# Numerical Optimization

## Unconstrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

## Unconstrained Minimization

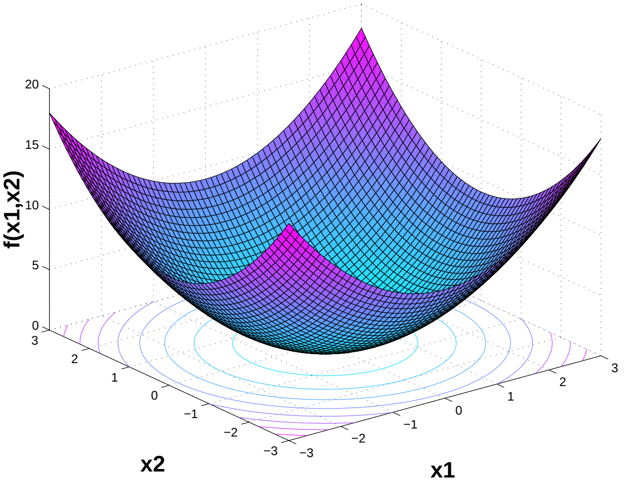
Let $f : \mathbb{R}^n \to \mathbb{R}$. Consider the optimization problem:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

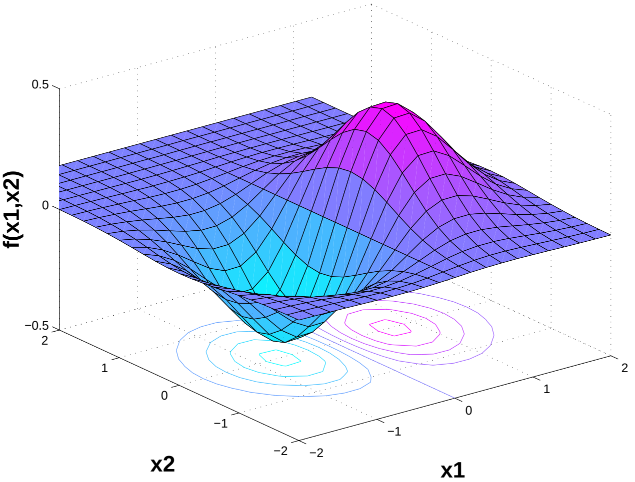
- Assumption: $f$ *is bounded below*.

**Definition**

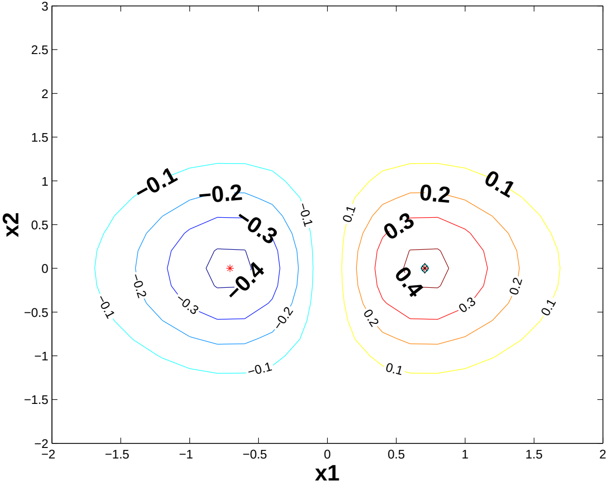
$\boldsymbol{x}^* \in \mathbb{R}^n$ is said to be a local minimum of $f$ if there is a $\delta > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad \forall\, \boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$.




Surface Plot : $f(\boldsymbol{x}) = x_1^2 + x_2^2$




Surface Plot : $f(\boldsymbol{x}) = x_1 \exp(-x_1^2 - x_2^2)$




Contour Plot : $f(\boldsymbol{x}) = x_1 \exp(-x_1^2 - x_2^2)$

The function value *does not decrease* in the local neighbourhood of a local minimum.

