Santiago Morales
201913369

1. Introducción al problema

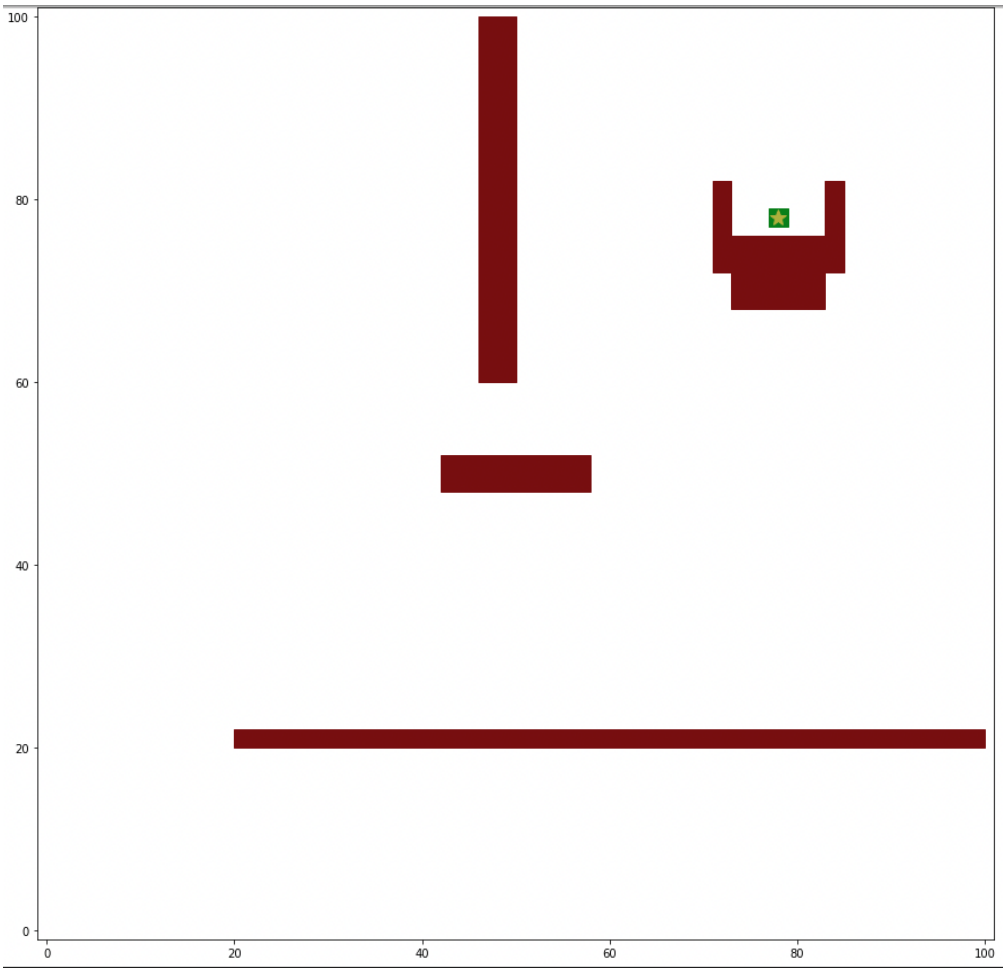

Figura 1: Retícula y zonas de peligro

Se busca implementar un método de iteración por política y un método de programación lineal para resolver el siguiente problema de decisión de Markov:

Considere un objeto volador que se mueve en una retícula de 100×100 bajo la influencia del viento. La meta del objeto volador es llegar a la región de la retícula con coordenadas $[77, 78] \times [77, 78]$ en el menor tiempo posible. En cada posición en el interior de la retícula el objeto tiene 4 acciones posibles $\mathrm{A} = \{u, d, l, r\}$ con probabilidades de transición dadas por:

$$Q((z,w)|(x,y),u) = \begin{cases} 0,3 & \text{if } z = x, \quad w = y+1 \\ 0,4 & \text{if } z = x, \quad w = y+2 \\ 0,2 & \text{if } z = x-1, \quad w = y+2 \\ 0,1 & \text{if } z = x-1, \quad w = y+1 \end{cases}$$

