: [2, 1, -1, 10]
194  s1       s2       s3       s4
195  Q(s,a) current
196  14.43    14.98    12.49    14.98
197  14.49    12.49    15.04    86.88
198  Q(s,a) new
199  14.49    15.04    12.53    15.04
200  14.53    12.53    15.08    88.19
201  ==========================
202  Iteration 21:
203  fr: [2, 1, -1, 10]
204  s1       s2       s3       s4
205  Q(s,a) current
206  14.49    15.04    12.53    15.04
207  14.53    12.53    15.08    88.19
208  Q(s,a) new
209  14.53    15.08    12.57    15.08
210  14.57    12.57    15.12    89.38
211  ==========================
212  Iteration 22:
213  fr: [2, 1, -1, 10]
214  s1       s2       s3       s4
```

```
215  Q(s,a) current
216  14.53    15.08    12.57    15.08
217  14.57    12.57    15.12    89.38
218  Q(s,a) new
219  14.57    15.12    12.60    15.12
220  14.60    12.60    15.14    90.44
221  ==========================
222  Iteration 23:
223  fr: [2, 1, -1, 10]
224  s1       s2       s3       s4
225  Q(s,a) current
226  14.57    15.12    12.60    15.12
227  14.60    12.60    15.14    90.44
228  Q(s,a) new
229  14.60    15.14    12.63    15.14
230  14.63    12.63    15.17    91.39
231  ==========================
232  Iteration 24:
233  fr: [2, 1, -1, 10]
234  s1       s2       s3       s4
235  Q(s,a) current
236  14.60    15.14    12.63    15.14
237  14.63    12.63    15.17    91.39
238  Q(s,a) new
239  14.63    15.17    12.65    15.17
240  14.65    12.65    15.18    92.25
241  ==========================
242  Iteration 25:
243  fr: [2, 1, -1, 10]
244  s1       s2       s3       s4
245  Q(s,a) current
246  14.63    15.17    12.65    15.17
247  14.65    12.65    15.18    92.25
248  Q(s,a) new
249  14.65    15.18    12.67    15.18
250  14.67    12.67    15.20    93.03
251  ==========================
252  Iteration 26:
253  fr: [2, 1, -1, 10]
254  s1       s2       s3       s4
255  Q(s,a) current
256  14.65    15.18    12.67    15.18
257  14.67    12.67    15.20    93.03
258  Q(s,a) new
259  14.67    15.20    12.68    15.20
260  14.68    12.68    15.21    93.73
261  ==========================
262  Iteration 27:
263  fr: [2, 1, -1, 10]
264  s1       s2       s3       s4
265  Q(s,a) current
266  14.67    15.20    12.68    15.20
```

```
267  14.68    12.68    15.21    93.73
268  Q(s,a) new
269  14.68    15.21    12.69    15.21
270  14.69    12.69    15.22    94.35
271  =========================
272  Iteration 28:
273  fr: [2, 1, -1, 10]
274  s1       s2       s3       s4
275  Q(s,a) current
276  14.68    15.21    12.69    15.21
277  14.69    12.69    15.22    94.35
278  Q(s,a) new
279  14.69    15.22    12.70    15.22
280  14.70    12.70    15.23    94.92
281  =========================
282  Iteration 29:
283  fr: [2, 1, -1, 10]
284  s1       s2       s3       s4
285  Q(s,a) current
286  14.69    15.22    12.70    15.22
287  14.70    12.70    15.23    94.92
288  Q(s,a) new
289  14.70    15.23    12.71    15.23
290  14.71    12.71    15.24    95.43
291  =========================
292  Iteration 30:
293  fr: [2, 1, -1, 10]
294  s1       s2       s3       s4
295  Q(s,a) current
296  14.70    15.23    12.71    15.23
297  14.71    12.71    15.24    95.43
298  Q(s,a) new
299  14.71    15.24    12.71    15.24
300  14.71    12.71    15.24    95.88
301  =========================
302  Iteration 31:
303  fr: [2, 1, -1, 10]
304  s1       s2       s3       s4
305  Q(s,a) current
306  14.71    15.24    12.71    15.24
307  14.71    12.71    15.24    95.88
308  Q(s,a) new
309  14.71    15.24    12.72    15.24
310  14.72    12.72    15.25    96.30
311  =========================
312  Iteration 32:
313  fr: [2, 1, -1, 10]
314  s1       s2       s3       s4
315  Q(s,a) current
316  14.71    15.24    12.72    15.24
317  14.72    12.72    15.25    96.30
318  Q(s,a) new
```

```
319  14.72    15.25    12.72    15.25
320  14.72    12.72    15.25    96.67
321  =========================
322  Iteration 33:
323  fr: [2, 1, -1, 10]
324  s1       s2       s3       s4
325  Q(s,a) current
326  14.72    15.25    12.72    15.25
327  14.72    12.72    15.25    96.67
328  Q(s,a) new
329  14.72    15.25    12.72    15.25
330  14.72    12.72    15.25    97.00
331  =========================
332  Iteration 34:
333  fr: [2, 1, -1, 10]
334  s1       s2       s3       s4
335  Q(s,a) current
336  14.72    15.25    12.72    15.25
337  14.72    12.72    15.25    97.00
338  Q(s,a) new
339  14.72    15.25    12.73    15.25
340  14.73    12.73    15.25    97.30
341  =========================
342  Iteration 35:
343  fr: [2, 1, -1, 10]
344  s1       s2       s3       s4
345  Q(s,a) current
346  14.72    15.25    12.73    15.25
347  14.73    12.73    15.25    97.30
348  Q(s,a) new
349  14.73    15.25    12.73    15.25
350  14.73    12.73    15.26    97.57
351  =========================
352  Iteration 36:
353  fr: [2, 1, -1, 10]
354  s1       s2       s3       s4
355  Q(s,a) current
356  14.73    15.25    12.73    15.25
357  14.73    12.73    15.26    97.57
358  Q(s,a) new
359  14.73    15.26    12.73    15.26
360  14.73    12.73    15.26    97.81
361  =========================
362  Iteration 37:
363  fr: [2, 1, -1, 10]
364  s1       s2       s3       s4
365  Q(s,a) current
366  14.73    15.26    12.73    15.26
367  14.73    12.73    15.26    97.81
368  Q(s,a) new
369  14.73    15.26    12.73    15.26
370  14.73    12.73    15.26    98.03
```

```
371  ===========================
372  Iteration 38:
373  fr: [2, 1, -1, 10]
374  s1       s2       s3       s4
375  Q(s,a) current
376  14.73    15.26    12.73    15.26
377  14.73    12.73    15.26    98.03
378  Q(s,a) new
379  14.73    15.26    12.73    15.26
380  14.73    12.73    15.26    98.23
381  ===========================
382  Iteration 39:
383  fr: [2, 1, -1, 10]
384  s1       s2       s3       s4
385  Q(s,a) current
386  14.73    15.26    12.73    15.26
387  14.73    12.73    15.26    98.23
388  Q(s,a) new
389  14.73    15.26    12.73    15.26
390  14.73    12.73    15.26    98.41
391  ===========================
392  Iteration 40:
393  fr: [2, 1, -1, 10]
394  s1       s2       s3       s4
395  Q(s,a) current
396  14.73    15.26    12.73    15.26
397  14.73    12.73    15.26    98.41
398  Q(s,a) new
399  14.73    15.26    12.73    15.26
400  14.73    12.73    15.26    98.56
401  ===========================
402  Iteration 41:
403  fr: [2, 1, -1, 10]
404  s1       s2       s3       s4
405  Q(s,a) current
406  14.73    15.26    12.73    15.26
407  14.73    12.73    15.26    98.56
408  Q(s,a) new
409  14.73    15.26    12.73    15.26
410  14.73    12.73    15.26    98.71
411  ===========================
412  Iteration 42:
413  fr: [2, 1, -1, 10]
414  s1       s2       s3       s4
415  Q(s,a) current
416  14.73    15.26    12.73    15.26
417  14.73    12.73    15.26    98.71
418  Q(s,a) new
419  14.73    15.26    12.73    15.26
420  14.73    12.73    15.26    98.84
421  ===========================
422  Iteration 43:
```

