Optimization Methods		Date:	*6th March, 2018*
Instructor: *Sujit Prakash Gujar*		Scribes:	Abhijith Nair Deepanshu Garg Vineeth Chelur

Lecture 16: Sufficient Conditions and Descent Methods

1 Recap

In lecture 15, the basics of non-linear program, optimality conditions and their proofs were covered

1.1 Non-linear Programs

Programs in which either the constraints or the objective function is non-linear.
Eg. Fermat-Weber Problem

1.2 Optimality Conditions

Consider a function $f : \Re^n \to \Re$

Global minimum: x^* is called a global minimum of f if $f(x^*) \leq f(x)\ \forall x \in \Re^n$

Local minimum: x^* is called a local minimum of f if $\exists \epsilon > 0$ s.t. $f(x^*) \leq f(x)\ \forall x \in ||x - x^*|| < \epsilon$

Global maximum: x^* is called a global maximum of f if $f(x^*) \geq f(x)\ \forall x \in \Re^n$

Local maximum: x^* is called a local maximum of f if $\exists \epsilon > 0$ s.t. $f(x^*) \geq f(x)\ \forall x \in ||x - x^*|| < \epsilon$

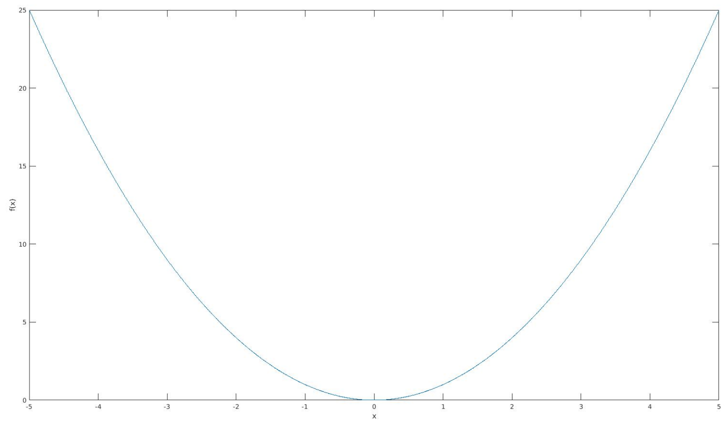
First Order Necessary Condition: If f is continuously differentiable in an open set S containing x^* and x^* is a local minimum, then $\nabla f(x^*) = 0$

Second Order Necessary Condition: If f is twice continuously differentiable in an open set S containing x^* and x^* is a local minimum, then $\nabla^2 f(x^*)$ is a positive semi-definite matrix

2 Necessary vs Sufficient Conditions

Consider the following examples where x^* is the solution of $\nabla f(x^*) = 0$

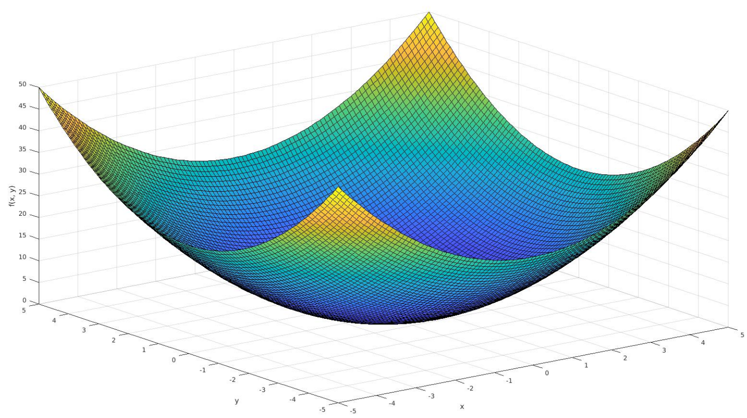
2.1 $f(x) = x^2$



$$\nabla f(x) = 2x \implies 2x^* = 0 \implies x^* = 0$$
$$\nabla^2 f(x) = 2 \implies \nabla^2 f(x^*) = 2$$

Global minimum occurs at 0, which satisfies both the necessary conditions.

