

**DİFERANSİYELLENEMEYEN OPTİMİZASYON
İÇİN BAZI SUBGRADYENT METOTLAR**

Emrah ÜNAL

**Yüksek Lisans Tezi
Matematik Anabilim Dalı
Doç. Dr. Murat SUBAŞI**

2012

Her hakkı saklıdır

ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

**DİFERANSİYELLENEMEYEN OPTİMİZASYON İÇİN BAZI
SUBGRADYENT METOTLAR**

Emrah ÜNAL

MATEMATİK ANABİLİM DALI

ERZURUM

2012

Her hakkı saklıdır



T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



TEZ ONAY FORMU

DİFERANSİYELLENEMEYEN OPTİMİZASYON İÇİN BAZI SUBGRADYENT
METOTLAR

Doç. Dr. Murat SUBAŞI danışmanlığında, Emrah ÜNAL tarafından hazırlanan bu çalışma 29.11.2012 tarihinde aşağıdaki jüri tarafından Uygulamalı Matematik Anabilim Dalı'nda Yüksek Lisans tezi olarak **oybirliği/oy çokluğu (3./3)** ile kabul edilmiştir.

Başkan : Doç.Dr.Murat SUBAŞI

İmza : 

Üye : Yrd.Doç.Dr.Arzu AYKUT

İmza : 

Üye : Doç.Dr.Alper Cihan KONYALIOĞLU

İmza : 

Yukarıdaki sonucu onaylıyorum



Prof. Dr. İhsan EFEOĞLU
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZET

Y.Lisans Tezi

DİFERANSİYELLENEMEYEN OPTİMİZASYON İÇİN BAZI SUBGRADYENT METOTLAR

Emrah ÜNAL

Atatürk Üniversitesi
Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı

Danışman: Doç. Dr. Murat SUBAŞI

Bu tezde ilk olarak subgradyent ve subdiferansiyel kavramları ele alınıp bazı örnekler verilmiştir. Daha sonra türevlenemeyen bir fonksiyonun minimumunun bulunmasına yardımcı olan subgradyent metot göz önüne alınarak, subgradyent metodun modifiye edilmiş halleri olan pür subgradyent metot, deflected subgradyent metot, koşullu subgradyent metot ve zikzak-free subgradyent metotları incelenmiştir. Son olarak yukarıdaki metotları kıyaslamamıza yardımcı olacak nümerik örnekler oluşturulmuştur.

2012, 72 Sayfa

Anahtar Kelimeler: Subgradyent, Subdiferansiyel, Subgradyent Metot

ABSTRACT

MS. Thesis

CERTAIN SUBGRADIENT METHODS FOR NONDIFFERENTIABLE OPTIMIZATION

Emrah ÜNAL

Atatürk University

Graduate School of Natural and Applied Sciences

Department of Mathematics

Supervisor: Assoc. Prof. Dr. Murat SUBAŞI

In this thesis firstly, taking into account the subjects subgradient and subdifferential, some examples have been given. After that, considering the subgradient method which helps to find out the minimum of an undifferentiable function, the modified versions of subgradient method, which are pur subgradient method, deflected subgradient method, conditional subgradient method and zigzag free subgradient method, have been dealt with. Finally, numerical examples are constituted which help to compare these methods.

2012, 72 Pages

Keywords: Subgradient, Subdifferential, Subgradient method

TEŐEKKÜR

Yüksek Lisans tezi olarak sunduđum bu alıŐma sırasında danıŐmanlıđımı üstlenip deđerli bilgi ve katkılarını benden esirgemeyen Atatürk Üniversitesi Fen Fakóltesi Matematik Bölümü öğretim üyelerinden Sayın Do. Dr. Murat SUBAŐI hocama en içten teşekkürlerimi sunmayı bor bilirim.

Ayrıca sađlamıŐ olduđu maddi destekten dolayı Tübitak kurumuna teşekkürlerimi sunarım.