```
423  fr: [2, 1, -1, 10]
424  s1       s2       s3       s4
425  Q(s,a) current
426  14.73    15.26    12.73    15.26
427  14.73    12.73    15.26    98.84
428  Q(s,a) new
429  14.73    15.26    12.74    15.26
430  14.74    12.74    15.26    98.95
431  ==========================
432  Iteration 44:
433  fr: [2, 1, -1, 10]
434  s1       s2       s3       s4
435  Q(s,a) current
436  14.73    15.26    12.74    15.26
437  14.74    12.74    15.26    98.95
438  Q(s,a) new
439  14.74    15.26    12.74    15.26
440  14.74    12.74    15.26    99.06
441  ==========================
442  Iteration 45:
443  fr: [2, 1, -1, 10]
444  s1       s2       s3       s4
445  Q(s,a) current
446  14.74    15.26    12.74    15.26
447  14.74    12.74    15.26    99.06
448  Q(s,a) new
449  14.74    15.26    12.74    15.26
450  14.74    12.74    15.26    99.15
451  Optimal Politic:
452  s1 = a1,        s2 = a1,        s3 = a2,        s4 = a2,
```

Which give out a result of:

	s1	s2	s3	s4
Q(s, a1)	**14.74**	**15.26**	12.74	15.26
Q(s, a2)	14.74	12.74	**15.26**	**99.15**
f_pi(s)	a1	a1	a2	a2

Below is a similarly plotted convergence plot for Q(s, a) values:
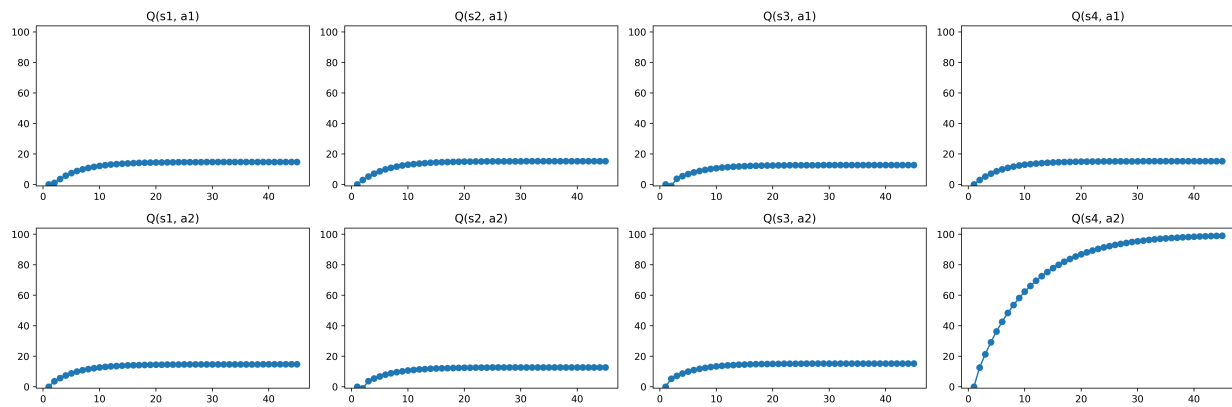
Figure 3: Image

2. Given the world defined by the Transition Function `Pmt(sf|s, a)`, the Reward Function `fr(s, a, sf)= fr(sf)` and `gamma = 0.6`:

$$f_{M_T}(s,a) = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{cc} s_f = s_1 \\ a_1 \quad a_2 \\ \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & 0 \\ 1 & 0.3 \end{bmatrix} \end{array} \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{cc} s_f = s_2 \\ a_1 \quad a_2 \\ \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \\ 0 & 0.6 \end{bmatrix} \end{array} \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{cc} s_f = s_3 \\ a_1 \quad a_2 \\ \begin{bmatrix} 0.1 & 0 \\ 0.5 & 1 \\ 0 & 0.1 \end{bmatrix} \end{array} \qquad f_R(s_f) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$
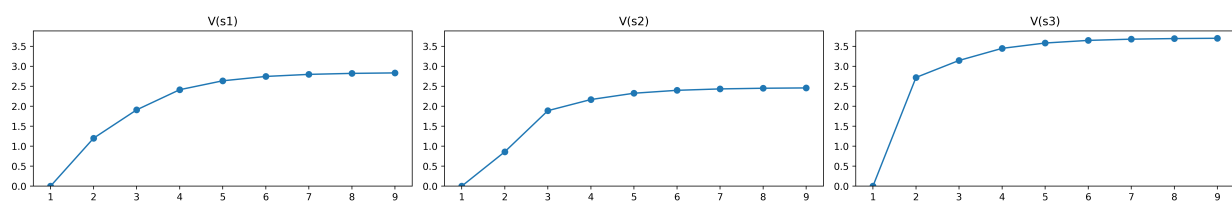
Figure 4: Image

a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

It was solved in 9 iterations, if we assume that **a change of less than 0.01 between iterations** means it converged:

```
1   ===========================
2   Iteration 1:
3   fr: [2, 1, -1]
4   s1        s2        s3
5   V(s) current
6   0.00      0.00      0.00
7   V(s) new
8   1.20      0.86      2.72
9   ===========================
10  Iteration 2:
11  fr: [2, 1, -1]
12  s1        s2        s3
13  V(s) current
14  1.20      0.86      2.72
15  V(s) new
16  1.91      1.89      3.15
17  ===========================
18  Iteration 3:
19  fr: [2, 1, -1]
20  s1        s2        s3
```

```
21  V(s) current
22  1.91      1.89       3.15
23  V(s) new
24  2.41      2.17       3.45
25  =========================
26  Iteration 4:
27  fr: [2, 1, -1]
28  s1        s2         s3
29  V(s) current
30  2.41      2.17       3.45
31  V(s) new
32  2.64      2.33       3.58
33  =========================
34  Iteration 5:
35  fr: [2, 1, -1]
36  s1        s2         s3
37  V(s) current
38  2.64      2.33       3.58
39  V(s) new
40  2.75      2.40       3.65
41  =========================
42  Iteration 6:
43  fr: [2, 1, -1]
44  s1        s2         s3
45  V(s) current
46  2.75      2.40       3.65
47  V(s) new
48  2.80      2.43       3.68
49  =========================
50  Iteration 7:
51  fr: [2, 1, -1]
52  s1        s2         s3
53  V(s) current
54  2.80      2.43       3.68
55  V(s) new
56  2.82      2.45       3.69
57  =========================
58  Iteration 8:
59  fr: [2, 1, -1]
60  s1        s2         s3
61  V(s) current
62  2.82      2.45       3.69
63  V(s) new
64  2.83      2.46       3.70
65  =========================
66  Iteration 9:
67  fr: [2, 1, -1]
68  s1        s2         s3
69  V(s) current
70  2.83      2.46       3.70
71  V(s) new
72  2.84      2.46       3.70
```

```
73  Optimal Politic:
74  s1 = a1,          s2 = a1,          s3 = a1
```

Which gives out a result of:

	s1	s2	s3
V(s)	2.84	2.46	3.70
f_pi(s)	a1	a1	a1

The convergence plot for each V(s) value is given below:

Figure 5: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s, a).

Solution

It was solved in 8 iterations, assuming **a change of less than 0.01 between iterations** means convergence:

```
1   ==========================
2   Iteration 1:
3   fr: [2, 1, -1]
4   s1        s2        s3
5   Q(s,a) current
6   0.00      0.00      0.00
7   0.00      0.00      0.00
8   Q(s,a) new
9   1.20      0.86      2.72
10  1.76      0.63      1.89
11  ==========================
12  Iteration 2:
13  fr: [2, 1, -1]
14  s1        s2        s3
15  Q(s,a) current
16  1.20      0.86      2.72
17  1.76      0.63      1.89
18  Q(s,a) new
19  2.04      1.93      3.23
20  2.37      0.94      2.41
21  ==========================
22  Iteration 3:
```

```
23  fr: [2, 1, -1]
24  s1       s2       s3
25  Q(s,a) current
26  2.04     1.93     3.23
27  2.37     0.94     2.41
28  Q(s,a) new
29  2.54     2.23     3.52
30  2.58     1.11     2.58
31  ==========================
32  Iteration 4:
33  fr: [2, 1, -1]
34  s1       s2       s3
35  Q(s,a) current
36  2.54     2.23     3.52
37  2.58     1.11     2.58
38  Q(s,a) new
39  2.70     2.37     3.62
40  2.66     1.17     2.66
41  ==========================
42  Iteration 5:
43  fr: [2, 1, -1]
44  s1       s2       s3
45  Q(s,a) current
46  2.70     2.37     3.62
47  2.66     1.17     2.66
48  Q(s,a) new
49  2.77     2.42     3.66
50  2.69     1.20     2.69
51  ==========================
52  Iteration 6:
53  fr: [2, 1, -1]
54  s1       s2       s3
55  Q(s,a) current
56  2.77     2.42     3.66
57  2.69     1.20     2.69
58  Q(s,a) new
59  2.81     2.44     3.69
60  2.71     1.21     2.71
61  ==========================
62  Iteration 7:
63  fr: [2, 1, -1]
64  s1       s2       s3
65  Q(s,a) current
66  2.81     2.44     3.69
67  2.71     1.21     2.71
68  Q(s,a) new
69  2.83     2.45     3.70
70  2.72     1.22     2.71
71  ==========================
72  Iteration 8:
73  fr: [2, 1, -1]
74  s1       s2       s3
```