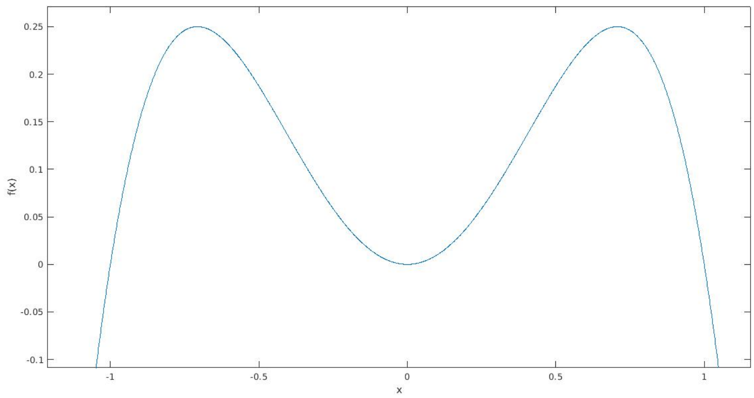
2.2 $f(x, y) = x^2 + y^2$



$$\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} \implies \begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \implies \nabla^2 f(x^*, y^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Global minimum occurs at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which satisfies both the necessary conditions.

2.3 $f(x) = x^2 - x^4$

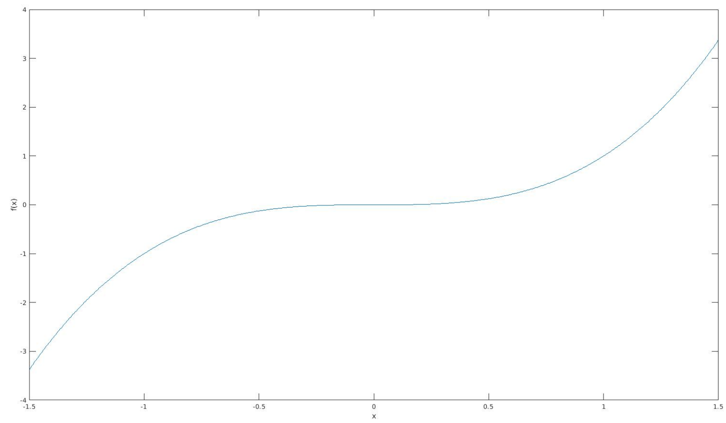


$$\nabla f(x) = 2x - 4x^3 \implies 2x^*(1 - 2(x^*)^2) = 0 \implies x^* = 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$$

$$\nabla^2 f(x) = 2 - 12x^2 \implies \nabla^2 f(x^*) = 2, -4, -4$$

For $x^* = 0$, $\nabla^2 f(x^*) \geq 0$, which satisfies both the necessary conditions. Notice that for $x^* = \frac{1}{\sqrt{2}}$ and $-\frac{1}{\sqrt{2}}$, $\nabla^2 f(x^*) \leq 0 \implies x^*$ is a local maximum.

2.4 $f(x) = x^3$



$$\nabla f(x) = 3x^2 \implies 3(x^*)^2 = 0 \implies x^* = 0$$

$$\nabla^2 f(x) = 6x \implies \nabla^2 f(x^*) = 0$$

Notice that $\nabla^2 f(x^*) \geq 0$. According to the necessary conditions, x^* should be a local minimum, but there are values lesser than $f(x^*)$ in its immediate neighbourhood and therefore is not a local minimum. Hence the above conditions are necessary but are not sufficient to get a local minimum of $f(x)$. A new definition is required to get the local minimum of f(x).
P.S. The point at $x^* = 0$ in the above equation is referred to as the saddle point.

3 Second Order sufficient condition

If f is twice differentiable at x^* and

1. $\nabla f(x^*) = 0$ and,
2. $\nabla^2 f(x^*)$ is positive definite

then x^* is a local minimum.
This condition is sufficient but not necessary.

3.1 Not necessary example

x^4 does not satisfy above condition, but still has a local minimum because it satisfies the 2^{nd} order necessary condition.

$$\begin{aligned} f(x) &= x^4 \\ \nabla f &= 4x^3 \\ \nabla^2 f &= 12x^2 \end{aligned}$$

Here, $12x^2$ is not positive definite at $x = 0$ (where $\nabla f(x) = 0$), but, $f(x)$ still has a local minimum at 0.

