In [9]:
```
import warnings ; warnings.filterwarnings('ignore')

import gymnasium as gym, gym_walk # Changed gym to gymnasium as gym is unmaint
import numpy as np

import random
import warnings

warnings.filterwarnings('ignore', category=DeprecationWarning)
np.set_printoptions(suppress=True)
random.seed(123); np.random.seed(123)
```

In [3]:
```
pip install gymnasium git+https://github.com/mimoralea/gym-walk#egg=gym-walk #
```

```
Collecting gym-walk
  Cloning https://github.com/mimoralea/gym-walk to /tmp/pip-install-61d6pn9g/gy
m-walk_2947cd5ce1c04fe6bc1314bd2ceea149
  Running command git clone --filter=blob:none --quiet https://github.com/mimor
alea/gym-walk /tmp/pip-install-61d6pn9g/gym-walk_2947cd5ce1c04fe6bc1314bd2ceea1
49
  Resolved https://github.com/mimoralea/gym-walk to commit b915b94cf2ad16f8833a
1ad92ea94e88159279f5
  Preparing metadata (setup.py) ... done
Requirement already satisfied: gymnasium in /usr/local/lib/python3.12/dist-pack
ages (1.2.0)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.12/dist-
packages (from gymnasium) (2.0.2)
Requirement already satisfied: cloudpickle>=1.2.0 in /usr/local/lib/python3.12/
dist-packages (from gymnasium) (3.1.1)
Requirement already satisfied: typing-extensions>=4.3.0 in /usr/local/lib/pytho
n3.12/dist-packages (from gymnasium) (4.15.0)
Requirement already satisfied: farama-notifications>=0.0.1 in /usr/local/lib/py
thon3.12/dist-packages (from gymnasium) (0.0.4)
Requirement already satisfied: gym in /usr/local/lib/python3.12/dist-packages
(from gym-walk) (0.25.2)
Requirement already satisfied: gym-notices>=0.0.4 in /usr/local/lib/python3.12/
dist-packages (from gym->gym-walk) (0.1.0)
Building wheels for collected packages: gym-walk
  Building wheel for gym-walk (setup.py) ... done
  Created wheel for gym-walk: filename=gym_walk-0.0.2-py3-none-any.whl size=537
7 sha256=98a907f73e0980c04b67aa7b5a2dc2f70c2f1dbfdf20f6bbf951d078d62315fa
  Stored in directory: /tmp/pip-ephem-wheel-cache-ye92yqv5/wheels/bf/23/e5/a94b
e4a90dd18f7ce958c21f192276cb01ef0daaf2bc66583b
Successfully built gym-walk
Installing collected packages: gym-walk
Successfully installed gym-walk-0.0.2
```

In [4]:
```
def print_policy(pi, P, action_symbols=('<', 'v', '>', '^'), n_cols=4, title='
    print(title)
    arrs = {k:v for k,v in enumerate(action_symbols)}
    for s in range(len(P)):
        a = pi(s)
        print("| ", end="")
        if np.all([done for action in P[s].values() for _, _, _, done in actio
            print("".rjust(9), end=" ")
        else:
            print(str(s).zfill(2), arrs[a].rjust(6), end=" ")
        if (s + 1) % n_cols == 0: print("|")
```

In [5]:

```
def print_state_value_function(V, P, n_cols=4, prec=3, title='State-value func
    print(title)
    for s in range(len(P)):
        v = V[s]
        print("| ", end="")
        if np.all([done for action in P[s].values() for _, _, _, done in actic
            print("".rjust(9), end=" ")
        else:
            print(str(s).zfill(2), '{}'.format(np.round(v, prec)).rjust(6), er
        if (s + 1) % n_cols == 0: print("|")
```

