Page :
Date :

⇒ Steepest Descent method

→ Consider the following unconstrained minimization problem:

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

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 have to search techniques or numerical optimization algorithm.

→ A common basic scheme is of the form:

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

x_k → current solution

d_k → 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).

Our goal: 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 ~~be~~ 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.

Page :
Date :

→ 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

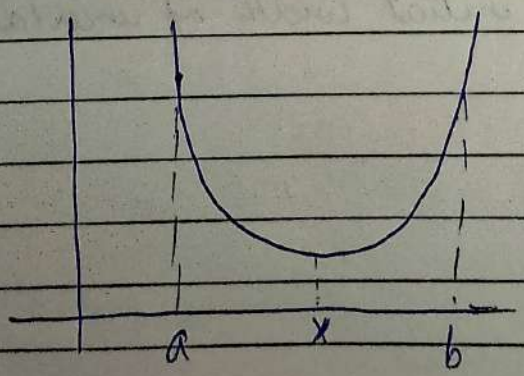
$$\lim_{k \to \infty} \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|^p} = \alpha$$

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

If $p = 1$ → sequence $\{x_k\}$ is linearly convergent
$p = 2$ → " " " quadratic " (order 2)
larger the value of p, faster the algorithm will converge.

→ Unimodal function: The function $f : [a,b] \to R$ is said to be a unimodal function if it has only one peak in the given interval $[a,b]$

→ Consider, a unimodal min function $f : [a,b] \to R$. Then, there exist $a \leq x \leq b$ such that
① f is strictly decreasing in $[a, x)$
② f is strictly increasing in $[x, b]$
similarly we can define unimodal max function.



unimodal min function

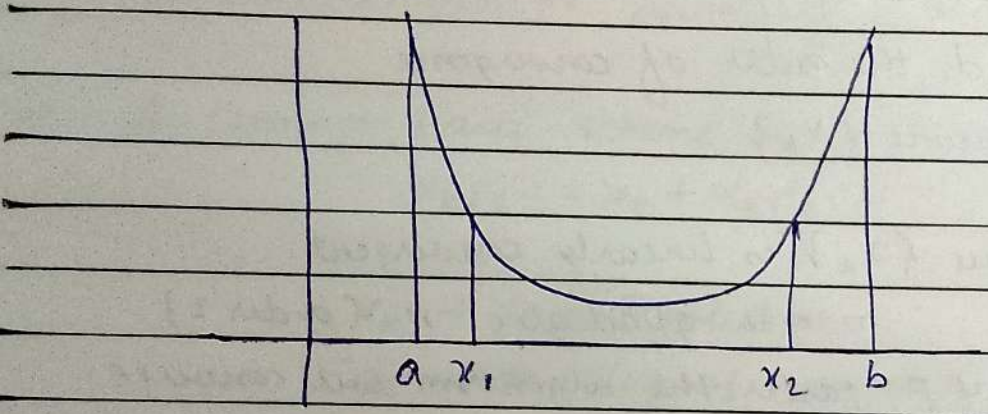
Page :
Date :

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

→ $f(x_1) < f(x_2) \Rightarrow x_{min} \in [a, x_2]$

→ $f(x_1) > f(x_2) \Rightarrow x_{min} \in [x_1, b]$

→ $f(x_1) = f(x_2) \Rightarrow x_{min} \in [x_1, x_2]$



→ Measure of effectiveness:

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

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

L_n → width of interval of uncertainity after n-experiments

L_0 → is the initial width of uncertainity

Page :
Date :

→ SD method:

Consider the following unconstrained minimization problem:

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

where f has continuous first order partial derivates in R^n.

Choose the starting points 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

$$\min_{\lambda_k} \{f(x_k + \lambda_k d_k)\}.$$

Stopping rule: $\| \nabla f(x_k) \| < \epsilon$

or

$$\| f(x_{k+1}) - f(x_k) \| < \epsilon'$$

Property of steepest Descent algorithm:

1. It is globally convergent.
2. It has order of convergence with $p=1$
3. It has descent property. $(f(x_{k+1}) < f(x_k))$ for all k

Page :

Date :

Ex: 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 initial point $X_1 = \left(1, \frac{1}{2}\right)^T$

Sol:

$\min f = x_1^2 - x_1 x_2 + x_2^2$ → To find

$X_1 = \left(1, \frac{1}{2}\right)^T$ → Given

$\nabla f = (2x_1 - x_2 \quad -x_1 + 2x_2)^T$ → Step (1)

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

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

So,

$$X_2 = X_1 + \lambda_1 d_1 = X_1 + \lambda_1 \left(-\frac{3}{2}, 0\right)^T$$

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

$$X_2 = \begin{pmatrix} 1 - \frac{3}{2}\lambda_1 \\ \frac{1}{2} \end{pmatrix}$$ → Step (4)

$$f(X_2) = \left(1 - \frac{3}{2}\lambda_1\right)^2 - \left(1 - \frac{3}{2}\lambda_1\right)\left(\frac{1}{2}\right) + \frac{1}{4}$$ → Step (5)

$$\frac{df}{d\lambda_1} = 0 \Rightarrow 2\left(1 - \frac{3}{2}\lambda_1\right)\left(-\frac{3}{2}\right) + \frac{3}{4} = 0$$