3.2 Saddle point example

$$\begin{aligned} f(x, y) &= y^2 - x^2 \\ \nabla f &= \begin{bmatrix} -2x \\ -2y \end{bmatrix} \\ \nabla^2 f &= \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} \end{aligned}$$

4 Minima in Convex Function

Strictly convex functions have a unique global minimum.
Proof: Let f be a strictly convex function. Let us assume x^* and x^{**} are two local minima and $x^* \neq x^{**}$ such that $f(x^*) < f(x^{**})$. Now, strict convexity implies that

$$\begin{aligned} & f(\lambda x^* + (1 - \lambda x^{**})) < \lambda f(x^*) + (1 - \lambda f(x^{**})) \\ & f(\lambda x^* + (1 - \lambda x^{**})) < \lambda f(x^{**}) + (1 - \lambda f(x^{**})) && \text{as } f(x^*) < f(x^{**}) \end{aligned}$$

If $\lambda \to 0$, we get $f(x^{**}) < f(x^{**})$, which is not possible.
$\Rightarrow$ We have a contradiction, therefore, our assumption that two local minima exist is wrong.
$\Rightarrow$ Strictly convex functions have a unique global minimum.

5 Arithmetic-Geometric Mean Inequality

Here we prove that geometric mean is lesser than or equal to the arithmetic mean.
To Prove:

$$x_1 x_2 \dots x_n \leq \frac{\sum_{i=1}^{n} x_i}{n}$$

for any set of positive number $x_i,\ i = 1, \dots, n$. By making change of variables

$$y_i = \ln(x_i), \quad i = 1, \dots, n,$$

we have $x_i = e^{y_i}$, so this inequality is equivalently written as

$$e^{\frac{y_1 + \cdots + y_n}{n}} \leq \frac{e^{y_1} + \cdots + e^{y_n}}{n}$$

which must be shown for all scalars $y_1, \dots, y_n$.

Now, we will use optimality conditions to prove this. Therefore, we will minimize

$$e^{y_1} + \cdots + e^{y_n},$$

over all $y = (y_1, \cdots, y_n)$ such that $y_1 + \cdots + y_n = s$ for an arbitary scalar s, and to show that the optimal value is greater than or equal to $ne^{\frac{s}{n}}$. We use the constraint $y_1 + \cdots + y_n = s$ to eliminate the variable y_n, therefore obtaining an unconstrained problem of minimizing

$$g(y_1, \cdots, y_n) = e^{y_1} + \cdots + e^{y_{n-1}} + e^{s-(y_1 + \cdots + y_{n-1})}$$

over y. The first order necessary conditions $\frac{\partial g}{\partial y_i} = 0, \forall i = 1, \dots, n-1$ yield the system of equations

$$e^{y_i} = e^{s-(y_1 + \cdots + y_{n-1})}, \quad \forall i = 1, \dots, n-1$$

Taking log both sides, we get

$$y_i = s - (y_1 + \cdots + y_{n-1}), \quad \forall i = 1, \dots, n-1$$

This system has only one solution, $y_i^* = \frac{s}{n},\ \forall i$. So, we see that at the minimum, the value of the arithmetic mean is equal to the value of the geometric mean. This is sufficient to show the inequality.

6 Step Size

There are multiple ways by which we can choose the step size α. Here we show some problems that can arise due to the method in which the step size is chosen.

1. Minimize the function $f = y^2 + 2x^2$ with starting point $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\alpha_k = ||\nabla f(x_k)||$. That is $x_{k+1} = x_k - \nabla f(x_k)$.

$$f(x, y) = y^2 + 2x^2 \implies \nabla f(x, y) = \begin{bmatrix} 4x \\ 2y \end{bmatrix}$$

Iteration 1

$$\begin{aligned} \nabla f(x_0) &= \begin{bmatrix} 4 \\ 2 \end{bmatrix} \\ x_1 &= x_0 - \nabla f(x_0) \\ x_1 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \end{bmatrix} \end{aligned}$$

Iteration 2

$$\begin{aligned} \nabla f(x_1) &= \begin{bmatrix} -12 \\ -2 \end{bmatrix} \\ x_2 &= x_1 - \nabla f(x_1) \\ x_2 &= \begin{bmatrix} -3 \\ -1 \end{bmatrix} - \begin{bmatrix} -12 \\ -12 \end{bmatrix} = \begin{bmatrix} 9 \\ 1 \end{bmatrix} \end{aligned}$$

Iteration 3

$$\begin{aligned} \nabla f(x_2) &= \begin{bmatrix} 36 \\ 2 \end{bmatrix} \\ x_3 &= x_2 - \nabla f(x_2) \\ x_3 &= \begin{bmatrix} 9 \\ 1 \end{bmatrix} - \begin{bmatrix} 36 \\ 2 \end{bmatrix} = \begin{bmatrix} -27 \\ -1 \end{bmatrix} \end{aligned}$$

So, if we do not choose step size smartly and keep moving in the same direction, we may end up bouncing between positive and negative gradients, not reaching zero.