In [20]:

```
def probability_success(env, pi, goal_state, n_episodes=100, max_steps=200):
    random.seed(123); np.random.seed(123) # Removed env.seed(123)
    results = []
    for _ in range(n_episodes):
        state, info = env.reset(seed=123) # Added seed here
        done, steps = False, 0
        while not done and steps < max_steps:
            state, _, done, _, h = env.step(pi(state)) # Added _ for truncatec
            steps += 1
        results.append(state == goal_state)
    return np.sum(results)/len(results)
```

In [21]:

```
def mean_return(env, pi, n_episodes=100, max_steps=200):
    random.seed(123); np.random.seed(123) # Removed env.seed(123)
    results = []
    for _ in range(n_episodes):
        state, info = env.reset(seed=123) # Added seed here
        done, steps = False, 0
        results.append(0.0)
        while not done and steps < max_steps:
            state, reward, done, _, _ = env.step(pi(state)) # Added _ for trur
            results[-1] += reward
            steps += 1
        results.append(0.0) # Added this line to match the original code's log
    return np.mean(results)
```

FrozenLake MDP

In [11]:

```
env = gym.make('FrozenLake-v1')
# P = env.env.P # Removed this line as it's no longer needed
init_state = env.reset()
goal_state = 15
LEFT, DOWN, RIGHT, UP = range(4)
```