↓ solve it

$\lambda_1 = \frac{1}{2}$ → Step (6)

$\frac{d^2 f}{d\lambda_1^2} > 0 \rightarrow$ It is minima

Page :
Date :

$\Rightarrow x_2 = \left(\frac{1}{4}, \frac{1}{2}\right)^T$ → step (7)

Now check for stopping ruling

$\|f(x_2) - f(x_1)\| = 0.75$

since $0.75 \not< 0.05$ which means we will proceed further

Next iteration

$x_3 = x_2 + \lambda_2 d_2$

$= \left(\frac{1}{4}, \frac{1}{2}\right)^T + \lambda_2 \left(0, -\frac{3}{4}\right)^T$

$d_2 = -\nabla f(x_2)$ → solving this and putting it in above you will get

$x_3 = \left(\frac{1}{4}, \frac{1}{2} - \frac{3}{4}\lambda_2\right)^T$

$f(x_3) = \frac{1}{16} - \left(\frac{1}{2} - \frac{3}{4}\lambda_2\right)\left(\frac{1}{4}\right) + \left(\frac{1}{2} - \frac{3}{4}\lambda_2\right)^2$

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

Hence $x_3 = \left(\frac{1}{4}, \frac{1}{8}\right)^T$. Also $|f(x_3) - f(x_2)| = \frac{9}{64} < 0.05$

Since the stopping criteria is met we will stop here say that x_3 is the optimal solution for our function $f(x_1, x_2)$

Page :
Date :

⇒ Other techniques:

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

$$y_{k+1} = y_k - \frac{g(y_k)}{g'(y_k)} \qquad g'(y_k) \neq 0$$

where y_k is the current iterate or the current approximation.

• Consider the following unconstrained minimization problem:

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

where $f: R^n \to R$ 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 Newton scheme (in numeric methods), we have

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

using taylor series: $f(x) \simeq f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)^T H_f(x_k)(x - x_k)$

now, $\nabla f(x) = 0$

$\Rightarrow \nabla f(x_k) + H_f(x_k)(x - x_k) = 0$

$\Rightarrow H_f(x_k)(x - x_k) = -\nabla f(x_k)$

$\Rightarrow x - x_k = -(H_f(x_k))^{-1} \nabla f(x_k)$

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

↓

Hessian Matrix which is invertible at $x = x_k$

Date :

It 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

Ex: $f(x_1, x_2) = x_1^2 - x_1x_2 + 3x_2^2 \qquad (x_1, x_2) \in R^2$

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

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

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

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

So, $x_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

similarly and $x_3 = x_2 - (H_f(x_2))^{-1} \nabla f(x_2)$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{11}\begin{pmatrix} 6 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is point of minima

This was for unconstrained optimization but what about constrained optimization problem.

Page :
Date :

→ Numerical optimization for constrained optimization:
The general form of a non-linear constrained optimization problem (CPP) is given as:

(NLP) Min $f(x)$

subject to $g_i(x) \leq 0$; $(i = 1, 2, \dots, m)$

where $F: R^n \to R$, $g_i : R^n \to R$; $i = 1, 2, \dots, m$ are differentiable convex functions.

A numerical optimization technique for constrained optimization problem, which can be solved using numerical techniques for unconstrained problems.

One such methods to convert the constrained NLP into an unconstrained problem is the introduction of a penalty function. Consider the following function:

$$\tilde{P}(x) = \begin{cases} 0; & x \in S \\ +\infty; & x \notin S \end{cases}$$

where $S = \{x \in R^n \mid g_i(x) \leq 0 \ (i = 1, 2, \dots, m)\}$ is the feasible set. Introduction of $\tilde{P}(x)$ in the objective function converts the NLP into equivalent unconstrained problem as:

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

Because $\tilde{P}(x)$ this function is not smooth or differentiable, so we have to make it smooth.

Page :
Date :

The penalty function method uses the concept of introduction of penalty function to convert the NLP into an unconstrained optimization problem. It uses a smooth penalty function (called quadratic loss function) defined as:

$$P(x) = \sum_{i=1}^{m} [\max(g_i(x), 0)]^2$$

and constructing a sequence of unconstrained problem defined as:

$$(UMP)_\alpha : \min_{x \in R^n} f(x) + \alpha P(x)$$

we do this to make our problem smooth.

(i) Choose a suitable penalty function. Here, we take

$$P(x) = \sum_{i=1}^{m} [\max(g_i(x), 0)]^2$$

(ii) Choose an increasing sequence of positive real numbers which tends to $+\infty$, i.e sequence $\{\alpha_k\}_{k=1}^{\infty}$ such that for each k, $\alpha_k > 0$, $\alpha_{k+1} > \alpha_k$ and $\{\alpha_k\} \to +\infty$.
In general, we take $\alpha_1 = 1$, $\alpha_2 = 10$, $\alpha_3 = 100$, $\alpha_4 = 1000$ and so on...

