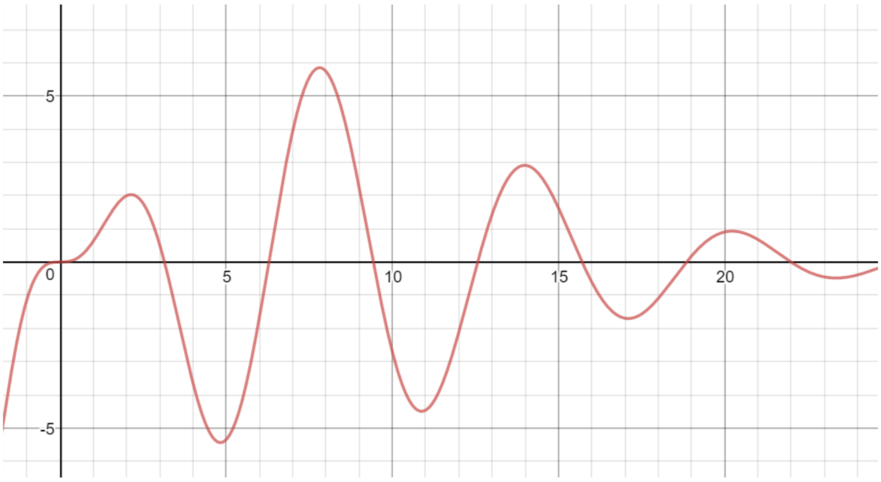
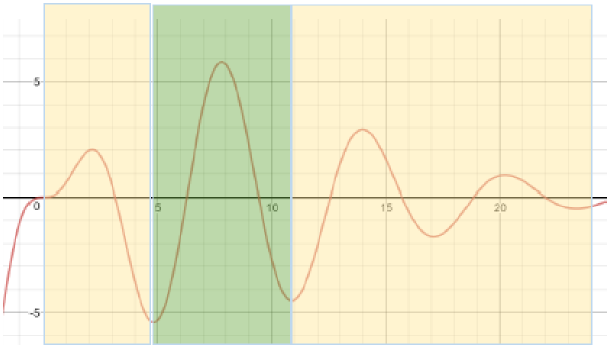
Local Search – Practice Questions



Consider above function in dimension 1

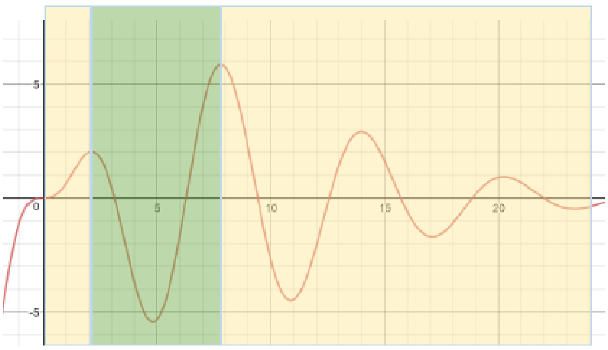
1. Indicate on the graph above which regions would lead to the global maximum, and which would lead to local maxima after a sufficient number of iterations.



2. Assuming each initial condition in [0,24] is equally likely, approximately how many restarts would it take if we implemented a hill climb with random restarts?

About 4.

3. Similarly, indicate which regions would lead to the global minimum, and which would lead to local minima after a sufficient number of iterations.



4. Assuming each initial condition in [0,24] is equally likely, approximately how many restarts would it take if we implemented a hill climb with random restarts?

About 4

5. As the computational time increases (towards infinity), is the standard hill climb guaranteed to find a solution?

No, the returned solution depends on the initial condition. There is also no guarantee that a solution exists.

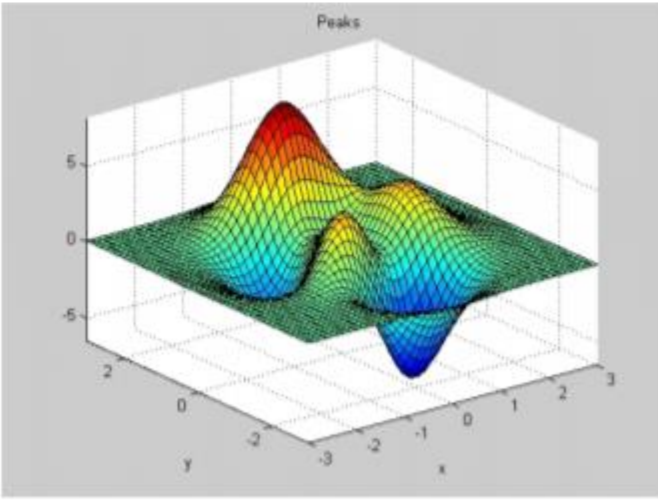
6. As the computational time increases (towards infinity), is the hill climb with random restarts guaranteed to find a solution?

   No, the returned solution depends on the initial condition. There is also no guarantee that a solution exists.

7. What is the difference between gradient ascent and hill climbing in this case? Which method do you prefer?

   They are similar in dimension one. Both methods will work.

Consider following function in dimension 2:



8. What is the main challenge using hill climbing on this function?

In dimension 1 there are only two directions: either increase x or decrease x. In dimension 2 there are infinitely many directions. With a fixed step size. Hill climbing will need to compute function value of sufficiently many points on the circle to determine the direction.

9. What is the difference between gradient ascent and hill climbing in this case? Which method do you prefer?

Gradient descent will give a explicit direction while hill climbing needs to compute many points to derive the optimal direction. In dimension 2 or higher we should use gradient descent if the function is differentiable.

Gradient Descent

In this problem we aim to minimize function $f(x) = (x-1)^2$ using gradient descent.

1. What is the gradient $\nabla f(x)$?

2(x-1)

2. What is the minimizer of this function (by inspection).?

   X=1

3. Using $x_0 = 0$ as initial point. Run 5 iterations of gradient descent using different step size η.(you can either do it manually or write a program). Give the value of $x_1 \sim x_5$ and $y_1 \sim y_5$.
   1) $\eta = 0.1$
   2) $\eta =$ 0.5
   3) η=1
   4) $\eta = 2$

What will happen if step size is too small? What will happen if step size is too big?

One can run following matlab code with different η value to derive the solution:

```
a=0%initialization
eta=0.1%step size
for i=1:5%number of iteation
   a=a-eta*(a-1);
   disp(['The ' num2str(i) '-th iterate is ' 'x=' num2str(a) ])
   disp(['The ' num2str(i) '-th function value is ' 'y=' num2str((a-1)^2) ])
end
```

If step size is too small like 1). It will take many iterations to reach optimal solution.
If step size is too large like 4) The oscillation will happen. One should decrease step size in this case.