OPTİMİZASYON TEKNİKLERİ-3. Hafta

ÇOK DEĞİŞKENLİ FONKSİYONLARIN OPTİMİZASYONUNU (Maks/Min bulunması)

Bu fonksiyonlar

$y = f(x_1, x_2, x_3 \ldots x_n)$ şeklinde gösterilir. Yani içerisinde birden fazla sayıda değişken bulunduruyor.

$z = f(x, y \ \ldots)$ şeklinde farklı karakterlerle de gösterebiliriz.

örneğin,

$f(x) = x_1^3 + x_2^2 + x_3^2 - 12x_1 - 12x_3 + 100$ ya da

$f(x, y) = x^2 - 4xy + y^2$ şeklinde olabilir.

D $y = f(x_1, x_2, x_3 \ldots x_n)$ şeklindeki çok değişkenli fonksiyonların ekstremumlara sahip olması için

Gerek Şart

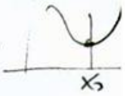
$$\nabla(x_1, x_2, \ldots x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots \frac{\partial f}{\partial x_n}\right) = 0$$

olması gerekir. Bu türevlerden elde edilen denklemlerin çözümü olan noktalara Sabit noktalar denir. ve her bir $x_1, x_2 \ldots x_n$ değişkeni yerine konulan sabit sayılar olmuş olur. örneğin 3 değişkenli ise

$X_0 = ((x_1)_0, (x_2)_0, (x_3)_0)$ şeklinde gösterebiliriz.

Sabit değer olan bir X_0 noktasının bir extremum noktası olması için yeter şart o noktadaki Hessian matrisinin değeri $H_f(X_0)$ belirler.

- $H_f(X_0) > 0 \Rightarrow X_0$ bir minimum noktadır.



- $H_f(X_0) < 0 \Rightarrow X_0$ bir maksimum noktadır.

- $H_f(X_0)$ tanımsız $\Rightarrow X_0$ büküm noktası olur.

Bunun anlamı şudur: Sabit X_0 noktasının minimum noktası olması için yeter şart Hessian Matrisinin bu noktada Pozitif tanımlı olmasıdır. Aynı yeter şart X_0'ın maksimum nokta olması için Hessian Matrisinin X_0 noktasında negatif tanımlı olması gerektiğini söyler.

Hessian matrisi nedir? Pozitif ve negatif tanımlı olması nasıl olmaktadır onlara bakalım.

(3)

Hessian Matrisinin tanımı şu şekildedir

$y = f(x_1, x_2, x_3 \ldots x_n)$ n değişkenli fonksiyon 2. mertebeden sürekli türevlerinden oluşturulan aşağıdaki gibi simetrik bir $n \times n$ matristir.

$$Hf(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}, & \frac{\partial^2 f}{\partial x_1 \partial x_2}, \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1}, & \frac{\partial^2 f}{\partial x_2^2}, \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1}, & \frac{\partial^2 f}{\partial x_n \partial x_2}, \ldots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{n \times n},$$

Simetrik bir matristir. ($\forall\, i \neq j \quad a_{ij} = a_{ji}$ olan matris)

Hessian matrisinin Alt matrislerini bulalım.

$$H = \begin{bmatrix} a_{11}, & a_{12} \ldots & a_{1n} \\ a_{21}, & a_{22} \ldots & a_{2n} \\ \hline a_{n1}, & a_{n2}, \ldots & a_{nn} \end{bmatrix}_{n \times n}$$

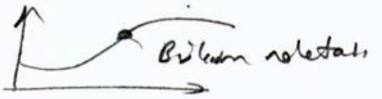
Hessian matrisinin alt matrisleri

$$A_1 = [\, a_{11} \,], \quad A_2 = \begin{bmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \end{bmatrix}, \ldots A_n = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2n} \\ a_{n1} & \ldots & a_{nn} \end{bmatrix}$$

şeklinde sol üst köşeden itibaren simetrik olacak şekilde her oluşturulan matris H matrisinin Alt matrisleridir.

(4) Bu alt matrislerin determinantlarının hepsi pozitif ise H (Hessian) matrisi pozitif tanımlı denektir. Yani;

- $\forall i$ için $\det A_i > 0$ ise H poz. tanımlı
  (her i için şeklinde okunur.)
- $\forall i$ için $(-1)^i \det A_i > 0$ ise H neg. tanımlı
- Diğer durumlarda tanımsızdır. Noktası bükülme noktasıdır.



Determinant neydi Kısaca bir hatırlayalım

Bir matrisin elemanlarını tek bir sayıya düşürme işlemidir. A matrisin determinant det A şeklinde yada |A| şeklinde gösterilir. determinantının alınabilmesi için kare matris olmalıdır. Yani satır sayısı ile sütun sayısı eşit olmalıdır.

Örnekler

$A = [-2] \Rightarrow$ tek boyutlu ise determinant aynıdır. $\det(A) = -2$ dir.

$A = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}_{2\times2}$ ise

$\det(A) = $ (birinci köşegenler çarpımı) $-$ (ikinci köşegen) $= 3\cdot2 - 4\cdot1 = 2.$

$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 4 & -2 & 1 \end{bmatrix}_{3\times3}$

$$\begin{array}{r|ccc|r} & 1 & 2 & 3 & \\ & 0 & -1 & 2 & \\ & 4 & -2 & 1 & \\ -12 & 1 & 2 & 3 & - \\ -4 & 0 & -1 & 2 & 0 \\ 0 & & & & 16 \\ \hline -16 & & & & 15 \end{array}$$

$\det(A) = 15 - (-16) = 31$

- ilk iki satırı aşağı yaz.
- Gösterilen çizgiler oluştur.
- Çapraz çizgiler oluştur. Çizgi üzerindeki sayıları çarp aşağıda topla. Sağdaki soldaki gibi.

