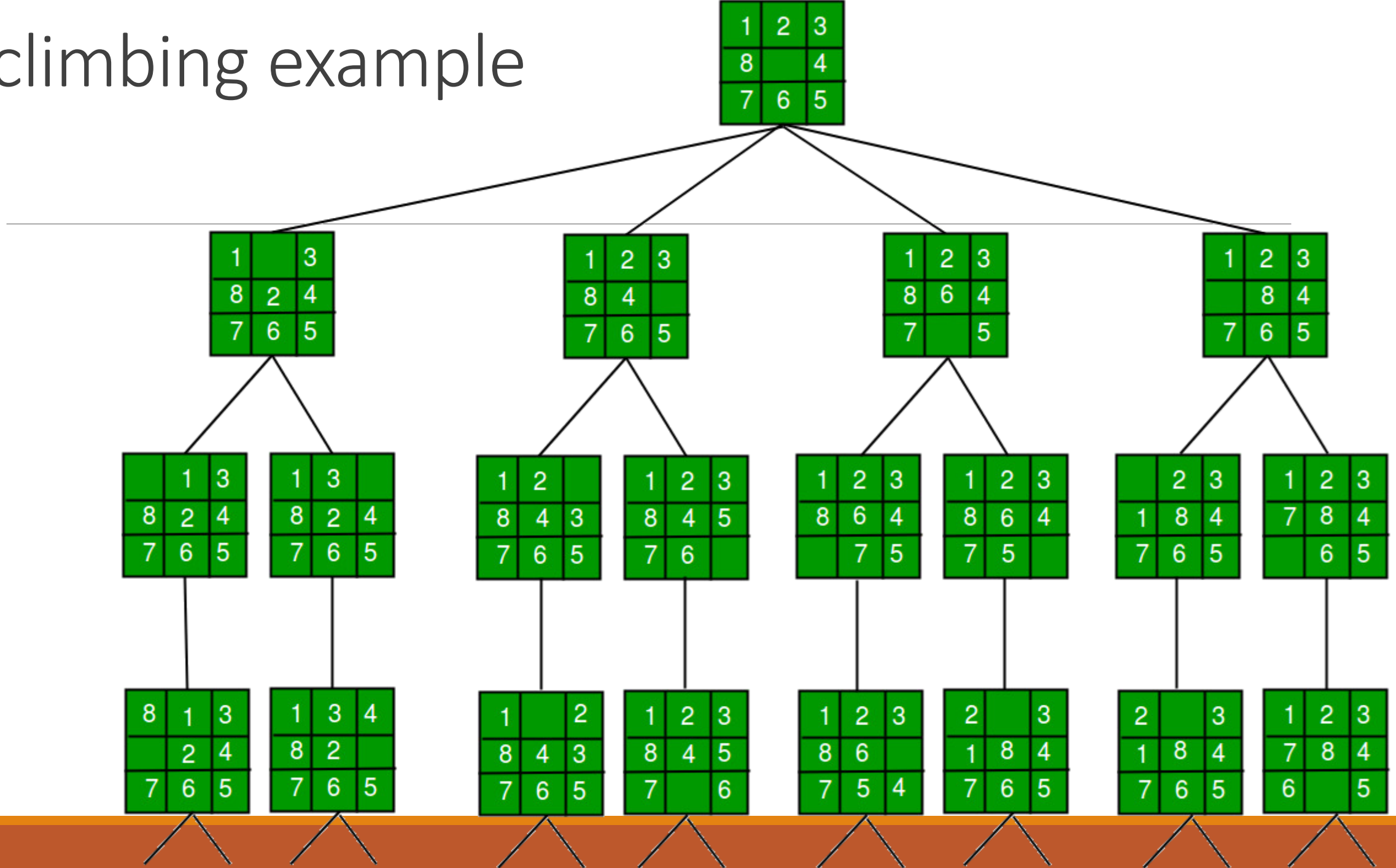


SEARCH IN COMPLEX ENVIRONMENTS

الدكتور مصطفى السيد

Hill climbing example



Manhattan Distance

Manhattan distance is a distance metric between two points in a N dimensional vector space

8	1	3
4		2
7	6	5

board

1	2	3
4	5	6
7	8	

goal

1	2	3	4	5	6	7	8
×	×	✓	✓	×	×	✓	×

Hamming = 5

1	2	3	4	5	6	7	8
1	2	0	0	2	2	0	3

Manhattan = 10
(1 + 2 + 2 + 2 + 3)