2. Minimize the function $f = x^2$ with $\alpha = 1$.

   (a) Starting point $x_0 = -5$

$$\begin{aligned} f(x) &= x^2 \\ \nabla f(x) &= 2x \\ x_{k+1} &= x_k - \frac{\nabla f(x_k)}{||\nabla f(x_k)||} \end{aligned}$$

Iteration 1

$$\begin{aligned} x_0 &= -5 \\ \nabla f(x_0) &= -10 \\ x_1 &= x_0 - \frac{\nabla f(x_0)}{||\nabla f(x_0)||} \\ x_1 &= -5 - \frac{-10}{10} = -4 \end{aligned}$$

Iteration 2

$$\begin{aligned} \nabla f(x_1) &= -8 \\ x_2 &= x_1 - \frac{\nabla f(x_1)}{||\nabla f(x_1)||} \\ x_2 &= -4 - \frac{-8}{8} = -3 \end{aligned}$$

Iteration 3

$$\begin{aligned} \nabla f(x_2) &= -6 \\ x_3 &= x_2 - \frac{\nabla f(x_2)}{||\nabla f(x_2)||} \\ x_3 &= -3 - \frac{-6}{6} = -2 \end{aligned}$$

Iteration 4

$$\begin{aligned} \nabla f(x_3) &= -4 \\ x_4 &= x_3 - \frac{\nabla f(x_3)}{||\nabla f(x_3)||} \\ x_4 &= -2 - \frac{-4}{4} = -1 \end{aligned}$$

Iteration 5

$$\nabla f(x_4) = -2$$
$$x_5 = x_4 - \frac{\nabla f(x_4)}{||\nabla f(x_4)||}$$
$$x_5 = -1 - \frac{-2}{2} = 0$$

Here we can see that the gradient decent has converged to the minima, which is $x_0 = 0$ when the starting point is $x_0 = -5$.

(b) Starting point $x_0 = -2.5$

Iteration 1

$$x_0 = -2.5$$
$$\nabla f(x_0) = -2.5$$
$$x_1 = x_0 - \frac{\nabla f(x_0)}{||\nabla f(x_0)||}$$
$$x_1 = -2.5 - \frac{-5}{5} = -1.5$$

Iteration 2

$$\nabla f(x_1) = -1.5$$
$$x_2 = x_1 - \frac{\nabla f(x_1)}{||\nabla f(x_1)||}$$
$$x_2 = -1.5 - \frac{-3}{3} = 1.5$$

Iteration 3

$$\nabla f(x_2) = 1.5$$
$$x_3 = x_2 - \frac{\nabla f(x_2)}{||\nabla f(x_2)||}$$
$$x_3 = -1.5 - \frac{1}{1} = -0.5$$

Iteration 4

$$\begin{aligned}\nabla f(x_3) &= -0.5 \\ x_4 &= x_3 - \frac{\nabla f(x_3)}{||\nabla f(x_3)||} \\ x_4 &= -0.5 - \frac{-1}{1} = 0.5\end{aligned}$$

Iteration 5

$$\begin{aligned}\nabla f(x_4) &= 0.5 \\ x_5 &= x_4 - \frac{\nabla f(x_4)}{||\nabla f(x_4)||} \\ x_5 &= 0.5 - \frac{1}{1} = -0.5\end{aligned}$$

Iteration 6

$$\begin{aligned}\nabla f(x_4) &= -0.5 \\ x_5 &= x_4 - \frac{\nabla f(x_4)}{||\nabla f(x_4)||} \\ x_5 &= -0.5 - \frac{-1}{1} = 0.5\end{aligned}$$

Here, on choosing x_0 to be a fractional values, it does not converge to the minima.
This shows that choosing a right step size is important for convergence. Hence, finding x for which $\nabla f(x) = 0$ is a difficult problem and cannot be solved without using the right α.

