A constrained optimization problem

The general form of a constrained optimization problem is given as:

$$\begin{array}{lll} \text{(P1)} & \text{Min} & f(x) \\ & \text{s/t} & g_i(x) \leq 0;\ i = 1, 2, ..., m \end{array}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R},\ g_i : \mathbb{R}^n \longrightarrow \mathbb{R};\ i = 1, 2, \ldots, m.$

Classification:

- If f and $g_i(\forall i)$ are linear, then the problem (P1) is called a **linear programming problem(LPP).**
- If the problem (P1) is not linear,then we call it a **non-linear programming problem(NLPP).**
- If in (P1), f is quadratic and $g_i(\forall i)$ are linear, then the problem is called **quadratic programming problem(QPP)**. It is a special class of a non-linear programming problem.

Convex Programming problem

If f and g_i $(i = 1, 2, ..., m)$ in (P1) are convex functions, then we (P1) is called a **convex programming problem(CPP).**

Different formats of CPP

Optimization problem	Conditions for (CPP)
Min $f(x)$ s/t $g_i(x) \leq 0;\ i = 1, 2, \ldots, m$	f and g_i $(\forall i)$ are convex.
Max $f(x)$ s/t $g_i(x) \leq 0;\ i = 1, 2, \ldots, m$	f is concave and g_i $(\forall i)$ are convex.

Min $f(x)$ s/t $g_i(x) \geq 0;\ i = 1, 2, \ldots, m$	f is convex and g_i $(\forall i)$ are concave.
Max $f(x)$ s/t $g_i(x) \geq 0;\ i = 1, 2, \ldots, m$	f and g_i $(\forall i)$ are concave.

Examples:

(P1) Min $x_1 + x_2$

subject to: $x_1^2 + x_2^2 \leq 4,$

$x_1^2 \leq x_2,$

$x_1, x_2 \geq 0.$

(P2) Max $2x_1 - x_2$

subject to: $x_1 + x_2 \leq 3,$

$x_1 x_2 \leq 1,$

$x_1, x_2 \geq 0.$

Here, (P1) is convex while (P2) is not a convex programming problem.

Theorem

Let g_i for each $i = 1, 2, \ldots.., m$ be a convex function. Then,

$$S = \{x \in R^n \ : g_i(x) \leq 0; \ i = 1, 2, \ldots, m\}$$

is a convex set.

(P1) $$\text{Min } f(x) = x_1 + x_2$$

$$\left.\begin{array}{l} \text{s/t} \quad x_1^2 + x_2^2 \leq 4 \\ \quad x_1^2 \leq x_2 \\ \quad x_1 \geq 0, \ x_2 \geq 0 \end{array}\right] \quad \begin{array}{l} \longrightarrow g_1 = x_1^2 + x_2^2 - 4 \leq 0 \\ \longrightarrow g_2 = x_1^2 - x_2 \leq 0 \\ g_3 = -x_1 \leq 0 \\ g_4 = -x_2 \leq 0 \end{array}$$

$$f = x_1 + x_2 \longrightarrow \nabla^2 f = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$g_1 = x_1^2 + x_2^2 - 4$

$$\nabla^2 g_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Minors of order 1×1: $2, 2$

Minors of order $2 \times 2 = |\nabla^2 g|$
$= 4$

$\Rightarrow \nabla^2 g_1$ is positive definite

$\Rightarrow g_1$ is convex.

$$g_2 = x_1^2 - x_2$$

$$\nabla^2 g_2 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

Minors of order $1\times 1 = 2, 0$

Minors of order 2×2

$= 0$

$\Rightarrow$ g_2 is a convex function on $\mathbb{R}^2$.

Quadratic Programming Problem

The general form of a quadratic programming problem is given as:

$$
\begin{aligned}
\text{(QPP)} \quad \text{Min} \quad & f(x) = c^T x + \frac{1}{2} x^T Q x \\
\text{s/t} \quad & Ax \leq b, \\
& x \geq 0
\end{aligned}
$$

where $c, x \in \mathbb{R}^n, b \in \mathbb{R}^m,$ A is a matrix of order $m \times n$ and Q is a symmetric matrix of order n.

Lemma

Let M be a symmetric positive semi-definite matrix of order n. Then, for any $x, y \in \mathbb{R}^n$,

$$x^T My \leq \frac{1}{2}[x^T Mx + y^T My]$$

Theorem

A quadratic programming problem is a **convex** programming problem when Q in the objective function is a symmetric **positive semi-definite matrix**.