```
75  Q(s,a)  current
76  2.83     2.45      3.70
77  2.72     1.22      2.71
78  Q(s,a)  new
79  2.84     2.46      3.70
80  2.72     1.22      2.72
81  Optimal  Politic:
82  s1 = a1,          s2 = a1,          s3 = a1
```

Which gives out a result of:

	s1	s2	s3
Q(s, a1)	**2.84**	**2.46**	**3.70**
Q(s, a2)	2.72	1.22	2.72
f_pi(s)	a1	a1	a1

The convergence plot for each Q(s,a) value is given below:
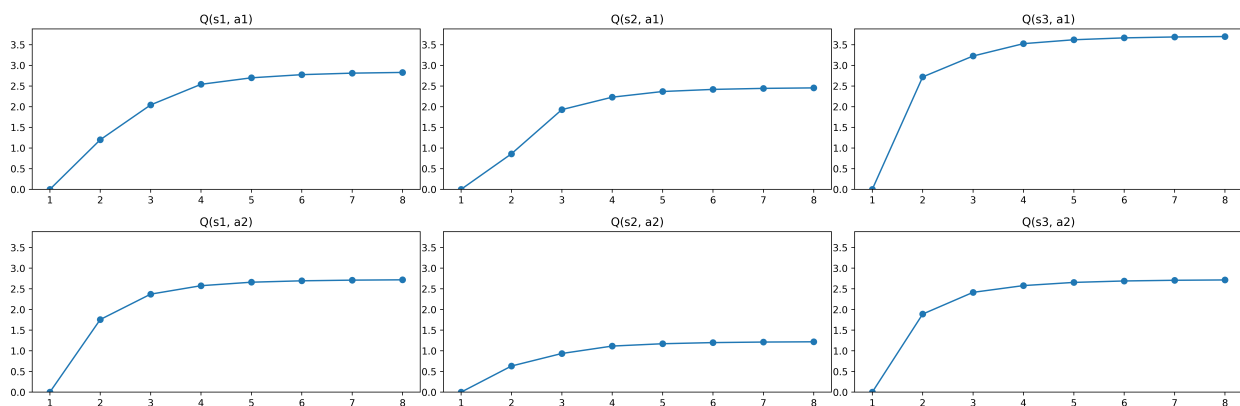
Figure 6: Image

3. Given the world defined by the graph, the Reward Function `fr(s,a,sf)` and `gamma = 0.9`:

$$f_R(s, a, s_f) = \begin{matrix} & \begin{matrix} s_f = s_1 \\ a_1 \quad a_2 \quad a_3 \end{matrix} & \begin{matrix} s_f = s_2 \\ a_1 \quad a_2 \quad a_3 \end{matrix} & \begin{matrix} s_f = s_3 \\ a_1 \quad a_2 \quad a_3 \end{matrix} & \begin{matrix} s_f = s_4 \\ a_1 \quad a_2 \quad a_3 \end{matrix} & \begin{matrix} s_f = s_5 \\ a_1 \quad a_2 \quad a_3 \end{matrix} \end{matrix}$$

$$f_R(s,a,s_f) = \begin{matrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{matrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 5 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
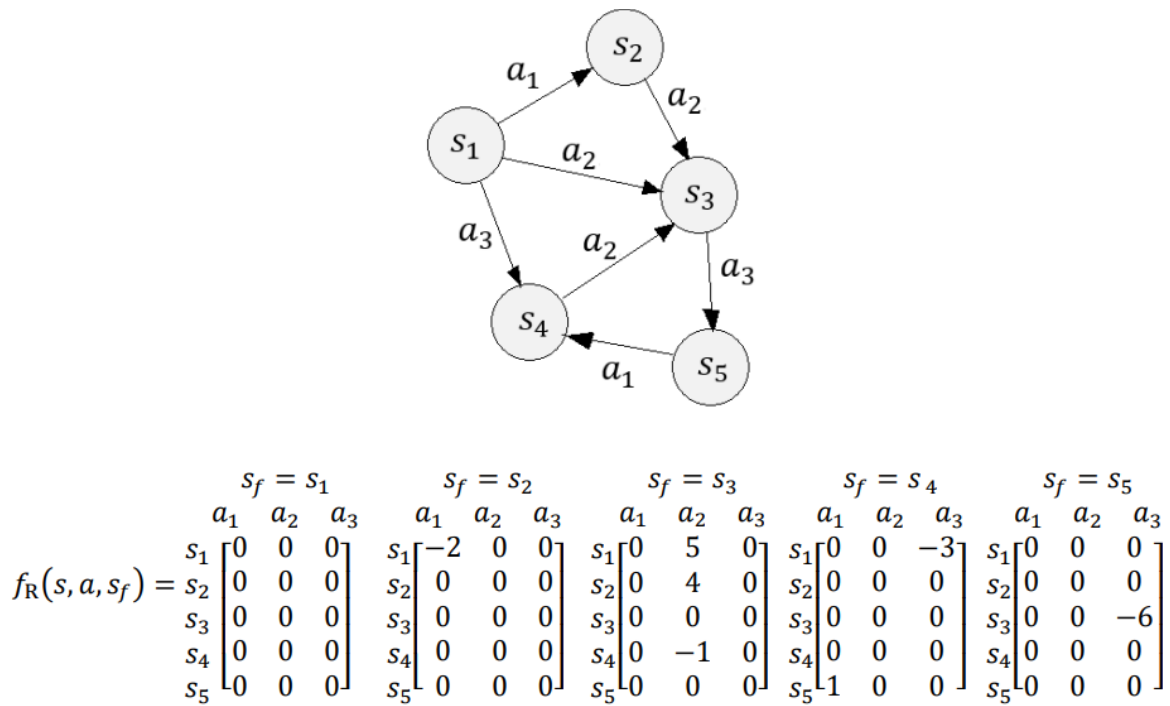
Figure 7: Image

a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

The Transition Function can therefore be defined as:

Figure 8: Image

Thus, the Bellman Optimality Equations were solved in 2 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1  ===========================
2  Iteration 1:
3  fr: [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
       0]], [[-2, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
       0]], [[0, 5, 0], [0, 4, 0], [0, 0, 0], [0, -1, 0], [0,
       0, 0]], [[0, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0], [1,
       0, 0]], [[0, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0, 0],
       [0, 0, 0]]]
4  s1        s2        s3        s4        s5
5  V(s) current
```

```
 6  0.00    0.00    0.00    0.00    0.00
 7  V(s) new
 8  5.00    4.00    0.00    0.00    1.00
 9  ==========================
10  Iteration 2:
11  fr: [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
        0]], [[-2, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
         0]], [[0, 5, 0], [0, 4, 0], [0, 0, 0], [0, -1, 0], [0,
        0, 0]], [[0, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0], [1,
         0, 0]], [[0, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0, 0],
        [0, 0, 0]]]
12  s1      s2      s3      s4      s5
13  V(s) current
14  5.00    4.00    0.00    0.00    1.00
15  V(s) new
16  5.00    4.00    0.00    0.00    1.00
17  Optimal Politic:
18  s1 = a2,        s2 = a2,        s3 = a1,        s4 = a1,
                s5 = a1
```

Which gives out a result of:

	s1	s2	s3	s4	s5
V(s)	5.00	4.00	0.00	0.00	1.00
f_pi(s)	a2	a2	a1	a1	a1

The convergence plot for each V(s) value is given below:
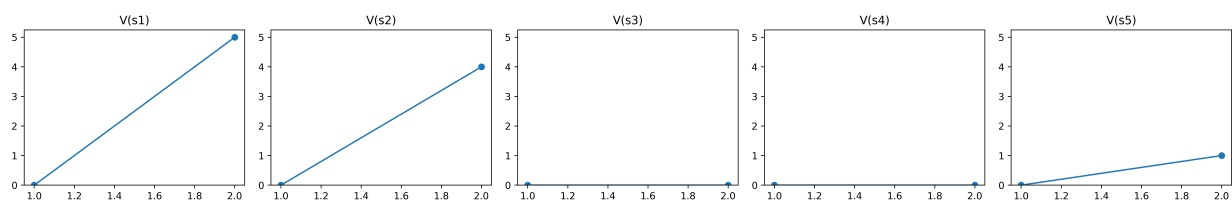
Figure 9: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s,a).

Solution

The Transition Function can therefore be defined as:

$$f_{MT}(s,a) = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{c} a_1 \; a_2 \; a_3 \\ \begin{bmatrix} s_2 & s_3 & s_4 \\ s_2 & s_3 & s_2 \\ s_2 & s_3 & s_5 \\ s_4 & s_3 & s_4 \\ s_4 & s_5 & s_5 \end{bmatrix} \end{array}$$

Figure 10: Image

Thus, the Bellman Optimality Equations were solved in 3 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1  ============================
2  Iteration 1:
3  fr: [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
      0]], [[-2, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
       0]], [[0, 5, 0], [0, 4, 0], [0, 0, 0], [0, -1, 0], [0,
      0, 0]], [[0, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0], [1,
       0, 0]], [[0, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0, 0],
      [0, 0, 0]]]
4  s1       s2        s3        s4        s5
5  Q(s,a) current
6  0.00     0.00      0.00      0.00      0.00
7  0.00     0.00      0.00      0.00      0.00
8  0.00     0.00      0.00      0.00      0.00
9  Q(s,a) new
10 -2.00    0.00      0.00      0.00      1.00
11 5.00     4.00      0.00      -1.00     0.90
12 -3.00    3.60      -5.10     0.00      0.90
13 ============================
14 Iteration 2:
15 fr: [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
      0]], [[-2, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
       0]], [[0, 5, 0], [0, 4, 0], [0, 0, 0], [0, -1, 0], [0,
      0, 0]], [[0, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0], [1,
       0, 0]], [[0, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0, 0],
      [0, 0, 0]]]
16 s1       s2        s3        s4        s5
17 Q(s,a) current
18 -2.00    0.00      0.00      0.00      1.00
19 5.00     4.00      0.00      -1.00     0.90
20 -3.00    3.60      -5.10     0.00      0.90
21 Q(s,a) new
22 1.60     3.60      0.00      0.00      1.00
23 5.00     4.00      0.00      -1.00     0.90
24 -3.00    3.60      -5.10     0.00      0.90
25 ============================
26 Iteration 3:
27 fr: [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
      0]], [[-2, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0,
       0]], [[0, 5, 0], [0, 4, 0], [0, 0, 0], [0, -1, 0], [0,
```

