Optimization Methods		Date:	*06/03/2018*
Instructor: *Sujit Prakash Gujar*		Scribes:	Anandaswarup Vadapalli (201050048) Moin Hussain Moti (201501066) Yash Goyal (201502181)

Lecture 16: Unconstrained Optimization

1 Recap (Necessary conditions for local extrema)

In this section, we recap some necessary conditions to find a local minimum (maximum) of a function $f : \mathbb{R}^n \to \mathbb{R}$.

1.1 Necessary conditions for local minimum (maximum)

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at x^* and

$$(i)\nabla f(x^*) = 0 \tag{1}$$

$$(ii)\nabla^2 f(x^*) \text{ is positive-semi definite (is negative-semi definite)} \tag{2}$$

then, x^* is a local minimum (maximum).
Equations 1 & 2 are called the First and Second order necessary conditions respectively.

Examples

- $f(x) = x^2$ (minimum at x = 0)

Figure 1: $f(x) = x^2$

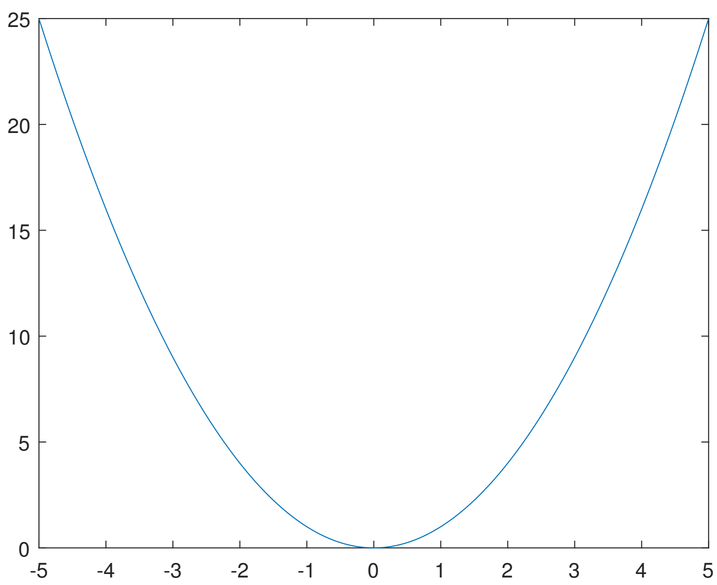
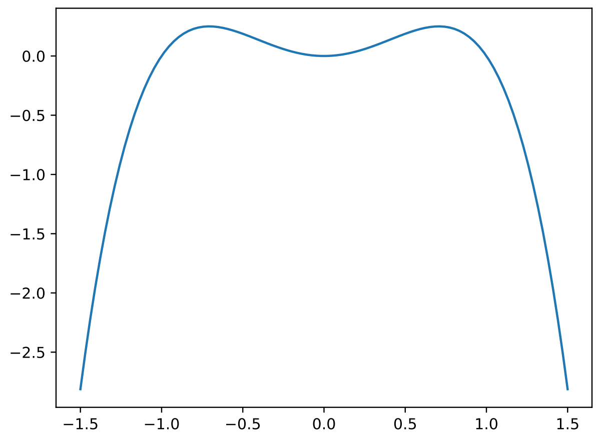


Figure 2: $f(x) = x^2 - x^4$



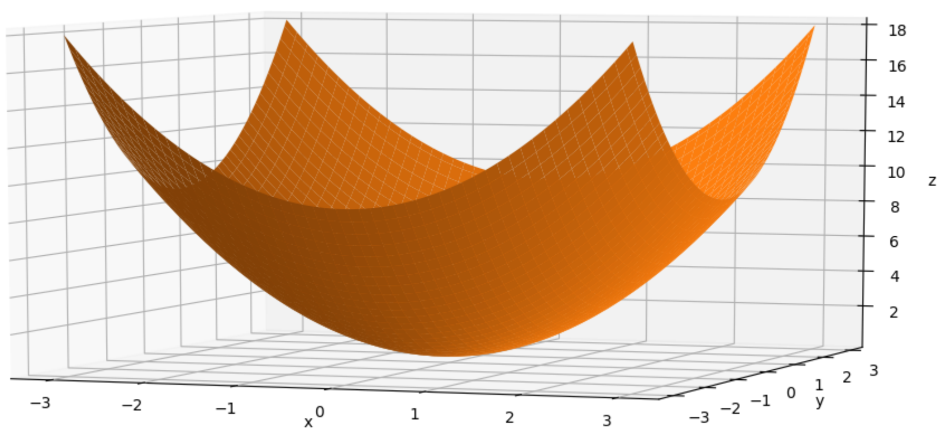
- $f(x) = x^2 - x^4$
  $f'(x) = 2x - 4x^3$
  $f'(x) = 0 \quad => \quad x = 0, -1/\sqrt{2}, 1/\sqrt{2}$