Emrah ÜNAL

Aralık 2012

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR.....	iii
ŞEKİLLER DİZİNİ	v
1. GİRİŞ	1
2. KURAMSAL TEMELLER	2
3. MATERYAL ve YÖNTEM	5
3.1. Subgradyent Kavramı	5
3.2. Subdiferansiyel Kavramı ve Bazı Fonksiyonların Subdiferansiyelinin Bulunması.....	8
3.3. Subgradyent metotlar	17
3.3.1. Pür Subgradyent metot	19
3.3.2. Deflected Subgradyent metot	25
3.3.3. Koşullu subgradyent metot.....	37
3.3.4. Zikzak-Free Subgradyent metot	51
3.4. Nümerik Uygulamalar	62
4. ARAŞTIRMA BULGULARI	70
5. SONUÇ	71
KAYNAKLAR	72
ÖZGEÇMİŞ	73

ŞEKİLLER DİZİNİ

Şekil 3.1. f fonksiyonunun bazı subgradyentleri.....	5
Şekil 3.2. $f = \max \{f_1, f_2\}$ fonksiyonunun grafiği.....	7
Şekil 3.3. $f(x) = x $ fonksiyonunun ve subdiferansiyelinin grafiği	9
Şekil 3.4. $f(x) = x $ fonksiyonunun subdiferansiyel kümesi	10
Şekil 3.5. $x = (x_1, x_2)$ olmak üzere $f(x) = \ x\ _1$ fonksiyonunun subdiferansiyel kümesi	11
Şekil 3.6. $(1, 0)$ noktası için $f(x) = \ x\ _1$ fonksiyonunun subdiferansiyel kümesi.....	12
Şekil 3.7. $f(x, y) = \max \{-x, x + y, x - 2y\}$ fonksiyonunun subdiferansiyel kümesi	14
Şekil 3.8. $(-t, -t)$ noktası için f fonksiyonunun subdiferansiyel kümesi olduğunu görürüz.....	16
Şekil 3.9. İterasyon noktalarının oluşturduğu zikzaklar	26
Şekil 3.10. Lemmanın ispatının şekil üzerinde gösterimi.....	29
Şekil 3.11. Pür subgradyent metotla oluşan ikinci tür zikzak.....	39
Şekil 3.12. $x \in bd \Omega$ noktasının tanjant ve normal konisi.....	40
Şekil 3.13. $\Omega = [a, b]$ kümesinin uç noktaları için subgradyent ve koşullu subgradyentler.....	43
Şekil 3.14. Pür subgradyentle optimal noktaya yaklaşma	65
Şekil 3.15. Deflected subgradyentle optimale yaklaşma	69

1. GİRİŞ

Klasik türev ve diferansiyel kavramları günümüz bilim ve teknolojisinin birçok alanlarında ortaya çıkan çoğu problemi çözmesine karşın bu kavramlar bir çoğunun çözümünde yetersiz kalmaktadır. Örneğin optimizasyon teorisi, kontrol teorisi ve oyun teorisi gibi bir çok alanda incelenen fonksiyonların çoğu diferansiyellenebilir olmayan fonksiyonlardır. Bu yüzden diferansiyellenemeyen fonksiyonlar üzerinde işlem yapmamızı sağlayacak subgradyent ve subdiferansiyel kavramlarına ihtiyaç duyulmuştur. Diferansiyellenemeyen konveks bir fonksiyonu minimalleştirmek için kullanılan subgradyent metot 1960 ve 1970 li yıllarda Shor tarafından Sovyetler Birliğinde geliştirilmiştir. (Akgül 1984; Shor 1985; Bertsekas 1999) çalışmaları subgradyent metotlar için iyi bir referanstır.

Bu çalışmamızda subgradyent ve subdiferansiyel kavramını tanıtıp, tanıttığımız bu kavramlardan faydalanarak konveks ve diferansiyellenemeyen bir fonksiyonu minimalleştirmede kullanılan subgradyent metodları inceleyeceğiz.

2. KURAMSAL TEMELLER

Bu bölümde (Dedetürk 2007), (Bayraktar 2010) çalışmalarından faydalandığımız bazı tanımlara yer verilecek.

Tanım 2.1: S , R^n nin açık bir alt kümesi ve $f : S \rightarrow R$ fonksiyonu verilmiş olsun. Ayrıca $x = (x_1, x_2, \dots, x_n) \in S$ ve $d = (d_1, d_2, \dots, d_n)$, R^n de birim vektör olsun. Bu takdirde

$$f'(x; d) = \lim_{t \rightarrow 0} \frac{f(x + td) - f(x)}{t}$$

değerine f nin x noktasında d yönünde yönlü türevi denir.

