IIIT Hyderabad **Scribed By: Harsh Bansal (2021101027)**
- **Abhinav Reddy Boddu (2021101034)**

Optimization Methods(CS1.404) **Lecture 6**
Instructor: Dr. Naresh Manwani 23^{th} **Jan 2025**

Outline. *This scribe outlines the conditions of FONC and SONC along with thier pitfalls. It also introduces the concepts of SOSC and HOSC*

1 Introduction

Our goal in optimization is to minimize a function $f(x)$ over a feasible region $S \subseteq \mathbb{R}^n$. This involves finding a point $x^* \in S$ such that $f(x^*) \leq f(x)$ for all $x \in S$. To determine whether a candidate solution x^* is a local minimum, we use necessary and sufficient conditions for optimality.

2 First order Necessary Condition (FONC)

2.1 Theorem 1 (FONC - General Case)

Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$ to be a continuously differentiable function where $f \in C^1(S)$, the class of all functions where their first partial derivatives are continuous. If x^* is a point on S that is a local minima of function f, then the following holds:

$$\nabla f(x^*)^\top d \geq 0$$

Proof of FONC - General Case

In order to model the behavior of f in a scalar step size $\alpha \geq 0$ in the direction vector $d \in \mathbb{R}^n$, let us define a univariate function $\phi : \mathbb{R} \to \mathbb{R}$:

$$\phi(\alpha) = f(x^* + \alpha d)$$

Using the first-order Taylor series expansion for $\phi(\alpha)$ around $\alpha = 0$, we write:

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \mathcal{O}(\alpha^2)$$

where $\mathcal{O}(\alpha^2)$ represents higher-order terms that vanish faster than α^2 as $\alpha \to 0$. From the definition of $\phi(\alpha) = f(x^* + \alpha d)$, we observe that:

$$\phi(0) = f(x^*).$$

Next, we compute $\phi'(\alpha)$, the derivative of $\phi(\alpha)$ with respect to α and applying the multivariate chain rule:

