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:

$$\min f(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n$$

- Assumption: $f$ is bounded below.

**Definition**

$x^* \in \mathbb{R}^n$ is said to be a local minimum of $f$ if there is a $\delta > 0$ such that $f(x^*) \leq f(x) \quad \forall x \in B(x^*, \delta)$. 
Surface Plot: \( f(x) = x_1^2 + x_2^2 \)
Surface Plot: \( f(x) = x_1 \exp(-x_1^2 - x_2^2) \)
Contour Plot: $f(x) = x_1 \exp(-x_1^2 - x_2^2)$

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

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

### Result

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

### Proof.

Given $g(\bar{x})^T d < 0$. Now, $f \in C^1 \Rightarrow g \in C^0$.

$\exists \delta > 0 \ \exists \ g(x)^T d < 0 \ \forall \ x \in LS(\bar{x}, \bar{x} + \delta d)$.

Choose any $\alpha \in (0, \delta)$. Using first order truncated Taylor series,

$$f(\bar{x} + \alpha d) = f(\bar{x}) + \alpha g(x)^T d \quad \text{where } x \in LS(\bar{x}, \bar{x} + \alpha d)$$

$\therefore f(\bar{x} + \alpha d) < f(\bar{x}) \ \forall \ \alpha \in (0, \delta)$

$\Rightarrow \quad d$ is a descent direction of $f$ at $\bar{x}$
First order approximation of Sat \(x^e\). 
\[
\{ (x^e) | \nabla \mathbf{g}(x^e) = 0 \}
\]
\[
\{ (x^e) | \mathbf{g}(x^e) = 0 \}
\]
\[
\{ (x^e) | \mathbf{f}(x^e) = 0 \}
\]
First Order Necessary Conditions (Unconstrained Minimization)

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

Proof.

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

\[
\therefore g(x^*)^T d = -g(x^*)^T g(x^*) < 0
\]

\( g(x^*)^T d < 0 \) \( \Rightarrow \) \( d \) is a descent direction of \( f \) at \( x^* \)

\( \Rightarrow x^* \) is not a local minimum, a contradiction.

Therefore, \( g(x^*) = 0 \).

Provides a stopping condition for an optimization algorithm
Example:

Consider the problem

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

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

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

\[ \min f(x) \triangleq x_1 \exp(-x_1^2 - x_2^2) \]

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

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

If \( g(x^*) = 0 \), then \( x^* \) is a stationary point.

Need higher order derivatives to confirm that a stationary point is a local minimum.
Second Order Necessary Conditions

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

Proof.

Let \( x^* \) be a local minimum of \( f \). From the first order necessary condition, \( g(x^*) = 0 \).

Assume \( H(x^*) \) is not positive semi-definite. So, \( \exists d \) such that \( d^TH(x^*)d < 0 \). Since \( H \) is continuous near \( x^* \), \( \exists \delta > 0 \) such that \( d^TH(x^* + \alpha d)d < 0 \forall \alpha \in (0, \delta) \).

Using second order truncated Taylor series around \( x^* \), we have for all \( \alpha \in (0, \delta) \),

\[
 f(x^* + \alpha d) = f(x^*) + \alpha g(x^*)^Td + \frac{1}{2}\alpha^2 d^TH(\bar{x})d \\
\] 

where \( \bar{x} \in LS(x^*, x^* + \alpha d) \)

\[
 \Rightarrow f(x^* + \alpha d) < f(x^*) \\
\]

\( \therefore x^* \) is not a local minimum, a contradiction.
Second Order Sufficient Conditions

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

Proof.

Since \( H \) is continuous and positive definite near \( x^* \), \( \exists \delta > 0 \) such that \( H(x) \) is positive definite for all \( x \in B(x^*, \delta) \).

Choose some \( x \in B(x^*, \delta) \). Using second order truncated Taylor series,

\[
f(x) = f(x^*) + g(x^*)^T (x - x^*) + \frac{1}{2} (x - x^*)^T H(\bar{x})(x - x^*)
\]

where \( \bar{x} \in LS(x, x^*) \).