Hill climbing example

1	2	3
8		4
7	6	5

0+0+0+0+2+2+2+2
+2=10



1		3
8	2	4
7	6	5

0+1+0+0+2
+2+2+2=9

1	2	3
8	4	
7	6	5

0+0+0+1
+2+2+2+
2=9

1	2	3
8	6	4
7		5

0+0+0+0
+2+1+2+
2=7

1	2	3
	8	4
7	6	5

0+0+0+0+2+2
+2+2=8

2+1+0+2
+1+0+2+
2+2=12

	1	3
8	2	4
7	6	5

1	3	
8	2	4
7	6	5

0+1+2+1
+0+2+2+
2=10

	2	
	4	3
7	6	5

1	2	3
8	4	5
7	6	

6

1	2	3
8	6	4
	7	5

1	2	3
8	6	4
7	5	

6

7

1	2	3
8	6	
7	5	4

2		3
1	8	4
7	6	5

6

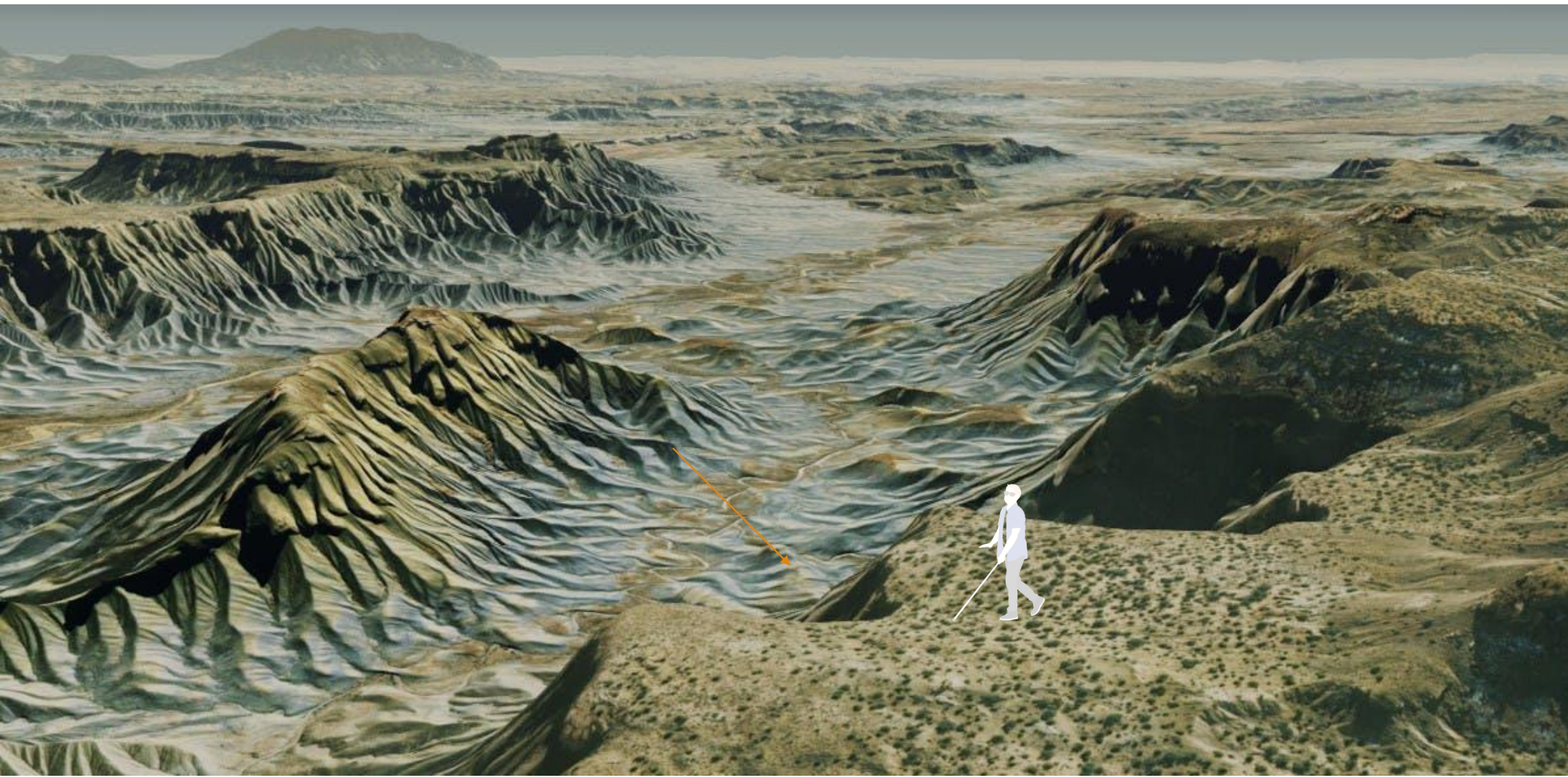
	2	3
1	8	4
7	6	5

1	2	3
7	8	4
	6	5

2		3
1	8	4
7	6	5

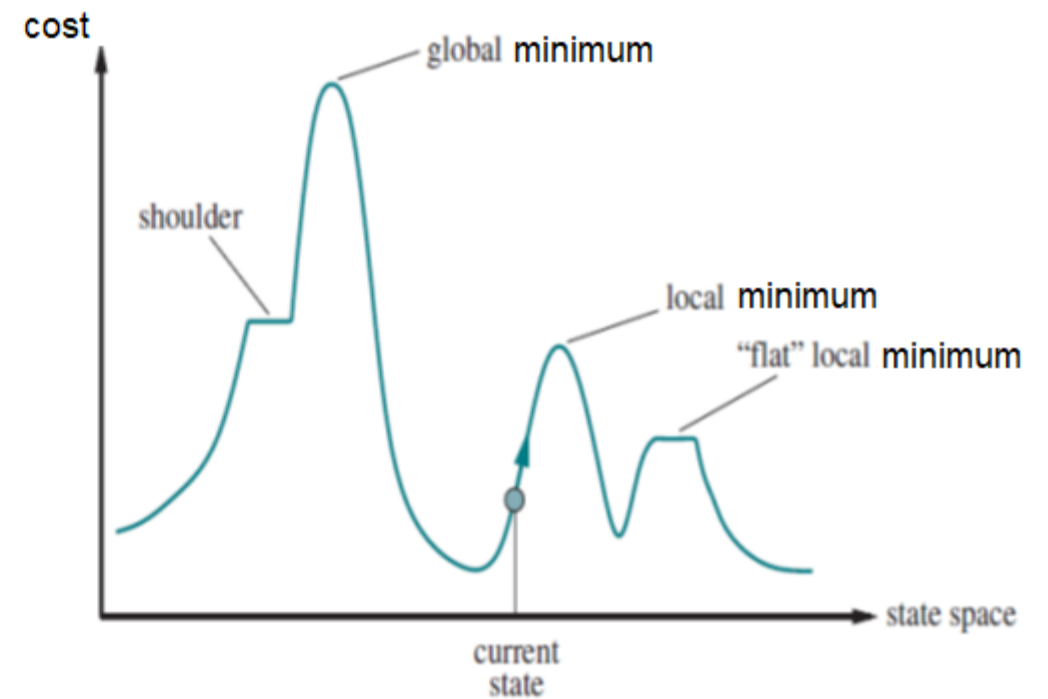
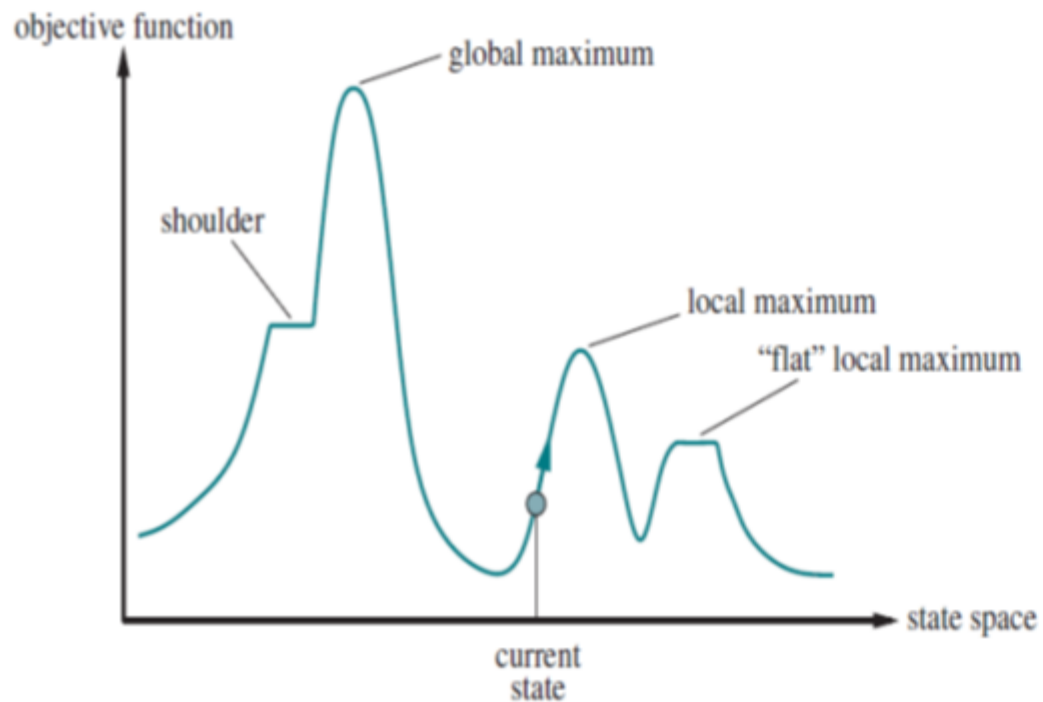
1	2	3
7	8	4
6		5





Local Search and Optimization Problems

Local search algorithms: operate by searching from a start state to neighboring states, without keeping track of the paths, nor the set of states that have been reached.