## Definition

Let $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. If there exists a direction $\boldsymbol{d} \in \mathbb{R}^n$ and $\delta > 0$ such that $f(\bar{\boldsymbol{x}} + \alpha\boldsymbol{d}) < f(\bar{\boldsymbol{x}})$ for all $\alpha \in (0, \delta)$, then $\boldsymbol{d}$ is said to be a descent direction of $f$ at $\bar{\boldsymbol{x}}$.

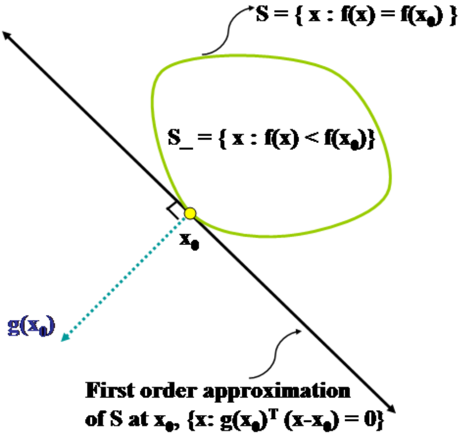
## Result

Let $f \in \mathcal{C}^1$ and $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. Let $\boldsymbol{g}(\bar{\boldsymbol{x}}) = \nabla f(\bar{\boldsymbol{x}})$. If $\boldsymbol{g}(\bar{\boldsymbol{x}})^T\boldsymbol{d} < 0$ then, $\boldsymbol{d}$ is a descent direction of $f$ at $\bar{\boldsymbol{x}}$.

## Proof.

Given $\boldsymbol{g}(\bar{\boldsymbol{x}})^T\boldsymbol{d} < 0$. Now, $f \in \mathcal{C}^1 \Rightarrow g \in \mathcal{C}^0$.
$\therefore \exists\, \delta > 0 \ni \boldsymbol{g}(\boldsymbol{x})^T\boldsymbol{d} < 0 \;\forall\, \boldsymbol{x} \in LS(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \delta\boldsymbol{d})$.
Choose any $\alpha \in (0, \delta)$. Using first order truncated Taylor series,

$$\begin{aligned} f(\bar{\boldsymbol{x}} + \alpha\boldsymbol{d}) &= f(\bar{\boldsymbol{x}}) + \alpha\boldsymbol{g}(\boldsymbol{x})^T\boldsymbol{d} \quad \text{where } \boldsymbol{x} \in LS(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \alpha\boldsymbol{d}) \\ \therefore f(\bar{\boldsymbol{x}} + \alpha\boldsymbol{d}) &< f(\bar{\boldsymbol{x}}) \;\forall\, \alpha \in (0, \delta) \\ &\Rightarrow \quad \boldsymbol{d} \text{ is a descent direction of } f \text{ at } \bar{\boldsymbol{x}} \end{aligned}$$



S = { x : f(x) = f(x0) }
S_ = { x : f(x) < f(x0)}
x0
g(x0)
First order approximation
of S at x0, {x: g(x0)T (x-x0) = 0}

## First Order Necessary Conditions (Unconstrained Minimization)

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$. If $\boldsymbol{x}^*$ is a local minimum of $f$, then $\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$.

### Proof.

Let $\boldsymbol{x}^*$ be a local minimum of $f$ and $\boldsymbol{g}(\boldsymbol{x}^*) \neq 0$.
Choose $\boldsymbol{d} = -\boldsymbol{g}(\boldsymbol{x}^*)$.

$$
\begin{aligned}
\therefore \boldsymbol{g}(\boldsymbol{x}^*)^T\boldsymbol{d} &= -\boldsymbol{g}(\boldsymbol{x}^*)^T\boldsymbol{g}(\boldsymbol{x}^*) < 0 \\
\boldsymbol{g}(\boldsymbol{x}^*)^T\boldsymbol{d} < 0 &\Rightarrow \boldsymbol{d} \text{ is a descent direction of } f \text{ at } \boldsymbol{x}^* \\
&\Rightarrow \boldsymbol{x}^* \text{ is not a local minimum, a contradiction.}
\end{aligned}
$$

