

Default State

# Policy Iteration Calculations

| $V_0 = 0$     | $V_1 = 0$    | $V_2 = 0$                      | <del><math>V_3</math></del> |
|---------------|--------------|--------------------------------|-----------------------------|
| $V_4 = -1000$ | $V_5 = 1000$ | <del><math>V_6</math></del>    | $V_7 = 0$                   |
| $V_8 = 0$     | $V_9 = 0$    | $V_{10} = 0$                   | $V_{11} = 0$                |
| $V_{12} = 0$  | $V_{13} = 0$ | <del><math>V_{14}</math></del> | $V_{15} = 0$                |

$$\begin{aligned}
 \uparrow V_0 &= -0.04 \uparrow V_9 = 711.8564 \\
 \downarrow V_1 &= -0.04 \leftarrow V_{10} = -0.04 \\
 \leftarrow V_2 &= -0.0436 \leftarrow V_{11} = -0.0724 \\
 \oplus V_4 &= 1000 \rightarrow V_{12} = -8.1436 \\
 V_5 &= 1000 \quad \uparrow V_{13} = 58.163684 \\
 \uparrow V_7 &= -0.04 \quad \downarrow V_{14} = -0.046516 \\
 \rightarrow V_8 &= -90.04 \quad \uparrow V_{15} = -0.046516
 \end{aligned}$$

First Iteration Evaluation:  $V_i = r(i) + \lambda V(S')$

$$\begin{aligned}
 V_0 &= -0.04 + 0.9[0.8(0) + 0.1(0) + 0.1(0)] \\
 &= \boxed{-0.04}
 \end{aligned}$$

$$\begin{aligned}
 V_1 &= -0.04 + 0.9[0.8(0) + 0.1(0) + 0.1(0)] \\
 &= \boxed{-0.04}
 \end{aligned}$$

$$\begin{aligned}
 V_2 &= -0.04 + 0.9[0.8(0) + 0.1(-0.04) + 0.1(0)] \\
 &= -0.04 + 0.9(-0.004) = \boxed{-0.0436}
 \end{aligned}$$

$$V_4 = \boxed{-1000}$$

$$V_5 = \boxed{1000}$$

$$\begin{aligned}
 V_7 &= -0.04 + 0.9[0.8(0) + 0.1(0) + 0.1(0)] \\
 &= \boxed{-0.04}
 \end{aligned}$$

$$\begin{aligned}
 V_8 &= -0.04 + 0.9[0.8(0) + 0.1(-1000) + 0.1(0)] \\
 &= -0.04 + 0.9[-100] = \boxed{-98.04}
 \end{aligned}$$

$$\begin{aligned}
 V_9 &= -0.04 + 0.9[0.8(1000) + 0.1(-98.04) + 0.1(0)] \\
 &= -0.04 + 0.9[790.956] = \boxed{711.8564}
 \end{aligned}$$

$$\begin{aligned}
 V_{10} &= -0.04 + 0.9[0.8(0) + 0.1(0) + 0.1(0)] \\
 &= \boxed{-0.04}
 \end{aligned}$$

$$\begin{aligned}
 V_{11} &\approx -0.04 + 0.9[0.8(-0.04) + 0.1(-0.04) + 0.1(0)] \\
 &= -0.04 + 0.9(-0.036) = \boxed{-0.0724}
 \end{aligned}$$

$$\begin{aligned}
 V_{12} &= -0.04 + 0.9[0.8(0) + 0.1(-90.04) + 0.1(0)] \\
 &= -0.04 + 0.9(-9.004) = \boxed{-8.1436}
 \end{aligned}$$

$$\begin{aligned}
 V_{13} &= -0.04 + 0.9[0.8(-8.1436) + 0.1(711.8564) + 0.1(0)] \\
 &= -0.04 + 0.9[64.67076] = \boxed{58.163684}
 \end{aligned}$$

$$\begin{aligned}
 V_{15} &= -0.04 + 0.9[0.8(0) + 0.1(-0.0724) + 0.1(0)] \\
 &= -0.04 + 0.9(-0.00724) = \boxed{-0.046516}
 \end{aligned}$$