$y=f(x_1, x_2, \dots x_n)$ n- değişkenli fonksiyonu 2. mertebeden sürekli kısmi türevlere sahipken bu fonksiyonun **hessian matrisi** simetrik bir matris (simetrik matris: $\forall i \neq j$ olmak üzere aij=aji olan matristir) olup aşağıdaki gibidir ;

$$Hf(x)=\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{nxn} = \begin{bmatrix} f_{11} & f_{12} & f_{1n} \\ f_{21} & f_{22} & f_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ f_{n1} & f_{n2} & f_{nn} \end{bmatrix}_{nxn}$$

$y=f(x_1, x_2, \dots x_n)$ fonksiyonların ekstremumlara sahip olması için;

i) Gerek şart:

$\forall(x_1, x_2, .., x_n) = grad f(x_1, x_2, .., x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n} \right) = 0$ olmasıdır. Bu denklemin çözümü olan noktalara **sabit noktalar** denir,

ii) Yeter şart: x_0 noktası $\nabla f(x) = 0$ şartını sağlayan nokta(**sabit nokta)** olsun.Buna göre;

1.Test :

i. **$Hf(x_0) > 0$ (Pozitif tanımlı) ise minimum noktasıdır**
ii. **$Hf(x_0) < 0$ (Negatif tanımlı)ise maximum noktasıdır**
iii. **$Hf(x_0)$ tanımsız ise x_0 büküm noktasıdır**

2.Test:

det(A-λI)=0 n.dereceden bir polinom denklem olup buna f fonksiyonunun **karakteristik polinomu** ,bu denklemin köklerine de **karakteristikler** veya **özdeğerler** denir.Buna göre ;

i. $\forall$i için $\lambda_i > 0$ **ise A pozitif tanımlıdır ($A=(a_{ij})_{nxn}$)**
ii. $\forall$i için $\lambda_i < 0$ **ise A negatif tanımlıdır ($A=(a_{ij})_{nxn}$)**
iii. **Diğer durumlarda tanımsızdır.**

A matrisinin tanımlılığı:nxn'lik bir A matrisinin tanımlılığını belirlemek için aşağıdaki test uygulanır.

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ . & . & . \\ . & . & . \\ a_{n1} & a_{n2} & a_{nn} \end{bmatrix}_{nxn}$$ olsun. A'nın uzanan alt matrisleri ;

$$A_1=[a_{11}] \;;\; A_2=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}_{2x2}, \dots, Aj = \begin{bmatrix} a_{11} & a_{12} & a_{1j} \\ a_{21} & a_{22} & a_{2j} \\ a_{j1} & a_{j2} & a_{jj} \end{bmatrix}_{jxj}, \dots An \begin{bmatrix} a_{11} & . & a_{1n} \\ a_{21} & . & a_{2n} \\ a_{n1} & . & a_{nn} \end{bmatrix}_{nxn}$$ olarak tanımlanır.

Bu matrislerin determinantlarının hepsi pozitif ise A matrisi pozitif tanımlıdır .Yani;

i. **$\forall i$ için det $A_i>0$ ise A pozitif tanımlıdır.**
ii. **$\forall i$ için $(-1)^i$det $A_i>0$ ise A negatif tanımlıdır.**
iii. **Diğer durumlarda tanımsızdır .Nokta büküm noktasıdır.**

(6) **Örnek** $f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2x_3 - x_1^2 - x_2^2 - x_3^2$

fonksiyonun ekstremumlarını inceleyin

Öncelikle fonksiyonun her bir değişkene göre türevlerini alalım. Bunları sıfıra eşitleyelim

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0$$

$$\frac{\partial f}{x_2} = x_3 - 2x_2 = 0$$

$$\frac{\partial f}{\partial x_3} = 2 + x_2 - 2x_3 = 0$$

Bu denklemleri çözelim.

$$1 - 2x_1 = 0 \Rightarrow$$

$$x_1 = \frac{1}{2}$$

$$x_3 - 2x_2 = 0 \Rightarrow$$

$$x_2 = \frac{x_3}{2}$$

$$2 + x_2 - 2x_3 = 0 \Rightarrow 2 + \frac{x_3}{2} - \frac{2 \cdot x_3}{1} = 0$$

$$2 + \frac{x_3 - 4x_3}{2} = 0 \Rightarrow 2 - \frac{3x_3}{2}$$

$$x_3 = \frac{4}{3} \Rightarrow x_2 = \frac{4/3}{2} = \frac{2}{3}$$

Buradan sabit noktalar

$x_1 = 1/2$, $x_2 = 2/3$, $x_3 = 4/3$ bulunmuş olur

Yani sabit olarak alacağımız değerler

$x_0 = (1/2,\ 2/3,\ 4/3)$ değerleri olur

7) Bu fonksiyonun Hessian matrisini oluşturalım. $n=3$ tür yani (x_1, x_2, x_3) vardır. matriste $n \times n = 3 \times 3$ olur.

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) = \frac{\partial}{\partial x_1}\left(\underbrace{1-2x_1}_{\frac{\partial f}{\partial x_1}}\right) = -2.$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2}\right) = \frac{\partial}{\partial x_1}\left(\underbrace{x_3 - 2x_2}_{\frac{\partial f}{\partial x_2}}\right) = 0$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_3} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_3}\right) = \frac{\partial}{\partial x_1}\left(2 + x_2 - 2x_3\right) = 0.$$

$\frac{\partial^2 f}{\partial x_2 \partial x_1}$ için $\frac{\partial}{\partial x_2}\left(1-2x_1\right) = 0$

$\frac{\partial^2 f}{\partial x_2^2}$ için $\frac{\partial}{\partial x_2}\left(x_3 - 2x_2\right) = -2$

$\frac{\partial^2 f}{\partial x_2 \partial x_3}$ için $\frac{\partial}{\partial x_2}\left(2 + x_2 - 2x_3\right) = 1$

$\frac{\partial^2 f}{\partial x_3 \partial x_1}$ için $\frac{\partial}{\partial x_3}\left(1-2x_1\right) = 0.$

$\frac{\partial^2 f}{\partial x_3 \partial x_2}$ için $\frac{\partial}{\partial x_3}\left(x_3 - 2x_2\right) = 1$

$\frac{\partial^2 f}{\partial x_3^2}$ için $\frac{\partial}{\partial x_3}\left(2 + x_2 - 2x_3\right) = -2.$

(8) Bu değerleri Hessian matrisi formatında yerine yazarsak

$$Hf(x) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}, & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}, & \dfrac{\partial^2 f}{\partial x_1 \partial x_3} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}, & \dfrac{\partial^2 f}{\partial x_2^2}, & \dfrac{\partial^2 f}{\partial x_2 \partial x_3} \\ \dfrac{\partial^2 f}{\partial x_3 \partial x_1}, & \dfrac{\partial^2 f}{\partial x_3 \partial x_2}, & \dfrac{\partial^2 f}{\partial x_3^2} \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