Tanım 2.2: $A \subset R^n$ bir küme olsun. $x, y \in A$ ve $\alpha \in [0, 1]$ iken

$$\alpha x + (1 - \alpha)y \in A$$

oluyorsa A kümesine konveks kümedir denir.

Tanım 2.3: $f : R^n \rightarrow R$ bir fonksiyon olsun. $\forall x, y \in R^n \times R^n$ ve $\forall \alpha \in [0, 1]$ için

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

eşitsizliği varsa bu f fonksiyonuna konveks fonksiyon denir.

Tanım 2.4: $S \subset R^n$ bir küme olsun. S kümesini kapsayan tüm konveks kümelerin kesişimine S kümesinin kapalı konveks zarfı denir ve $co S$ ile gösterilir.

Tanım 2.5 (Cauchy-Schwartz Eşitsizliği): X bir iç çarpım uzayı olmak üzere,

$\forall x, y \in X$ için

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

eşitsizliği geçerlidir.

Tanım 2.6: f, R^n de konveks bir fonksiyon olmak üzere

$$\emptyset \neq \text{dom}f = \{x \in R^n : f(x) < \infty\}$$

kümesine f fonksiyonunun tanım kümesi denir.

Tanım 2.7: $x = (x_1, x_2, \dots, x_n)$ olmak üzere 1 norm $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$ yani;

$$\|x\|_1 = \sum_{k=1}^n |x_k|$$

şeklindedir.

Tanım 2.8: $x = (x_1, x_2, \dots, x_n)$ olmak üzere 2 norm $\|x\|_2 = \left[x_1^2 + x_2^2 + \dots + x_n^2 \right]^{1/2}$

yani;

$$\|x\|_2 = \left[\sum_{k=1}^n x_k^2 \right]^{\frac{1}{2}}$$

şeklindedir.

Tanım 2.9: $x = (x_1, x_2, \dots, x_n)$ olmak üzere $\|x\|_\infty = \max |x_i| : i = 1, 2, \dots, n$ dir.

Tanım 2.10: $\Omega \in R^n$ kapalı konveks bir küme ve x_0 , Ω 'nın sınırında bir nokta olsun.

$y \neq 0$, R^n uzayında bir vektör olmak üzere her $x \in \Omega$ için

$$\langle y, x \rangle \leq \langle y, x_0 \rangle$$

ise x_0 noktası için destek hiperdüzlem kümesi

$$H = \{ x \in R^n : \langle y, x \rangle = \langle y, x_0 \rangle \}$$

şeklinde olur.

3. MATERYAL ve YÖNTEM

3.1. Subgradyent Kavramı

$f : R^n \rightarrow R$ konveks fonksiyonu verilsin. $x \in \text{dom}f$ noktasında, her $y \in \text{dom}f$ için

$$f(y) \geq f(x) + s^T (y - x)$$

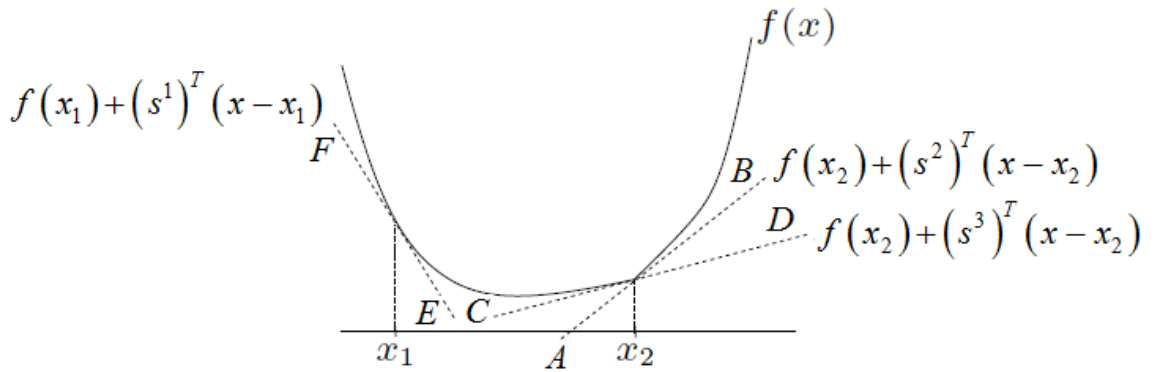
oluyorsa $s \in R^n$ elemanına f fonksiyonunun x noktasında bir subgradyenti denir. Konkav fonksiyon için subgradyent, eşitsizliğin yönü değiştirilerek tanımlanabilir.