Examples:

$$\text{(QP1)} \quad \text{Min} \quad f(x) = x_1^2 + 2x_2^2 + 3x_3^2 + 4x_1x_3 + 2x_2x_3 + x_1 - x_2 + 2x_3$$
$$\text{subject to} \quad x_1 + x_2 + x_3 \leq 10,$$
$$x_1, x_2, x_3 \geq 0.$$

$$\text{(QP2)} \quad \text{Min} \quad f(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 6x_1 - 8x_2$$
$$\text{subject to} \quad 2x_1 - x_2 \leq 13,$$
$$x_1, x_2 \geq 0.$$

Here, (QP2) is a convex QPP while (QP1) is not a convex QPP.

(P1) $$Q = \nabla^2 f = \begin{pmatrix} 2 & 0 & 4 \\ 0 & 4 & 2 \\ 4 & 2 & 6 \end{pmatrix}$$

Minors of order 1×1: $2, 4, 6$

Minors of order 2×2: $\begin{vmatrix} 2 & 0 \\ 0 & 4 \end{vmatrix} = 8$

$$\begin{vmatrix} 4 & 2 \\ 2 & 6 \end{vmatrix} = 24 - 4 = 20$$

$$\begin{vmatrix} 2 & 4 \\ 4 & 6 \end{vmatrix} = 12 - 16 = \underline{\underline{-4}}$$

Constrained Optimization Problem

The general form of a constrained optimization is given as:

$$\begin{array}{lll} \text{(P2)} & \text{Min} & f(x) \\ \text{s/t} & g_i(x) \leq 0; & i = 1, 2, ..., m. \end{array}$$

where f and $g_i(\forall i) : \mathbb{R}^n \to \mathbb{R}$ are defined and continuously differentiable functions.

Necessary Condition

Let $\bar{x}$ be a local min point of the problem at which basic constraint qualification holds. Then there exist multipliers (called KKT-multipliers) $\bar{\lambda}_i, i = 1, 2, ..., m$ such that the following conditions hold:

1. $\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0,$
2. $g_i(\bar{x}) \leq 0, i = 1, 2, ..., m,$
3. $\bar{\lambda}_i g_i(\bar{x}) = 0, i = 1, 2, ..., m,$
4. $\bar{\lambda}_i \geq 0$ for all i.

These conditions are called KKT-conditions.

Sufficient Condition

Let $(\bar{x}, \bar{\lambda}_1, \bar{\lambda}_2, ..., \bar{\lambda}_m)$ satisfy the KKT-conditions $(1) - (4)$. Let f and $g_i (\forall i)$ be differentiable convex functions. Then $\bar{x}$ is a global min point of the problem (P2).

Remark:

Without the convexity assumptions on f and g_i, the KKT conditions are not sufficient for a point $\bar{x}$ to be a local min/global min point.

For example:

$$\begin{aligned} \text{Min} \quad & -x_2 \\ \text{subject to:} \quad & x_1^2 + x_2^2 \leq 4 \\ & -x_1^2 + x_2 \leq 0. \end{aligned}$$

The point **(0,0)** satisfy KKT-conditions but it is not a local/global min point.

$$\text{Min} \quad -x_2$$
$$\text{s/t} \quad g_1 = x_1^2 + x_2^2 - 4 \le 0$$
$$g_2 = -x_1^2 + x_2 \le 0$$

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \Rightarrow \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2}\right) + \lambda_1 \left(\frac{\partial g_1}{\partial x_1} \quad \frac{\partial g_1}{\partial x_2}\right)$$
$$+ \lambda_2 \left(\frac{\partial g_2}{\partial x_1} \quad \frac{\partial g_2}{\partial x_2}\right)$$
$$= (0, 0)$$

$$\Rightarrow (0 \quad -1) + \lambda_1 (2x_1 \quad 2x_2)$$
$$+ \lambda_2 (-2x_1 \quad 1) = (0,0)$$

$$\Rightarrow \left.\begin{aligned} 0 + 2x_1\lambda_1 - 2x_1\lambda_2 &= 0 \\ -1 + 2x_2\lambda_1 + \lambda_2 &= 0. \end{aligned}\right\}$$

Problems

1. Show that $(3/2, 9/4)^T$ is a unique global optimal solution for the following problem:

$$
\begin{aligned}
\text{(P1)} \quad \text{Min} \quad & f(x) = \left(x_1 - \frac{9}{4}\right)^2 + (x_2 - 2)^2 \\
\text{subject to} \quad & x_1^2 \le x_2, \\
& x_1 + x_2 \le 6, \\
& x_1, x_2 \ge 0.
\end{aligned}
$$