$$Q((z,w)|(x,y),d) = \begin{cases} 0,3 & \text{if } z = x, \quad w = y \\ 0,3 & \text{if } z = x, \quad w = y-1 \\ 0,2 & \text{if } z = x-1, \quad w = y \\ 0,2 & \text{if } z = x-1, \quad w = y-1 \end{cases}$$

$$Q((z,w)|(x,y),l) = \begin{cases} 0,3 & \text{if } z = x-1, \quad w = y+1 \\ 0,2 & \text{if } z = x-1, \quad w = y \\ 0,3 & \text{if } z = x-2, \quad w = y \\ 0,2 & \text{if } z = x-2, \quad w = y+1 \end{cases}$$

$$Q((z,w)|(x,y),r) = \begin{cases} 0,3 & \text{if } z = x+1, \quad w = y \\ 0,4 & \text{if } z = x+1, \quad w = y+1 \\ 0,2 & \text{if } z = x, \quad w = y \\ 0,1 & \text{if } z = x, \quad w = y+1 \end{cases}$$

Para definir las probabilidades de los bordes, si una coordenada no está definida; es decir, si se sale de la retícula, la probabilidad de llegar a esa retícula se le suma a la casilla más cercana a esta que sí está la retícula. También se definió que el objeto volador se queda en la meta si llega. Los posibles estados son las posibles coordenadas (discretas) de la retícula y las recompensas están dadas por:

$$r(x,y) = \begin{cases} 1 & \text{if } (x,y) \quad \text{está en la meta} \\ 0 & \text{if } (x,y) \quad \text{no está en la meta} \\ -5 & \text{if } (x,y) \quad \text{está en una zona de peligro} \end{cases}$$

Pues resulta que en la zona de circulación del objeto volador hay algunos lugares contaminados por radiación, así que no es conveniente que pase mucho tiempo en esos lugares.

2. Implementación y resultados

2.1. Implementación por iteración por política

Para iteración por política se implementó el algoritmo exactamente como se describe en la página 175 del libro. La política inicial era ir siempre a la derecha y para la inversión de la matriz $(I - \lambda P_{d_n})$ se utilizó el paquete **scipy.sparse** para optimizar el tiempo de inversión. El algoritmo tomaba entre 2 y 3min.

Si $\lambda = 0{,}8$, al simular 100 trayectorias del objeto volador que inician en (95, 5) y siguen la política encontrada, se obtuvo el resultado de la Figura 2. Se observa que el avión evita todos los obstáculos y también parece la política hace que los aviones lleguen en el menor tiempo posible, teniendo en cuenta que se deben evitar las zonas con accidentes nucleares. También se observa que la política tiene en cuenta el viento hacia la izquierda y arriba que se observa en las probabilidades de transición. En la Figura 3 se observa que la política para las trayectorias que inician en (40,90) no funciona muy bien, esto sugiere que $\lambda = 0{,}8$ tampoco es suficiente para que funcione para las trayectorias que inician en (40, 90).

Si $\lambda = 0{,}5$, al simular 100 trayectorias del objeto volador que inician en (95, 5) y siguen la política encontrada, se obtuvo el resultado de las Figuras 3 y 4. Se observa que las trayectorias ya solo evitan los obstáculos y no se dirigen a la meta, se infiere que un lambda pequeño hacen que las recompensas de llegar al objetivo sean tan pequeñas que ya el computador no puede diferenciar estos valores, esto se puede mejorar si aumentamos lambda o la recompensa de llegar a la meta.

Si $\lambda = 0{,}95$, se simularon 100 trayectorias del objeto volador que inician en (95, 5) y (40, 90) y siguen la política encontrada. Se observa que las trayectorias buscan llegar a la meta y

evitan la mayoría de los obstáculos, se infiere que en algunos casos da más recompensa pasar rápidamente por los obstáculos y llegar más rápido a la meta, esto se debe a la manera en que definí las recompensas, pues vale la pena llegar a la meta y quedarse ahí para ganar la recompensa de los siguientes tiempos.