(iii) choose an arbitrary starting point $x_0 \in R^n$.
Construct the following unconstrained minimization problem:

$$(UMP)_{\alpha_1} : \min_{x \in R^n} f(x) + \alpha_1 P(x)$$

and solve it using a suitable unconstrained minimization technique, starting with x_0. let x_1 be the optimal solution of $(UMP)_{\alpha_1}$ set $k = 1$

Date :

(IV) Construct $(UMP)_{\alpha_{k+1}}$ as

$$(UMP)_{\alpha_{k+1}} : \min_{x \in R^n} q(x, \alpha_{k+1}) = f(x) + \alpha_k P(x)$$

and solve it using a suitable unconstrained minimization technique, starting with x_k where x_k is the optimal solution of $(UMP)_{\alpha_k}$.

(V) Stopping Criteria:

Continue iterations of Step (4) till $\alpha_k P(x_k) < \epsilon$ or $q(x_k, \alpha_k) - f(x_k) < \epsilon$ for some tolerance level $\epsilon > 0$

What's the proof that it will converge? we have some criteria that we can use to check:

lema 1 Let $\bar{x}_k$ denote the optimal solution of $(UMP)_{\alpha_k}$ i.e $q(\bar{x}_k, \alpha_k) = \min_{x \in R^n} q(x, \alpha_k)$. where

$q(x, \alpha_k) = f(x) + \alpha_k P(x)$, $\alpha_k > 0$ then,

(I) $q(\bar{x}_k, \alpha_k) \leq q(\bar{x}_{k+1}, \alpha_{k+1})$

(II) $P(\bar{x}_k) \geq P(\bar{x}_{k+1})$

(III) $f(\bar{x}_k) \leq f(\bar{x}_{k+1})$

lema 2: Let $\bar{x}$ be an optimal solution of the given non-linear programming problem (NLP) Then for each k:

$$f(\bar{x}) \geq q(\bar{x}_{k+1}, \alpha_k) \geq f(\bar{x}_k)$$

Proof of the lema 1 and lemma 2 is not required in ML

Page:
Date:

Ex: Solve the following NLP using Penalty function method, starting with $x_0 = (2,2)^T$ and $\epsilon = 0.001$

(P) min $3x_1^2 + 2x_2^2 + 2x_1x_2 - 20x_1 - 16x_2$

subject to $x_1 + x_2 = 5$

Sol: The constraint $x_1 + x_2 = 5$ can be written as $x_1 + x_2 - 5 \leq 0$ and $-x_1 - x_2 + 5 \leq 0$

Thus, the penalty function is given as:

$P(x)$ or $P(x_1, x_2) = [Max((x_1 + x_2 - 5), 0)]^2 + [Max((-x_1 - x_2 + 5), 0)]^2$

$= (x_1 + x_2 - 5)^2$ ← ~~In both cases we will get~~ this because if $x_1 + x_2 - 5$ is -ve other will be positive vice versa

Now our unconstrained min problem will be:

$g(x) = \min f(x) + \alpha_k (x_1 + x_2 - 5)^2$

If $k = 1$ and $x_1 = 1$, then:

~~$g(x) = f(x) + (x_1 + x_2 - 5)^2$~~

$\Rightarrow (3x_1^2 + 2x_2^2 + 2x_1x_2 - 20x_1 - 16x_2) + (x_1^2 + x_2^2 + 25 - 10x_1 - 10x_2 + 2x_1x_2)$

$\Rightarrow 4x_1^2 + 3x_2^2 + 4x_1x_2 - 30x_1 - 26x_2 + 25$

$X_0 = (2,2)^T$, $\nabla g = \begin{pmatrix} 8x_1 + 4x_2 - 30 \\ 6x_2 + 4x_1 - 26 \end{pmatrix}$

$\nabla g(2,2) = \begin{pmatrix} -6 \\ -6 \end{pmatrix}$

$X_1 = X_0 + \alpha_0 \begin{pmatrix} 6 \\ 6 \end{pmatrix}$ ← (Plus because we take min of gradient)

$= \begin{pmatrix} 2 + 6\alpha_0 \\ 2 + 6\alpha_0 \end{pmatrix}$

Page :
Date :

Similarly we will keeep searching with other values of k and alpha for our problem the optimum lies at k = 3 and $\alpha_k = 100$

for this

$x_k = (2.334, 2.669)$

$f(x_k) = -46.3353$

$g(x_k, \lambda_k) = -46.3344$

$P(x_k) = 0.000000009$

$\alpha_k P(x_k) = 0.0009$ which is less than our ϵ value

so our optimum solution using our stopping criteria is :

$(x_1, x_2) = (2.334, 2.669)$