$$\phi'(\alpha) = \frac{d}{d\alpha} f(x^* + \alpha d).$$

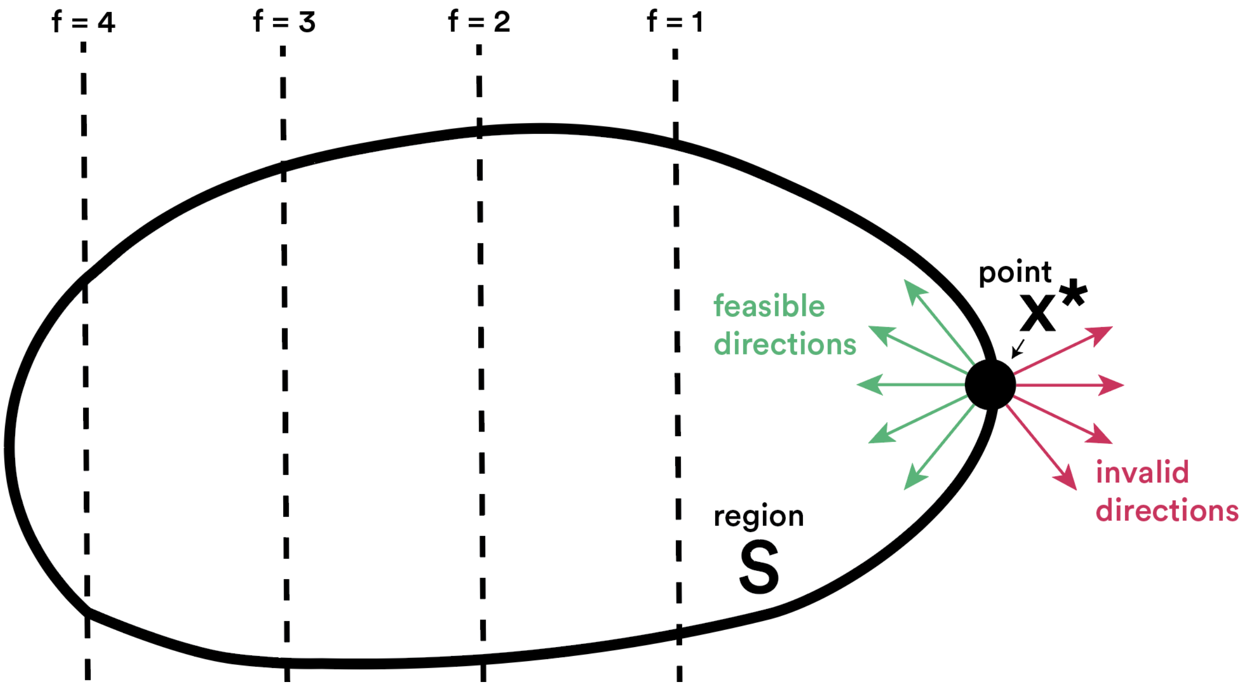


Figure 1: Direction 2D Visualization on region S

$$\phi'(\alpha) = \nabla f(x^* + \alpha d)^\top d,$$

At $\alpha = 0$, the point becomes x^*, and the gradient simplifies to:

$$\phi'(0) = \nabla f(x^*)^\top d.$$

This quantity represents the directional derivative. We can thus determine that the first-order Taylor series approximation becomes:

$$\phi(\alpha) \approx f(x^*) + \nabla f(x^*)^\top d \cdot \alpha.$$

Assuming that x^* is a local minima of f, if we move in an adequately small step size α along any feasible direction d, it is expected that the change in function value is non-negative and there $\exists \alpha_0$ such that:

$$\phi(\alpha) \geq \phi(0) \Rightarrow f(x^* + \alpha d) \geq f(x^*) \quad \forall \alpha \in (0, \alpha_0]$$

This implies for all directions d:

$$\phi'(0) = \nabla f(x^*)^\top d. \geq 0. \tag{1}$$

2.2 Theorem 2 (FONC - Interior Case)

Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$ to be a continuously differentiable function where $f \in C^1(S)$, the class of all functions where their first partial derivatives are continuous. If x^* is an interior point on S that is a local minima of function f, then the following holds:

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

Proof of FONC - Interior Case

If x^* is an interior point of the feasible region S, then any direction d is a feasible direction. Moreover, if d is a feasible direction, then $-d$ is also a feasible direction because $x^* + \alpha d \in S$ implies $x^* - \alpha d \in S$ for small enough $\alpha > 0$.

Using (1), for all feasible directions $-d$, we have:

$$\nabla f(x^*)^\top (-d). \geq 0.$$

$$\nabla f(x^*)^\top d. \leq 0. \tag{2}$$

If x^* is a local minimum of f, then $\nabla f(x^*)^\top d \geq 0$ for all feasible directions d, and $\nabla f(x^*)^\top d \leq 0$ for the reverse direction $-d$. Combining conditions (1) and (2), we obtain:

$$\nabla f(x^*)^\top d = 0$$

As x^* is an interior point, d can represent any direction in $\mathbb{R}^n$, this implies:

$$\nabla f(x^*) = 0.$$

Example 1: $f(x_1, x_2) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$

Consider the function $f(x_1, x_2) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$. The gradient is:

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2 + 3 \end{bmatrix}.$$

At $[x_1, x_2] = [0, 3]$, we evaluate the gradient:

$$\nabla f(0, 3) = \begin{bmatrix} 2(0) \\ 3 + 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}.$$

Since $\nabla f(0, 3) \neq 0$, the first-order necessary condition (FONC) for an interior point is not satisfied. Next, we analyze feasible directions at $[0, 3]$. For a direction $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$, the condition for feasible directions is:

$$\nabla f(0, 3)^\top d = 6d_2 \geq 0.$$

- If $d_2 > 0$, $6d_2 > 0$, indicating f increases.
- If $d_2 < 0$, $6d_2 < 0$, indicating f decreases.

Hence, $[0, 3]$ is not a local minimum of $f(x_1, x_2) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$ because:

- The FONC fails ($\nabla f(0, 3) \neq 0$).
- The feasible direction analysis shows that f decreases in the $-x_2$ direction.

2.3 Pitfalls of FONC

While $\nabla f(x^*) = 0$ is a necessary condition for x^* to be a local minimum, it is not sufficient. Satisfying FONC merely indicates that x^* is a *critical point*, which could correspond to:

1. A local minimum,
2. A local maximum,
3. A saddle point.

Example 1: Local Minimum

Consider $f(x) = (x-2)^2$. The gradient is:

$$\nabla f(x) = 2x - 4.$$

At $x^* = 2$, we have $\nabla f(x^*) = 0$. Evaluating $f(x)$ around x^*, we see that $f(x) \geq f(0)$ for all x near 0, indicating $x^* = 2$ is a local minimum.