Bulduğumuz bu Hessian matrisinin Alt matrislerini oluşturalım.

$$A_1 = [-2], \quad A_2 = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -2, & 0, & 0 \\ 0, & -2, & 1 \\ 0, & 1, & -2 \end{bmatrix}$$

Bu alt matrislerin determinantlarını Hesaplayalım. Önce det

$\boxed{\det A_1 = -2.}$

$\det A_2 = (-2).(-2) - 0$

$\boxed{\det A_2 = 4}$

$$A_3 = \begin{bmatrix} -2, & 0, & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

$$\begin{matrix} 0 & -2 & 0 & 0 \\ -2 & 0 & -2 & 1 \\ 0 & & & \end{matrix}$$

$$\frac{+}{-2} \qquad \frac{+}{-8}$$

$-8 - (-2) = -6$

$\boxed{\det A_3 = -6}$

$(-1)^i \det A_i > 0$ ise neg. tanımlıdır

$i=1 \quad \det A_1 = -2 \Rightarrow (-1)^1 (-2) = 2$

$i=2 \quad \det A_2 = 4 \Rightarrow (-1)^2 \, (4) = 4$

$i=3 \quad \det A_3 = -6 \Rightarrow (-1)^3 \, (-6) = 6$

$(-1)^i \det A_i > 0$ kuralını hepsi sağladığı için H matrisi negatif tanımlıdır.

Tüm determinantlar pozitif değil

dolayısıyla $X_0 = ((x_1)_0, (x_2)_0, (x_3)_0) = (1/2, \ 2/3, \ 4/3)$ noktasında fonksiyon maksimuma sahiptir. Yani fonksiyon konkavdır.

Örnek : $f(x_1,x_2,x_3)=x_1+2x_3+x_2x_3-x_1^2-x_2^2-x_3^2$ fonksiyonun ekstremumlarını ve konveks yada konkavlığını inceleyiniz

1)Gerek şart: $\nabla f(x)=0$ olmalıdır

$$\frac{\partial f}{\partial x_1}=1-2x_1=0 \quad , \quad \frac{\partial f}{\partial x_2}=x_3-2x_2=0 \quad , \quad \frac{\partial f}{\partial x_3}=2+x_2-2x_3=0$$

Buradan; $x_1=1/2$, $x_2=2/3$, $x_3=4/3$ olduğu görülür . $x_0=(1/2,2/3,4/3)$ sabit noktadır.

2)Yeter şart:Bu fonksiyona ait Hessian matrisini oluşturalım

$$Hf(x_0)=\begin{bmatrix} f_{x1fx1} & f_{x1x2} & f_{x1x3} \\ f_{x2x1} & f_{x2x2} & f_{x2x3} \\ f_{x3x1} & f_{x3x2} & f_{x3x3} \end{bmatrix} \text{ ise } Hf(x_0)=\begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

$H_1=[-2]$ olduğundan det $H_1=-2$

$H_2=\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$ olduğundan det $H_2=4$

$H_3\begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ olduğundan det $H_3=-6$

$H_f(x)$ negatif tanımlıdır $x_0=(1/2,2/3,4/3)$ maksimum noktadır

Not:yeter şart için II. Metot şu şekildedir.

$$\text{Det }(H-I\lambda)=\begin{vmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{vmatrix} - \lambda\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} -2-\lambda & 0 & 0 \\ 0 & -2-\lambda & 1 \\ 0 & 1 & -2-\lambda \end{vmatrix} = 0 \text{ ve buradan;}$$

$p(\lambda)=(-2-\lambda)[(-2-\lambda)^2-1]=0$ ise $p(\lambda)=-\lambda^3-6\lambda^2-11\lambda-6=0$ olur.

$p(-\lambda)=\lambda^3-6\lambda^2+11\lambda-6=0$ 'dır (3 tane negatif kök vardır , fonksiyon konkavdır yani x_0 noktasımaksimum noktasıdır)

Örnek : $f(x,y)=x^2-4xy+y^2$ fonksiyonunun ekstremumlarını ve konveks yada konkavlığını inceleyin.

Çözüm:

1)Gerek şart: $\nabla f(x)=0$ olmalıdır

$$\frac{\partial f}{\partial x}=2x-4y=0 \, , \, \frac{\partial f}{\partial y}=-4x+2y=0$$

Buradan; x=0, y=0 olduğu görülür. $x_0=(0,0)$ sabit noktadır.

2)Yeter şart:$Hf(x_0)=\begin{bmatrix} fxx & fxy \\ fyx & fyy \end{bmatrix}=\begin{bmatrix} 2 & -4 \\ -4 & 2 \end{bmatrix}$

$H_1=[2]$ ise det $H_1=2>0$

$H_2=\begin{bmatrix} 2 & -4 \\ -4 & 2 \end{bmatrix}$ ise det H_2 = -12<0

H_f tanımsız olduğundan f fonksiyonu ne konveks nede konkavdır, $x_0$büküm noktasıdır.

Örnek: $f(x,y,z)=4x^2-6y^2-2xy+3xz-2y-4yz+1$ fonksiyonunun ekstremumlarını bulunuz

Çözüm:

Gerek Şart: $\nabla f(x)=0$ olmalıdır.Buna göre;

$f_x=8x-2y+3z=0$

$f_y=-12y-2x-4z-2=0$

$f_z=3x-4y=0$ olup buradan $x_0(-6/7,-9/14,13/7)$ bulunur.