$f : R^n \rightarrow R$ konkav bir fonksiyon $x \in \text{dom}f$ noktasında, her $y \in \text{dom}f$ için

$$f(y) \leq f(x) + s^T (y - x)$$

oluyorsa $s \in R^n$ ye f de x in bir subgradyenti denir.

$n = 1$ için aşağıdaki şekli inceleyelim.



Şekil 3.1. f fonksiyonunun bazı subgradyentleri

f fonksiyonunun x_1 noktasındaki subgradyenti EF doğrusunun eğimi olan s^1 sayısı olup tektir. Çünkü bu noktada f fonksiyonu türevlenebilirdir. Bunun ispatı Teorem 3.1 de yapılmıştır. f fonksiyonun x_2 noktasındaki subgradyentlerinin iki tanesi sırasıyla AB ve CD doğrularının eğimi olan s^2 ve s^3 sayılarıdır.

Bundan sonraki işlemler konveks fonksiyonlar için yapılacaktır ama aynı işlemler benzer şekilde konkav fonksiyonlar için de yapılabilir.

Eğer f konveks ve türevlenebilir bir fonksiyon ise x noktasında gradyent aynı zamanda subgradyenttir.

Teorem 3.1: $f : R^n \rightarrow R$ konveks ve türevlenebilen bir fonksiyon ve x noktasında f fonksiyonunun bir gradyenti $\nabla f(x)$ vektörü olsun. O halde $\forall x \in R^n$ için $\nabla f(x)$ bir subgradyenttir.

İspat: $\forall x \in R^n$ noktasında $z = x$ için

$$\nabla f(x) \cdot (z - x) \leq f(z) - f(x)$$

eşitsizliğinin sağlandığı açıktır. Sadece $z \neq x$ durumunun gösterilmesi yeterlidir. f fonksiyonu türevlenebildiği için x noktasında $z - x$ yönünde f 'nin yönlü türevi

$$\lim_{t \rightarrow 0^+} \frac{f(x + t(z - x)) - f(x)}{t}$$

ifadesi tarafından verilir ve bu ifade $\nabla f(x) \cdot (z - x)$ e eşittir. f konveks olduğundan $t \in (0,1)$ için

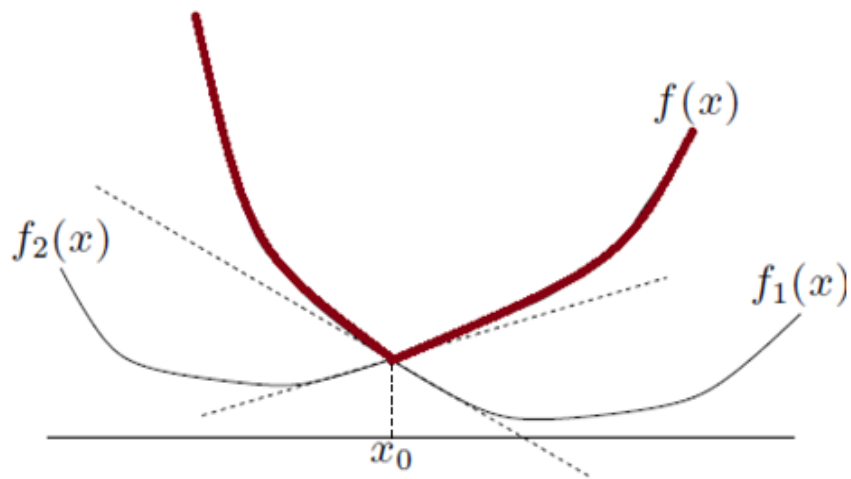
$$\begin{aligned}
f(z) - f(x) &= \frac{t f(z) + (1-t) f(x) - f(x)}{t} \\
&\geq \frac{f(tx + (1-t)x) - f(x)}{t} \\
&= \frac{f(x + t(z-x)) - f(x)}{t} \\
f(z) - f(x) &\geq \lim_{t \rightarrow 0^+} \frac{f(x + t(z-x)) - f(x)}{t} = \nabla f(x) \cdot (z-x)
\end{aligned}$$

olup istenilen

$$\nabla f(x) \cdot (z-x) \leq f(z) - f(x)$$

sonucu elde edilmiştir. Böylece ispat tamamlanmıştır.