In [15]:

```
P
```

```
Out[15]: {0: {0: [(0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 4, 0.0, False)],
           1: [(0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 4, 0.0, False),
            (0.3333333333333333, 1, 0.0, False)],
           2: [(0.3333333333333333, 4, 0.0, False),
            (0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 0, 0.0, False)],
           3: [(0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 0, 0.0, False)]},
          1: {0: [(0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 5, 0.0, True)],
           1: [(0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 5, 0.0, True),
            (0.3333333333333333, 2, 0.0, False)],
           2: [(0.3333333333333333, 5, 0.0, True),
            (0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 1, 0.0, False)],
           3: [(0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 0, 0.0, False)]},
          2: {0: [(0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 6, 0.0, False)],
           1: [(0.3333333333333333, 1, 0.0, False),
            (0.3333333333333333, 6, 0.0, False),
            (0.3333333333333333, 3, 0.0, False)],
           2: [(0.3333333333333333, 6, 0.0, False),
            (0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 2, 0.0, False)],
           3: [(0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 1, 0.0, False)]},
          3: {0: [(0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 7, 0.0, True)],
           1: [(0.3333333333333333, 2, 0.0, False),
            (0.3333333333333333, 7, 0.0, True),
            (0.3333333333333333, 3, 0.0, False)],
           2: [(0.3333333333333333, 7, 0.0, True),
            (0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 3, 0.0, False)],
           3: [(0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 3, 0.0, False),
            (0.3333333333333333, 2, 0.0, False)]},
          4: {0: [(0.3333333333333333, 0, 0.0, False),
            (0.3333333333333333, 4, 0.0, False),
            (0.3333333333333333, 8, 0.0, False)],
           1: [(0.3333333333333333, 4, 0.0, False),
            (0.3333333333333333, 8, 0.0, False),
            (0.3333333333333333, 5, 0.0, True)],
 2: [(0.3333333333333333, 8, 0.0, False),
  (0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 0, 0.0, False)],
 3: [(0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 0, 0.0, False),
  (0.3333333333333333, 4, 0.0, False)]},
5: {0: [(1.0, 5, 0, True)],
 1: [(1.0, 5, 0, True)],
 2: [(1.0, 5, 0, True)],
 3: [(1.0, 5, 0, True)]},
6: {0: [(0.3333333333333333, 2, 0.0, False),
  (0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 10, 0.0, False)],
 1: [(0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 10, 0.0, False),
  (0.3333333333333333, 7, 0.0, True)],
 2: [(0.3333333333333333, 10, 0.0, False),
  (0.3333333333333333, 7, 0.0, True),
  (0.3333333333333333, 2, 0.0, False)],
 3: [(0.3333333333333333, 7, 0.0, True),
  (0.3333333333333333, 2, 0.0, False),
  (0.3333333333333333, 5, 0.0, True)]},
7: {0: [(1.0, 7, 0, True)],
 1: [(1.0, 7, 0, True)],
 2: [(1.0, 7, 0, True)],
 3: [(1.0, 7, 0, True)]},
8: {0: [(0.3333333333333333, 4, 0.0, False),
  (0.3333333333333333, 8, 0.0, False),
  (0.3333333333333333, 12, 0.0, True)],
 1: [(0.3333333333333333, 8, 0.0, False),
  (0.3333333333333333, 12, 0.0, True),
  (0.3333333333333333, 9, 0.0, False)],
 2: [(0.3333333333333333, 12, 0.0, True),
  (0.3333333333333333, 9, 0.0, False),
  (0.3333333333333333, 4, 0.0, False)],
 3: [(0.3333333333333333, 9, 0.0, False),
  (0.3333333333333333, 4, 0.0, False),
  (0.3333333333333333, 8, 0.0, False)]},
9: {0: [(0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 8, 0.0, False),
  (0.3333333333333333, 13, 0.0, False)],
 1: [(0.3333333333333333, 8, 0.0, False),
  (0.3333333333333333, 13, 0.0, False),
  (0.3333333333333333, 10, 0.0, False)],
 2: [(0.3333333333333333, 13, 0.0, False),
  (0.3333333333333333, 10, 0.0, False),
  (0.3333333333333333, 5, 0.0, True)],
 3: [(0.3333333333333333, 10, 0.0, False),
  (0.3333333333333333, 5, 0.0, True),
  (0.3333333333333333, 8, 0.0, False)]},
10: {0: [(0.3333333333333333, 6, 0.0, False),
  (0.3333333333333333, 9, 0.0, False),
  (0.3333333333333333, 14, 0.0, False)],
 1: [(0.3333333333333333, 9, 0.0, False),
   (0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 11, 0.0, True)],
  2: [(0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 11, 0.0, True),
   (0.3333333333333333, 6, 0.0, False)],
  3: [(0.3333333333333333, 11, 0.0, True),
   (0.3333333333333333, 6, 0.0, False),
   (0.3333333333333333, 9, 0.0, False)]},
 11: {0: [(1.0, 11, 0, True)],
  1: [(1.0, 11, 0, True)],
  2: [(1.0, 11, 0, True)],
  3: [(1.0, 11, 0, True)]},
 12: {0: [(1.0, 12, 0, True)],
  1: [(1.0, 12, 0, True)],
  2: [(1.0, 12, 0, True)],
  3: [(1.0, 12, 0, True)]},
 13: {0: [(0.3333333333333333, 9, 0.0, False),
   (0.3333333333333333, 12, 0.0, True),
   (0.3333333333333333, 13, 0.0, False)],
  1: [(0.3333333333333333, 12, 0.0, True),
   (0.3333333333333333, 13, 0.0, False),
   (0.3333333333333333, 14, 0.0, False)],
  2: [(0.3333333333333333, 13, 0.0, False),
   (0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 9, 0.0, False)],
  3: [(0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 9, 0.0, False),
   (0.3333333333333333, 12, 0.0, True)]},
 14: {0: [(0.3333333333333333, 10, 0.0, False),
   (0.3333333333333333, 13, 0.0, False),
   (0.3333333333333333, 14, 0.0, False)],
  1: [(0.3333333333333333, 13, 0.0, False),
   (0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 15, 1.0, True)],
  2: [(0.3333333333333333, 14, 0.0, False),
   (0.3333333333333333, 15, 1.0, True),
   (0.3333333333333333, 10, 0.0, False)],
  3: [(0.3333333333333333, 15, 1.0, True),
   (0.3333333333333333, 10, 0.0, False),
   (0.3333333333333333, 13, 0.0, False)]},
 15: {0: [(1.0, 15, 0, True)],
  1: [(1.0, 15, 0, True)],
  2: [(1.0, 15, 0, True)],
  3: [(1.0, 15, 0, True)]}}
```