Por último, si $\lambda = 0{,}999$, se simularon 100 trayectorias del objeto volador que inician en (95, 5) y (40, 90) y siguen la política encontrada. Se observa que las trayectorias buscan llegar a la meta y no evitan los obstáculos, se infiere que en todos los casos da más recompensa pasar rápidamente por los obstáculos y llegar más rápido a la meta.

3. Conclusión

El método de iteración por política resultó efectivo para la solución del problema, aunque como se puede observar en los resultados, las políticas dependen fuertemente de λ y de las recompensas, por lo que si se definen de otra manera se pueden obtener distintos resultados, aunque los usados ilustran bastante bien cómo funcionan las soluciones con diferentes valores de λ.

También es importante mencionar que los paquetes de matrices *sparse* hacen que la solución de este problema se obtenga mucho más rápido, optimizando más algunas definiciones y operaciones de matrices se puede optimizar el código usado para que funcione para mayores n.

Santiago Morales
201913369

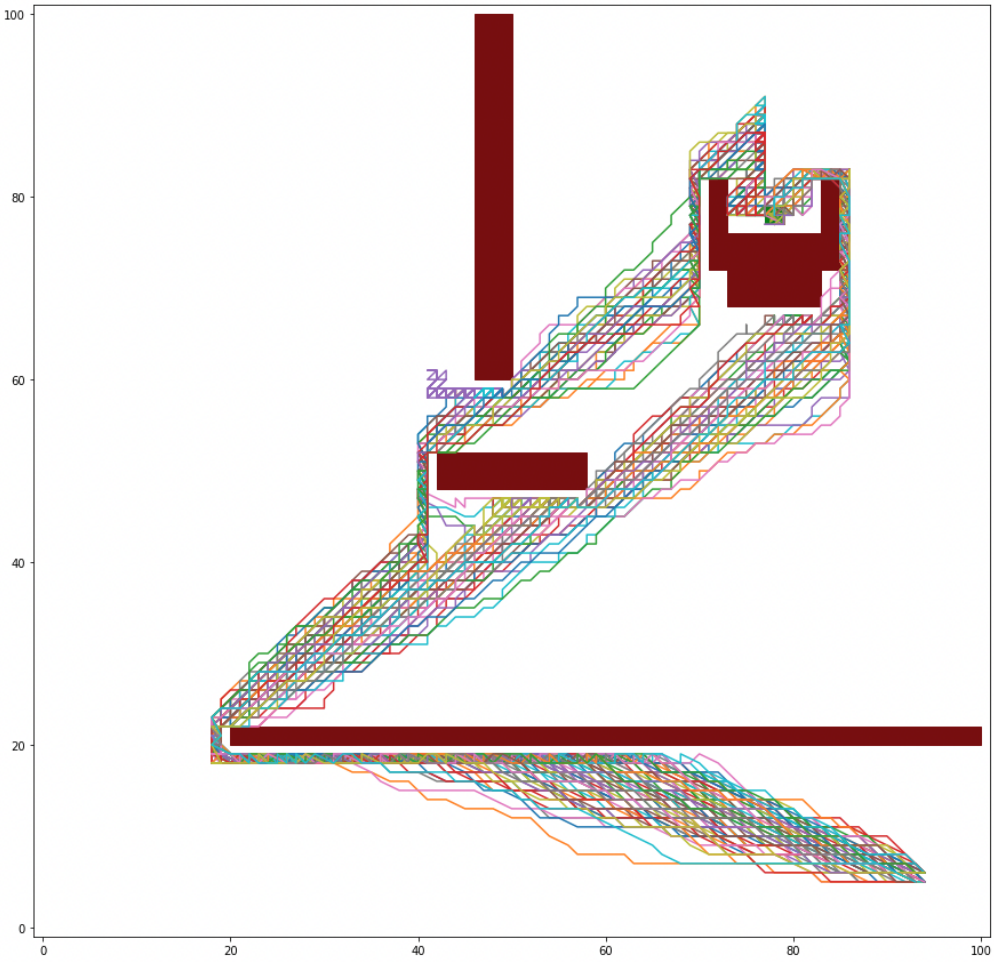

Figura 2: Iteración por política y $\lambda = 0{,}8$

Santiago Morales
201913369

Figura 3: Iteración por política y $\lambda = 0{,}8$ o $\lambda = 0{,}5$ (Sucede algo muy similar)