First Policy Improvement:  $L = \text{Left}, R = \text{Right}, U = V_p, D = \text{Down}$

$$\begin{aligned}
 V_0 &= R \quad \boxed{V_0} \\
 V_1 &= R \quad \boxed{V_0} \quad \boxed{V_4} \\
 L &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-1000) \\
 &= -100.324
 \end{aligned}$$

$$\begin{aligned}
 V_2 &= R \quad \boxed{V_1} = 0.8(-0.0436) + 0.1(-0.0436) \\
 &= \boxed{-0.04072}
 \end{aligned}$$

$$\begin{aligned}
 V_3 &= R \quad \boxed{V_2} \quad \boxed{V_4} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-1000) \\
 &= -100.324
 \end{aligned}$$

$$\begin{aligned}
 V_4 &= R \quad \boxed{V_3} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0436) \\
 &= -0.0436
 \end{aligned}$$

$$\begin{aligned}
 V_5 &= R \quad \boxed{V_4} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.04) \\
 &= -0.04
 \end{aligned}$$

$$\begin{aligned}
 V_6 &= R \quad \boxed{V_5} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0436) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_7 &= R \quad \boxed{V_6} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.04) \\
 &= -0.04
 \end{aligned}$$

$$\begin{aligned}
 V_8 &= R \quad \boxed{V_7} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_9 &= R \quad \boxed{V_8} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{10} &= R \quad \boxed{V_9} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{11} &= R \quad \boxed{V_{10}} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{12} &= R \quad \boxed{V_{11}} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{13} &= R \quad \boxed{V_{12}} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{14} &= R \quad \boxed{V_{13}} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{15} &= R \quad \boxed{V_{14}} \quad \boxed{V_1} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$\begin{aligned}
 V_{16} &= R \quad \boxed{V_{15}} \quad \boxed{V_2} \\
 R &= 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724) \\
 &= -0.04324
 \end{aligned}$$