In [13]:

```
P = env.unwrapped.P
```

In [14]:

```
init_state
```

Out[14]:

```
(0, {'prob': 1})
```

In [17]:

```
state, reward, terminated, truncated, info = env.step(RIGHT)
print("state:{0} - reward:{1} - terminated:{2} - truncated:{3} - info:{4}".for
state:1 - reward:0.0 - terminated:False - truncated:False - info:{'prob': 0.333
3333333333333}
```

In [18]:

```
# Adversarial Policy
pi_frozenlake1 = lambda s: {
    0: RIGHT,
    1: RIGHT,
    2: RIGHT,
    3: RIGHT,
    4: RIGHT,
    5: RIGHT,
    6: RIGHT,
    7: RIGHT,
    8: RIGHT,
    9: RIGHT,
    10:RIGHT,
    11:RIGHT,
    12:RIGHT,
    13:RIGHT,
    14:RIGHT,
    15:RIGHT #Stop
}[s]
print("Name:Jeevanesh")
print("Register Number:         ")
print_policy(pi_frozenlake1, P, action_symbols=('<', 'v', '>', '^'), n_cols=4)
```

```
Name:Jeevanesh
Register Number:
Policy:
| 00      > | 01      > | 02      > | 03      > |
| 04      > |           | 06      > |           |
| 08      > | 09      > | 10      > |           |
|           | 13      > | 14      > |           |
```

In [22]:

```
print('Reaches goal {:.2f}%. Obtains an average undiscounted return of {:.4f}.
    probability_success(env, pi_frozenlake1, goal_state=goal_state)*100,
    mean_return(env, pi_frozenlake1)))
```

```
Reaches goal 0.00%. Obtains an average undiscounted return of 0.0000.
```

Policy Evaluation

In [23]:

```
def policy_evaluation(pi, P, gamma=1.0, theta=1e-10):
    prev_V = np.zeros(len(P), dtype=np.float64)
    while True:
        V = np.zeros(len(P), dtype=np.float64)
        for s in range(len(P)):
            for prob, next_state, reward, done in P[s][pi(s)]:
                V[s] += prob * (reward + gamma * prev_V[next_state] * (not dor
        if np.max(np.abs(prev_V - V)) < theta:
            break
        prev_V = V.copy()
    return V
```

In [24]:

```
# Code to evaluate the adversarial policy
V1 = policy_evaluation(pi_frozenlake1, P)
print("Name:JEEVANESH S")
print("Register Number:212222243002")
print_state_value_function(V1, P, n_cols=4, prec=5)
```

```
Name:JEEVANESH S
Register Number:212222243002
State-value function:
| 00 0.0315 | 01 0.02381 | 02 0.04762 | 03    0.0 |
| 04 0.03919 |           | 06 0.09524 |           |
| 08 0.08608 | 09 0.21905 | 10 0.2381 |           |
|           | 13 0.41905 | 14 0.61905 |           |
```

In [41]:

```
print('Reaches goal with {:.2f}% probability. Obtains a {:.4f} average undiscc
    probability_success(env, pi_frozenlake1, goal_state=goal_state) * 100,
    mean_return(env, pi_frozenlake1)))
```

```
Reaches goal with 0.00% probability. Obtains a 0.0000 average undiscounted retu
rn.
```

Policy Improvement

In [32]:

```
def policy_improvement(V, P, gamma=1.0):
    Q = np.zeros((len(P), len(P[0])), dtype=np.float64)
    # Write your code here to improve the given policy
    for s in range(len(P)):
      for a in range(len(P[s])):
        for prob,next_state,reward,done in P[s][a]:
          Q[s][a]+=prob*(reward+gamma*V[next_state]*(not done))
          new_pi=lambda s:{s:a for s, a in enumerate(np.argmax(Q,axis=1))}[s]
    return new_pi
```