Santiago Morales
201913369

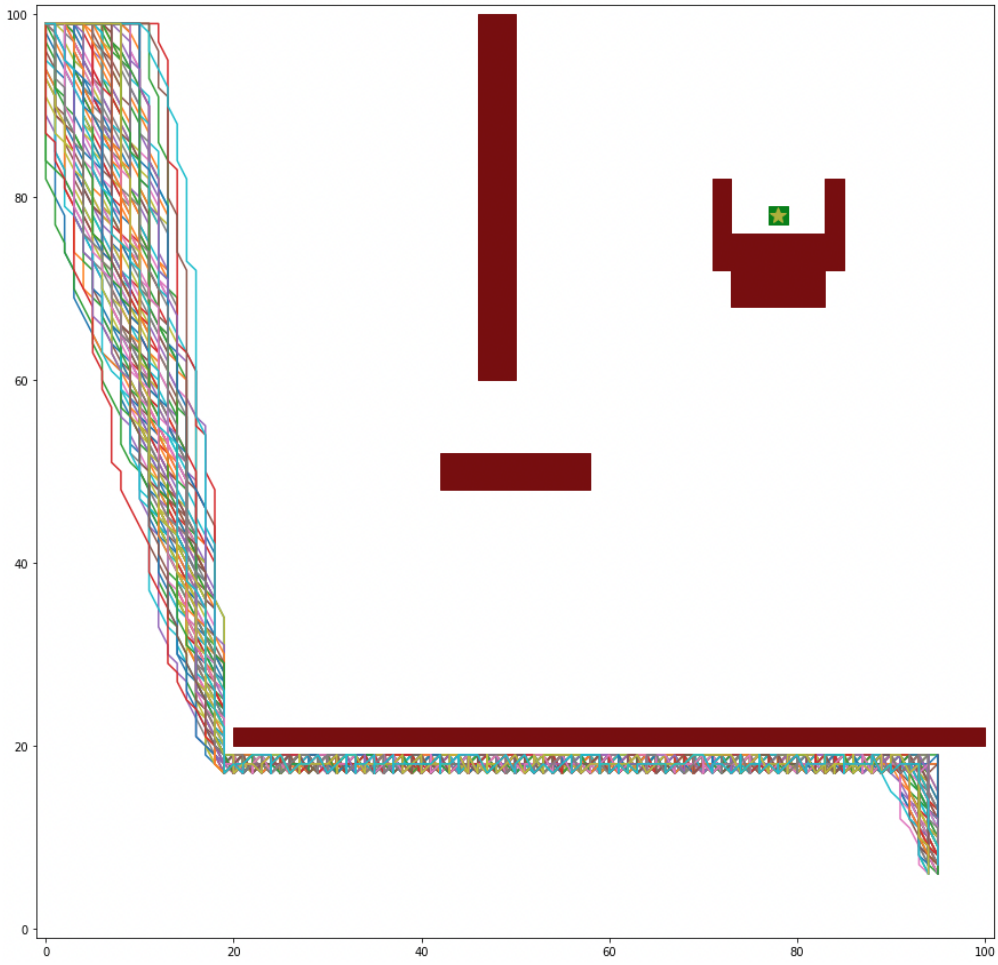

Figura 4: Iteración por política y $\lambda = 0{,}5$

Santiago Morales
201913369