  $f'(x) = 2 - 12x^2$

  Thus, x = 0 gives local minimum.
- $f(x, y) = x^2 + y^2$

  $$\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Figure 3: $f(x, y) = x^2 + y^2$

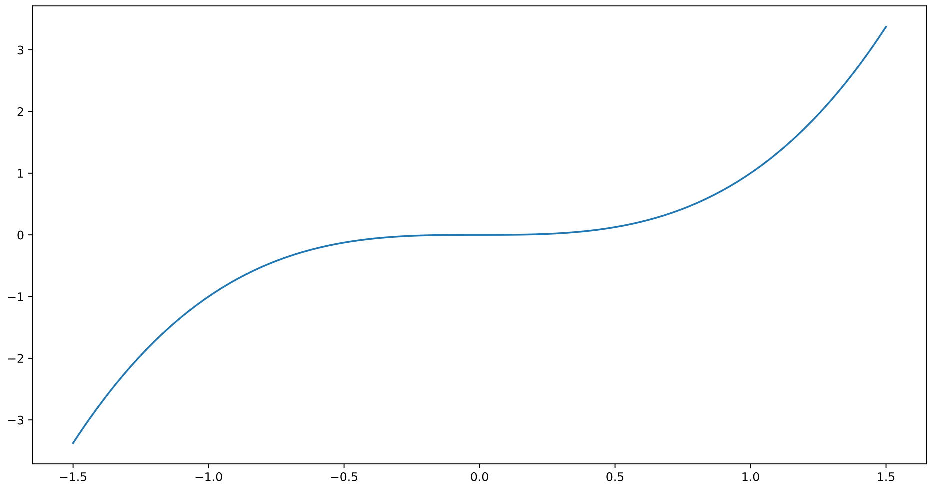


$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
Since $\nabla f(0,0) = 0$ and $\nabla^2 f(0,0)$ is positive semi-definite, (x,y) = (0,0) gives local minimum.

- $f(x) = x^3$

  $f'(x) = 3x^2$

Figure 4: $f(x) = x^3$



$f''(x) = 6x$
at $x = 0, f'(x) = 0$ and $f''(x) >= 0$
but still, it's not local minimum, which shows that the above conditions are necessary but not sufficient.

1.2 Saddle points

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at x^* and

$$(i) \nabla f(x^*) = 0 \tag{3}$$

$$(ii) \nabla^2 f(x^*) \text{ is neither positive semi- definite nor negative semi-definite} \tag{4}$$

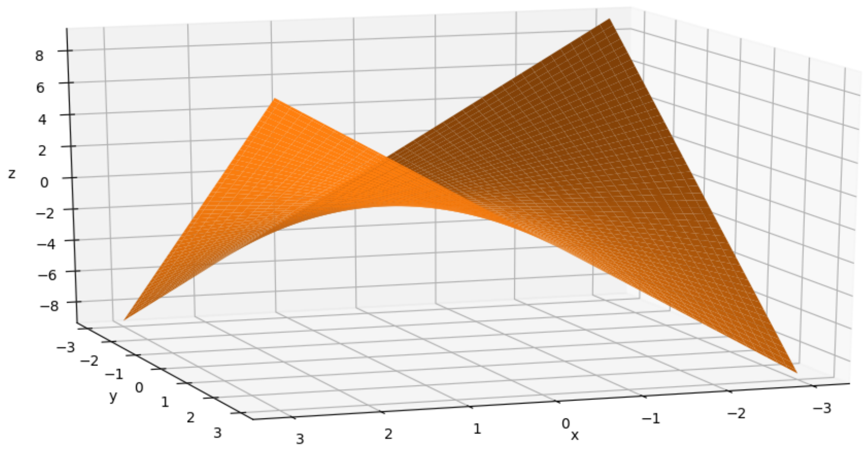
then, x^* is a saddle point.