Örnek 3.1: f_1, f_2 konveks ve türevlenebilir fonksiyonlar olmak üzere $f = \max\{f_1, f_2\}$ fonksiyonu aşağıdaki şekilde gibidir.



Şekil 3.2. $f = \max\{f_1, f_2\}$ fonksiyonunun grafiği

$f_1(x) > f_2(x)$ şartını sağlayan yerlerde f fonksiyonu türevlenebilirdir ve subgradyent $s = \nabla f_1(x)$ olup tektir. Eğer $f_1(x) < f_2(x)$ şartı sağlanıyorsa subgradyent $s = \nabla f_2(x)$ olup tektir. Eğer $f_1(x_0) = f_2(x_0)$ olursa bu noktada f fonksiyonu türevlenebilir değildir ve subgradyent $s \in [\nabla f_1(x_0), \nabla f_2(x_0)]$ olup bu aralıktaki her reel sayı bir subgradyenttir.

3.2. Subdiferansiyel Kavramı ve Bazı Fonksiyonların Subdiferansiyeli

Belirli bir x noktasında diferansiyellenemeyen konveks (konkav) bir f fonksiyonunun bütün subgradyentlerinin kümesine, f fonksiyonunun x noktasındaki subdiferansiyeli denir ve $\partial f(x)$ şeklinde gösterilir.

Yani;

$$\partial f(x) = \{s : f(z) \geq f(x) + s(z-x), x \in \text{dom}f \text{ ve } z \in \text{dom}f\}$$

şeklindedir.

Şimdi türevlenebilir fonksiyonların maksimumlarından oluşan bir fonksiyonun subdiferansiyel kümesinin nasıl bulunacağını görelim. Fonksiyonumuz

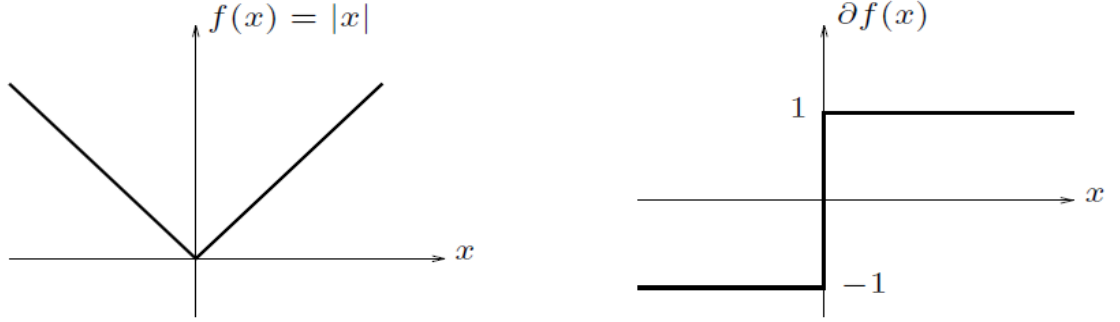
$$f = \max_{i=1, \dots, n} f_i$$

şeklinde ise bu fonksiyonun subdiferansiyel kümesi

$$\partial f(x) = \text{co} \{ \nabla f_i(x) : f_i(x) = f(x) \}$$

şeklinde bulunur.

Örnek 3.2: $f(x) = |x|$ fonksiyonunu ele alalım;



Şekil 3.3. $f(x) = |x|$ fonksiyonunun ve subdiferansiyelinin grafiği

İlk önce bu fonksiyonu maksimum fonksiyon cinsinden yazmalıyız. O halde

$$f(x) = \max\{-x, x\}$$

şeklinde yazabiliriz. Bu şekilde yazılınca $f_i(x)$ fonksiyonları türevlenebilir fonksiyonlara dönüşür. O halde

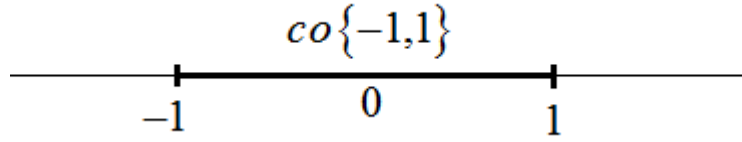
$$\nabla f_i(x) : f_i(x) = f(x)$$

kümesini teşkil edebiliriz. $f(x) = \max\{-x, x\}$, $f_1(x) = -x$ ve $f_2(x) = x$ olur.