Therefore, $\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$. □

Provides a stopping condition for an optimization algorithm

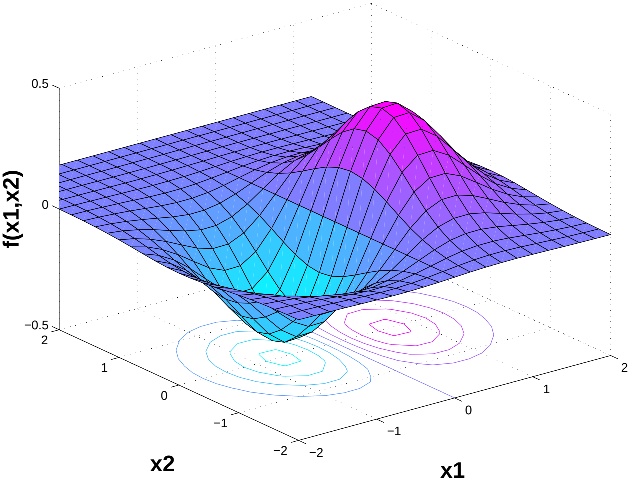
Example:

- Consider the problem

$$\min\ f(\boldsymbol{x}) \stackrel{\Delta}{=} x_1 \exp(-x_1^2 - x_2^2)$$

- $$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} \exp(-x_1^2 - x_2^2)(1 - 2x_1^2) \\ \exp(-x_1^2 - x_2^2)(-2x_1x_2) \end{pmatrix}.$$

- $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ at $(\frac{1}{\sqrt{2}}, 0)^T$ and $(-\frac{1}{\sqrt{2}}, 0)^T$.




The function has a local minimum at $(-\frac{1}{\sqrt{2}}, 0)^T$ and a local maximum at $(\frac{1}{\sqrt{2}}, 0)^T$

- Consider the problem

$$\min\ f(\boldsymbol{x}) \triangleq x_1 \exp(-x_1^2 - x_2^2)$$

- $$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} \exp(-x_1^2 - x_2^2)(1 - 2x_1^2) \\ \exp(-x_1^2 - x_2^2)(-2x_1x_2) \end{pmatrix}.$$

- $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ at $(\frac{1}{\sqrt{2}}, 0)^T$ and $(-\frac{1}{\sqrt{2}}, 0)^T$.
- If $\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$, then $\boldsymbol{x}^*$ is a *stationary point*.
- Need higher order derivatives to confirm that a stationary point is a local minimum

## Second Order Necessary Conditions

Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^2$. If $\boldsymbol{x}^*$ is a local minimum of $f$, then $\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\boldsymbol{H}(\boldsymbol{x}^*)$ is positive semi-definite.

### Proof.

Let $\boldsymbol{x}^*$ be a local minimum of $f$. From the first order necessary condition, $\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$.
Assume $\boldsymbol{H}(\boldsymbol{x}^*)$ is not positive semi-definite. So, $\exists\, \boldsymbol{d}$ such that $\boldsymbol{d}^T\boldsymbol{H}(\boldsymbol{x}^*)\boldsymbol{d} < 0$. Since $H$ is continuous near $\boldsymbol{x}^*$, $\exists\, \delta > 0$ such that $\boldsymbol{d}^T\boldsymbol{H}(\boldsymbol{x}^* + \alpha\boldsymbol{d})\boldsymbol{d} < 0 \;\forall\, \alpha \in (0, \delta)$.
Using second order truncated Taylor series around $\boldsymbol{x}^*$, we have for all $\alpha \in (0, \delta)$,

$$
\begin{aligned}
f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) &= f(\boldsymbol{x}^*) + \alpha\boldsymbol{g}(\boldsymbol{x}^*)^T\boldsymbol{d} + \frac{1}{2}\alpha^2\boldsymbol{d}^T\boldsymbol{H}(\bar{\boldsymbol{x}})\boldsymbol{d} \\
&\quad \text{where } \bar{\boldsymbol{x}} \in LS(\boldsymbol{x}^*, \boldsymbol{x}^* + \alpha\boldsymbol{d}) \\
\Rightarrow f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) &< f(\boldsymbol{x}^*)
\end{aligned}
$$

$\therefore\, \boldsymbol{x}^*$ is not a local minimum, a contradiction.