7 Solving for Step Size

$$g(\alpha) = f\left(x_k - \alpha \frac{\nabla f(x_k)}{||\nabla f(x_k)||}\right)$$

We can solve for $g'(x) = 0$ and choose α_k as argmin $g(\alpha)$.

1. Minimize the function $f = x^2$ with starting point $x_0 = -5$.

$$
\begin{aligned}
f(x) &= x^2 \\
x_0 &= -5 \\
\nabla f(x) &= 2x \\
\therefore \nabla f(x_0) &= -10 \\
g(\alpha) &= f(-5+\alpha) = (\alpha-5)^2 \\
g^{'}(\alpha) &= 2\alpha - 10 \\
\text{Set } g^{'}(\alpha) &= 0 \\
\therefore 2\alpha - 10 &= 0 \\
\Rightarrow \alpha &= 5
\end{aligned}
$$

2. Minimize the function $f = y^2 + 2x^2$ with starting point $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$$
\begin{aligned}
f(x,y) &= y^2 + 2x^2 \\
\nabla f(x,y) &= \begin{bmatrix} 4x \\ 2y \end{bmatrix}
\end{aligned}
$$

Iteration 1

$$
\begin{aligned}
x_0 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
\nabla f(x_0) &= \begin{bmatrix} 4 \\ 2 \end{bmatrix} \\
g(\alpha) &= f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} 4 \\ 2 \end{bmatrix}\right) \\
&= (1-2\alpha)^2 + 2(1-4\alpha)^2 \\
&= 36\alpha^2 - 20\alpha + 3 \\
g^{'}(\alpha) &= 0 \\
\Rightarrow 72\alpha = 20 \Rightarrow \alpha &= \frac{20}{72} = \frac{5}{18} \\
x_1 &= \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{5}{18}\left(\begin{bmatrix} 4 \\ 2 \end{bmatrix}\right)\right) \\
x_1 &= \begin{bmatrix} \frac{-1}{9} \\ \frac{4}{9} \end{bmatrix}
\end{aligned}
$$

Iteration 2

$$x_1 = \begin{bmatrix} \frac{-1}{9} \\ \frac{4}{9} \end{bmatrix}$$

$$\nabla f(x_1) = \begin{bmatrix} \frac{-4}{9} \\ \frac{8}{9} \end{bmatrix}$$

$$g(\alpha) = f\left(\begin{bmatrix} \frac{-1}{9} \\ \frac{4}{9} \end{bmatrix} - \alpha \begin{bmatrix} \frac{-4}{9} \\ \frac{8}{9} \end{bmatrix}\right)$$

$$g^{'}(\alpha) = 0$$

$$\alpha = \frac{5}{12}$$

$$x_2 = \left(\begin{bmatrix} \frac{-1}{9} \\ \frac{4}{9} \end{bmatrix} - \frac{5}{12} \begin{bmatrix} \frac{-4}{9} \\ \frac{8}{9} \end{bmatrix}\right)$$

$$x_2 = \begin{bmatrix} \frac{2}{27} \\ \frac{2}{27} \end{bmatrix}$$

Iteration 3

$$x_2 = \begin{bmatrix} \frac{2}{27} \\ \frac{2}{27} \end{bmatrix}$$

$$\nabla f(x_2) = \begin{bmatrix} \frac{8}{27} \\ \frac{4}{27} \end{bmatrix}$$

$$g(\alpha) = f\left(\begin{bmatrix} \frac{2}{27} \\ \frac{2}{27} \end{bmatrix} - \alpha \begin{bmatrix} \frac{8}{27} \\ \frac{4}{27} \end{bmatrix}\right)$$

$$g^{'}(\alpha) = 0$$

$$\alpha = \frac{5}{18}$$

$$x_3 = \left(\begin{bmatrix} \frac{2}{27} \\ \frac{2}{27} \end{bmatrix} - \frac{5}{18} \begin{bmatrix} \frac{8}{27} \\ \frac{4}{27} \end{bmatrix}\right)$$

$$x_3 = \begin{bmatrix} \frac{-2}{243} \\ \frac{8}{243} \end{bmatrix}$$

The above steps can be viewed as the figure below. The goal of minimization is to reach the red point, i.e. $(0, 0, 0)$. We will get very close to the minimum but reach it only after infinite steps. Hence, using this method, we may take infinite steps to reach the minimum.