Definitions

Ridges: a sequence of local maxima that is very difficult for greedy algorithms to navigate

Plateaus: A plateau is a flat area of the state-space landscape

Hill climbing example

1	2	3
8		4
7	6	5

$$0+0+0+0+2+2+2+2+2=10$$



Local minimum

1		3
8	2	4
7	6	5

$$0+1+0+0+2+2+2+2=9$$

1	2	3
8	4	
7	6	5

$$0+0+0+1+2+2+2+2=9$$

1	2	3
8	6	4
7		5

$$0+0+0+0+2+1+2+2=7$$

1	2	3
	8	4
7	6	5

$$0+0+0+0+2+2+2+2=8$$

Ridges

$$2+1+0+2+1+0+2+2=12$$

	1	3
8	2	4
7	6	5

1	3	
8	2	4
7	6	5

$$0+1+2+1+0+2+2+2=10$$

	2	
4	3	
7	6	5

1	2	3
8	4	5
7	6	

Plateaus

6

6

1	2	3
8	6	4
	7	5

1	2	3
8	6	4
7	5	

7

6

1	2	3
8	6	
7	5	4

2		3
1	8	4
7	6	5

shoulder

8

8

	2	3
1	8	4
7	6	5

1	2	3
7	8	4
	6	5

2		3
1	8	4
7	6	5

1	2	3
7	8	4
6		5

Pseudo code (greedy local search)