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

\[
f(x) > f(x^*) \ \forall \ x \in B(x^*, \delta).
\]

This implies that \( x^* \) is a strict local minimum.
Example:

Consider the problem

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

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

$g(x) = 0$ at $x_1^* = \left( \frac{1}{\sqrt{2}}, 0 \right)^T$ and $x_2^* = \left( -\frac{1}{\sqrt{2}}, 0 \right)^T$.

$H(x_2^*) = \begin{pmatrix} 2\sqrt{2} \exp(-\frac{1}{2}) & 0 \\ 0 & \sqrt{2} \exp(-\frac{1}{2}) \end{pmatrix}$ is positive definite $\Rightarrow x_2^*$ is a strict local minimum

$H(x_1^*) = \begin{pmatrix} -2\sqrt{2} \exp(-\frac{1}{2}) & 0 \\ 0 & -\sqrt{2} \exp(-\frac{1}{2}) \end{pmatrix}$ is negative definite $\Rightarrow x_1^*$ is a strict local maximum.
Example:

- Consider the problem

\[ \min f(x) \triangleq (x_2 - x_1^2)^2 + x_1^5 \]

- \( g(x) = \begin{pmatrix} 5x_1^4 - 4x_1(x_2 - x_1^2) \\ 2(x_2 - x_1^2) \end{pmatrix} \).

- 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 \(f(x)\).
Example:

- Consider the problem

\[
\min f(x) \triangleq x_1^2 + \exp(x_1 + x_2)
\]

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

- Need an iterative method to solve \( g(x) = 0 \).
An iterative optimization algorithm generates a sequence \( \{x^k\}_{k \geq 0} \), which converges to a local minimum.

**Unconstrained Minimization Algorithm**

1. Initialize \( x^0, k := 0 \).
2. **while** stopping condition is not satisfied at \( x^k \)
   
   (a) Find \( x^{k+1} \) such that \( f(x^{k+1}) < f(x^k) \).
   
   (b) \( k := k + 1 \)

**endwhile**

**Output**: \( x^* = x^k \), a local minimum of \( f(x) \).
Unconstrained Minimization Algorithm

(1) Initialize $x^0, k := 0$.

(2) while stopping condition is not satisfied at $x^k$

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

   (b) $k := k + 1$

endwhile

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

- How to find $x^{k+1}$ 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 $x^0$?
Stopping Conditions for a minimization problem:

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

**Practical Stopping conditions**

Assumption: There are no *stationary* points

- \( \|g(x^k)\| \leq \epsilon \)

- \( \|g(x^k)\| \leq \epsilon(1 + |f(x^k)|) \)

- \( \frac{f(x^k) - f(x^{k+1})}{|f(x^k)|} \leq \epsilon \)
Speed of Convergence

Assume that an optimization algorithm generates a sequence \( \{x^k\} \), which converges to \( x^* \).

How *fast* does the sequence converge to \( x^* \)?

**Definition**

The sequence \( \{x^k\} \) converges to \( x^* \) with order \( p \) if

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

- Asymptotically, \( \|x^{k+1} - x^*\| = \beta \|x^k - x^*\|^p \)
- Higher the value of \( p \), faster is the convergence.
(1) \( p = 1, 0 < \beta < 1 \) (Linear Convergence)

Some Examples:
- \( \beta = 0.1, \|x^0 - x^*\| = 0.1 \)
  
  Norms of \( \|x^k - x^*\| : 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, \ldots \)
- \( \beta = 0.9, \|x^0 - x^*\| = 0.1 \)
  
  Norms of \( \|x^k - x^*\| : 10^{-1}, 0.09, 0.081, 0.0729, \ldots \)

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

Example:
- \( \beta = 1, \|x^0 - x^*\| = 0.1 \)
  
  Norms of \( \|x^k - x^*\| : 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, \ldots \)

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

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

then this convergence is called superlinear convergence

Shirish Shevade  |  Numerical Optimization