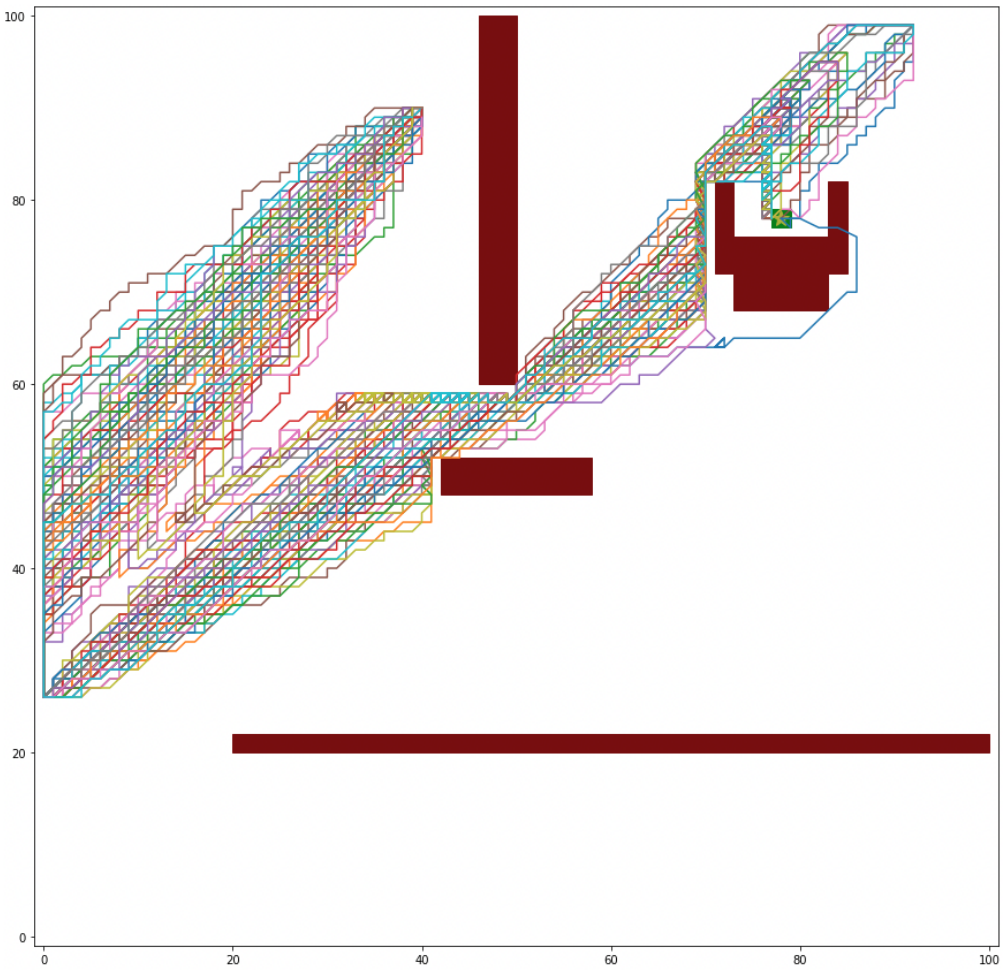

Figura 5: Iteración por política y $\lambda = 0{,}95$

Santiago Morales
201913369

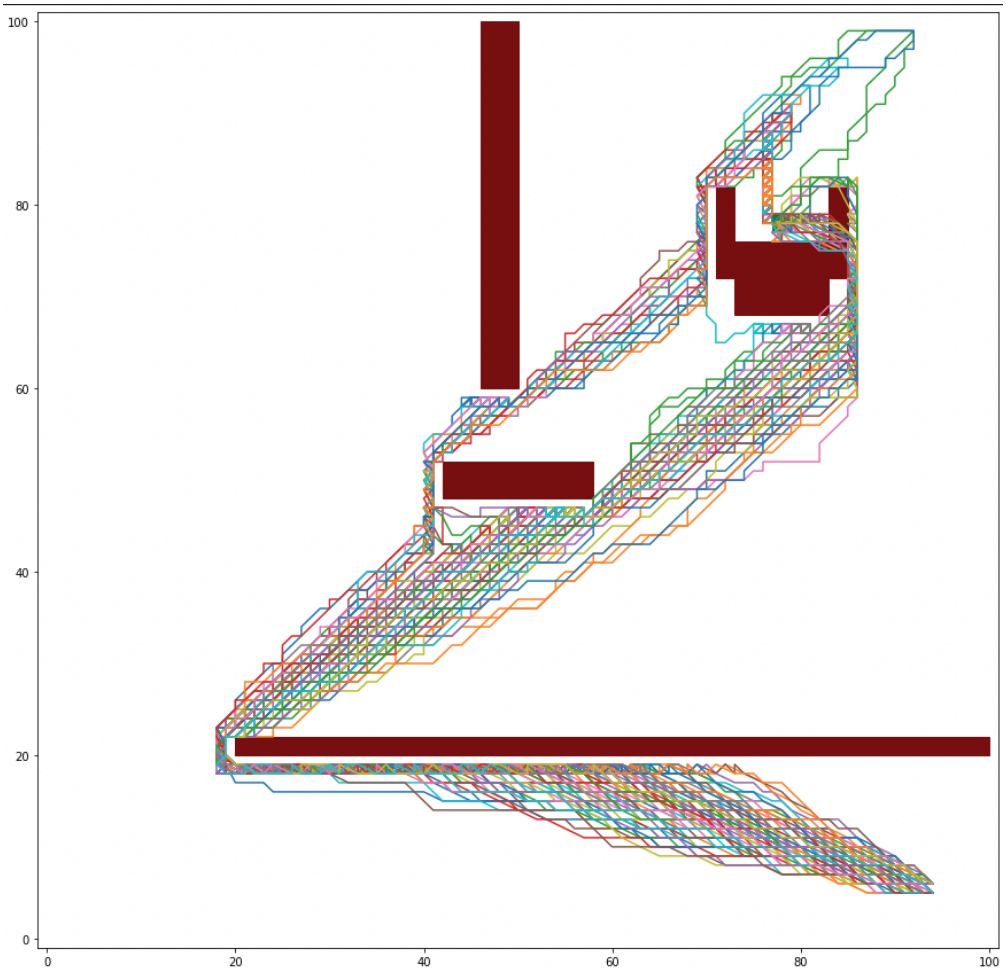