## Second Order Sufficient Conditions

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in \mathcal{C}^2$. If $\boldsymbol{g}(\boldsymbol{x}^*) = 0$ and $\boldsymbol{H}(\boldsymbol{x}^*)$ is positive definite, then $\boldsymbol{x}^*$ is a strict local minimum of $f$.

## Proof.

Since $\boldsymbol{H}$ is continuous and positive definite near $\boldsymbol{x}^*$, $\exists\, \delta > 0$ such that $\boldsymbol{H}(\boldsymbol{x})$ is positive definite for all $\boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$.
Choose some $\boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$. Using second order truncated Taylor series,

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}^*) + \boldsymbol{g}(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x} - \boldsymbol{x}^*) \\ & \quad \text{where } \bar{\boldsymbol{x}} \in LS(\boldsymbol{x}, \boldsymbol{x}^*). \end{aligned}$$

Since $(\boldsymbol{x} - \boldsymbol{x}^*)^T \boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x} - \boldsymbol{x}^*) > 0 \;\forall\, \boldsymbol{x} \in B(\boldsymbol{x}^*, \delta)$,

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*) \;\forall\, \boldsymbol{x} \in B(\boldsymbol{x}^*, \delta).$$

This implies that $\boldsymbol{x}^*$ is a strict local minimum. $\square$

Example:

- Consider the problem

$$\min\ f(\boldsymbol{x}) \triangleq x_1 \exp(-x_1^2 - x_2^2)$$

- $$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} \exp(-x_1^2 - x_2^2)(1 - 2x_1^2) \\ \exp(-x_1^2 - x_2^2)(-2x_1 x_2) \end{pmatrix}.$$

- $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ at ${\boldsymbol{x}_1^*}^T = (\frac{1}{\sqrt{2}}, 0)^T$ and ${\boldsymbol{x}_2^*}^T = (-\frac{1}{\sqrt{2}}, 0)^T$.

- $\boldsymbol{H}(\boldsymbol{x}_2^*) = \begin{pmatrix} 2\sqrt{2}\exp(-\frac{1}{2}) & 0 \\ 0 & \sqrt{2}\exp(-\frac{1}{2}) \end{pmatrix}$ is positive definite $\Rightarrow$ $\boldsymbol{x}_2^*$ is a strict local minimum

- $\boldsymbol{H}(\boldsymbol{x}_1^*) = \begin{pmatrix} -2\sqrt{2}\exp(-\frac{1}{2}) & 0 \\ 0 & -\sqrt{2}\exp(-\frac{1}{2}) \end{pmatrix}$ is negative definite $\Rightarrow$ $\boldsymbol{x}_1^*$ is a strict local maximum

Example:

- Consider the problem

$$\min\ f(\boldsymbol{x}) \triangleq (x_2 - x_1^2)^2 + x_1^5$$

- $\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 5x_1^4 - 4x_1(x_2 - x_1^2) \\ 2(x_2 - x_1^2) \end{pmatrix}.$
- Stationary Point: $(0,0)^T$
- Hessian matrix at $(0,0)^T$ :

$$\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

- Hessian is positive semi-definite at $(0,0)^T$; $(0,0)^T$ is neither a local minimum nor a local maximum of $f(\boldsymbol{x})$.

Example:

- Consider the problem

$$\min\ f(\boldsymbol{x}) \triangleq x_1^2 + \exp(x_1 + x_2)$$

- $$\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 2x_1 + \exp(x_1 + x_2) \\ \exp(x_1 + x_2) \end{pmatrix}.$$

- Need an iterative method to solve $\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$ .