Examples

- $f(x,y) = xy$

  $\nabla f(x,y) = \begin{bmatrix} y \\ x \end{bmatrix}$
  $\nabla^2 f(x,y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
  $\nabla^2 f(x,y)$ is neither positive semi-definite nor negative semi-definite,
  Thus, (0,0) is a saddle point.

Figure 5: $f(x, y) = xy$



2 Sufficient conditions for local minimum

Let us consider a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, and it's linear approximation at a point x^* (i.e. $f(x^* + \alpha d)$). Applying Taylor's series expansion we get,

$$f(x^* + \alpha d) - f(x^*) = \alpha^\intercal \nabla f(x^*) + \frac{\alpha^2}{2} d^\intercal \nabla^2 f(x^*) d \tag{5}$$

If point x^* is a local minimum, then $f(x^* + \alpha d) - f(x^*) > 0$ (sufficient condition). That is,

$$\alpha^\intercal \nabla f(x^*) + \frac{\alpha^2}{2} d^\intercal \nabla^2 f(x^*) d > 0 \tag{6}$$

Now from the first order necessary condition (Equation 1), we know that $\nabla f(x^*) = 0$. Substituting in Equation 6 we get

$$\frac{\alpha^2}{2} d^\intercal \nabla^2 f(x^*) d > 0 \tag{7}$$

$$\Rightarrow \quad d^\intercal \nabla^2 f(x^*) d > 0 \tag{8}$$

i.e., $\nabla^2 f(x^*)$ is positive definite

2.1 Sufficient conditions for local minimum (maximum)

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice differentiable at x^* and

$$(i) \nabla f(x^*) = 0 \tag{9}$$

$$(ii) \nabla^2 f(x^*) \text{ is positive definite (is negative definite)} \tag{10}$$

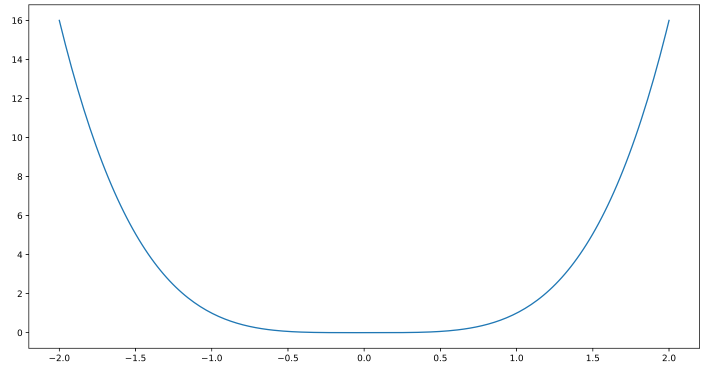
then, x^* is a local minimum (maximum).

These conditions are called second order sufficient conditions. These conditions are sufficient but not necessary.

Example

- $f(x) = x^4$
  $f'(x) = 4x^3$

Figure 6: $f(x) = x^4$



  $f''(x) = 12x^2$
  x = 0 doesn't satisfy the above mentioned conditions,
  but still it x = 0 is a local minimum

- $f(x, y) = y^2 - x^2$

  $\nabla f(x, y) = \begin{bmatrix} -2x \\ 2y \end{bmatrix}$
  $\nabla^2 f(x, y) = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$
  Since $\nabla f(0, 0) = 0$ and $\nabla^2 f(0, 0)$ is neither positive definite nor negative definite, (x,y) = (0,0) is a saddle point.

3 Relationship between convexity and extrema

Lemma 3.1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex, then it has a unique global minimum*

Proof. We proceed with proof by contradiction.
Suppose x^* and x^{**} are two local minima of f and $x^* \neq x^{**}$ and $f(x^*) < f(x^{**})$.
Since, $f(x)$ is strictly convex,

$$f(\lambda x^* + (1 - \lambda)x^{**}) < \lambda f(x^*) + (1 - \lambda)f(x^{**}) \tag{11}$$