```
         0, 0]], [[0, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0], [1,
          0, 0]], [[0, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0, 0],
         [0, 0, 0]]]
28  s1        s2        s3        s4        s5
29  Q(s,a)  current
30  1.60     3.60      0.00      0.00      1.00
31  5.00     4.00      0.00      -1.00     0.90
32  -3.00    3.60      -5.10     0.00      0.90
33  Q(s,a)  new
34  1.60     3.60      0.00      0.00      1.00
35  5.00     4.00      0.00      -1.00     0.90
36  -3.00    3.60      -5.10     0.00      0.90
37  Optimal Politic:
38  s1 = a2,          s2 = a2,            s3 = a1,           s4 = a1,
                s5 = a1,
```

Which gives out a result of:

	s1	s2	s3	s4	s5
Q(s, a1)	1.60	3.60	**0.00**	**0.00**	**1.00**
Q(s, a2)	**5.00**	**4.00**	0.00	-1.00	0.90
Q(s, a3)	-3.00	3.60	-5.10	0.00	0.90
f_pi(s)	a2	a2	a1	a1	a1

The convergence plot for each Q(s,a) value is given below:
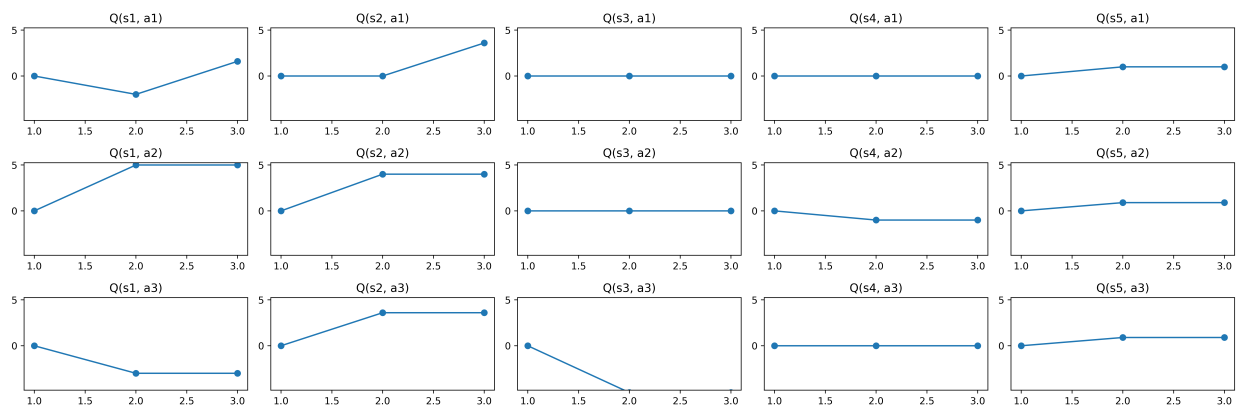
Figure 11: Image

4. Given the world defined by the graph and with the following Reward Function `fr(s,a, sf)` and `gamma = 0.7`:

$$f_R(s,a,s_f) = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \begin{array}{ccc} a_1 & a_2 & a_3 \\ 9 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \\ 2 & 0 & 0 \end{array} \quad \begin{array}{ccc} a_1 & a_2 & a_3 \\ -2 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{ccc} a_1 & a_2 & a_3 \\ 0 & 5 & 0 \\ 0 & 4 & -1 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{ccc} a_1 & a_2 & a_3 \\ 1 & 0 & -3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{array} \quad \begin{array}{ccc} a_1 & a_2 & a_3 \\ 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}$$

with headers $s_f = s_1$, $s_f = s_2$, $s_f = s_3$, $s_f = s_4$, $s_f = s_5$.

Figure 12: Image

a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

The Transition Model would be then:

Figure 13: Image

Thus, the Bellman Optimality Equations were solved in 2 iterations if we assume that
a change smaller than 0.01 between iterations means convergence.

```
1  ============================
2  Iteration 1:
3  fr: [[[9, 0, 0], [0, 0, 0], [0, 1, 0], [0, 0, -2], [2, 0,
       0]], [[-2, 2, 0], [0, 0, 0], [0, 0, 0], [0, -3, 0], [0,
       0, 0]], [[0, 5, 0], [0, 4, -1], [0, 0, 0], [0, -1, 0],
       [0, 0, 0]], [[1, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0],
       [1, -1, 0]], [[3, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0,
       0], [0, 0, 0]]]
4  s1       s2        s3       s4        s5
```

```
5  V(s) current
6  0.00      0.00      0.00      0.00      0.00
7  V(s) new
8  2.60      0.00      0.00      0.00      0.00
9  ===========================
10 Iteration 2:
11 fr: [[[9, 0, 0], [0, 0, 0], [0, 1, 0], [0, 0, -2], [2, 0,
       0]], [[-2, 2, 0], [0, 0, 0], [0, 0, 0], [0, -3, 0], [0,
       0, 0]], [[0, 5, 0], [0, 4, -1], [0, 0, 0], [0, -1, 0],
       [0, 0, 0]], [[1, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0],
        [1, -1, 0]], [[3, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0,
       0], [0, 0, 0]]]
12 s1       s2       s3       s4       s5
13 V(s) current
14 2.60      0.00      0.00      0.00      0.00
15 V(s) new
16 2.60      0.00      0.00      0.00      0.00
17 Optimal Politic:
18 s1 = a1,        s2 = a1,        s3 = a1,        s4 = a1,
            s5 = a1
```

Which gives out a result of:

	s1	s2	s3	s4	s5
V(s)	2.60	0.00	0.00	0.00	0.00
f_pi(s)	a1	a1	a1	a1	a1

The convergence plot for each V(s) value is given below:

Figure 14: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s,a).

Solution

The Transition Function can therefore be defined as:

Figure 15: Image

Thus, the Bellman Optimality Equations were solved in 2 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1  ===========================
2  Iteration 1:
3  fr: [[[9, 0, 0], [0, 0, 0], [0, 1, 0], [0, 0, -2], [2, 0,
       0]], [[-2, 2, 0], [0, 0, 0], [0, 0, 0], [0, -3, 0], [0,
       0, 0]], [[0, 5, 0], [0, 4, -1], [0, 0, 0], [0, -1, 0],
       [0, 0, 0]], [[1, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0],
        [1, -1, 0]], [[3, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0,
       0], [0, 0, 0]]]
4  s1       s2       s3       s4       s5
5  Q(s,a) current
6  0.00     0.00     0.00     0.00     0.00
7  0.00     0.00     0.00     0.00     0.00
8  0.00     0.00     0.00     0.00     0.00
9  Q(s,a) new
10 2.60     0.00     0.00     0.00     0.00
11 2.00     0.00     0.00     -1.40    -1.00
12 1.82     -0.40    0.00     0.00     0.00
13 ===========================
14 Iteration 2:
15 fr: [[[9, 0, 0], [0, 0, 0], [0, 1, 0], [0, 0, -2], [2, 0,
       0]], [[-2, 2, 0], [0, 0, 0], [0, 0, 0], [0, -3, 0], [0,
       0, 0]], [[0, 5, 0], [0, 4, -1], [0, 0, 0], [0, -1, 0],
       [0, 0, 0]], [[1, 0, -3], [0, 0, 0], [0, 0, 0], [0, 0, 0],
        [1, -1, 0]], [[3, 0, 0], [0, 0, 0], [0, 0, -6], [0, 0,
       0], [0, 0, 0]]]
16 s1       s2       s3       s4       s5
17 Q(s,a) current
18 2.60     0.00     0.00     0.00     0.00
19 2.00     0.00     0.00     -1.40    -1.00
20 1.82     -0.40    0.00     0.00     0.00
21 Q(s,a) new
22 2.60     0.00     0.00     0.00     0.00
23 2.00     0.00     0.00     -1.40    -1.00
24 1.82     -0.40    0.00     0.00     0.00
25 Optimal Politic:
26 s1 = a1,           s2 = a1,           s3 = a1,           s4 = a1,
             s5 = a1,
```

Which gives out a result of:

	s1	s2	s3	s4	s5
Q(s, a1)	**2.60**	**0.00**	**0.00**	**0.00**	**0.00**
Q(s, a2)	2.00	0.00	0.00	-1.40	-1.00
Q(s, a3)	1.82	-0.40	0.00	0.00	0.00
f_pi(s)	a1	a1	a1	a1	a1

The convergence plot for each Q(s,a) value is given below:
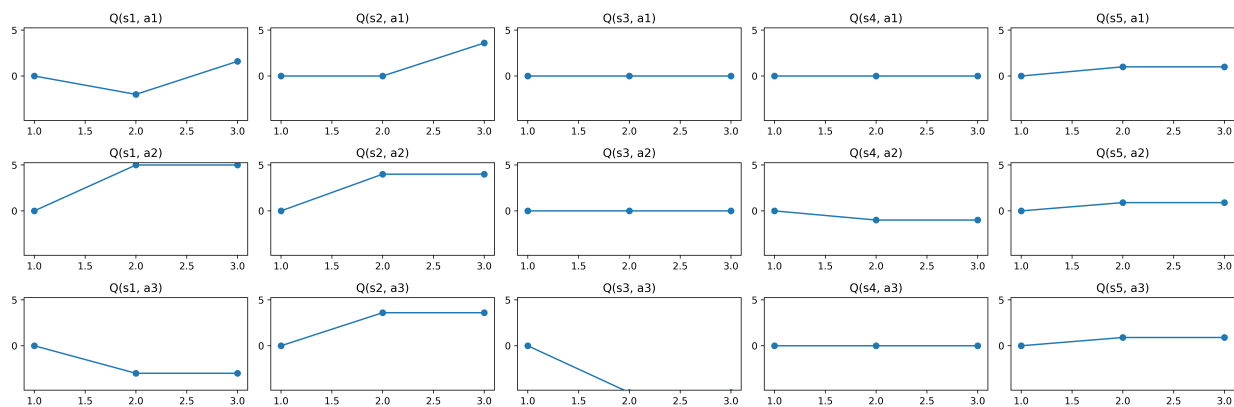
Figure 16: Image

5. The world has the states set S = {sf1, s1, s2, s3, sf2} where s1 = initial state and, sf1 and sf2 are terminal states:
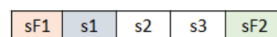
Figure 17: Image

The world has the set of actions A = {->, <-} where:

- -> = agent moves to the right one cell

- <- = agent moves to the left one cell

The Reward Function `fr(s,a,sf)= fr(sf)` only depends on the state to which the agent arrives:

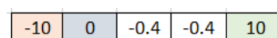

Figure 18: Image

Note: use gamma = 0.8.