Yeter şart : $f_{xx}=-2$,$f_{xy}=3$,$f_{yy}=-12$,$f_{yz}=-4$,$f_{zz}=0$ olup Hessian matrisi;

$$H=\begin{bmatrix} 8 & -2 & 3 \\ -2 & -12 & -4 \\ 3 & -4 & 0 \end{bmatrix} \text{elde edilir.}$$

1.test : det [8] = 8 >0 ve det $\begin{bmatrix} 8 & -2 \\ -2 & -12 \end{bmatrix}=-100<0$ ($H_f(x)$ tanımsız , x_0 büküm noktası , fonksiyon ne konveks ne de konkavdır)

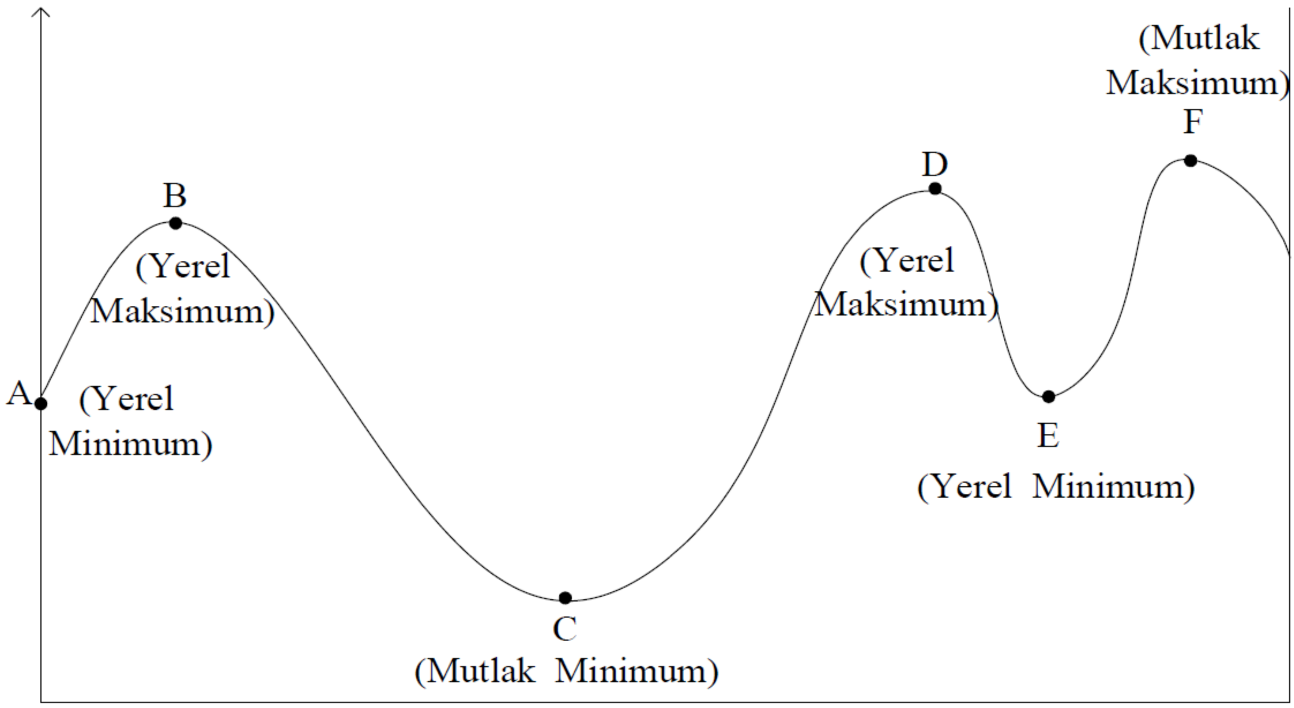
Kısıtlı doğrusal olmayan problemlerin ekstremumlarını yerleştirmek için geliştirilen Klasik Optimizasyon Teorisi sayısal hesaplamalar için uygun değildir. Fakat bu teoriler hesap algoritmalarını geliştirmede temel teşkil ederler. Öyle ki Kuadratik Programlama Karush – Kuhn – Tucker gerek ve yeter şartlarını kullanan mükemmel bir örnektir.

ÇOK DEĞİŞKENLİ MAK/MİN ARAMA ALGORİTMALARI

Gradient Metodu:n - değişkenli bir $f(x)=f(x_1,x_2,......,x_n)$ fonksiyonunun yerel optimumunu bulmak için 140 yılı aşkın bir süredir çok çeşitli algoritmalar geliştirilmiştir.

Bu algoritmaların çoğu şu esasa göre tasarlanmıştır.

$f(x)=f(x_1,x_2,......,x_n)$ n - değişkenli fonksiyonunun grafiğini tepe ve vadilerden oluşan bir sıradağlar topluluğu olarak gözönüne alırsak, vadilerin en alt tarafındaki bölgeler fonksiyonun minimumunu gösterirken tepelerin en üstündeki bölgelerde fonksiyonun maksimumunu temsil eder.Eğer minimum noktada değil isek bulunduğumuz noktadan aşağıya doğru minimumu buluncaya kadar hareketimize devam ederiz. Maksimum noktada değil isekte bulunduğumuz noktadan yukarıya doğru maksimumu buluncaya kadar hareketimize devam ederiz(2). (Şekil12)



Şekil 12

Bir gradientin yönü en dik çıkış (steepest ascent) yönüdür. Tersi ise en dik iniş (steepest descent)yönüdür. Yani en dik çıkış yönü $\nabla f(x)$ yönünde iken en dik iniş yönüde - $\nabla f(x)$ yönündedir(1).

n-değişkenli bir $f(x) = f(x_1, x_2, ..., x_n)$ fonksiyonunu gözönüne alalım.

Burada, $(x_1, x_2, ..., x_n)$ n - boyutlu öklidyen uzayda;

$$X = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}$$ sütun vektörü ile temsil edilir.

f(x) fonksiyonunun gradienti ise grad f(x) veya $\nabla f(x)$ ile gösterilir ki;

$$\text{grad } f(x) = (f_1, f_2, \ldots\ldots, f_n) \quad (12) \text{ veya}$$

$$\nabla f(x) = \frac{\partial f}{\partial x_k} \qquad (k = 1, 2, \ldots\ldots, n) \qquad (1)$$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \ldots\ldots\ldots, \frac{\partial f}{\partial x_n} \right) \text{ şeklindedir.}$$

ifadesinde kısmi türevler $f_k = \dfrac{\partial f}{\partial x_k}$ $k = 1, 2, \ldots\ldots, n$ x'de belirlenir(1).

Şimdi $X = (x_1, x_2, \ldots\ldots, x_n)$ uygun değerleri üzerinde herhangi bir kısıtlama olmaksızın $f(X) = (x_1, x_2, \ldots\ldots, x_n)$ konkav fonksiyonunun maksimizasyon problemini gözönüne alalım.