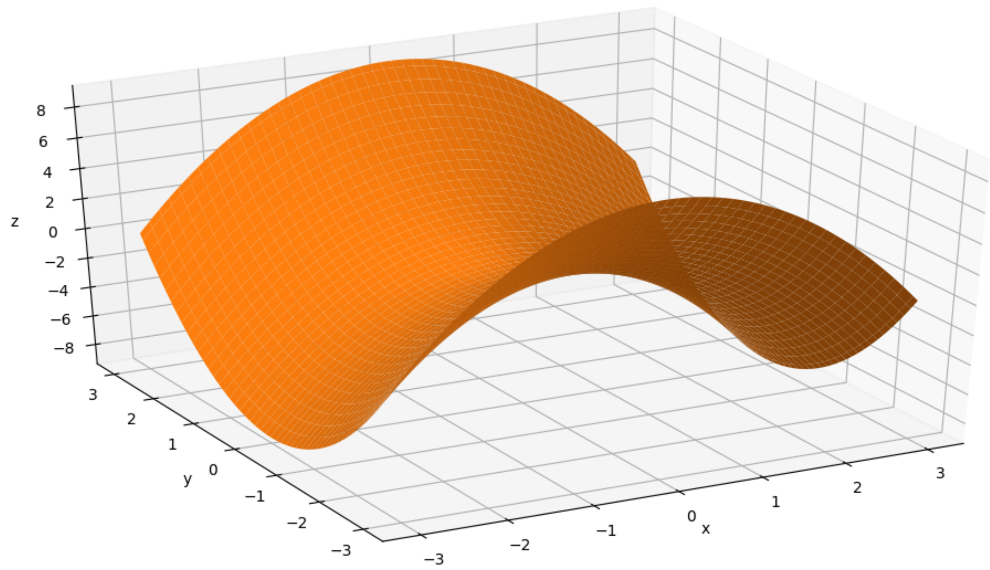
from the definition of strict convexity.
Now since $f(x^*) < f(x^{**})$ (from our initial assumption), Equation 11 becomes

$$f(\lambda x^* + (1 - \lambda)x^{**}) < f(x^{**}) \tag{12}$$

As $\lambda \to 0$, this results in $f(x^{**}) < f(x^{**})$, which is a contradiction.
$\therefore$ If $f : \mathcal{R}^n \to \mathcal{R}$ is strictly convex, it has a unique global minumum.

$\square$

Figure 7: $f(x, y) = y^2 - x^2$



4 Descent methods

As we have seen in the previous sections, to find local minima of a function $f : \mathcal{R}^n \to \mathcal{R}$, we find that x for which $\nabla f(x) = 0$.

However in most cases in the real world, solving $\nabla f(x) = 0$ analytically may be computationally expensive/not feasible. In such cases, we use a class of methods called **descent methods** to reach some local minimum.

In descent methods, we start with some value x_0 and keep moving with steps proportional to the negative of the gradient of the function at the current point, in the hope of reaching some local minima.

Examples

- $f(x, y) = x + y$

  $$\nabla f(x, y) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

  move towards (-1,-1)

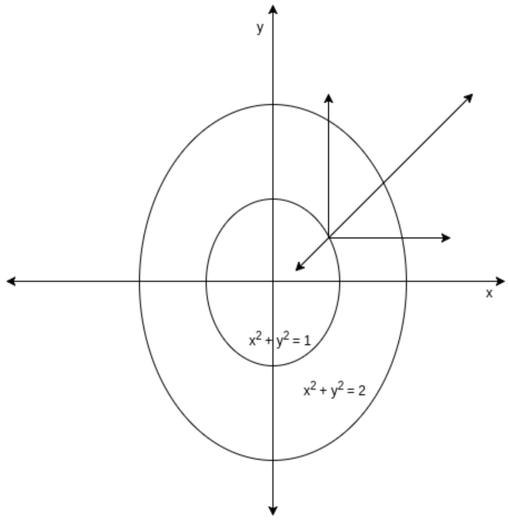
- $f(x, y) = x^2 + y^2$

  $$\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

  move towards $(-2x, -2y)$ where x, y is the current point.
  As can be seen from Figure 8, the value of $f(x, y)$ is increasing in both x and y directions but the maximum increase is in the centrifugal direction, i.e. the direction of $\nabla f(x)$. Hence, to find the local minima of $f(x, y)$ using descent methods, we move opposite to the direction of maximum increase in gradient, i.e. in the direction of $-\nabla f(x)$.

More formally, to find the local minima of a function $f : \mathbb{R}^n \to \mathbb{R}$, the update step is

Figure 8: $f(x,y) = x^2 + y^2$



$$x_{k+1} = x_k - \alpha \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|} \tag{13}$$

where x_{k+1} and x_k are the value of x at the k^{th} and $k+1^{th}$ step respectively. We stop the update when $x_{k+1} - x_k < \epsilon$ where ϵ is some small value.