$$V8R: L = 0.8(-90.04) + 0.1(-1000) + 0.1(-8.1436)$$

$$= -172.84636$$

$$R = 0.8(711.8564) + 0.1(-1000) + 0.1(-8.1436)$$

$$= \boxed{462.67076}$$

$$U = 0.8(-1000) + 0.1(-90.04) + 0.1(711.8564)$$

$$= -737.81836$$

$$D = 0.8(-8.1436) + 0.1(-90.04) + 0.1(711.8564)$$

$$= 55.66676$$

$$V_{10}R: L = 0.8(711.8564) + 0.1(-90.04) + 0.1(-8.1436)$$

$$= \boxed{562.47712}$$

$$R = 0.8(-0.0724) + 0.1(-0.0724) + 0.1(-0.0724)$$

$$= -0.6592$$

$$U = 0.8(-0.04) + 0.1(-711.8564) + 0.1(-0.0724)$$

$$= 71.1464$$

$$D = 0.8(-0.04) + 0.1(711.8564) + 0.1(-0.0724)$$

$$= 71.1464$$

$$V_{11}R: L = 0.8(-8.1436) + 0.1(-90.04) + 0.1(-8.1436)$$

$$= -16.33324$$

$$V_8$$

$$V_{12}$$

$$R = 0.8(58.163684) + 0.1(-90.04) + 0.1(-8.1436)$$

$$= \boxed{36.7125872}$$

$$V_{12}$$

$$V_{13}$$

$$U = 0.8(-90.04) + 0.1(-8.1436) + 0.1(58.163684)$$

$$= -14.682676$$

$$D = 0.8(-8.1436) + 0.1(-8.1436) + 0.1(58.163684)$$

$$= -1.9128716$$

$$V_{12}R: L = 0.8(-0.046516) + 0.1(-0.0724) + 0.1(-0.046516)$$

$$= -0.0491044$$

$$R = 0.8(-0.046516) + 0.1(-0.0724) + 0.1(-0.046516)$$

$$= -0.0491044$$

$$U = 0.8(-0.0724) + 0.1(-0.046516) + 0.1(-0.046516)$$

$$= -0.0672232$$

$$D = 0.8(-0.046516) + 0.1(-0.046516) + 0.1(-0.046516)$$

$$= -0.046516$$

$$V9R: L = 0.8(-90.04) + 0.1(1000) + 0.1(58.163684)$$

$$= 33.7843684$$

$$R = 0.8(-0.04) + 0.1(1000) + 0.1(58.163684)$$

$$= 105.7843684$$

$$U = 0.8(-1000) + 0.1(-90.04) + 0.1(-0.04)$$

$$= \boxed{179.992}$$

$$D = 0.8(58.163684) + 0.1(-90.04) + 0.1(-0.04)$$

$$= 37.5229472$$

$$V_{13}R: L = 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.046516)$$

$$= -0.0406516$$

$$R = 0.8(-0.0724) + 0.1(-0.04) + 0.1(-0.046516)$$

$$= -0.0665716$$

$$U = 0.8(-0.04) + 0.1(-0.04) + 0.1(-0.0724)$$

$$= -0.33124$$

$$D = 0.8(-0.046516) + 0.1(-0.04) + 0.1(-0.0724)$$

$$= -0.0484528$$

$$V_{13}R: L = 0.8(-8.1436) + 0.1(711.8564) + 0.1(58.163684)$$

$$= 70.4871284$$

$$R = 0.8(58.163684) + 0.1(711.8564) + 0.1(58.163684)$$

$$= 123.5329556$$

$$U = 0.8(711.8564) + 0.1(-8.1436) + 0.1(58.163684)$$

$$= \boxed{157.4871284}$$

$$D = 0.8(58.163684) + 0.1(-8.1436) + 0.1(58.163684)$$

$$= 51.5329556$$

### Post-First Iteration State

|                                |                               |  |   |
|--------------------------------|-------------------------------|--|---|
| $\uparrow V_8 = -0.04$         | $\downarrow V_1 = -0.04$      | <del><math>\leftarrow V_2 = -0.0436</math></del> | <del><math>\rightarrow V_3 = -0.04</math></del> |
| $V_4 = -1000$                  | $V_9 = 1000$                  | <del><math>V_5 = -0.04</math></del>              | <del><math>V_7 = -0.04</math></del>             |
| $\rightarrow V_8 = -90.04$     | $\uparrow V_9 = 711.8564$     | $\leftarrow V_{10} = -0.04$                      | <del><math>V_{11} = -0.0724</math></del>        |
| $\rightarrow V_{12} = -8.1436$ | $\uparrow V_{13} = 58.163684$ | <del><math>V_{14} = -0.04</math></del>           | $\downarrow V_{15} = -0.046516$                 |

## Second Iteration Evaluation

$$V_0 = -0.04 + 0.9[0.8(-0.04) + 0.1(-0.04) + 0.1(-0.04)]$$

$$= -0.04 + 0.9[-0.04] = \boxed{-0.076}$$

$$V_1 = -0.04 + 0.9[0.8(1900) + 0.1(-0.076) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9[799,98804] = \boxed{719,949236}$$

$$V_2 = -0.04 + 0.9[0.8(719,949236) + 0.1(-0.0436) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9[518,31560192] = \boxed{518,31560192}$$

$$V_3 = -1000 + 0.9(-1000) = \boxed{-1900}$$

$$V_4 = 1000 + 0.9(1000) = \boxed{1900}$$

$$V_5 = -0.04 + 0.9[0.8(-0.04) + 0.1(-0.04) + 0.1(-0.04)]$$

$$= -0.04 + 0.9[-0.04] = \boxed{-0.076}$$

$$V_6 = -0.04 + 0.9[0.8(711,8564) + 0.1(-0.076) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9(549,67026) = \boxed{494,663684}$$

$$V_7 = -0.04 + 0.9[0.8(-0.04) + 0.1(-0.04) + 0.1(-0.04)]$$

$$= -0.04 + 0.9[-0.04] = \boxed{-0.076}$$

$$V_8 = -0.04 + 0.9[0.8(711,8564) + 0.1(-0.076) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9(549,67026) = \boxed{494,663684}$$

$$V_9 = -0.04 + 0.9[0.8(1900) + 0.1(494,663684) + 0.1(-0.076)]$$

$$= -0.04 + 0.9[1569,4623684] = \boxed{1412,97613156}$$

$$V_{10} = -0.04 + 0.9[0.8(711,8564) + 0.1(-0.076) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9[1129,97290525] = \boxed{1016,93561472}$$