Example 2: Local Maximum

Consider $f(x) = -(x-2)^2$. The gradient is:

$$\nabla f(x) = -2x + 4.$$

At $x^* = 2$, we have $\nabla f(x^*) = 0$. Evaluating $f(x)$ around x^*, we see that $f(x) \leq f(0)$ for all x near 0, indicating $x^* = 2$ is a local maximum.

Example 3: Saddle Point for $f(x) = x^3$

Consider the function $f(x) = x^3$. The derivative is:

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

At $x^* = 0$, we have $f'(x^*) = 0$. However, evaluating $f(x)$ around $x^* = 0$ reveals that the function behaves differently on either side:

$$f(x) > f(0) \quad \text{for } x > 0, \quad \text{and} \quad f(x) < f(0) \quad \text{for } x < 0.$$

This shows that $x^* = 0$ is a saddle point because $f'(x^*) = 0$ satisfies the first-order necessary condition (FONC), but x^* is neither a local minimum nor a local maximum.

3 Second order Necessary Condition (SONC)

3.1 General Case

Let $U \subseteq \mathbb{R}^n$ and $f : U \to \mathbb{R}$ be a twice continuously differentiable function on U (i.e., $f \in C^2(U)$). Suppose x^* is a local minima of f over U, and let $\mathbf{d}$ represent a feasible direction at x^*. If the first-order condition satisfies *"flatness"* in the direction $\mathbf{d}$, i.e., $\nabla f(x^*)^T \mathbf{d} = 0$, then the second-order condition in that direction must hold:

$$\mathbf{d}^T \nabla^2 f(x^*) \mathbf{d} \geq 0$$

where $\nabla^2 f(x^*)$ denotes the Hessian matrix of f at x^*.

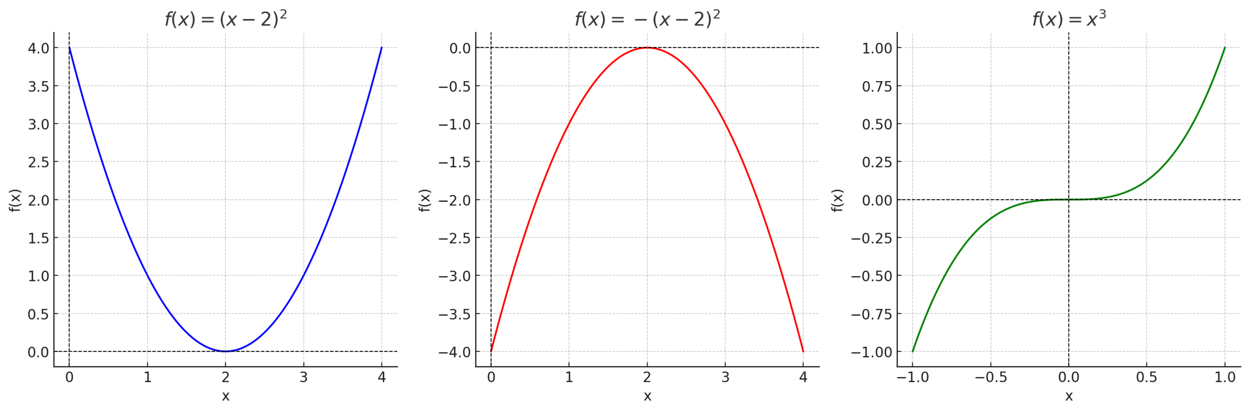


Figure 2: Example Graphs that satisfy FONC

Proof

We prove this result by contradiction. Assume x^* is a local minima, $\mathbf{d}$ is a feasible direction at x^*, $\nabla f(x^*)^T\mathbf{d} = 0$, but contrary to the theorem, $\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} < 0$.
Consider the function $\phi(\alpha) = f(x^* + \alpha\mathbf{d})$. Using the second-order Taylor expansion of $\phi(\alpha)$ around $\alpha = 0$, we have:

$$\phi(\alpha) = \phi(0) + \alpha\phi'(0) + \frac{\alpha^2}{2}\phi''(0) + o(\alpha^2)$$

At $\alpha = 0$, we calculate the following terms:

- $\phi(0) = f(x^*)$
- $\phi'(0) = \nabla f(x^*)^T\mathbf{d} = 0 \quad$ (by assumption)
- $\phi''(0) = \mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} < 0 \quad$ (by contradiction hypothesis)

Substituting these into the Taylor expansion yields:

$$f(x^* + \alpha\mathbf{d}) = f(x^*) + \frac{\alpha^2}{2}\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} + o(\alpha^2)$$

For sufficiently small $\alpha > 0$, the term $\frac{\alpha^2}{2}\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d}$ dominates $o(\alpha^2)$, and since $\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} < 0$, we have:

$$f(x^* + \alpha\mathbf{d}) < f(x^*)$$

This contradicts the assumption that x^* is a local minima. Hence, our assumption that $\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} < 0$ must be false. We conclude:

$$\mathbf{d}^T\nabla^2 f(x^*)\mathbf{d} \geq 0.$$

3.2 Interior Case

Suppose $U \subseteq \mathbb{R}^n$ and $f : U \to \mathbb{R}$ is twice continuously differentiable on U (i.e., $f \in C^2(U)$). Let x^* be a local minima of f over U such that x^* lies in the interior of U. Then, the following conditions hold:

1. $\nabla f(x^*) = 0 \quad$ (Stationarity)
2. $\nabla^2 f(x^*)$ is positive semi-definite, i.e., $\mathbf{d}^T \nabla^2 f(x^*) \mathbf{d} \geq 0$ for all directions $\mathbf{d} \in \mathbb{R}^n$.

This implies that for any direction $\mathbf{d}$, the quadratic form associated with the Hessian matrix at x^* must be non-negative. Equivalently, all eigenvalues of $\nabla^2 f(x^*)$ are non-negative

Proof

Let $x^* \in \text{int}(U)$ be a local minima. From the first-order necessary condition (FONC) for interior points, we know that $\nabla f(x^*) = 0$.
Now, for any direction $\mathbf{d} \in \mathbb{R}^n$, since x^* is an interior point, $\mathbf{d}$ is a feasible direction. Since $\nabla f(x^*)^T \mathbf{d} = 0$ (as $\nabla f(x^*) = 0$), we can apply the SONC for the general case. By SONC, we have:

$$\mathbf{d}^T \nabla^2 f(x^*) \mathbf{d} \geq 0 \quad \forall \mathbf{d} \in \mathbb{R}^n$$

This implies that the Hessian matrix $\nabla^2 f(x^*)$ is positive semi-definite. Alternatively, consider the function $\phi(\alpha) = f(x^* + \alpha \mathbf{d})$. From the FONC, $\phi'(0) = 0$, and $\phi(0) = f(x^*)$. Suppose, for contradiction, that $\nabla^2 f(x^*)$ is not positive semi-definite. Then, there exists a direction $\mathbf{d}$ such that $\mathbf{d}^T \nabla^2 f(x^*) \mathbf{d} < 0$, i.e., $\phi''(0) < 0$.
Since $f \in C^2(U)$, ϕ is C^2 and $o(\alpha)$ is continuous. Because $\phi''(0) < 0$, there exists $\delta > 0$ such that $\phi''(\xi) < 0$ for all $\xi \in (0, \delta)$. Using the second-order Taylor expansion with remainder for some $\xi \in (0, \delta)$:

$$\phi(\alpha) = \phi(0) + \alpha \phi'(0) + \frac{\alpha^2}{2} \phi''(\xi)$$

Since $\phi'(0) = 0$ and $\phi''(\xi) < 0$ for $\xi \in (0, \delta)$, we obtain:

$$\phi(\delta) = \phi(0) + \frac{\delta^2}{2} \phi''(\xi) < \phi(0)$$

Thus, $f(x^* + \delta \mathbf{d}) < f(x^*)$, which contradicts the local minimality of x^*. Therefore, $\nabla^2 f(x^*)$ must be positive semi-definite.

Example 1: $f(x) = (x-2)^2$

Consider the function $f(x) = (x-2)^2$. The derivative is:

$$f'(x) = 2(x-2).$$

At $x^* = 2$, we have $f'(x^*) = 0$. This satisfies the first-order necessary condition (FONC).
Next, let us examine the second derivative of $f(x)$:

$$f''(x) = 2.$$

At $x^* = 2$, we find that $f''(x^*) = 2 > 0$. Since the second derivative is positive, the second-order necessary condition (SONC) is satisfied.
Thus, $x^* = 2$ is a local minimum of $f(x)$. Additionally, evaluating $f(x)$ around $x^* = 2$ confirms this:

$$f(x) > f(2) \quad \text{for all } x \neq 2.$$

This example demonstrates that SONC is sufficient to conclude that $x^* = 2$ is a local minimum for $f(x) = (x-2)^2$.