In [28]:

```
pi_2 = policy_improvement(V1, P)
print("Name: JEEVANESH S ")
print("Register Number: 212222243002")
print_policy(pi_2, P, action_symbols=('<', 'v', '>', '^'), n_cols=4)
```

```
Name: JEEVANESH S
Register Number: 212222243002
Policy:
| 00      < | 01      ^ | 02      < | 03      < |
| 04      < |           | 06      < |           |
| 08      ^ | 09      v | 10      < |           |
|           | 13      > | 14      v |           |
```

In [29]:

```
print('Reaches goal {:.2f}%. Obtains an average undiscounted return of {:.4f}.
    probability_success(env, pi_2, goal_state=goal_state)*100,
    mean_return(env, pi_2)))
```

```
Reaches goal 100.00%. Obtains an average undiscounted return of 0.5000.
```

In [30]:
```
V2 = policy_evaluation(pi_2, P)
print("Name:212222243002")
print("Register Number:212222243002")
print_state_value_function(V2, P, n_cols=4, prec=5)
```

```
Name:212222243002
Register Number:212222243002
State-value function:
| 00 0.78049 | 01 0.65854 | 02 0.53659 | 03 0.26829 |
| 04 0.78049 |           | 06 0.41463 |           |
| 08 0.78049 | 09 0.78049 | 10 0.70732 |           |
|           | 13 0.85366 | 14 0.92683 |           |
```

In [31]:
```
# comparing the initial and the improved policy
if(np.sum(V1>=V2)==16):
  print("The Adversarial policy is the better policy")
elif(np.sum(V2>=V1)==16):
  print("The Improved policy is the better policy")
else:
  print("Both policies have their merits.")
```

```
The Improved policy is the better policy
```

Policy Iteration

In [33]:
```
def policy_iteration(P, gamma=1.0, theta=1e-10):
   random_actions=np.random.choice(tuple(P[0].keys()),len(P))
   pi = lambda s: {s:a for s, a in enumerate(random_actions)}[s]
   while True:
    old_pi = {s:pi(s) for s in range(len(P))}
    V = policy_evaluation(pi, P,gamma,theta)
    pi = policy_improvement(V,P,gamma)
    if old_pi == {s:pi(s) for s in range(len(P))}:
      break
   return V, pi
```

In [34]:
```
optimal_V, optimal_pi = policy_iteration(P)
```

In [36]:
```
print("Name:JEEVANESH")
print("Register Number:212222243002")
print('Optimal policy and state-value function (PI):')
print_policy(optimal_pi, P, action_symbols=('<', 'v', '>', '^'), n_cols=4)
```

```
Name:JEEVANESH
Register Number:212222243002
Optimal policy and state-value function (PI):
Policy:
| 00      < | 01      ^ | 02      ^ | 03      ^ |
| 04      < |           | 06      < |           |
| 08      ^ | 09      v | 10      < |           |
|           | 13      > | 14      v |           |
```

In [37]:

```
print('Reaches goal {:.2f}%. Obtains an average undiscounted return of {:.4f}.
    probability_success(env, optimal_pi, goal_state=goal_state)*100,
    mean_return(env, optimal_pi)))
```

```
Reaches goal 100.00%. Obtains an average undiscounted return of 0.5000.
```

In [38]:

```
print("Name:JEEVANESH")
print("Register Number:212222243002        ")
print_state_value_function(optimal_V, P, n_cols=7, prec=5)
```

```
Name:JEEVANESH
Register Number:212222243002
State-value function:
| 00 0.82353 | 01 0.82353 | 02 0.82353 | 03 0.82353 | 04 0.82353 |           |
06 0.52941 |
|           | 08 0.82353 | 09 0.82353 | 10 0.76471 |           |           | 13
0.88235 |
| 14 0.94118 |
```