$f_1(x) = f(x)$ ise $f(x) = -x$ olup $\nabla f_1(x) = -1$ olur. $f_2(x) = f(x)$ ise $f(x) = x$ olup

$\nabla f_2(x) = 1$ olur. Böylece bu küme

şeklinde oluşur. $-1,1$ kümesinin konveks zarfı da $co\{-1,1\} = [-1,1]$ şeklindedir.



Şekil 3.4. $f(x) = |x|$ fonksiyonunun subdiferansiyel kümesi

Örnek 3.3. $f(x) = \|x\|_1$ ve $x = (x_1, x_2)$ olsun. f fonksiyonunun subdiferansiyel kümesini bulalım.

$f(x) = |x_1| + |x_2|$ şeklindedir. O halde bu fonksiyonu maksimum şeklinde yazarsak

$$f(x) = \max \{ (x_1 + x_2), (-x_1 + x_2), (x_1 - x_2), (-x_1 - x_2) \}$$

olur. $f_1(x) = x_1 + x_2$, $f_2(x) = -x_1 + x_2$, $f_3(x) = x_1 - x_2$, $f_4(x) = -x_1 - x_2$ denirse

$$\nabla f_i(x) : f_i(x) = f(x)$$

kümesi oluşturulabilir.

$$f_1(x) = f(x) \text{ ise } f(x) = x_1 + x_2 \text{ olup } \nabla f_1(x) = (1, 1)$$

$$f_2(x) = f(x) \text{ ise } f(x) = -x_1 + x_2 \text{ olup } \nabla f_2(x) = (-1, 1)$$

$$f_3(x) = f(x) \text{ ise } f(x) = x_1 - x_2 \text{ olup } \nabla f_3(x) = (1, -1)$$

$$f_4(x) = f(x) \text{ ise } f(x) = -x_1 - x_2 \text{ olup } \nabla f_4(x) = (-1, -1)$$

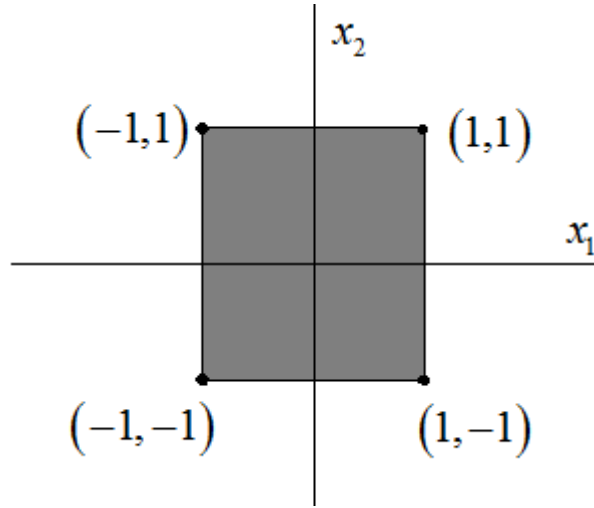
olacaktır. Böylece bu küme

$$1,1, -1,1, 1,-1, -1,-1$$

şeklinde olacaktır. Bu kümenin konveks zarfı alınırsa

$$\partial f \ x = \text{co} \ 1,1, -1,1, 1,-1, -1,-1$$

olur. Bunu şekil üzerinde gösterip $\partial f \ x$ kümesini bulalım.



Şekil 3.5. $x = x_1, x_2$ olmak üzere $f \ x = \|x\|_1$ fonksiyonunun subdiferansiyel kümesi

Yani $\partial f \ x = x_1, x_2 : \|x_1, x_2\|_\infty \leq 1$ dir. Şimdi de özel bir nokta için subdiferansiyel kümesini bulalım. Örneğin $(1,0)$ noktasını alalım. Bu nokta için yukarıdaki dört alt fonksiyondan en yüksek değeri alanlar seçilir. O halde bu fonksiyon