Example 2: $f(x) = x^3$

Consider the function $f(x) = x^3$. The derivative is:

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

At $x^* = 0$, we have $f'(x^*) = 0$. However, evaluating $f(x)$ around $x^* = 0$ reveals that the function behaves differently on either side of x^*:

$$f(x) > f(0) \quad \text{for } x > 0, \quad \text{and} \quad f(x) < f(0) \quad \text{for } x < 0.$$

This indicates that $x^* = 0$ is a saddle point. Although $f'(x^*) = 0$ satisfies the first-order necessary condition (FONC), x^* is neither a local minimum nor a local maximum.
Now, let us examine the second-order necessary condition (SONC). The second derivative of $f(x)$ is:

$$f''(x) = 6x.$$

At $x^* = 0$, we find that $f''(x^*) = 0$. According to SONC, if $f''(x^*) \geq 0$, x^* could potentially be a local minimum. However, since $f''(x^*) = 0$, SONC does not provide conclusive information.
In this case, $x^* = 0$ is clearly not a local minimum, as $f(x) = x^3$ exhibits a saddle point at $x^* = 0$. This example demonstrates that while SONC is necessary, it is not sufficient for determining whether a critical point is a local minimum. Additional analysis or higher-order conditions may be required to accurately classify the nature of the critical point.

4 Second Order Sufficient Condition (SOSC)

Theorem 1 (Second Order Sufficient Condition (SONC))**.** *consider the domain set $S \subseteq \mathbb{R}^n$. Consider a function $f : S \to \mathbb{R}$ such that $f \in \mathcal{C}^2(S)$. Let $\overline{X}^*$ be an interior point of S. Suppose $\nabla f(\overline{X}^*) = \overline{0}$ and $H(\overline{X}^*) = \nabla^2 f(\overline{X}^*)$ is symmetric and positive definite, then $\overline{X}^*$ is a local minima of function f*

As we have seen earlier, for the case of $f(x) = x^3$, *SONC* is not sufficient to identify local minima. So we introduce Second Order Sufficient Condition (*SOSC*).
Proof:
If $H(\overline{X}^*)$ is positive definite, it has n eigen-vectors $\overline{d_1}, \overline{d_2}, \cdots, \overline{d_n}$, and corresponding eigen-values $\overline{\lambda_1}, \overline{\lambda_2}, \cdots \overline{\lambda_n}$ where all the λs are > 0.
Any direction $\overline{d} \in \mathbb{R}^n$ can be expressed as linear combination of eigen vectors as

$$\overline{d} = \Sigma_{i=1}^n \alpha_i \overline{d_i}$$

Now, consider the expression $\overline{d}^T H(\overline{X}^*)\overline{d}$:

$$
\begin{aligned}
&= \overline{d}^T H(\overline{X}^*)\overline{d} \\
&= \left(\sum_{i=1}^{n} \alpha_i \overline{d_i}\right)^T \cdot H(\overline{X}^*) \cdot \left(\sum_{j=1}^{n} \alpha_j \overline{d_j}\right) \\
&= \left(\sum_{i=1}^{n} \alpha_i \overline{d_i}^T\right) \cdot H(\overline{X}^*) \cdot \left(\sum_{j=1}^{n} \alpha_j \overline{d_j}\right) \\
&= \left(\sum_{i=1}^{n} \alpha_i \overline{d_i}^T\right) \cdot \left(\sum_{j=1}^{n} \alpha_j H(\overline{X}^*) \cdot \overline{d_j}\right) \\
&= \left(\sum_{i=1}^{n} \alpha_i \overline{d_i}^T\right) \cdot \left(\sum_{j=1}^{n} \alpha_j \lambda_j \overline{d_j}\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \lambda_j \overline{d_i}^T \cdot \overline{d_j} \\
&= \sum_{i=1}^{n} \alpha_i^2 \lambda_j \overline{d_i}^T \cdot \overline{d_i} \quad \big(\text{Since eigen-vectors are orthogonal, } \overline{d_i} \cdot \overline{d_j} = 0 \text{ if } i \neq j\big) \\
&= \sum_{i=1}^{n} \alpha_i^2 \lambda_j \|\overline{d_i}\|^2 \\
&\geq \lambda_{min} \sum_{i=1}^{n} \alpha_i^2 \|\overline{d_i}\|^2 \\
&\geq \lambda_{min} \|\overline{d}\|^2 \\
&> 0 \quad \big(\lambda_{min} > 0 \text{ for positive definite H, and } \|\overline{d}\|^2 > 0\big)
\end{aligned}
$$