Do the following:

a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

The Transition Function would be:

$$f_{MT}(s,a) = \begin{array}{c} s_{F_1} \\ s_1 \\ s_2 \\ s_3 \\ s_{F_2} \end{array} \begin{array}{c} \rightarrow \quad \leftarrow \\ \left[\begin{array}{cc} s_1 & s_{F_1} \\ s_2 & s_{F_1} \\ s_3 & s_1 \\ s_{F_2} & s_2 \\ s_{F_2} & s_3 \end{array} \right] \end{array}$$

Figure 19: Image

Thus, the Bellman Optimality Equations were solved in 32 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1   ==========================
2   Iteration 1:
3   fr: [-10, 0, -0.4, -0.4, 10]
4   sf1     s1      s2      s3      sf2
5   V(s) current
6   0.00    0.00    0.00    0.00    0.00
7   V(s) new
8   0.00    -0.40   -0.32   10.00   10.00
9   ==========================
10  Iteration 2:
11  fr: [-10, 0, -0.4, -0.4, 10]
12  sf1     s1      s2      s3      sf2
13  V(s) current
14  0.00    -0.40   -0.32   10.00   10.00
15  V(s) new
16  -0.32   -0.66   7.60    18.00   18.00
17  ==========================
18  Iteration 3:
19  fr: [-10, 0, -0.4, -0.4, 10]
20  sf1     s1      s2      s3      sf2
21  V(s) current
22  -0.32   -0.66   7.60    18.00   18.00
23  V(s) new
24  -0.52   5.68    14.00   24.40   24.40
25  ==========================
26  Iteration 4:
27  fr: [-10, 0, -0.4, -0.4, 10]
28  sf1     s1      s2      s3      sf2
29  V(s) current
30  -0.52   5.68    14.00   24.40   24.40
31  V(s) new
32  4.54    10.80   19.12   29.52   29.52
```

```
33  =========================
34  Iteration 5:
35  fr: [-10, 0, -0.4, -0.4, 10]
36  sf1      s1       s2       s3       sf2
37  V(s) current
38  4.54     10.80    19.12    29.52    29.52
39  V(s) new
40  8.64     14.90    23.22    33.62    33.62
41  =========================
42  Iteration 6:
43  fr: [-10, 0, -0.4, -0.4, 10]
44  sf1      s1       s2       s3       sf2
45  V(s) current
46  8.64     14.90    23.22    33.62    33.62
47  V(s) new
48  11.92    18.17    26.49    36.89    36.89
49  =========================
50  Iteration 7:
51  fr: [-10, 0, -0.4, -0.4, 10]
52  sf1      s1       s2       s3       sf2
53  V(s) current
54  11.92    18.17    26.49    36.89    36.89
55  V(s) new
56  14.54    20.79    29.11    39.51    39.51
57  =========================
58  Iteration 8:
59  fr: [-10, 0, -0.4, -0.4, 10]
60  sf1      s1       s2       s3       sf2
61  V(s) current
62  14.54    20.79    29.11    39.51    39.51
63  V(s) new
64  16.64    22.89    31.21    41.61    41.61
65  =========================
66  Iteration 9:
67  fr: [-10, 0, -0.4, -0.4, 10]
68  sf1      s1       s2       s3       sf2
69  V(s) current
70  16.64    22.89    31.21    41.61    41.61
71  V(s) new
72  18.31    24.57    32.89    43.29    43.29
73  =========================
74  Iteration 10:
75  fr: [-10, 0, -0.4, -0.4, 10]
76  sf1      s1       s2       s3       sf2
77  V(s) current
78  18.31    24.57    32.89    43.29    43.29
79  V(s) new
80  19.66    25.91    34.23    44.63    44.63
81  =========================
82  Iteration 11:
83  fr: [-10, 0, -0.4, -0.4, 10]
84  sf1      s1       s2       s3       sf2
```

```
 85   V(s) current
 86   19.66    25.91    34.23    44.63    44.63
 87   V(s) new
 88   20.73    26.99    35.31    45.71    45.71
 89   =========================
 90   Iteration 12:
 91   fr: [-10, 0, -0.4, -0.4, 10]
 92   sf1      s1       s2       s3       sf2
 93   V(s) current
 94   20.73    26.99    35.31    45.71    45.71
 95   V(s) new
 96   21.59    27.84    36.16    46.56    46.56
 97   =========================
 98   Iteration 13:
 99   fr: [-10, 0, -0.4, -0.4, 10]
100   sf1      s1       s2       s3       sf2
101   V(s) current
102   21.59    27.84    36.16    46.56    46.56
103   V(s) new
104   22.28    28.53    36.85    47.25    47.25
105   =========================
106   Iteration 14:
107   fr: [-10, 0, -0.4, -0.4, 10]
108   sf1      s1       s2       s3       sf2
109   V(s) current
110   22.28    28.53    36.85    47.25    47.25
111   V(s) new
112   22.82    29.08    37.40    47.80    47.80
113   =========================
114   Iteration 15:
115   fr: [-10, 0, -0.4, -0.4, 10]
116   sf1      s1       s2       s3       sf2
117   V(s) current
118   22.82    29.08    37.40    47.80    47.80
119   V(s) new
120   23.26    29.52    37.84    48.24    48.24
121   =========================
122   Iteration 16:
123   fr: [-10, 0, -0.4, -0.4, 10]
124   sf1      s1       s2       s3       sf2
125   V(s) current
126   23.26    29.52    37.84    48.24    48.24
127   V(s) new
128   23.62    29.87    38.19    48.59    48.59
129   =========================
130   Iteration 17:
131   fr: [-10, 0, -0.4, -0.4, 10]
132   sf1      s1       s2       s3       sf2
133   V(s) current
134   23.62    29.87    38.19    48.59    48.59
135   V(s) new
136   23.90    30.15    38.47    48.87    48.87
```

```
137  =========================
138  Iteration 18:
139  fr: [-10, 0, -0.4, -0.4, 10]
140  sf1      s1       s2       s3       sf2
141  V(s) current
142  23.90   30.15   38.47   48.87   48.87
143  V(s) new
144  24.12   30.38   38.70   49.10   49.10
145  =========================
146  Iteration 19:
147  fr: [-10, 0, -0.4, -0.4, 10]
148  sf1      s1       s2       s3       sf2
149  V(s) current
150  24.12   30.38   38.70   49.10   49.10
151  V(s) new
152  24.30   30.56   38.88   49.28   49.28
153  =========================
154  Iteration 20:
155  fr: [-10, 0, -0.4, -0.4, 10]
156  sf1      s1       s2       s3       sf2
157  V(s) current
158  24.30   30.56   38.88   49.28   49.28
159  V(s) new
160  24.45   30.70   39.02   49.42   49.42
161  =========================
162  Iteration 21:
163  fr: [-10, 0, -0.4, -0.4, 10]
164  sf1      s1       s2       s3       sf2
165  V(s) current
166  24.45   30.70   39.02   49.42   49.42
167  V(s) new
168  24.56   30.82   39.14   49.54   49.54
169  =========================
170  Iteration 22:
171  fr: [-10, 0, -0.4, -0.4, 10]
172  sf1      s1       s2       s3       sf2
173  V(s) current
174  24.56   30.82   39.14   49.54   49.54
175  V(s) new
176  24.66   30.91   39.23   49.63   49.63
177  =========================
178  Iteration 23:
179  fr: [-10, 0, -0.4, -0.4, 10]
180  sf1      s1       s2       s3       sf2
181  V(s) current
182  24.66   30.91   39.23   49.63   49.63
183  V(s) new
184  24.73   30.98   39.30   49.70   49.70
185  =========================
186  Iteration 24:
187  fr: [-10, 0, -0.4, -0.4, 10]
188  sf1      s1       s2       s3       sf2
```

```
189  V(s) current
190  24.73    30.98    39.30    49.70    49.70
191  V(s) new
192  24.79    31.04    39.36    49.76    49.76
193  =========================
194  Iteration 25:
195  fr: [-10, 0, -0.4, -0.4, 10]
196  sf1      s1       s2       s3       sf2
197  V(s) current
198  24.79    31.04    39.36    49.76    49.76
199  V(s) new
200  24.84    31.09    39.41    49.81    49.81
201  =========================
202  Iteration 26:
203  fr: [-10, 0, -0.4, -0.4, 10]
204  sf1      s1       s2       s3       sf2
205  V(s) current
206  24.84    31.09    39.41    49.81    49.81
207  V(s) new
208  24.87    31.13    39.45    49.85    49.85
209  =========================
210  Iteration 27:
211  fr: [-10, 0, -0.4, -0.4, 10]
212  sf1      s1       s2       s3       sf2
213  V(s) current
214  24.87    31.13    39.45    49.85    49.85
215  V(s) new
216  24.90    31.16    39.48    49.88    49.88
217  =========================
218  Iteration 28:
219  fr: [-10, 0, -0.4, -0.4, 10]
220  sf1      s1       s2       s3       sf2
221  V(s) current
222  24.90    31.16    39.48    49.88    49.88
223  V(s) new
224  24.93    31.18    39.50    49.90    49.90
225  =========================
226  Iteration 29:
227  fr: [-10, 0, -0.4, -0.4, 10]
228  sf1      s1       s2       s3       sf2
229  V(s) current
230  24.93    31.18    39.50    49.90    49.90
231  V(s) new
232  24.95    31.20    39.52    49.92    49.92
233  =========================
234  Iteration 30:
235  fr: [-10, 0, -0.4, -0.4, 10]
236  sf1      s1       s2       s3       sf2
237  V(s) current
238  24.95    31.20    39.52    49.92    49.92
239  V(s) new
240  24.96    31.22    39.54    49.94    49.94
```