Figura 6: Iteración por política y $\lambda = 0{,}95$

Santiago Morales
201913369

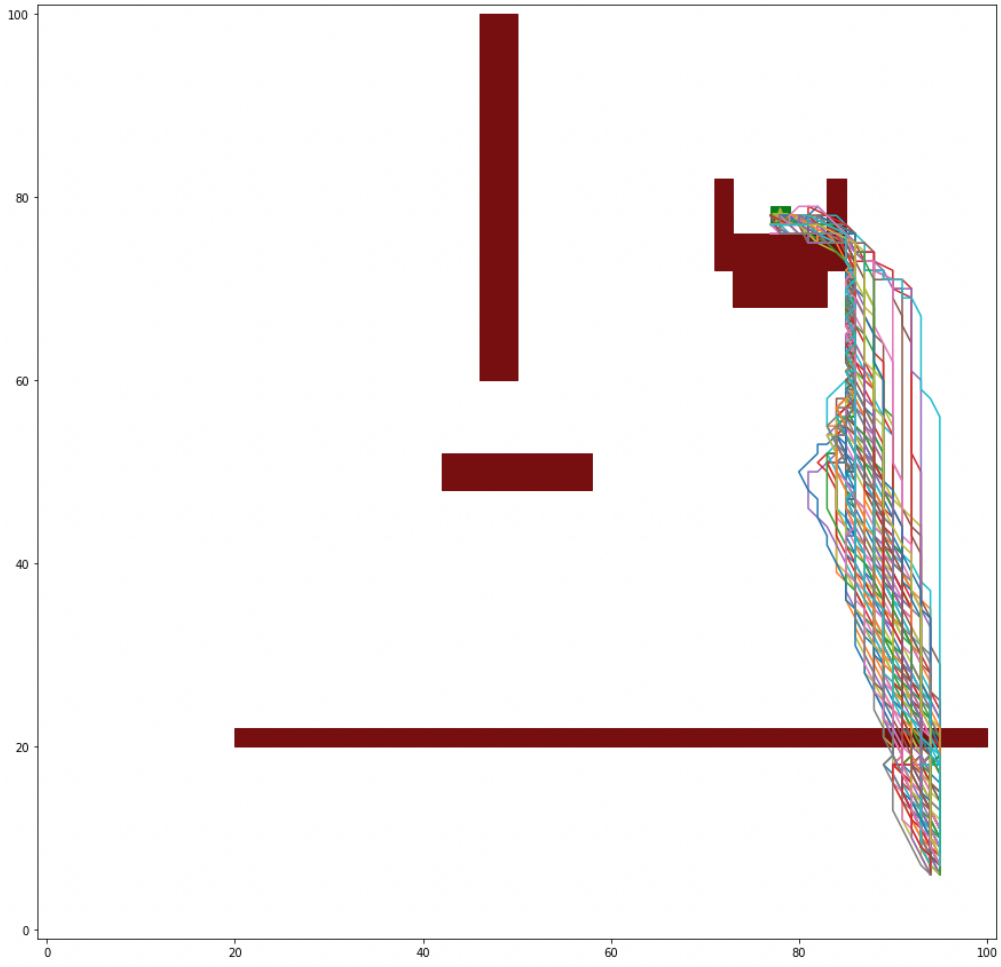