$$f \ x = \max \ x_1 + x_2, x_1 - x_2$$

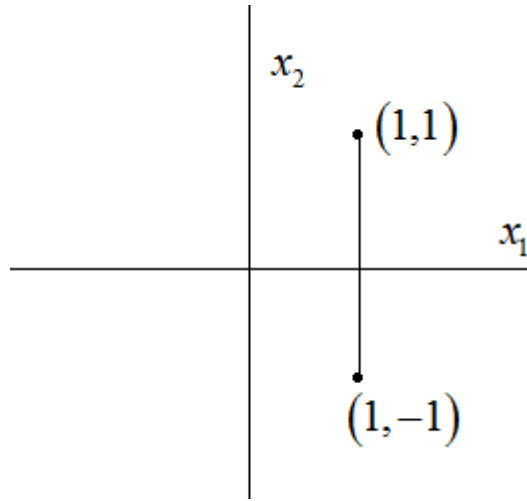
şeklindedir $f_1 \ x = x_1 + x_2, f_2 \ x = x_1 - x_2$ olup konveks zarfı alınacak küme

$$1,1, 1,-1$$

kümesidir. Yani

$$\partial f_{1,0} = \text{co} \{1,1, 1,-1\}$$

elde edilir. Bu kümeyi aşağıdaki gibi şekil üzerinde gösterebiliriz.



Şekil 3.6. $1,0$ noktası için $f(x) = \|x\|_1$ fonksiyonunun subdiferansiyel kümesi

Yani $\partial f_{1,0} = \{x_1, x_2 : x_1 = 1, -1 \leq x_2 \leq 1\}$ olur. Şimdi de $1,1$ noktası için subdiferansiyel kümesini bulalım. $1,1$ noktası için en yüksek değer $f(x) = x_1 + x_2$ için elde edilir. O halde $\nabla f(x) = 1,1$ olur. Konveks zarfı alınacak küme de $1,1$ kümesi olur. O halde

$$\partial f_{1,1} = \text{co} \{1,1\} = 1,1$$

olur.

Örnek 3.4. $f(x) = \|x\|_1$ ve $x = (x_1, x_2, \dots, x_n)$ olsun. f fonksiyonunun subdiferansiyelini bulalım.

$f(x) = \|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$ olup bu fonksiyonu maksimum şeklinde yazarsak

$$f(x) = \max_{s^T x_i : s_i \in \{-1, 1\}}$$

olur. O halde

$$\nabla f_i(x) : f_i(x) = f(x)$$

kümesi oluşturulmalıdır. $s^T x_i$ türevi olan bir fonksiyon olup

$$\nabla f_i(x) = \begin{cases} 1 & x_i > 0 \\ -1 & x_i < 0 \\ -1 \text{ yada } 1 & x_i = 0 \end{cases}$$

olur. Yani x_1, x_2, \dots, x_n sayılarından hangileri sıfırdan büyük ise 1 değerini alır. Hangileri sıfırdan küçükse -1 değerini alır. x_i sayılarından herhangi biri sıfır olursa -1 yada 1 değerini alır. O halde

$$\partial f(x) = \text{co} \nabla f_i(x) : f_i(x) = f(x)$$

$$\partial f(x) = \{x : \|x\|_\infty \leq 1\}$$

olur.

Örnek 3.5. $f(x, y) = \max\{-x, x+y, x-2y\}$ fonksiyonu için subdiferansiyel kümesini bulalım.

$f_1(x, y) = -x$, $f_2(x, y) = x+y$ ve $f_3(x, y) = x-2y$ olup hepsi türevlenebilirdir. O halde $\nabla f_i(x, y) : f_i(x, y) = f(x, y)$ kümesini oluşturalım.

$$f_1(x, y) = -x \text{ olup } \nabla f_1(x, y) = (-1, 0)$$

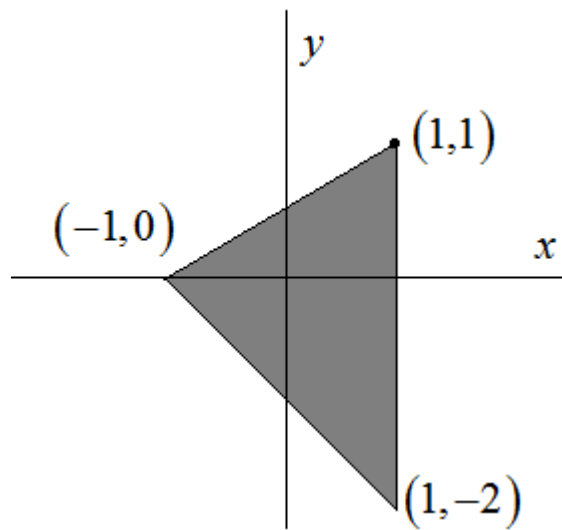
$$f_2(x, y) = x+y \text{ olup } \nabla f_2(x, y) = (1, 1)$$

$$f_3(x, y) = x-2y \text{ olup } \nabla f_3(x, y) = (1, -2)$$

olur. O halde bu fonksiyonun subdiferansiyel kümesi

$$\partial f(x, y) = \text{co}\{-1, 0, 1, 1, 1, -2\}$$

olur. Bunu şekil üzerinde gösterelim.