2. Solve the following problem:

$$
\begin{aligned}
\text{(P2)} \quad \text{Min} \quad & f(x) = x_1^2 + x_2^2 - 6x_1 - 4x_2 + 13 \\
\text{subject to} \quad & x_1^2 + x_2^2 \le 52, \\
& x_1, x_2 \ge 0.
\end{aligned}
$$

$$\text{Min } f = \left(x_1 - 9/4\right)^2 + \left(x_2 - 2\right)^2$$

$$\text{s/t } \quad g_1 = x_1^2 - x_2 \le 0$$

$$g_2 = x_1 + x_2 - 6 \le 0$$

$$g_3 = -x_1 \le 0$$

$$g_4 = -x_2 \le 0$$

$$\nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \longrightarrow \text{positive definite}$$

$$\Rightarrow f \text{ is convex}$$

$$\nabla^2 g_1 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \longrightarrow \text{positive semi-definite}$$

$$\Rightarrow g_1 \text{ is convex.}$$

$$\nabla f(x) + \sum_{i=1}^{4} \lambda_i \nabla g_i(x) = 0$$

$$\begin{pmatrix} 2(x_1 - 9/4) & 2(x_2-2) \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x_1 & -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 & 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} -1 & 0 \end{pmatrix} + \lambda_4 \begin{pmatrix} 0 & -1 \end{pmatrix} = (0,0)$$

$$\Rightarrow \left.\begin{aligned} 2(x_1 - 9/4) + 2x_1\lambda_1 + \lambda_2 - \lambda_3 &= 0 \\ 2(x_2 - 2) - \lambda_1 + \lambda_2 - \lambda_4 &= 0 \end{aligned}\right\} \longrightarrow \begin{aligned} &2\left(\tfrac{3}{2} - \tfrac{9}{4}\right) + \cancel{2} \times \tfrac{3}{\cancel{2}} \lambda_1 = 0 \\ &2\left(-\tfrac{3}{4}\right) + 3\lambda_1 = 0 \\ &\Rightarrow \boxed{\lambda_1 = \tfrac{1}{2}} \end{aligned}$$

$$\lambda_1 (x_1^2 - x_2) = 0 = \lambda_2 (x_1 + x_2 - 6) = \lambda_3 x_1 = \lambda_4 x_2$$

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0.$$

$$x_1 = 3/2, \; x_2 = 9/4 \Rightarrow \lambda_3 = 0, \; \lambda_4 = 0, \; \lambda_2 = 0$$

$$2\left(\tfrac{9}{4} - 2\right) - \tfrac{1}{2} = \tfrac{9}{2} - 4 - \tfrac{1}{2} = \tfrac{9}{2} - \tfrac{9}{2} = 0$$

Lagrange multiplier.

max/min $Z = f(x,y)$.

s.t

$g(x,y) = c$

Let λ be the Lagrange multiplier

let $F(x,y,\lambda) = f(x,y) - \lambda[g(x,y) - c]$

$$F'_x = f'_x - \lambda g'_x = 0 \Rightarrow f'_x = \lambda g'_x$$
$$F'_y = f'_y - \lambda g'_y = 0 \Rightarrow f'_y = \lambda g'_y.$$
$$F'_\lambda = -[g(x,y) - c] = 0 \Rightarrow g(x,y) = c$$

λ

$$f'_x = \lambda g'_x$$
$$f'_y = \lambda g'_y$$

Unconstrained optimization problems

Consider the following unconstrained minimization problem:

$$(P) \qquad \min_{x \in R^n} f(x).$$

The question arises how to find a point $\bar{x} \in R^n$ which solves (or atleast approximately solves) (P). Because in general, our analytical approach may not work for all types of optimization problems. So, we move to search techniques.

Basic scheme

A common basic scheme is of the form:

$$x_{k+1} = x_k + \alpha_k d_k$$

where x_k is the current solution, d_k is the direction of movement from x_k and $\alpha_k > 0$ is the step size (distance upto which we move from x_k in the direction d_k). How to find α_k and d_k to find next iteration x_{k+1} such that we move to the solution of (P) in an efficient manner?

Descent property

An algorithm for solving (P) is said to have a descent property if $f(x_{k+1}) < f(x_k)$ for all k. That is, as we proceed, the value of objective function should decrease.

Order of convergence

Let a sequence $\{x_k\}$ converge to a point $\bar{x}$ and let $x_k \neq \bar{x}$ for sufficiently large k. The quantity $||x_k - \bar{x}||$ is called the error of the k^{th} iteration. Suppose there exist p and $0 < \alpha < \infty$ such that

$$\lim_{k \longrightarrow \infty} \frac{||x_{k+1} - \bar{x}||}{||x_k - \bar{x}||^p} = \alpha,$$

then p is called the order of convergence of the sequence $\{x_k\}$.