```
241  ===========================
242  Iteration 31:
243  fr: [-10, 0, -0.4, -0.4, 10]
244  sf1      s1       s2       s3       sf2
245  V(s) current
246  24.96   31.22   39.54   49.94   49.94
247  V(s) new
248  24.97   31.23   39.55   49.95   49.95
249  ===========================
250  Iteration 32:
251  fr: [-10, 0, -0.4, -0.4, 10]
252  sf1      s1       s2       s3       sf2
253  V(s) current
254  24.97   31.23   39.55   49.95   49.95
255  V(s) new
256  24.98   31.24   39.56   49.96   49.96
257  Optimal Politic:
258  sf1 = ->,        s1 = ->,            s2 = ->,          s3 = ->,
              sf2 = ->
```

Which gives out a result of:

	sf1	s1	s2	s3	sf2
V(s)	24.98	31.24	39.56	49.96	49.96
f_pi(s)	->	->	->	->	->

The convergence plot for each V(s) value is given below:
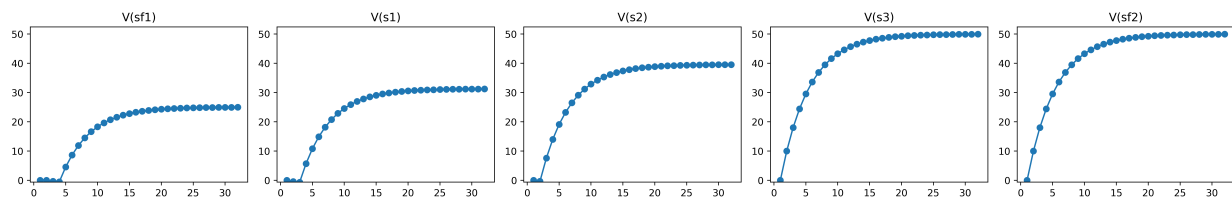
Figure 20: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s,a).

Solution

The Transition Function would be:

$$f_{MT}(s,a) = \begin{array}{c} sf_1 \\ s_1 \\ s_2 \\ s_3 \\ sf_2 \end{array} \begin{array}{cc} \rightarrow & \leftarrow \\ \left[\begin{array}{cc} s_1 & sf_1 \\ s_2 & sf_1 \\ s_3 & s_1 \\ sf_2 & s_2 \\ sf_2 & s_3 \end{array}\right] \end{array}$$

Figure 21: Image

Thus, the Bellman Optimality Equations were solved in 32 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
 1  ==========================
 2  Iteration 1:
 3  fr: [-10, 0, -0.4, -0.4, 10]
 4  sf1      s1       s2        s3        sf2
 5  Q(s,a) current
 6  0.00     0.00     0.00      0.00      0.00
 7  0.00     0.00     0.00      0.00      0.00
 8  Q(s,a) new
 9  0.00     -0.40    -0.40     10.00     10.00
10  -10.00   -10.00   -0.32     -0.66     7.60
11  ==========================
12  Iteration 2:
13  fr: [-10, 0, -0.4, -0.4, 10]
14  sf1      s1       s2        s3        sf2
15  Q(s,a) current
16  0.00     -0.40    -0.40     10.00     10.00
17  -10.00   -10.00   -0.32     -0.66     7.60
18  Q(s,a) new
19  -0.32    -0.66    7.60      18.00     18.00
20  -10.26   -10.26   -0.52     5.68      14.00
21  ==========================
22  Iteration 3:
23  fr: [-10, 0, -0.4, -0.4, 10]
24  sf1      s1       s2        s3        sf2
25  Q(s,a) current
26  -0.32    -0.66    7.60      18.00     18.00
27  -10.26   -10.26   -0.52     5.68      14.00
28  Q(s,a) new
29  -0.52    5.68     14.00     24.40     24.40
30  -10.42   -10.42   4.54      10.80     19.12
31  ==========================
32  Iteration 4:
33  fr: [-10, 0, -0.4, -0.4, 10]
34  sf1      s1       s2        s3        sf2
35  Q(s,a) current
36  -0.52    5.68     14.00     24.40     24.40
37  -10.42   -10.42   4.54      10.80     19.12
38  Q(s,a) new
39  4.54     10.80    19.12     29.52     29.52
```

```
40  -6.36    -6.36    8.64     14.90    23.22
41  ==========================
42  Iteration 5:
43  fr: [-10, 0, -0.4, -0.4, 10]
44  sf1      s1       s2       s3       sf2
45  Q(s,a) current
46  4.54     10.80    19.12    29.52    29.52
47  -6.36    -6.36    8.64     14.90    23.22
48  Q(s,a) new
49  8.64     14.90    23.22    33.62    33.62
50  -3.09    -3.09    11.92    18.17    26.49
51  ==========================
52  Iteration 6:
53  fr: [-10, 0, -0.4, -0.4, 10]
54  sf1      s1       s2       s3       sf2
55  Q(s,a) current
56  8.64     14.90    23.22    33.62    33.62
57  -3.09    -3.09    11.92    18.17    26.49
58  Q(s,a) new
59  11.92    18.17    26.49    36.89    36.89
60  -0.47    -0.47    14.54    20.79    29.11
61  ==========================
62  Iteration 7:
63  fr: [-10, 0, -0.4, -0.4, 10]
64  sf1      s1       s2       s3       sf2
65  Q(s,a) current
66  11.92    18.17    26.49    36.89    36.89
67  -0.47    -0.47    14.54    20.79    29.11
68  Q(s,a) new
69  14.54    20.79    29.11    39.51    39.51
70  1.63     1.63     16.64    22.89    31.21
71  ==========================
72  Iteration 8:
73  fr: [-10, 0, -0.4, -0.4, 10]
74  sf1      s1       s2       s3       sf2
75  Q(s,a) current
76  14.54    20.79    29.11    39.51    39.51
77  1.63     1.63     16.64    22.89    31.21
78  Q(s,a) new
79  16.64    22.89    31.21    41.61    41.61
80  3.31     3.31     18.31    24.57    32.89
81  ==========================
82  Iteration 9:
83  fr: [-10, 0, -0.4, -0.4, 10]
84  sf1      s1       s2       s3       sf2
85  Q(s,a) current
86  16.64    22.89    31.21    41.61    41.61
87  3.31     3.31     18.31    24.57    32.89
88  Q(s,a) new
89  18.31    24.57    32.89    43.29    43.29
90  4.65     4.65     19.66    25.91    34.23
91  ==========================
```

```
 92  Iteration 10:
 93  fr: [-10, 0, -0.4, -0.4, 10]
 94  sf1       s1        s2        s3        sf2
 95  Q(s,a) current
 96  18.31    24.57    32.89    43.29    43.29
 97  4.65     4.65     19.66    25.91    34.23
 98  Q(s,a) new
 99  19.66    25.91    34.23    44.63    44.63
100  5.72     5.72     20.73    26.99    35.31
101  =========================
102  Iteration 11:
103  fr: [-10, 0, -0.4, -0.4, 10]
104  sf1       s1        s2        s3        sf2
105  Q(s,a) current
106  19.66    25.91    34.23    44.63    44.63
107  5.72     5.72     20.73    26.99    35.31
108  Q(s,a) new
109  20.73    26.99    35.31    45.71    45.71
110  6.58     6.58     21.59    27.84    36.16
111  =========================
112  Iteration 12:
113  fr: [-10, 0, -0.4, -0.4, 10]
114  sf1       s1        s2        s3        sf2
115  Q(s,a) current
116  20.73    26.99    35.31    45.71    45.71
117  6.58     6.58     21.59    27.84    36.16
118  Q(s,a) new
119  21.59    27.84    36.16    46.56    46.56
120  7.27     7.27     22.28    28.53    36.85
121  =========================
122  Iteration 13:
123  fr: [-10, 0, -0.4, -0.4, 10]
124  sf1       s1        s2        s3        sf2
125  Q(s,a) current
126  21.59    27.84    36.16    46.56    46.56
127  7.27     7.27     22.28    28.53    36.85
128  Q(s,a) new
129  22.28    28.53    36.85    47.25    47.25
130  7.82     7.82     22.82    29.08    37.40
131  =========================
132  Iteration 14:
133  fr: [-10, 0, -0.4, -0.4, 10]
134  sf1       s1        s2        s3        sf2
135  Q(s,a) current
136  22.28    28.53    36.85    47.25    47.25
137  7.82     7.82     22.82    29.08    37.40
138  Q(s,a) new
139  22.82    29.08    37.40    47.80    47.80
140  8.26     8.26     23.26    29.52    37.84
141  =========================
142  Iteration 15:
143  fr: [-10, 0, -0.4, -0.4, 10]
```

```
144  sf1      s1       s2       s3       sf2
145  Q(s,a) current
146  22.82   29.08    37.40    47.80    47.80
147  8.26     8.26    23.26    29.52    37.84
148  Q(s,a) new
149  23.26   29.52    37.84    48.24    48.24
150  8.61     8.61    23.62    29.87    38.19
151  =========================
152  Iteration 16:
153  fr: [-10, 0, -0.4, -0.4, 10]
154  sf1      s1       s2       s3       sf2
155  Q(s,a) current
156  23.26   29.52    37.84    48.24    48.24
157  8.61     8.61    23.62    29.87    38.19
158  Q(s,a) new
159  23.62   29.87    38.19    48.59    48.59
160  8.89     8.89    23.90    30.15    38.47
161  =========================
162  Iteration 17:
163  fr: [-10, 0, -0.4, -0.4, 10]
164  sf1      s1       s2       s3       sf2
165  Q(s,a) current
166  23.62   29.87    38.19    48.59    48.59
167  8.89     8.89    23.90    30.15    38.47
168  Q(s,a) new
169  23.90   30.15    38.47    48.87    48.87
170  9.12     9.12    24.12    30.38    38.70
171  =========================
172  Iteration 18:
173  fr: [-10, 0, -0.4, -0.4, 10]
174  sf1      s1       s2       s3       sf2
175  Q(s,a) current
176  23.90   30.15    38.47    48.87    48.87
177  9.12     9.12    24.12    30.38    38.70
178  Q(s,a) new
179  24.12   30.38    38.70    49.10    49.10
180  9.30     9.30    24.30    30.56    38.88
181  =========================
182  Iteration 19:
183  fr: [-10, 0, -0.4, -0.4, 10]
184  sf1      s1       s2       s3       sf2
185  Q(s,a) current
186  24.12   30.38    38.70    49.10    49.10
187  9.30     9.30    24.30    30.56    38.88
188  Q(s,a) new
189  24.30   30.56    38.88    49.28    49.28
190  9.44     9.44    24.45    30.70    39.02
191  =========================
192  Iteration 20:
193  fr: [-10, 0, -0.4, -0.4, 10]
194  sf1      s1       s2       s3       sf2
195  Q(s,a) current
```

```
196  24.30     30.56     38.88     49.28     49.28
197  9.44      9.44      24.45     30.70     39.02
198  Q(s,a) new
199  24.45     30.70     39.02     49.42     49.42
200  9.56      9.56      24.56     30.82     39.14
201  =========================
202  Iteration 21:
203  fr: [-10, 0, -0.4, -0.4, 10]
204  sf1       s1        s2        s3        sf2
205  Q(s,a) current
206  24.45     30.70     39.02     49.42     49.42
207  9.56      9.56      24.56     30.82     39.14
208  Q(s,a) new
209  24.56     30.82     39.14     49.54     49.54
210  9.65      9.65      24.66     30.91     39.23
211  =========================
212  Iteration 22:
213  fr: [-10, 0, -0.4, -0.4, 10]
214  sf1       s1        s2        s3        sf2
215  Q(s,a) current
216  24.56     30.82     39.14     49.54     49.54
217  9.65      9.65      24.66     30.91     39.23
218  Q(s,a) new
219  24.66     30.91     39.23     49.63     49.63
220  9.72      9.72      24.73     30.98     39.30
221  =========================
222  Iteration 23:
223  fr: [-10, 0, -0.4, -0.4, 10]
224  sf1       s1        s2        s3        sf2
225  Q(s,a) current
226  24.66     30.91     39.23     49.63     49.63
227  9.72      9.72      24.73     30.98     39.30
228  Q(s,a) new
229  24.73     30.98     39.30     49.70     49.70
230  9.78      9.78      24.79     31.04     39.36
231  =========================
232  Iteration 24:
233  fr: [-10, 0, -0.4, -0.4, 10]
234  sf1       s1        s2        s3        sf2
235  Q(s,a) current
236  24.73     30.98     39.30     49.70     49.70
237  9.78      9.78      24.79     31.04     39.36
238  Q(s,a) new
239  24.79     31.04     39.36     49.76     49.76
240  9.83      9.83      24.84     31.09     39.41
241  =========================
242  Iteration 25:
243  fr: [-10, 0, -0.4, -0.4, 10]
244  sf1       s1        s2        s3        sf2
245  Q(s,a) current
246  24.79     31.04     39.36     49.76     49.76
247  9.83      9.83      24.84     31.09     39.41
```

```
248  Q(s,a) new
249  24.84    31.09    39.41    49.81    49.81
250  9.87     9.87     24.87    31.13    39.45
251  =========================
252  Iteration 26:
253  fr: [-10, 0, -0.4, -0.4, 10]
254  sf1      s1       s2       s3       sf2
255  Q(s,a) current
256  24.84    31.09    39.41    49.81    49.81
257  9.87     9.87     24.87    31.13    39.45
258  Q(s,a) new
259  24.87    31.13    39.45    49.85    49.85
260  9.90     9.90     24.90    31.16    39.48
261  =========================
262  Iteration 27:
263  fr: [-10, 0, -0.4, -0.4, 10]
264  sf1      s1       s2       s3       sf2
265  Q(s,a) current
266  24.87    31.13    39.45    49.85    49.85
267  9.90     9.90     24.90    31.16    39.48
268  Q(s,a) new
269  24.90    31.16    39.48    49.88    49.88
270  9.92     9.92     24.93    31.18    39.50
271  =========================
272  Iteration 28:
273  fr: [-10, 0, -0.4, -0.4, 10]
274  sf1      s1       s2       s3       sf2
275  Q(s,a) current
276  24.90    31.16    39.48    49.88    49.88
277  9.92     9.92     24.93    31.18    39.50
278  Q(s,a) new
279  24.93    31.18    39.50    49.90    49.90
280  9.94     9.94     24.95    31.20    39.52
281  =========================
282  Iteration 29:
283  fr: [-10, 0, -0.4, -0.4, 10]
284  sf1      s1       s2       s3       sf2
285  Q(s,a) current
286  24.93    31.18    39.50    49.90    49.90
287  9.94     9.94     24.95    31.20    39.52
288  Q(s,a) new
289  24.95    31.20    39.52    49.92    49.92
290  9.96     9.96     24.96    31.22    39.54
291  =========================
292  Iteration 30:
293  fr: [-10, 0, -0.4, -0.4, 10]
294  sf1      s1       s2       s3       sf2
295  Q(s,a) current
296  24.95    31.20    39.52    49.92    49.92
297  9.96     9.96     24.96    31.22    39.54
298  Q(s,a) new
299  24.96    31.22    39.54    49.94    49.94
```

```
300  9.97      9.97      24.97     31.23     39.55
301  ==========================
302  Iteration 31:
303  fr: [-10, 0, -0.4, -0.4, 10]
304  sf1       s1        s2        s3        sf2
305  Q(s,a) current
306  24.96     31.22     39.54     49.94     49.94
307  9.97      9.97      24.97     31.23     39.55
308  Q(s,a) new
309  24.97     31.23     39.55     49.95     49.95
310  9.98      9.98      24.98     31.24     39.56
311  ==========================
312  Iteration 32:
313  fr: [-10, 0, -0.4, -0.4, 10]
314  sf1       s1        s2        s3        sf2
315  Q(s,a) current
316  24.97     31.23     39.55     49.95     49.95
317  9.98      9.98      24.98     31.24     39.56
318  Q(s,a) new
319  24.98     31.24     39.56     49.96     49.96
320  9.99      9.99      24.99     31.25     39.57
321  Optimal Politic:
322  sf1 = ->,        s1 = ->,           s2 = ->,            s3 = ->,
                  sf2 = ->
```