function HILL-CLIMBING(*problem*) **returns** a state that is a local maximum

current \leftarrow *problem*.INITIAL

while *true* **do**

neighbor \leftarrow a highest-valued successor state of *current*

if VALUE(*neighbor*) \leq VALUE(*current*) **then return** *current*

current \leftarrow *neighbor*

Local search and optimization

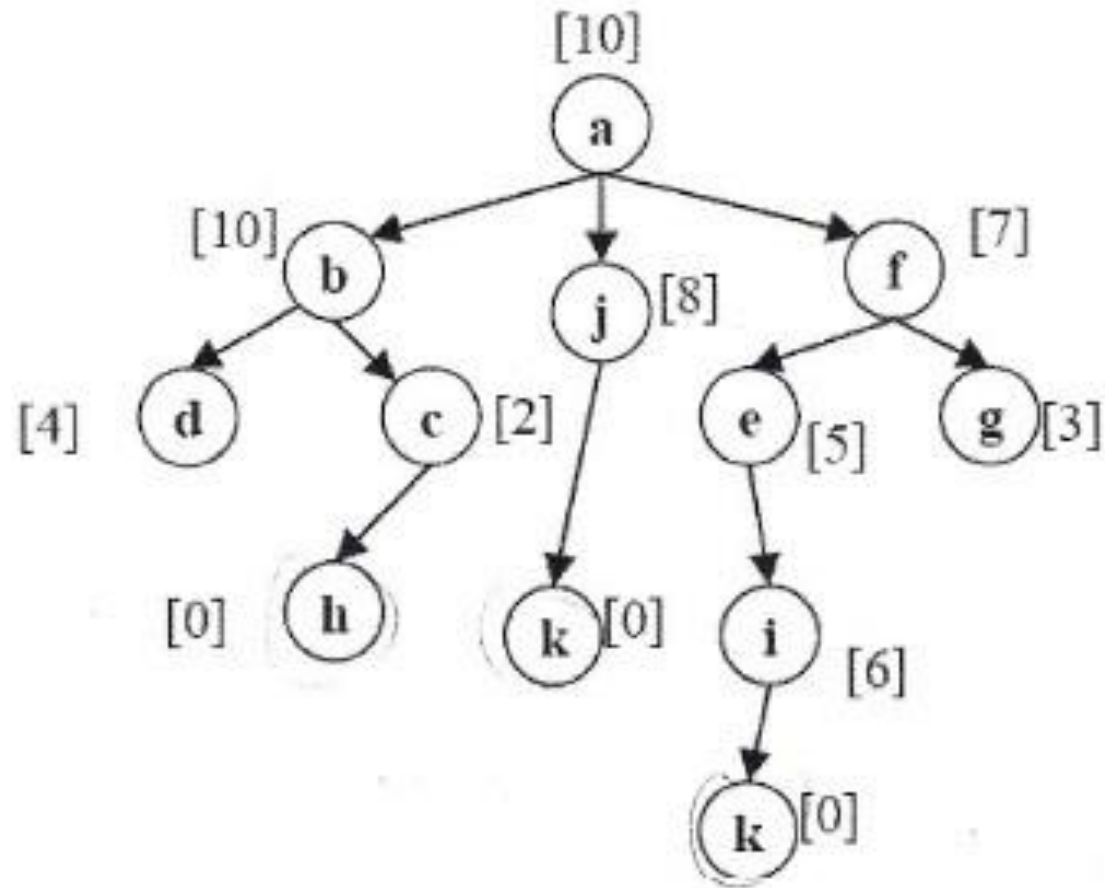
Local search

- Keep track of single current state
- Move only to neighboring states
- Ignore paths
- Use little space

Advantages:

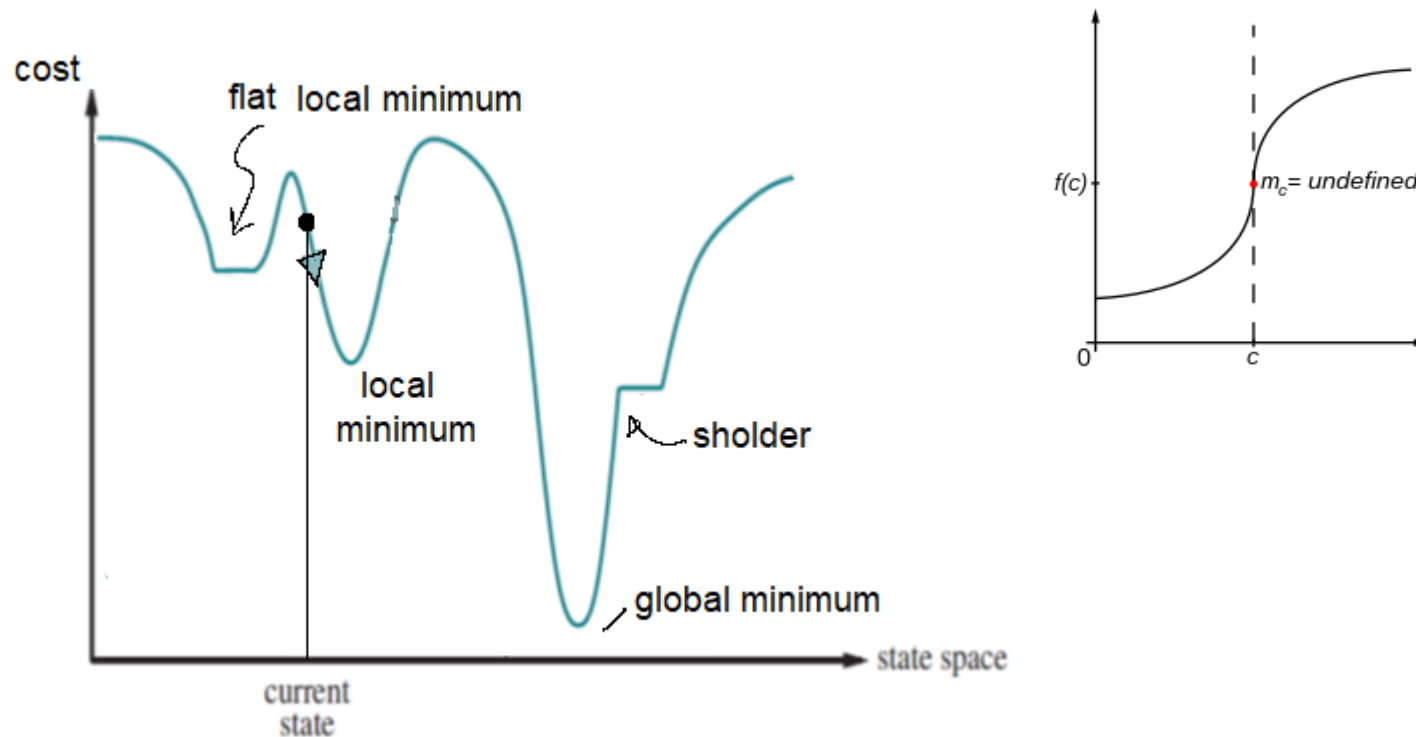
- Use very little memory
- Can often find reasonable solutions in large or infinite (continuous) state spaces.

Example

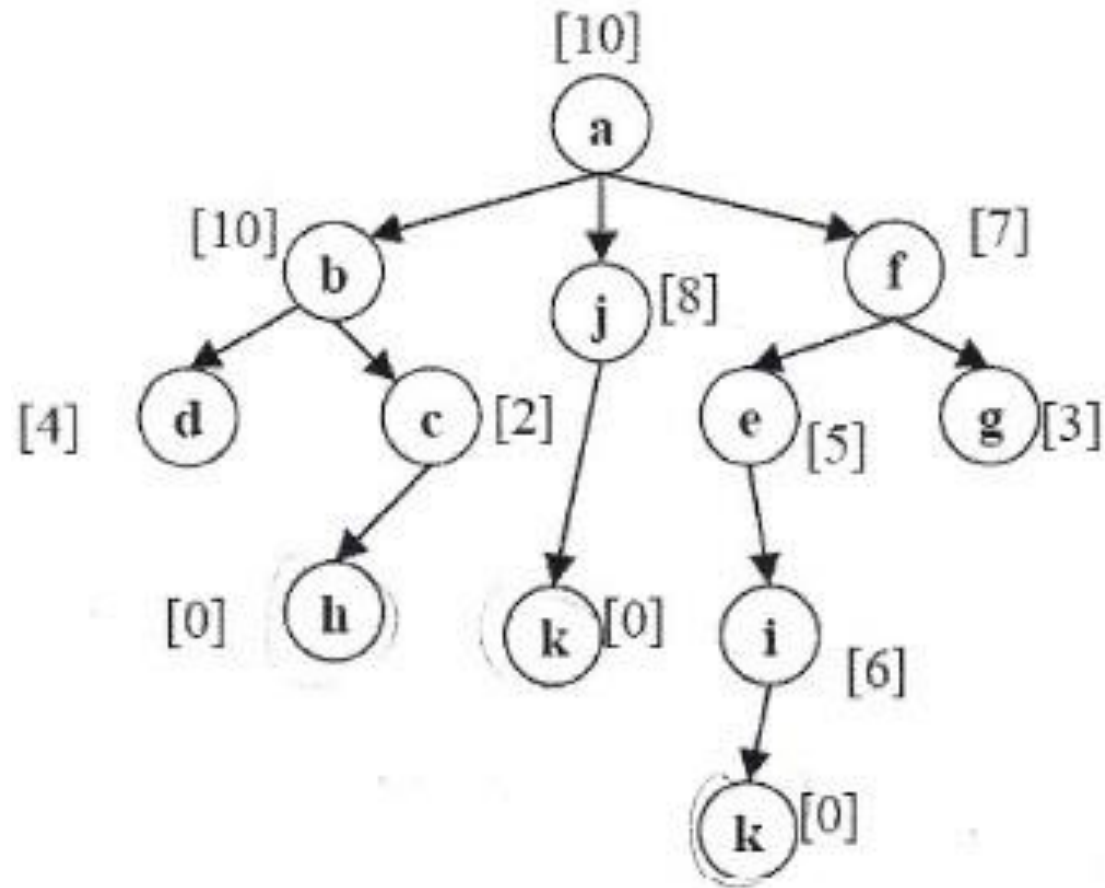


Definitions

gradient descent (also often called **steepest descent**) is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function.