$$V_{11} = -0.04 + 0.9[0.8(1900) + 0.1(-0.076) + 0.1(-0.0436)]$$

$$= -0.04 + 0.9(813,537848179) = \boxed{732,145856161}$$

$$V_{12} = -0.04 + 0.9[0.8(58,163684) + 0.1(494,663684) + 0.1(-0.076)]$$

$$= -0.04 + 0.9(95,1829556) = \boxed{85,6246004}$$

$$V_{13} = -0.04 + 0.9[0.8(1900) + 0.1(58,163684) + 0.1(-0.076)]$$

$$+ 0.1(58,163684) = -0.04 + 0.9(1144,35973965)$$

$$= \boxed{1029,88376569}$$

$$V_{14} = -0.04 + 0.9[0.8(-0.046516) + 0.1(-0.046516) + 0.1(-0.046516)]$$

$$= -0.04 + 0.9(-0.046516) = \boxed{-0.0818644}$$

## Second Policy Improvement

$$L = 0.8(-0.076) + 0.1(-0.076) + 0.1(-0.076)$$

$$= \boxed{-190,9684}$$

$$(R) = 0.8(719,949236) + 0.1(-0.076) + 0.1(-1900)$$

$$= \boxed{381,9517888}$$

$$U = 0.8(-0.076) + 0.1(-0.076) + 0.1(719,949236)$$

$$= \boxed{71,9265236}$$

$$D = 0.8(-1900) + 0.1(-0.076) + 0.1(719,949236)$$

$$= \boxed{-1448,059364}$$

$$V_2 = L = 0.8(719,949236) + 0.1(518,31560192) + 0.1(518,31560192)$$

$$= \boxed{679,622597184}$$

$$R = 0.8(518,31560192) + 0.1(518,31560192) + 0.1(518,31560192)$$

$$= \boxed{518,31560192}$$

$$U = 0.8(518,31560192) + 0.1(719,949236) + 0.1(518,31560192)$$

$$= \boxed{538,478965728}$$

$$D = 0.8(518,31560192) + 0.1(719,949236) + 0.1(518,31560192)$$

$$= \boxed{538,478965728}$$

$$L = 0.8(-0.076) + 0.1(719,949236) + 0.1(-1900)$$

$$= \boxed{261,9341236}$$

$$R = 0.8(518,31560192) + 0.1(719,949236) + 0.1(1900)$$

$$= \boxed{676,642409136}$$

$$U = 0.8(719,949236) + 0.1(-0.076) + 0.1(518,31560192)$$

$$= \boxed{627,783348992}$$

$$D = 0.8(-1900) + 0.1(-0.076) + 0.1(518,31560192)$$

$$= \boxed{1571,82396019}$$

$$V_3 = L = 0.8(-0.076) + 0.1(-0.076) + 0.1(732,145856161)$$

$$= \boxed{73,1461856161}$$

$$R = 0.8(-0.076) + 0.1(-0.076) + 0.1(732,145856161)$$

$$= \boxed{73,1461856161}$$

$$U = 0.8(-0.076) + 0.1(-0.076) + 0.1(-0.076)$$

$$= \boxed{-0.076}$$

$$D = 0.8(-732,145856161) + 0.1(-0.076) + 0.1(-0.076)$$

$$= \boxed{585,655884929}$$

$V_8 = 0.8(494,663684) + 0.1(1900) + 0.1(85,6246004)$   
 $V_8 \quad V_4 \quad V_{12}$   
 $\approx 214,293413204$

$R = 0.8(1412,47613156) + 0.1(1900) + 0.1(85,6246004)$   
 $V_9 \quad V_4 \quad V_{12}$   
 $\approx 1948,943371252$

$V = 0.8(-1900) + 0.1(494,663684) + 0.1(1412,47613156)$   
 $V_8 \quad V_9$   
 $\approx -1329,286018447$

$D = 0.8(85,62466004) + 0.1(494,663684)$   
 $V_{12} \quad V_8$   
 $+ 0.1(1412,47613156) = 259,213709588$

---

$V_{10} \quad D = 0.8(1412,47613156) + 0.1(1016,93561472)$   
 $V_9 \quad V_{10}$   
 $+ 0.1(1016,93561472)$   
 $\approx 1333,36802819$

$R = 0.8(732,145856161) + 0.1(1016,93561472)$   
 $V_{11} \quad V_{10}$   
 $+ 0.1(1016,93561472) = 789,103807873$

$U = 0.8(1016,93561472) + 0.1(1412,47613156)$   
 $V_{10} \quad V_9$   
 $+ 0.1(732,145856161) = 1028,01069055$

$D = 0.8(1016,93561472) + 0.1(1412,47613156)$   
 $V_{10} \quad V_9$   
 $+ 0.1(732,145856161) = 1028,01069055$

---

$V_{12} \quad L \approx 0.8(85,62466004) + 0.1(494,663684)$   
 $V_8 \quad V_9$   
 $+ 0.1(85,62466004) = 72,062194036$

$R = 0.8(1029,88376569) + 0.1(494,663684)$   
 $V_{13} \quad V_8$   
 $+ 0.1(85,62466004) = 881,235846756$

$V = 0.8(494,663684) + 0.1(85,62466004)$   
 $V_{12} \quad V_{12}$   
 $+ 0.1(1029,88376569) = 507,281789773$

$D = 0.8(85,62466004) + 0.1(85,62466004)$   
 $V_{12} \quad V_{12}$   
 $+ 0.1(1029,88376569) = 180,090570605$