- An iterative optimization algorithm generates a sequence $\{\boldsymbol{x}^k\}_{k\geq 0}$, which converges to a local minimum.

---

**Unconstrained Minimization Algorithm**

---

(1) Initialize $\boldsymbol{x}^0$, $k := 0$.

(2) **while** *stopping condition is not satisfied at* $\boldsymbol{x}^k$

  (a) Find $\boldsymbol{x}^{k+1}$ such that $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$.

  (b) $k := k + 1$

**endwhile**

**Output :** $\boldsymbol{x}^* = \boldsymbol{x}^k$, a local minimum of $f(\boldsymbol{x})$.

---

## Unconstrained Minimization Algorithm

(1) Initialize $\boldsymbol{x}^0, k := 0$.

(2) **while** *stopping condition is not satisfied at* $\boldsymbol{x}^k$

  (a) Find $\boldsymbol{x}^{k+1}$ such that $f(\boldsymbol{x}^{k+1}) < f(\boldsymbol{x}^k)$.

  (b) $k := k + 1$

**endwhile**

**Output :** $\boldsymbol{x}^* = \boldsymbol{x}^k$, a local minimum of $f(\boldsymbol{x})$.

- How to find $\boldsymbol{x}^{k+1}$ in Step 2(a) of the algorithm?
- Which *stopping condition* can be used?
- Does the algorithm converge? If yes, how fast does it converge?
- Does the convergence and its speed depend on $\boldsymbol{x}^0$?

Stopping Conditions for a minimization problem:

- $\|\boldsymbol{g}(\boldsymbol{x}^k)\| = 0$ and $\boldsymbol{H}(\boldsymbol{x}^k)$ is positive semi-definite

## Practical Stopping conditions

Assumption: There are no *stationary* points

- $$\|\boldsymbol{g}(\boldsymbol{x}^k)\| \leq \epsilon$$
- $$\|\boldsymbol{g}(\boldsymbol{x}^k)\| \leq \epsilon(1 + |f(\boldsymbol{x}^k)|)$$
- $$\frac{f(\boldsymbol{x}^k) - f(\boldsymbol{x}^{k+1})}{|f(\boldsymbol{x}^k)|} \leq \epsilon$$

## Speed of Convergence

- Assume that an optimization algorithm generates a sequence $\{\boldsymbol{x}^k\}$, which converges to $\boldsymbol{x}^*$.
- How *fast* does the sequence converge to $\boldsymbol{x}^*$?

### Definition

The sequence $\{\boldsymbol{x}^k\}$ converges to $\boldsymbol{x}^*$ with order $p$ if

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^{p}} = \beta, \quad \beta < \infty$$

- Asymptotically, $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| = \beta\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^{p}$
- Higher the value of $p$, faster is the convergence.

(1) $p = 1, 0 < \beta < 1$ (Linear Convergence)

Some Examples:

- $\beta = .1, \|\boldsymbol{x}^0 - \boldsymbol{x}^*\| = .1$
  Norms of $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| : 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, \ldots$
- $\beta = .9, \|\boldsymbol{x}^0 - \boldsymbol{x}^*\| = .1$
  Norms of $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| : 10^{-1}, .09, .081, .0729, \ldots$

(2) $p = 2, \beta > 0$ (Quadratic Convergence)

Example:

- $\beta = 1, \|\boldsymbol{x}^0 - \boldsymbol{x}^*\| = .1$
  Norms of $\|\boldsymbol{x}^k - \boldsymbol{x}^*\| : 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, \ldots$

(3) Suppose an algorithm generates a convergent sequence $\{\boldsymbol{x}^k\}$ such that

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^k - \boldsymbol{x}^*\|} = 0 \text{ and } \lim_{k\to\infty} \frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^2} = \infty$$

then this convergence is called superlinear convergence