Which gives out a final result of:

	s1	s2	s3	s4	s5
Q(s, ->)	**24.98**	**31.24**	**39.56**	**49.96**	**49.96**
Q(s, <-)	9.99	9.99	24.99	31.25	39.57
f_pi(s)	->	->	->	->	->

The convergence plot for each Q(s,a) value is given below:
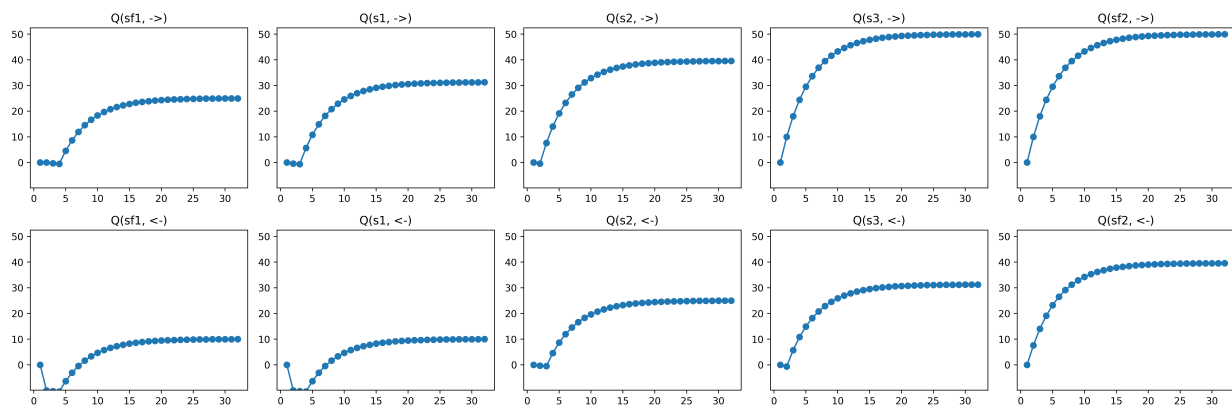
Figure 22: Image

6. The world has the states set S = {sf1, s1, s2, s3, sf2} where s1 = initial state and, sf1 and sf2 are terminal states:
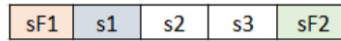

| sF1 | s1 | s2 | s3 | sF2 |

Figure 23: Image

The world has the set of actions A = {->, <-} where:

- -> = agent moves to the right one cell with probability 0.8, and one cell to the left with probability 0.2.

- <- = agent moves to the left one cell with probability 0.8, and one cell to the right with probability 0.2.

The Reward Function `fr(s,a,sf)= fr(sf)` only depends on the state to which the agent arrives:

| -10 | 0 | -0.4 | -0.4 | 10 |
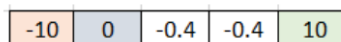
Figure 24: Image

Note: use `gamma` = 0.6.

Do the following:

a. Solve the Bellman Optimality Equations by Value Iteration for V(s).

Solution

The Transition Function would be:

Figure 25: Image

Thus, the Bellman Optimality Equations were solved in 13 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1   ==========================
2   Iteration 1:
3   fr: [-10, 0, -0.4, -0.4, 10]
```

```
 4  sf1      s1       s2       s3       sf2
 5  V(s) current
 6  0.00     0.00     0.00     0.00     0.00
 7  V(s) new
 8  -2.00    -2.56    -0.63    7.84     8.86
 9  ==========================
10  Iteration 2:
11  fr: [-10, 0, -0.4, -0.4, 10]
12  sf1      s1       s2       s3       sf2
13  V(s) current
14  -2.00    -2.56    -0.63    7.84     8.86
15  V(s) new
16  -3.47    -3.04    3.08     12.54    13.68
17  ==========================
18  Iteration 3:
19  fr: [-10, 0, -0.4, -0.4, 10]
20  sf1      s1       s2       s3       sf2
21  V(s) current
22  -3.47    -3.04    3.08     12.54    13.68
23  V(s) new
24  -3.87    -1.31    5.54     15.15    16.30
25  ==========================
26  Iteration 4:
27  fr: [-10, 0, -0.4, -0.4, 10]
28  sf1      s1       s2       s3       sf2
29  V(s) current
30  -3.87    -1.31    5.54     15.15    16.30
31  V(s) new
32  -3.09    -0.03    6.95     16.58    17.74
33  ==========================
34  Iteration 5:
35  fr: [-10, 0, -0.4, -0.4, 10]
36  sf1      s1       s2       s3       sf2
37  V(s) current
38  -3.09    -0.03    6.95     16.58    17.74
39  V(s) new
40  -2.39    0.73     7.73     17.36    18.52
41  ==========================
42  Iteration 6:
43  fr: [-10, 0, -0.4, -0.4, 10]
44  sf1      s1       s2       s3       sf2
45  V(s) current
46  -2.39    0.73     7.73     17.36    18.52
47  V(s) new
48  -1.94    1.16     8.15     17.79    18.94
49  ==========================
50  Iteration 7:
51  fr: [-10, 0, -0.4, -0.4, 10]
52  sf1      s1       s2       s3       sf2
53  V(s) current
54  -1.94    1.16     8.15     17.79    18.94
55  V(s) new
```