Figura 7: Iteración por política y $\lambda = 0{,}99$

Santiago Morales
201913369

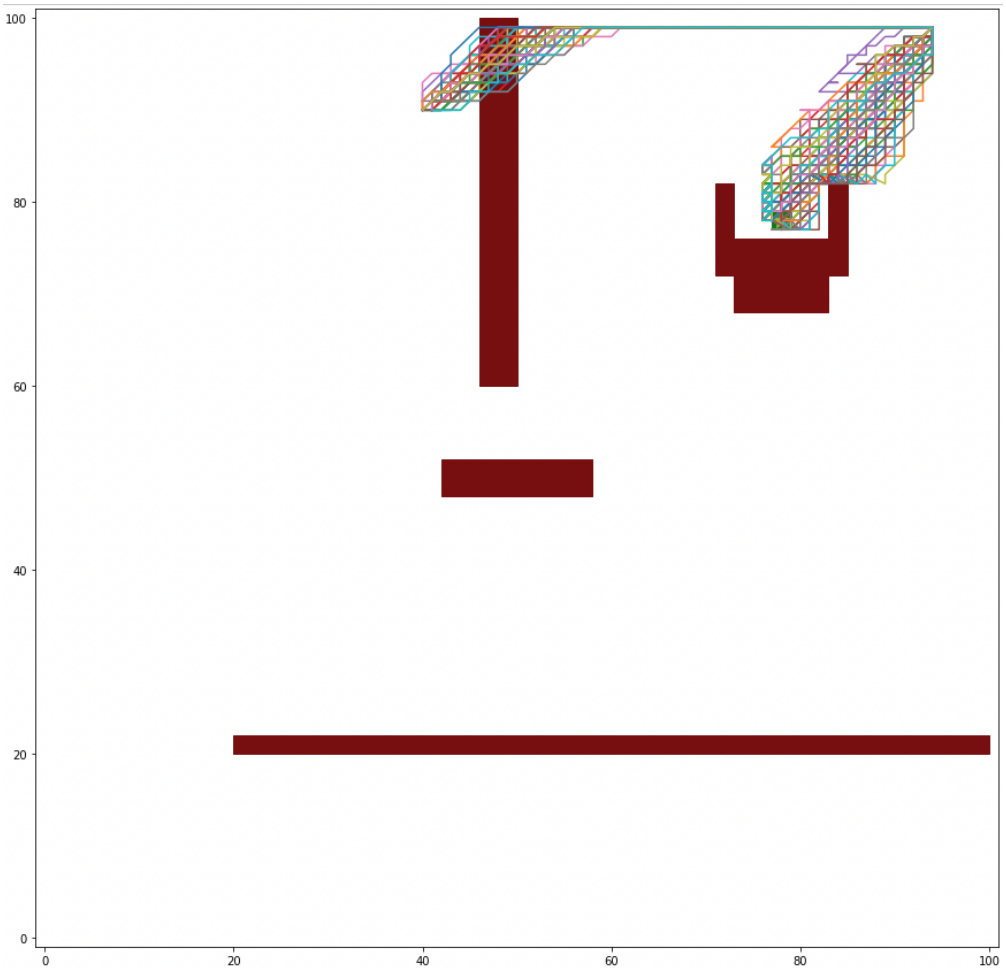

Figura 8: Iteración por política y $\lambda = 0{,}99$