Bir boyutlu arama işlemini çok boyutlu problemimiz için genişletmeye çalışalım. Bir boyutlu problemlerde bayağı türev bir veya iki olası yolu yani x'in azalış veya x'in artış yönünü seçmek için kullanılır. Amaç bir noktayı araştırmaktır ki bu noktada türev sıfırdır. $\left[f'(x) = 0 \right]$

Bu özellik çok değişkenli bir fonksiyon için şu şekilde genişletilebilir. Sabit noktaları araştırmak amaç iken, bu noktalarda bütün kısmi türevler sıfıra eşittir(7).

Bu nedenle bir boyutlu aramanın çok boyutlu aramaya genişletilebilmesi kısmi türevleri kullanarak hareket yönünde özel bir yön seçmeyi gerektirir. İşte bu gereksinim amaç fonksiyonunun gradientini kullanmayı kapsar.

f(x) amaç fonksiyonu yüksek mertebeden sürekli kısmi türevlere sahip olduğunda herbir x noktasında f(x) , $\nabla f(x)$ ile gösterilen bir gradiente sahiptir(7).

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots\ldots\ldots, \frac{\partial f}{\partial x_n} \right)$$

Özel bir $x = x'$ noktasındaki gradient ise elemanları kısmi türevler olan ve x' noktasında değeri olan bir vektördür.

$$\nabla f(x') = \left(\frac{\partial f}{\partial x_1}, \ldots\ldots\ldots, \frac{\partial f}{\partial x_n} \right) \Bigg|_{x = x'}$$

Gradient arama işlemi kısıtsız bir problemin çözümü için etkili bir arama işlemidir. Bu işlem gradientin yönündeki hareketi koruyan ve x^* optimal çözümünü araştıran bir yöntemdir.

Bu nedenle normal olarak $\nabla f(x)$'in yönünde x'i sürekli olarak değiştirmek pratik degildir. Çünkü bu değişikliklerin serisi sürekli olarak $\frac{\partial f}{\partial x_j}$'lerin yeniden belirlenmesini ve her ikisinin yönünde değişiklik yapmayı gerektirir.

Bundan dolayı en iyi yaklaşım geçerli aşikar çözümden sabit bir yöndeki hareketi koruyan ve f(x) fonksiyonunun artışı durana dek devam eden bir yaklaşımdır(19).

f(x)'in artışının bittiği nokta ise son aşikar çözümdür ve bu noktada gradienthareketin yeni yönünü belirlemek için yeniden hesaplanacaktır. Bu yaklaşımla herbir ardışık işlem x' geçerli aşikar çözümün değişikliğini kapsar. Şöyle ki;

$$x' = x' + t^* \nabla f(x') \text{ alalım.} \qquad \text{(I)}$$

Burada, t^*, $f\left[x' + t\nabla f(x')\right]$'yi maksimum eden pozitif (t) değeridir.

$$f\left[x' + t^* \nabla f(x')\right] = \max f\left[x' + t\nabla f(x')\right] \qquad (t \geq 0)$$

(I) ifadesini genişletecek olursak;

$$x_j = x_j' + t\left(\frac{\partial f}{\partial x_j}\right)\Bigg|_{x = x'} \qquad j = (1, 2, \ldots\ldots, n)$$

elde edilir. Bu son ifade x_j için yalnızca sabitleri ve t'yi içerir. Yani f(x) t'nin bir fonksiyonudur. Bu gradient arama işleminin iterasyonları küçük bir ε toleransı ile $\nabla f(x) = 0$ olana dek sürdürülür. Yani;

$$\left|\frac{\partial f}{\partial x_j}\right| \leq \varepsilon \qquad j = (1, 2, \ldots\ldots, n)$$

Şimdi bu değindiğimiz özelliklere göre gradient işlemini basit bir algoritma ile tanımlayalım.

Başlangıç Adımı: İstenilen hassasiyet ε ve başlangıç aşikar çözüm (x')'yi alalım.

Ardışık Adımlar:

1) $x_j = x_j^{'} + t\left(\frac{\partial f}{\partial x_j}\right)\Bigg|_{x=x^{'}}$, $j = (1, 2, \ldots\ldots\ldots, n)$

Kümesi oluşturarak t'nin bir fonksiyonu olan $f\left[x^{'} + t\,\nabla f(x^{'})\right]$ ifadesi oluşturulur ve f(x) yerine bu ifadeler konulur. Son adıma gidilir.

2) $t \geq 0$ için bir boyutlu arama işlemi kullanılarak $f\left[x^{'} + t\,\nabla f(x^{'})\right]$'yi maksimum kılan $t = t^{*}$ bulunur.

3) $x^{'} = x^{'} + t^{*}\,\nabla f(x^{'})$ oluşturulur ve son adıma geçilir.

Son Adım: $x = x^{'}$ de $\nabla f(x^{'})$ hesaplanır. Daha sonra $\left|\frac{\partial f}{\partial x_j}\right| \leq \varepsilon$'un sağlanıp sağlanmadığı kontrol edilir. Eğer bu kontrol sağlanıyorsa x^{*} optimal çözümünün yaklaşık beklenen değeri için $x^{'}$ optimal çözüm alınır. Aksi halde ilk adıma dönülüp işlem tekrarlanır(7).

Bütün gradient arama tekniklerinin ortak özelliği

$\nabla f = g = \left[\frac{\partial f}{\partial f_1}, \ldots\ldots\ldots, \frac{\partial f}{\partial f_n}\right]$'i kullanmalarıdır.

Ayrıca; $x_{k=1} = x_k - t\,\nabla f(x_k)$ iterasyon işlemini kullanmalarıdır.

Örnek: Gradient arama işleminin nasıl yapıldığını göstermek için konkav bir fonksiyon için iki değişkenli kısıtsız optimizasyon problemini göz önüne alalım.