Example



Variants Of Hill Climbing

- 1. Stochastic hill climbing** :chooses at Stochastic hill climbing random from among the uphill moves;
- 2. First-choice hill climbing** : implements stochastic First-choice hill climbing hill climbing by generating successors randomly until one is generated that is better than the current state.
 - For each restart: run until termination vs. run for a fixed time
 - Run a fixed number of restarts or run indefinitely
- 3. Random-restart Hill Climbing**,: If at first you Random-restart hill climbing don't succeed, try, try again." It conducts a series of hill-climbing searches from randomly generated initial states, until a goal is found.

Simulated Annealing

Combine hill climbing with a random walk in a way that yields both efficiency and completeness.

function SIMULATED-ANNEALING(*problem, schedule*) **returns** a solution state

current \leftarrow *problem*.INITIAL

for $t = 1$ **to** ∞ **do**

$T \leftarrow$ *schedule*(t)

if $T = 0$ **then return** *current*

next \leftarrow a randomly selected successor of *current*

$\Delta E \leftarrow$ VALUE(*current*) – VALUE(*next*)

if $\Delta E > 0$ **then** *current* \leftarrow *next*

else *current* \leftarrow *next* only with probability $e^{\Delta E/T}$

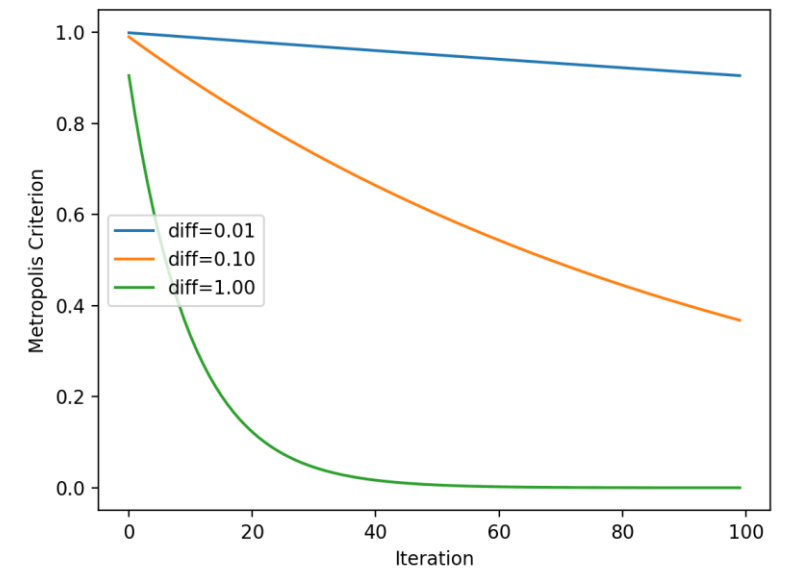
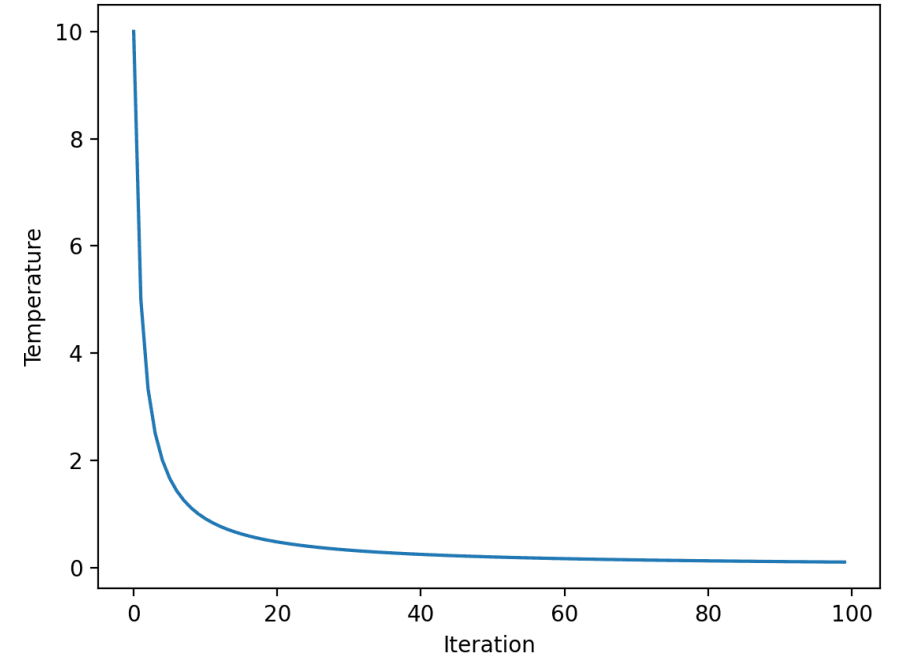
Instead of picking the best move, however, it picks a random move