Examples

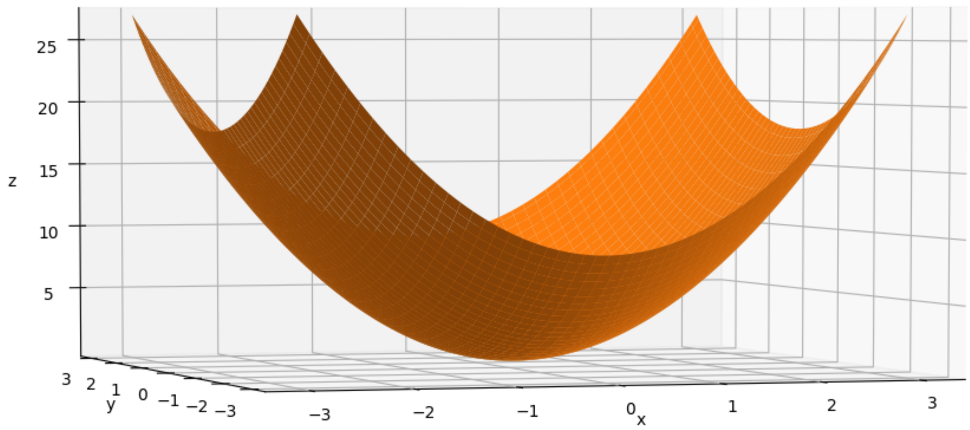
- $f(x,y) = 2x^2 + y^2$

  $\nabla f(x,y) = \begin{bmatrix} 4x \\ 2y \end{bmatrix}$
  $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
  $\nabla f(x_0) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$
  $\alpha = \|\nabla f(x_k)\|$
  $x_1 = x_0 - \nabla f(x_0)$
  $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$
  $x_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix} - \begin{bmatrix} -12 \\ -2 \end{bmatrix} = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$
  $x_3 = \begin{bmatrix} 9 \\ 1 \end{bmatrix} - \begin{bmatrix} 36 \\ 2 \end{bmatrix} = \begin{bmatrix} -27 \\ -1 \end{bmatrix}$
  Thus, $\alpha = \|\nabla f(x_k)\|$ is not converging.

- $f(x) = x^2$

  $\nabla f(x) = 2x$
  $x_0 = -5$
  $\alpha = 1$
  $x_1 = -5 - 2(-5)/10 = -4$
  $x_2 = -4 - 2(-4)/8 = -3$
  $x_3 = -2$

Figure 9: $f(x, y) = 2x^2 + y^2$



$x_4 = -1$
$x_5 = 0$
Now gradient is zero, thus we've reached the minimum.
If $x_0 = -2.5$
$x_1 = -1.5$
$x_2 = -0.5$
$x_3 = 0.5$
$x_4 = -0.5$
$x_5 = 0.5$
and so on.
Thus fixing α is not a good solution.

$$g(\alpha) = f(x_k - \alpha \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}) \tag{14}$$

Solving $g'(\alpha) = 0$
$\alpha_k = argmin_\alpha g(\alpha)$

Examples

- $f(x) = x^2$

  $\nabla f(x) = 2x$
  $x_0 = -5$
  $g(\alpha) = f(-5 + \alpha)$
  $g(\alpha) = (-5 + \alpha)^2$
  $g'(\alpha) = 0$
  $\alpha = 5$
  Thus, $x_1 = x_0 - \alpha \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$
  $x_1 = -5 - (5)(-10)/10$

$x_1 = 0$

- $f(x, y) = 2x^2 + y^2$

  $\nabla f(x, y) = \begin{bmatrix} 4x \\ 2y \end{bmatrix}$
  $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
  $g(\alpha) = f(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} 4 \\ 2 \end{bmatrix})$
  $g(\alpha) = (1 - 2\alpha)^2 + 2(1 - 4\alpha)^2$
  $g'(\alpha) = 0$
  $\alpha = 5/18$

  $x_1 = \begin{bmatrix} -1/9 \\ 4/9 \end{bmatrix}$
  $x_2 = \begin{bmatrix} 1/81 \\ 16/81 \end{bmatrix}$
  $x_3 = \begin{bmatrix} -1/729 \\ -16/729 \end{bmatrix}$

Figure 10: $f(x, y) = 2x^2 + y^2$