Max. $f(x) = 2x_1x_2 + 28x_1 - x_1^2 - x_1^4 + 4x_2 - x_2^2$

Çözüm:x_1 ve x_2 değişkenlerine göre kısmi türevlerini alırsak;

$$\frac{df}{dx_1} = 2x_2 + 28 - 2x_1 - 4x_1^3$$

$\frac{df}{dx_2} = 2x_1 + 4 - 2x_2$ olarak bulunur. Buna göre f(x) in grandienti

$$\nabla f(x) = \text{grad } f(x) = \left(\frac{df}{dx_1}, \frac{df}{dx_2}\right) = (2x_2 + 28 - 2x_1 - 4x_1^3,\ 2x_1 + 4 - 2x_2)\text{ dır.}$$

grad f(x) = 0 şartından; $(2x_2 + 28 - 2x_1 - 4x_1^3,\ 2x_1 + 4 - 2x_2) = (0, 0)$

Böylece, $2x_2 + 28 - 2x_1 - 4x_1^3 = 0$

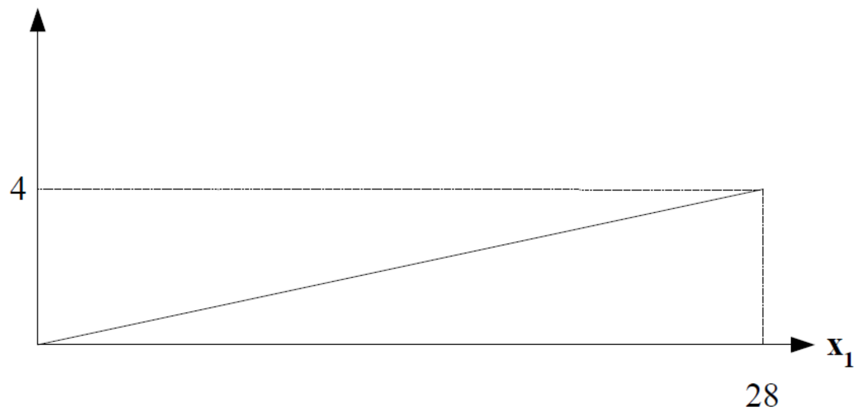
$2x_1 + 4 - 2x_2 = 0$

Bu denklemlerin aynı anda çözümü ile optimal çözüm $(x_1, x_2) = (2, 4)$ bulunur. Buna göre; $f(2,4) = 2.(2).(4) + 28.(2) - (2)^2 - (2)^4 + 4.(4) - (4)^2 = 52$ ve grad $f(2,4) = 8 + 28 - 4 - 32, 4 + 4 - 8) = (0,0)$ bulunur.

Şimdi f(x) ve grad f(x) göz önüne alınarak gradient arama işleminin nasıl elde edilebileceğini görelim. Bu işleme başlamak için bir başlangıç aşikar çözüme ihtiyacımız vardır ki; bu noktada $f(x_1, x_2) = (0, 0)$ dır. Böylece $(x_1, x_2) = (0, 0)$ başlangıç aşikar çözüm olarak alınabilir.

Bu nokta için gradient; grad f (0 , 0) = (28.4) Bu şu anlama gelir x = (0 , 0) noktasında f(x) deki max. artma oranı (0 , 0) dan (0 , 0) + (28.4) e hareket edilerek bulunur. Buna göre (0 , 0) ve (28.4) noktaları arasındaki doğru denklemi;(Şekil.13)

(0 , 0) + [(28.4) – (0 , 0)] = (28t , 4t) dır.Burada $t > 0$ olmalıdır.



Şekil.13

(0 , 0) dan (28,4) noktalarını birleştiren doğru boyunca ne kadar hareket etmeliyiz. Bir başka deyişle; (28t , 4t) doğrusu için t sıfırdan ne kadar arttırılmalıdır.

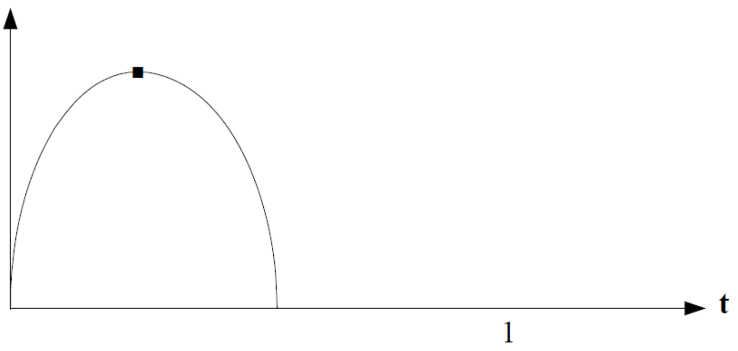
Bu soruların cevabı şudur: f(x) ‘in artması durana kadar bu doğru boyunca hareket edilir. Yani $x_1 = 28t$ ve $x_2 = 4t$ değerleri fonksiyonda yerine yazılır.

$f(28t, 4t) = 2(28t)(4t) + 28(28t) - (28t)^2 - (28t)^4 + 4(4t) - (4t)^2$

$f(28t, 4t) = 800t - 576t^2 - 614.656t^4$

f (28t , 4t) ‘ nin değeri max. olana kadar t’nin arttırılması gerekir. Bunu gerçekleştirmek için bir boyutlu arama işlemi ile t’nin maksimum değeri; $t^* = 0.0665$ olarak bulunur. (Şekil.14)

$$f(28t,4t) = 800t - 576t^2 - 614.656t^4$$



Şekil.14

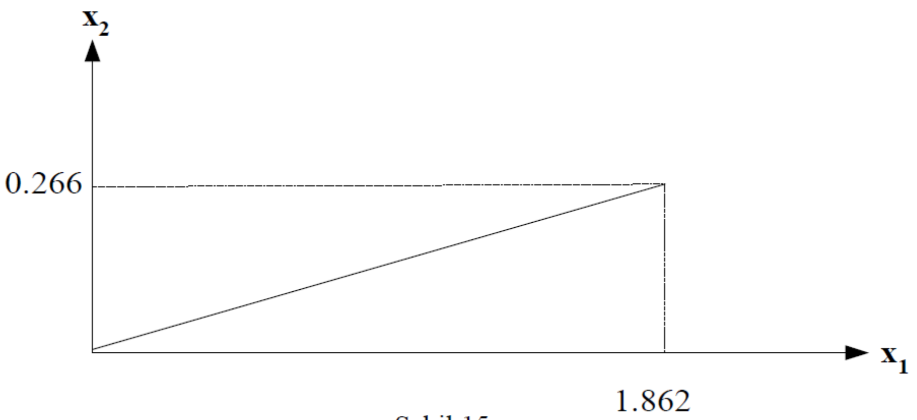
Buna göre yeni aşikar çözüm;$(x_1, x_2) = (28t^*, 4t^*) = (1.862\ ,\ 0.266\ )$ olur.