Boltzmann distribution

Simulated Annealing

if $\Delta E > 0$ **then** $current \leftarrow next$
else $current \leftarrow next$ only with probability $e^{\Delta E/T}$


high T: probability of “locally bad” move is higher
low T: probability of “locally bad” move is lower

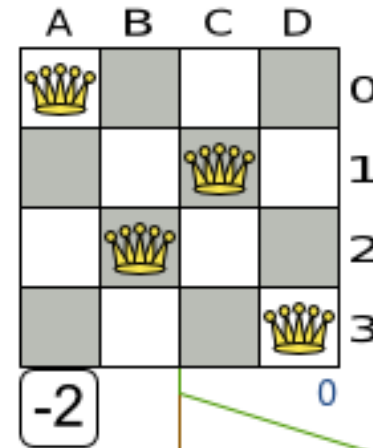


Simulated Annealing (Time Gradient aware)

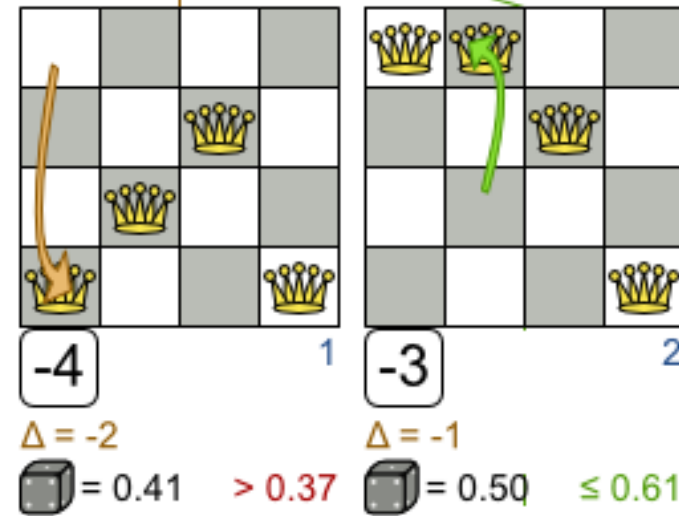
N queens (n = 4, startingTemperature = 2)