Therefore,

$$
\overline{d}^T H(\overline{X}^*)\overline{d} > 0 \tag{1}
$$

Now, consider the point $\overline{X}^* + \alpha\overline{d}$ in the neighbourhood of $\overline{X}^*$ along direction $\overline{d}$. Taking the 2nd order tailor series expansion for f at this point gives:

$$
\begin{aligned}
&f(\overline{X}^* + \alpha\overline{d}) = f(\overline{X}^*) + \alpha \nabla f(\overline{X}^*)^T \cdot \overline{d} + \frac{\alpha^2}{2} \overline{d}^T \nabla^2 f(\overline{X}^*)\overline{d} + o(\alpha^2 \|\overline{d}\|^2) \\
&f(\overline{X}^* + \alpha\overline{d}) = f(\overline{X}^*) + \frac{\alpha^2}{2} \overline{d}^T \nabla^2 f(\overline{X}^*)\overline{d} + o(\alpha^2 \|\overline{d}\|^2) \quad (\text{Since by FONC, } \nabla f(\overline{X}^*)^T \cdot \overline{d} = 0) \\
&f(\overline{X}^* + \alpha\overline{d}) > f(\overline{X}^*) + o(\alpha^2 \|\overline{d}\|^2) \quad (\text{Since from equation 1, } \overline{d}^T \nabla^2 f(\overline{X}^*)\overline{d} > 0) \\
&f(\overline{X}^* + \alpha\overline{d}) > f(\overline{X}^*)
\end{aligned}
$$

Therefore, $f(\overline{X}^* + \alpha\overline{d}) > f(\overline{X}^*)$. i.e $\overline{X}^*$ is a local minima.

5 Higher Order Sufficient Condition (HOSC)

Can SOSC still go wrong? **YES**. Lets take an example:

Problem 1. *Is $x = 0$ a local minima for the function $f(x) = x^4$?*

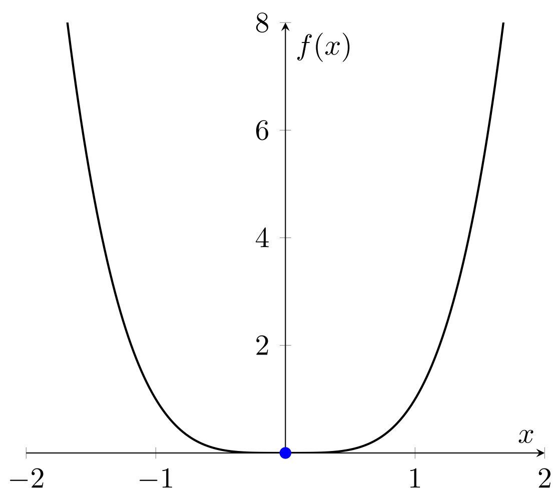
Soln.
Given function $f(x) = x^4$.
Considering the 1^{st} derivative:

$$f'(x) = 4x^3$$
$$f'(0) = 0$$

Considering the 2^{nd} derivative:

$$f''(x) = 12x^2$$
$$f''(0) = 0$$

So the $x = 0$ satisfies FONC and SONC. It does not satisfy SOSC. Yet, we know that $x = 0$ is a local minima for the function $f(x) = x^4$



To handle such cases, we extend the sufficient optimality condition.

Theorem 2 (Higher Order Sufficient Optimality Condition (HOSC))**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $f \in \mathcal{C}^\infty$ and f is not a constant function. Let $f^{(k)}(x)$ the k^{th} order derivative of f. And, let $\min\limits_x f(x)$ be the optimization problem. $\overline{X}^*$ is a local minima of f if and only if (iff) first non-zero element of $\{f^{(k)}(\overline{X}^*)\}_{k=1}^{\infty}$ is +ve and occurs for even value of k.*