$f(1.862\ ,\ 0.266\ ) = 38.63$

Bu ilk iterasyonu;

Adım	x’	grad f(x’)	x’ + [grad f(x’)]	t^*	x’+ t^* [grad f(x’)]
1	(0,0)	(28.4)	(0+ 28t , 0+ 4t)	0.067	(1.862,0.266)

şeklinde bir genel tablo ile ifade edebiliriz.(Şekil.15)



Şekil.15

Şimdi ikinci iterasyonu ilk iterasyonun sonucuna göre tekrarlayalım. Geçerli aşikar çözüm $(x_1\ ,\ x_2) = (\ 1.862, 0.266\ )$ noktasıyla $(\ 1.862, 0.266\ ) + (0.832\ ,\ 7.459\ )$ noktasını birleştiren doğru boyunca olacaktır.

Bu doğrunun denklemi;

$(\ 1.862, 0.266\ ) + (-1.028\ , 7.193) = (1.862 - 1.028t\ ,\ 0.266 + 7.193t)$

bu doğru boyunca;

$$x_1 = 1.862 - 1.028\ t$$

$$x_2 = 0.266 + 7.193\ t \quad \text{dır.}$$

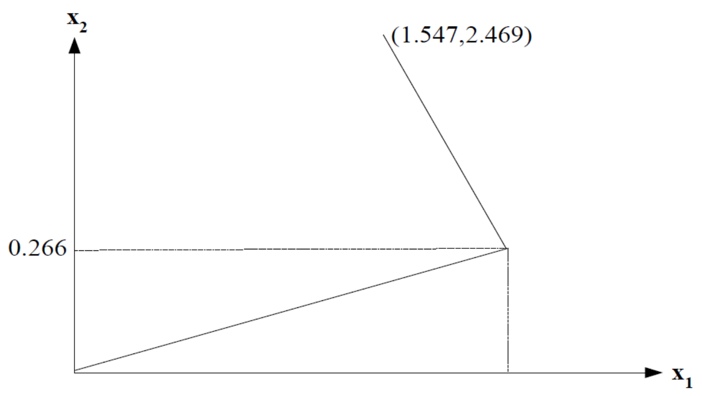
Böylece; $f(x_1, x_2) = 38.6 + 52.8\ t - 89.7\ t^2 + 8.14\ t^3 - 1.13\ t^4$ ve bir boyutlu arama işlemi gerçekleştirilirse; $t^* = 0.306$ olarak bulunur. Buna göre yeni aşikar çözüm;

$(x_1\ ,\ x_2) = (1.862 - 1.028t\ ,\ 0.266 + 7.193t) = (1.547\ ,\ 2.469\ )$

Şimdi bu iki iterasyonu

Adım	x’	grad f(x’)	x’ + [grad f(x’)]	t^*	x’+ t^* [grad f(x’)]
1	(0,0)	(28.4)	(0+ 28t , 0+ 4t)	0.067	(1.862,0.266)
2	(1.862,0.266)	(-1.03,7.193)	(1.86-1.03t,0.27+7.19t)	0.306	(1.547,2.469)

şeklinde bir genel tablo ile gösterilebilir. A aşikar çözümlerin optimal noktaya doğru hareketlerini de yandaki şekille gösterebilir. (Şekil.16)



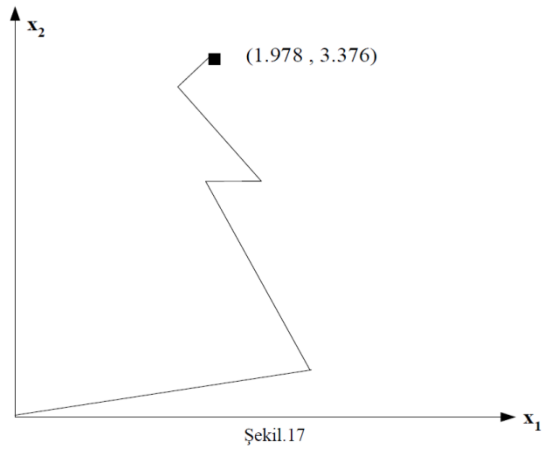
Şekil.16

Bundan sonraki iterasyonlarıda genel tablomuzla aşağıdaki şekilde verebiliriz.

Adım	x’	grad f(x’)	x’ + [grad f(x’)]	t^*	x’+ t^* [grad f(x’)]
1	(0,0)	(28.4)	(0+ 28t , 0+ 4t)	0.067	(1.862,0.266)
2	(1.862,0.266)	(-1.03,7.193)	(1.86 – 1.03t,0.27 + 7.19t)	0.306	(1.547,2.469)
3	(1.547,2.469)	(15.09,2.155)	(1.55+15.1t ,2.47 + 2.19t)	0.027	(1.948,2.526)
4	(1.948,2.526)	(-0.41,2.843)	(1.95+0.41t ,2.53 + 2.84t)	0.291	(1.978,3.354)
5	(1.829,3.354)	(6.59,0.99)	(1.83+6.59t ,3.35+ 0.99t)	0.023	(1.978,3.376)

İterasyonların sayısı istenilen hassasiyete bağlı olarak belirlenir. Eğer iterasyona devam edilecek olursa belli bir iterasyon sonunda tam optimal çözüme erişilebilir.

Bu örneğimizde 0.1 hata toleransı söz konusu iken tam optimal çözüm olan $(x_1 , x_2) = (2,4)$ noktasını elde etmek için yapılan iterasyonlar yukarıdaki tabloda sunulmuştur. Bu optimal çözüme ulaşıncaya kadar elde edilen bütün aşikar çözümler birleştirildiğinde son aşikar çözüm olan $(x_1 , x_2) = (2,4)$ (optimal çözüm) noktasına doğru zigzag bir grafik ortaya çıkar (Şekil.17).



Şekil.17

3.4. UYGULAMALAR VE YORUMLARI

3.4.1. Seçilen Doğrusal Olmayan Programlama Tekniklerinin Karşılaştırmalı Genel Yorumu

Bu çalışmamızda seçtiğimiz Doğrusal Olmayan Programlama Tekniklerini tek-çok değişkenli, kısıtlı-kısıtsız, direkt-indirekt olmak üzere üç ana başlık altında tasnif edebiliriz. Buna göre Newton-Raphson Metodu tek değişkenli , indirekt(türev kullanan), kısıtsız bir ardışık işlem metodudur.