---

$V_{15} \quad L = 0.8(-0.0818644) + 0.1(732,145856161)$   
 $V_{15} \quad V_{11}$   
 $+ 0.1(-0.0818644) = 73,1409076561$

$R = 0.8(-0.0818644) + 0.1(732,145856161)$   
 $V_{15} \quad V_{11}$   
 $+ 0.1(-0.0818644) = 73,1409076561$

$U = 0.8(732,145856161) + 0.1(-0.0818644)$   
 $V_{15} \quad V_{15}$   
 $+ 0.1(-0.0818644) = 585,200312049$

$D = 0.8(-0.0818644) + 0.1(-0.0818644)$   
 $V_{15} \quad V_{15}$   
 $+ 0.1(-0.0818644) = -0.0818644$

---

$V_9 \quad L = 0.8(494,663684) + 0.1(1900) + 0.1(1029,88376569)$   
 $V_8 \quad V_5 \quad V_{13}$   
 $\approx 688,719,323,769$

$R = 0.8(1016,93561472) + 0.1(1900) + 0.1(1029,88376569)$   
 $V_{10} \quad V_5 \quad V_{13}$   
 $\approx 1106,53225747$

$U = 0.8(1900) + 0.1(494,663684) + 0.1(1016,93561472)$   
 $V_5 \quad V_8$   
 $\approx 1671,15992987$

$D = 0.8(1029,88376569) + 0.1(494,663684) + 0.1(1016,93561472)$   
 $V_{13} \quad V_8$   
 $\approx 975,066942424$

---

$V_{11} \quad U = 0.8(1016,93561472) + 0.1(-0.076) + 0.1(-0.0818644)$   
 $V_{10} \quad V_7 \quad V_5$   
 $\approx 1813,532205336$

$R = 0.8(732,145856161) + 0.1(-0.076) + 0.1(-0.0818644)$   
 $V_{11} \quad V_7 \quad V_5$   
 $\approx 585,700898489$

$U = 0.8(-0.076) + 0.1(1016,93561472) + 0.1(732,145856161)$   
 $V_7 \quad V_{10} \quad V_{11}$   
 $\approx 174,847347088$

$D = 0.8(-0.0818644) + 0.1(1016,93561472) + 0.1(732,145856161)$   
 $V_{15} \quad V_{10} \quad V_{11}$   
 $\approx 174,842695568$

---

$V_{13} \quad U = 0.8(85,62466004) + 0.1(1412,47613156)$   
 $V_{12} \quad V_9$   
 $+ 0.1(1029,88376569) = 312,735717257$

$R = 0.8(85,62466004) + 0.1(1412,47613156)$   
 $V_{13} \quad V_9$   
 $+ 0.1(-0.079)$

$R = 0.8(1029,88376569) + 0.1(1412,47613156)$   
 $V_{13} \quad V_9$   
 $+ 0.1(1029,88376569) = 1068,14300228$

$U = 0.8(1412,47613156) + 0.1(85,62466004)$   
 $V_{13} \quad V_9$   
 $+ 0.1(1029,88376569) = 1241,53174782$

$D = 0.8(1029,88376569) + 0.1(85,62466004)$   
 $V_{13} \quad V_{12}$   
 $+ 0.1(1029,88376569) = 935,457855125$

# First Iteration

|                    |                    |                  |                    |
|--------------------|--------------------|------------------|--------------------|
| $V_0 = -0.04$      | $V_1 = -0.04$      | $V_2 = -0.0436$  |                    |
| $V_4 = -1900$      | $V_5 = 1900$       |                  | $V_7 = -0.04$      |
| $V_8 = -90.04$     | $V_9 = 711.8564$   | $V_{10} = -0.04$ | $V_{11} = -0.0724$ |
| $V_{12} = -8.1436$ | $V_{13} = 58.1637$ |                  | $V_{15} = -0.0465$ |

# Second Iteration

|   |                      |                      |                     |
|---|----------------------|----------------------|---------------------|
| $V_0 = -0.076$  | $V_1 = 719.9492$     | $V_2 = 518.3156$     |                     |
| $V_4 = -1900$   | $V_5 = 1900$         |                      | $V_7 = -0.076$      |
| $V_8 = 494.6637$  | $V_9 = 1412.4761$    | $V_{10} = 1016.9396$ | $V_{11} = 732.1459$ |
| <del><math>V_{12} = -891.1</math></del><br>$V_{12} = 85.6247$ | $V_{13} = 1029.8838$ |                      | $V_{15} = -0.0819$  |