```
56  -1.68    1.39     8.38     18.02     19.17
57  =========================
58  Iteration 8:
59  fr: [-10, 0, -0.4, -0.4, 10]
60  sf1      s1       s2       s3       sf2
61  V(s) current
62  -1.68    1.39     8.38     18.02     19.17
63  V(s) new
64  -1.53    1.52     8.51     18.15     19.30
65  =========================
66  Iteration 9:
67  fr: [-10, 0, -0.4, -0.4, 10]
68  sf1      s1       s2       s3       sf2
69  V(s) current
70  -1.53    1.52     8.51     18.15     19.30
71  V(s) new
72  -1.45    1.59     8.58     18.21     19.37
73  =========================
74  Iteration 10:
75  fr: [-10, 0, -0.4, -0.4, 10]
76  sf1      s1       s2       s3       sf2
77  V(s) current
78  -1.45    1.59     8.58     18.21     19.37
79  V(s) new
80  -1.41    1.63     8.62     18.25     19.41
81  =========================
82  Iteration 11:
83  fr: [-10, 0, -0.4, -0.4, 10]
84  sf1      s1       s2       s3       sf2
85  V(s) current
86  -1.41    1.63     8.62     18.25     19.41
87  V(s) new
88  -1.39    1.65     8.64     18.27     19.43
89  =========================
90  Iteration 12:
91  fr: [-10, 0, -0.4, -0.4, 10]
92  sf1      s1       s2       s3       sf2
93  V(s) current
94  -1.39    1.65     8.64     18.27     19.43
95  V(s) new
96  -1.37    1.66     8.65     18.28     19.44
97  =========================
98  Iteration 13:
99  fr: [-10, 0, -0.4, -0.4, 10]
100 sf1      s1       s2       s3       sf2
101 V(s) current
102 -1.37    1.66     8.65     18.28     19.44
103 V(s) new
104 -1.37    1.67     8.66     18.29     19.45
105 Optimal Politic:
106 sf1 = ->,        s1 = ->,         s2 = ->,          s3 = ->,
              sf2 = ->
```

Which gives out a result of:

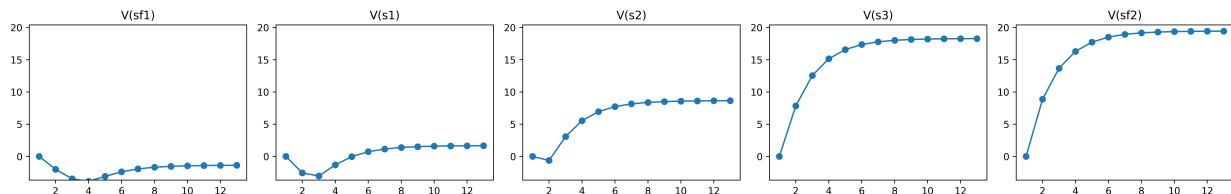|	sf1	s1	s2	s3	sf2
V(s)	-1.37	1.67	8.66	18.29	19.45
f_pi(s)	->	->	->	->	->

The convergence plot for each V(s) value is given below:

Figure 26: Image

b. Solve the Bellman Optimality Equations by Value Iteration for Q(s,a).

Solution

The Transition Function would be:

Figure 27: Image

Thus, the Bellman Optimality Equations were solved in 13 iterations if we assume that **a change smaller than 0.01 between iterations** means convergence.

```
1   ===========================
2   Iteration 1:
3   fr: [-10, 0, -0.4, -0.4, 10]
4   sf1      s1       s2       s3       sf2
5   Q(s,a)  current
6   0.00     0.00     0.00     0.00     0.00
7   0.00     0.00     0.00     0.00     0.00
8   Q(s,a)  new
9   -2.00    -2.32    -0.32    7.92     8.87
10  -8.00    -9.04    -0.24    2.63     6.55
11  ===========================
12  Iteration 2:
13  fr: [-10, 0, -0.4, -0.4, 10]
```

```
14  sf1      s1       s2       s3       sf2
15  Q(s,a) current
16  -2.00    -2.32    -0.32    7.92     8.87
17  -8.00    -9.04    -0.24    2.63     6.55
18  Q(s,a) new
19  -3.35    -2.84    3.14     12.55    13.68
20  -9.95    -9.31    0.06     4.83     9.35
21  ==========================
22  Iteration 3:
23  fr: [-10, 0, -0.4, -0.4, 10]
24  sf1      s1       s2       s3       sf2
25  Q(s,a) current
26  -3.35    -2.84    3.14     12.55    13.68
27  -9.95    -9.31    0.06     4.83     9.35
28  Q(s,a) new
29  -3.77    -1.26    5.55     15.16    16.31
30  -9.96    -9.22    1.13     6.30     10.91
31  ==========================
32  Iteration 4:
33  fr: [-10, 0, -0.4, -0.4, 10]
34  sf1      s1       s2       s3       sf2
35  Q(s,a) current
36  -3.77    -1.26    5.55     15.16    16.31
37  -9.96    -9.22    1.13     6.30     10.91
38  Q(s,a) new
39  -3.06    -0.02    6.95     16.58    17.74
40  -9.47    -8.71    1.90     7.15     11.77
41  ==========================
42  Iteration 5:
43  fr: [-10, 0, -0.4, -0.4, 10]
44  sf1      s1       s2       s3       sf2
45  Q(s,a) current
46  -3.06    -0.02    6.95     16.58    17.74
47  -9.47    -8.71    1.90     7.15     11.77
48  Q(s,a) new
49  -2.38    0.73     7.73     17.36    18.52
50  -9.05    -8.29    2.35     7.61     12.24
51  ==========================
52  Iteration 6:
53  fr: [-10, 0, -0.4, -0.4, 10]
54  sf1      s1       s2       s3       sf2
55  Q(s,a) current
56  -2.38    0.73     7.73     17.36    18.52
57  -9.05    -8.29    2.35     7.61     12.24
58  Q(s,a) new
59  -1.93    1.16     8.15     17.79    18.94
60  -8.79    -8.03    2.61     7.87     12.49
61  ==========================
62  Iteration 7:
63  fr: [-10, 0, -0.4, -0.4, 10]
64  sf1      s1       s2       s3       sf2
65  Q(s,a) current
```

```
66  -1.93    1.16     8.15      17.79     18.94
67  -8.79    -8.03    2.61      7.87      12.49
68  Q(s,a) new
69  -1.68    1.39     8.38      18.02     19.17
70  -8.64    -7.88    2.75      8.01      12.63
71  =========================
72  Iteration 8:
73  fr: [-10, 0, -0.4, -0.4, 10]
74  sf1      s1       s2        s3        sf2
75  Q(s,a) current
76  -1.68    1.39     8.38      18.02     19.17
77  -8.64    -7.88    2.75      8.01      12.63
78  Q(s,a) new
79  -1.53    1.52     8.51      18.15     19.30
80  -8.55    -7.79    2.83      8.08      12.71
81  =========================
82  Iteration 9:
83  fr: [-10, 0, -0.4, -0.4, 10]
84  sf1      s1       s2        s3        sf2
85  Q(s,a) current
86  -1.53    1.52     8.51      18.15     19.30
87  -8.55    -7.79    2.83      8.08      12.71
88  Q(s,a) new
89  -1.45    1.59     8.58      18.21     19.37
90  -8.51    -7.75    2.87      8.12      12.75
91  =========================
92  Iteration 10:
93  fr: [-10, 0, -0.4, -0.4, 10]
94  sf1      s1       s2        s3        sf2
95  Q(s,a) current
96  -1.45    1.59     8.58      18.21     19.37
97  -8.51    -7.75    2.87      8.12      12.75
98  Q(s,a) new
99  -1.41    1.63     8.62      18.25     19.41
100 -8.48    -7.72    2.89      8.15      12.77
101 =========================
102 Iteration 11:
103 fr: [-10, 0, -0.4, -0.4, 10]
104 sf1      s1       s2        s3        sf2
105 Q(s,a) current
106 -1.41    1.63     8.62      18.25     19.41
107 -8.48    -7.72    2.89      8.15      12.77
108 Q(s,a) new
109 -1.39    1.65     8.64      18.27     19.43
110 -8.47    -7.71    2.90      8.16      12.78
111 =========================
112 Iteration 12:
113 fr: [-10, 0, -0.4, -0.4, 10]
114 sf1      s1       s2        s3        sf2
115 Q(s,a) current
116 -1.39    1.65     8.64      18.27     19.43
117 -8.47    -7.71    2.90      8.16      12.78
```

```
118  Q(s,a) new
119  -1.37    1.66     8.65      18.28     19.44
120  -8.46   -7.70     2.91       8.16     12.79
121  =========================
122  Iteration 13:
123  fr: [-10, 0, -0.4, -0.4, 10]
124  sf1      s1       s2        s3       sf2
125  Q(s,a) current
126  -1.37    1.66     8.65      18.28     19.44
127  -8.46   -7.70     2.91       8.16     12.79
128  Q(s,a) new
129  -1.37    1.67     8.66      18.29     19.45
130  -8.46   -7.70     2.92       8.17     12.79
131  Optimal Politic:
132  sf1 = ->,        s1 = ->,          s2 = ->,          s3 = ->,
                 sf2 = ->
```

Which gives out a result of:

	sf1	s1	s2	s3	sf2
Q(s, ->)	**-1.37**	**1.67**	**8.66**	**18.29**	**19.45**
Q(s, <-)	-8.46	-7.70	2.92	8.17	12.79
f_pi(s)	->	->	->	->	->

The convergence plot for each Q(s,a) value is given below:

Figure 28: Image