Continued...

- For $p = 1$, we say that the sequence $\{x_k\}$ is linearly convergent.
- For $p = 2$, the sequence $\{x_k\}$ is called quadratic convergent.
- Larger the value of p , faster the algorithm will converge.

Unimodal function

The function $f : [a, b] \longrightarrow R$ is said to be a unimodal function if it has only one peak in the given interval $[a, b]$.
Consider, a unimodal min function $f : [a, b] \longrightarrow R$. Then there exists $a \leq x \leq b$, such that

1. f is strictly decreasing in $[a, x)$.
2. f is strictly increasing in $[x, b]$.

Similarly, we can define for a unimodal max function.

Continued...

Let $f(x)$ be the unimodal min function on the interval of uncertainty $[a, b]$. Take two distinct points (called experiments) x_1 and x_2 such that $x_1 < x_2$, then the following cases may arise

- $f(x_1) < f(x_2) \implies x_{min} \in [a, x_2]$
- $f(x_1) > f(x_2) \implies x_{min} \in [x_1, b]$
- $f(x_1) = f(x_2) \implies x_{min} \in [x_1, x_2]$.

Measure of effectiveness

The measure of effectiveness of any search technique, α is defined as

$$\alpha = \frac{L_n}{L_0}$$

where, L_n is the width of interval of uncertainty after $n-$experiments and L_0 is the initial width of uncetainty.

Steepest Descent method

Consider the following unconstrained minimization problem:

$$\min_{x \in R^n} f(x)$$

where f has continuous first order partial derivatives in R^n.
Choose the starting point as X_1 and move toward the optimal point according to the following rule:

$$X_{k+1} = X_k + \lambda_k d_k$$

where $d_k = -\nabla f(X_k)$ and λ_k is the optimal step size which can be obtained by $\min\{f(X_k + \lambda_k d_k)\}$.
Stopping rule: $||\nabla f(X_k)|| < \epsilon$ or $||f(X_{k+1}) - f(X_k)|| < \epsilon'$.

Steepest Descent algorithm

- is globally convergent.
- has order of convergence unity.
- has descent property.

Example

Use the steepest descent method to minimize $f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2$ such that $|f(X_{k+1}) - f(X_k)| < 0.05$. Take $X_1 = \left(1, \frac{1}{2}\right)^T$.

$$\text{Min } f = x_1^2 - x_1 x_2 + x_2^2$$

$$X_1 = \left(1, \tfrac{1}{2}\right)^T \qquad \nabla f = \left(2x_1 - x_2 \quad -x_1 + 2x_2\right)^T$$

$$\nabla f(X_1) = \left(\tfrac{3}{2} \quad 0\right)^T$$

$$d_1 = -\nabla f(X_1) = \left(-\tfrac{3}{2}, 0\right)^T$$

$$X_2 = X_1 + \alpha_1 d_1 = X_1 + \alpha_1 \left(-\tfrac{3}{2}, 0\right)^T$$

$$= \left(1 \quad \tfrac{1}{2}\right)^T + \alpha_1 \left(-\tfrac{3}{2} \quad 0\right)^T$$

$$X_2 = \begin{pmatrix} 1 - \frac{3}{2}\alpha_1 \\ \frac{1}{2} \end{pmatrix}$$

$$f(x_2) = \left(1 - \frac{3}{2}\alpha_1\right)^2 - \left(1 - \frac{3}{2}\alpha_1\right)\left(\frac{1}{2}\right) + \frac{1}{4}$$

$$\frac{df}{d\alpha_1} = 0 \Rightarrow \cancel{2}\left(1 - \frac{3}{2}\alpha_1\right)\left(-\frac{3}{\cancel{2}}\right) + \frac{3}{4} = 0$$

$$\frac{d^2f}{d\alpha_1^2} > 0 \rightarrow \underline{\text{minima}}$$

$$-3 + \frac{9}{2}\alpha_1 + \frac{3}{4} = 0$$

$$\Rightarrow \frac{9}{2}\alpha_1 = 3 - \frac{3}{4} = \frac{9}{4}$$

$$\Rightarrow \alpha_1 = \frac{1}{2}$$

$$x_2 = \left(\frac{1}{4} \quad \frac{1}{2}\right)^T$$

$$\| f(x_2) - f(x_1) \|$$

Solution

$f(X_1) = f\left(1, \frac{1}{2}\right) = \frac{3}{4}, \quad \nabla f(x_1, x_2) = (2x_1 - x_2, -x_1 + 2x_2)^T$

and $\nabla f(X_1) = \left(\frac{3}{2}, 0\right)^T = -d_1$

$X_2 = X_1 + \lambda_1 d_1$

$= \left(1, \frac{1}{2}\right)^T + \lambda_1 \left(-\frac{3}{2}, 0\right)^T = \left(1 - \frac{3}{2}\lambda_1, \frac{1}{2}\right)^T.$