Temperature decreases for each step

Step 0		
t	Δ	max 
2.0	≥ 0	any
	-1	0.61
	-2	0.37
	-3	0.22
	-4	0.14




$$\max \text{cube} = e^{\Delta/t}$$

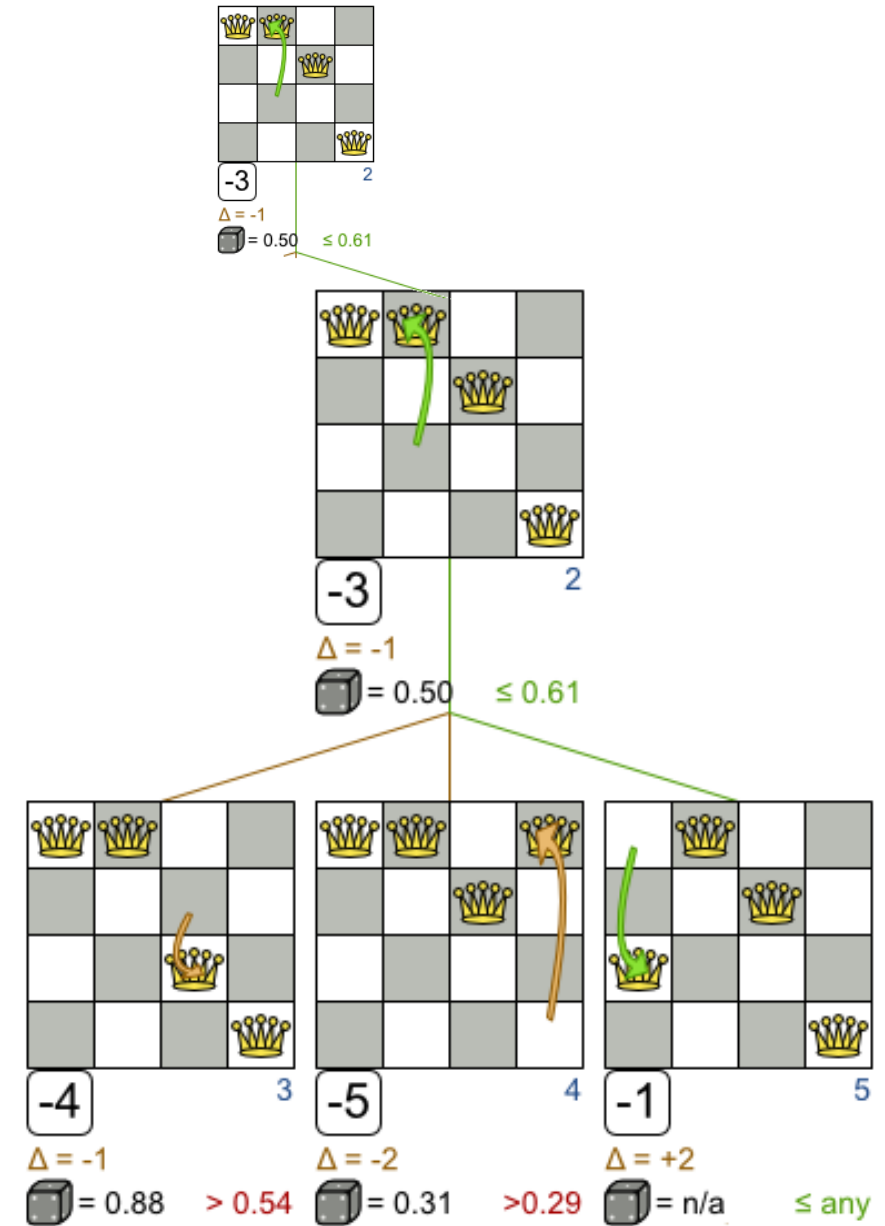


Simulated Annealing (Time Gradient aware)

N queens (n = 4, startingTemperature = 2)

Step 1


t	Δ	max 
1.6	≥ 0	any
	-1	0.54
	-2	0.29
	-3	0.15
	-4	0.08

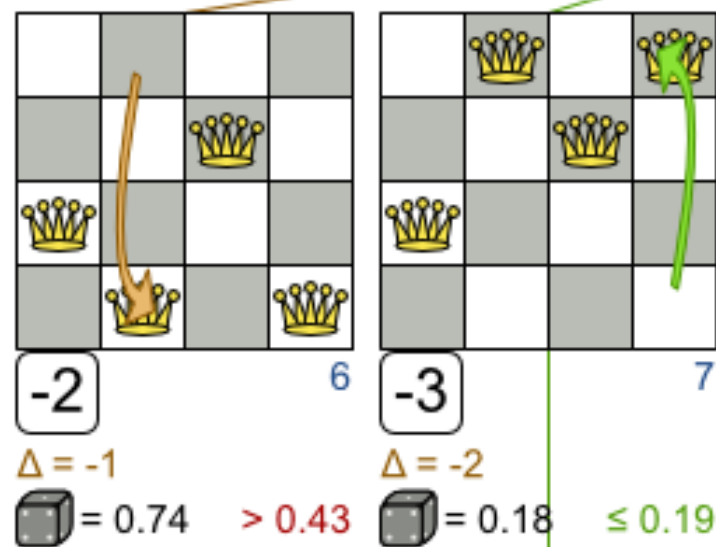


Simulated Annealing (Time Gradient aware)

N queens (n = 4, startingTemperature = 2)

Step 2


t	Δ	max 
1.2	≥ 0	any
	-1	0.43
	-2	0.19
	-3	0.08
	-4	0.04

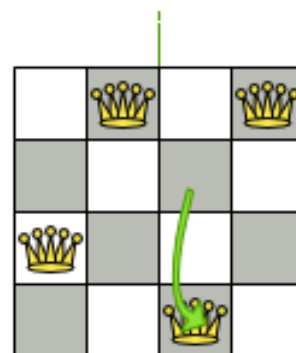


Simulated Annealing (Time Gradient aware)

N queens (n = 4, startingTemperature = 2)

Step 3


t	Δ	max 
0.8	≥ 0	any
	-1	0.29
	-2	0.08
	-3	0.02
	-4	0.01



-1


8

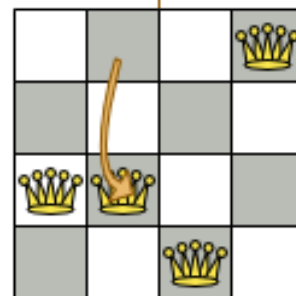
$\Delta = +2$

 = n/a

\leq any

Step 4


t	Δ	max 
0.4	≥ 0	any
	-1	0.08
	-2	0.01
	-3	0.00
	-4	0.00

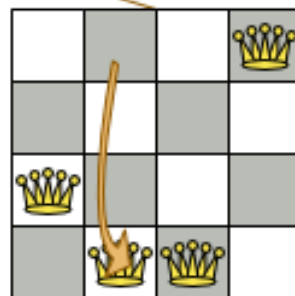


-3

9

$\Delta = -2$

 = 0.97 > 0.01

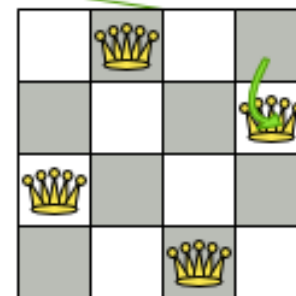


-2

10

$\Delta = -1$

 = 0.11 > 0.08



0

11

$\Delta = +1$

 = n/a \leq any

Local beam search

Idea: Keeping only **one** node in memory is an extreme reaction to memory problems.

Local beam save n nodes in stack: $k = 1 \rightarrow$ Hill climbing , $k = \infty \rightarrow$ Best first search

variant called **Stochastic Beam Search**

Example