Nelder - Mead Medodu ise çok değişkenli kısıtsız ve direkt (türev kullanmayan) bir arama metodudur. Gradient metodu ise adından da anlaşılacağı gibi çok değişkenli ,kısıtsız ve indirekt bir arama metodudur. Frank-Wolfe Algoritması ise çok değişkenli , doğrusal kısıtlı ve indirekt bir Gradient arama işlemidir. SUMT Algoritması ise çok değişkenli, doğrusal ve doğrusal olmayan eşitlik ve eşitsizlik kısıtlı ve indirekt olan bir Gradient arama işlemidir.

Bu tasnife göre karşılaştırmalı yorumu kısıtlı metodları kendi aralarında , kısıtsız metodları kendi aralarında olmak üzere ayıralım.

Newton-Raphson Metodu y = f(x) tek değişkenli fonksiyonunun sıfırlarını bulan metodlarının arasında gerçek köke en hızlı yakınsayan bir ardaşık arama işlemidir. Bu nedenle çalışmada tek değişkenli fonksiyonları temsilen seçilmiştir.

Nelder - Mead Medodu ile Gradient Metodu çok değişkenli kısıtsız fonksiyonların optimizasyonunu bulan iki metod olmasına rağmen Nelder ve Mead Medodu türev kullanmayan (direkt) bir arama işlemi , Gradient Metodu ise türevleri kullanan (indirekt) arama işlemidir.

Gradient arama işleminin en önemli özelliği optimum çözüme steepest ascent (en dik çıkış), steepest descent (en dik iniş) veya zigzag şeklinde bir grafik çizerek yakınsamasıdır. Bu Gradient arama işleminin en karakteristik özelliğidir. Bu nedenle uygun bölgede seçilen her başlangıç aşikar çözüm optimum çözüme hızlı bir şekilde yakınsar. Uygulama 4 ve Uygulama 5 bunu açıkca ortaya koymaktadır. Fakat bu metodun bir dezavantajı adım uzunluğunun küçülmesi durumunda bir sonraki iterasyona geçememesidir. Bu bilgisayarlardaki alt taşması , üs taşması gibi hatalarından kaynaklanmaktadır. Bu nedenle ancak uygun başlangıç aşikar çözümlerin, seçimi ile tam optimum çözüme ulaşabiliriz.

Nelder – Mead tarafından geliştirilen metod ise bir simpleks metodudur ve çok değişkenli kısıtsız fonksiyonların optimizasyonunu bulmakta oldukça etkilidir. İki boyutta simpleks bir üçgen olduğundan, ilk üçgenin üç köşe noktası başlanğıç vektörler alınarak işleme başlanır.

Bu köşe noktalarda maksimum problem için en büyük, minimum problem için en küçük fonksiyon değerine sahip noktaya Best Vertex (En iyi köşe nokta), sonraki en iyi noktaya Good Vertex (Sonraki en iyi köşe nokta) ve maksimum problemlerde en küçük, minimum problemlerde en büyük fonksiyon değerine sahip noktaya da Worst Vertex (En kötü köşe nokta) denir. Bu işlem her iterasyonda tekrarlanarak iki boyutlu problmlerde üçgenlerin bir dizisi ile optimum çözüme yaklaşılır. B, G ve W ile isimlendirdiğimiz üçgenin bu köşe noktaları birbirlerine eşit iken bir noktayı gösterirler ki işte bu nokta aranılan optimumu noktadır. Bu nedenle başlangış vektörlerinin iyi seçimi ile bu metod Gradient Metodundan daha etkili bir arama işlemi haline gelir.

Buna göre, Gradient Algoritması her iterasyonda kısmi türevlerin hesabı gerektiğinden Nelder - Mead Metoduna göre daha çok işlem süresine sahip bir arama işlemidir. Gradient arama işleminde adım uzunlukları arasındaki fark yakınsaklığın durumu göstermesi açısından önemli iken Nelder - Mead Metodunda bu ölçüt her iterasyondaki üçgenin bir öncekine göre daha küçülmesi yani B, G ve W noktalarının birbirine yaklaşması ile belirlenebilir.

Şimdi de Frank-Wolfe ve SUMT Algoritmalarının kendi aralarında karşılaştırmalı yorumlarını yapalım. Bu iki algoritmada kısıtlı çok değişkenli fonksiyonların optimizasyonunu bulurlar. İkisinin ortak özelliği işlem adımları esnasında Gradient arama işlemini kullanmalarıdır. Yani her ikisi de İndirekt (türev kullanan) arama işlemidir.

Frank-Wolfe Algoritması doğrusal kısıtlı problemlerin optimizasyonunu bulurken, SUMT Algoritması hem doğrusal hem doğrusal olmayan kısıtlı problemlerin optimizasyonunu bulur. Frank-Wolfe Algoritmasının genel karakteristiği doğrusal kısıtlı , doğrusal amaç fonksiyonu olan bir doğrusal programlama problemine dönüşür . Frank-Wolfe Algoritmasının diğer önemli bir karakteristiği ise iterasyonların optimum çözüme doğru iki farklı yönden yaklaşmasıdır. Bu nedenle Frank-Wolfe Algoritmasında optimum çözüme yakınsama SUMT Algoritmasına göre daha yavaş bir yakınsamadır.

SUMT Algoritmasının en genel karekteristiği ise ceza fonksiyonlarını kullanarak kısıtlı optimizasyon problemlerini kısıtsız optimizasyon problemlerinin bir dizisi ile çözmesidir. Bu nedenle SUMT Algoritması uygun bölge içinde seçilen bir başlangıç aşikar çözüm ile optimum çözüme çok hızlı bir şekilde yakınsayan ve $r = 0.000000001$ değeri için;

$P(x,r) = f(x) - rB(x)$ fonksiyonunun amaç değerini bulan bir gradient arama işlemidir.

Bu özellikleri ile SUMT Algoritması gerek doğrusal kısıtlı gerekse doğrusal olmayan kısıtlı fonksiyonların optimizasyonunda çok etkili bir algoritmadır.