Now, to determine, λ_1,

$f(X_2) = f\left(1 - \frac{3}{2}\lambda_1, \frac{1}{2}\right) = \left(\frac{2 - 3\lambda_1}{2}\right)^2 - \left(\frac{2 - 3\lambda_1}{4}\right) + \frac{1}{4}$

$\frac{df(X_2)}{d\lambda_1} = 0 \implies \lambda_1 = \frac{1}{2}.$

Therefore, $X_2 = \left(\frac{1}{4}, \frac{1}{2}\right)^T$. Since $|f(X_2) - f(X_1)| = 0.75 \not< 0.05$

Continued...

So find the next iteration,

Now,

$$
\begin{aligned}
X_3 &= X_2 + \lambda_2 d_2 \\
&= \left(\frac{1}{4}, \frac{1}{2}\right)^T + \lambda_2 \left(0, -\frac{3}{4}\right)^T, d_2 = -\nabla f(X_2) \\
&= \left(\frac{1}{4}, \frac{1}{2} - \frac{3}{4}\lambda_2\right)^T
\end{aligned}
$$

$$f(X_3) = \frac{1}{16} - \left(\frac{1}{2} - \frac{3}{4}\lambda_2\right)\left(\frac{1}{4}\right) + \left(\frac{1}{2} - \frac{3}{4}\lambda_2\right)^2$$

$$\frac{df}{d\lambda_2} = 0 \implies \lambda_2 = \frac{1}{2}.$$

Hence, $X_3 = \left(\frac{1}{4}, \frac{1}{8}\right)^T$. Also, $|f(X_3) - f(X_2)| = \frac{9}{64} < 0.05$.

Newton's method

Basic Scheme

Newton's method is an iterative method used for finding real roots of the equation $g(y) = 0, \; y \in \mathbb{R}$. The iterative formula for finding roots is given as:

$$y_{k+1} = y_k - \frac{g(y_k)}{g'(y_k)}$$

where y_k is the current iterate or the current approximation.

For unconstrained optimization

Consider the following unconstrained minimization problem:

$$(P) \quad \min_{x \in R^n} f(x)$$

where $f : R^n \longrightarrow R$ is a differentiable function. For solving (P), we have to find $\bar{x} \in R^n$ such that $\nabla f(\bar{x}) = 0$. So, by the Newton scheme (in numerical methods), we have

$$x_{k+1} = x_k - (H_f(x_k))^{-1} \nabla f(x_k). \tag{1}$$

Proof

The quadratic approximation the function f in (P), in a neighbourhood of x_k by the Taylor series is given as:

$$f(x) \approx f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T H_f(x_k)(x - x_k).$$

For minimization, $\nabla f(x) = 0$. This implies,

$$\begin{aligned}
& \nabla f(x_k) + H_f(x_k)(x - x_k) = 0 \\
\implies & H_f(x_k)(x - x_k) = -\nabla f(x_k) \\
\implies & x - x_k = -\left(H_f(x_k)\right)^{-1} \nabla f(x_k) \\
\text{or} \quad & x_{k+1} = x_k - \left(H_f(x_k)\right)^{-1} \nabla f(x_k).
\end{aligned}$$

This method has order of convergence, $p = 2$ and it has descent property. For solving quadratic functions (involving positive definite quadratic form), it will take exactly one iteration to find the optimal solution.

Example

Use Newton's method to minimize

$$f(x_1, x_2) = x_1^2 - x_1 x_2 + 3x_2^2, \ (x_1, x_2) \in R^2.$$

Take initial approximation $x_1 = (1, 2)^T$.

Solution

$x_{k+1} = x_k - (H_f(x_k))^{-1} \nabla f(x_k)$.

$$H_f(x) = \begin{bmatrix} 2 & -1 \\ -1 & 6 \end{bmatrix}, \nabla f(x) = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 6x_2 \end{bmatrix}$$

$$(H_f(x))^{-1} = \frac{1}{11} \begin{bmatrix} 6 & 1 \\ 1 & 2 \end{bmatrix}, \nabla f(x_1) = (0, 11)^T$$