Şekil 3.7. $f(x, y) = \max\{-x, x+y, x-2y\}$ fonksiyonunun subdiferansiyel kümesi

Bu fonksiyonla alakalı olarak özel bir nokta için subdiferansiyel kümesini bulalım. Bu noktamız $t > 0$ olmak şartıyla $-t, -t$ noktası olsun. Bu noktayı $f_1(x, y)$, $f_2(x, y)$ ve $f_3(x, y)$ fonksiyonlarında yerine yazılıp maksimum değeri veren fonksiyon yada fonksiyonlar alınacaktır.

$$f_1(x, y) = -x \text{ olup } f_1(-t, -t) = t$$

$$f_2(x, y) = x + y \text{ olup } f_2(x, y) = -2t$$

$$f_3(x, y) = x - 2y \text{ olup } f_3(x, y) = t$$

olur. Bu taktirde $\nabla f_i(x) : f_i(x) = f(x)$ kümesi için

$$f_1(x, y) = -x \text{ olup } \nabla f_1(x, y) = -1, 0$$

$$f_3(x, y) = x - 2y \text{ olup } \nabla f_3(x, y) = 1, -2$$

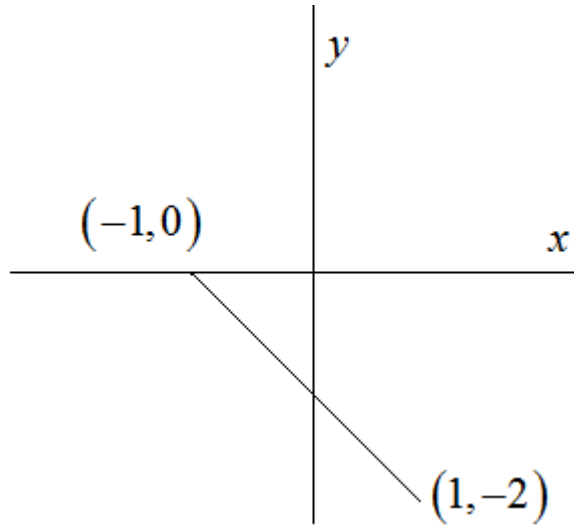
bulunur. O halde $-t, -t$ noktası için subdiferansiyel kümesi

$$\partial f(-t, -t) = \text{co} \{-1, 0, 1, -2\}$$

olur. Bu ise iki noktayı birleştiren doğru parçasıdır. Yani

$$\partial f(-t, -t) = \{x, y : y = -x - 1, -1 < x < 1, -2 < y < 0\}$$

kümesidir. Şekil üzerinde şöyle gösterilir,



Şekil 3.8. $-t, -t$ noktası için f fonksiyonunun subdiferansiyel kümesi

Teorem 3.2. $x \in R^n$ noktasında f fonksiyonunun $\partial f(x)$ subdiferansiyel kümesi konveks bir kümedir.

İspat: $s^1, s^2 \in \partial f(x)$ olduğunu varsayalım. O zaman $\forall z \in R^n$ için

$$\begin{aligned} s^1 \cdot (z-x) &\leq f(z) - f(x) \\ s^2 \cdot (z-x) &\leq f(z) - f(x) \end{aligned}$$

olur. Herhangi bir $\alpha \in [0, 1]$ için ve $z \in R^n$ için

$$\begin{aligned} [\alpha s^1 + (1-\alpha) s^2] \cdot (z-x) &= \alpha s^1 \cdot (z-x) + (1-\alpha) s^2 \cdot (z-x) \\ &\leq \alpha [f(z) - f(x)] + (1-\alpha) [f(z) - f(x)] \\ &= f(z) - f(x) \end{aligned}$$

olup subgradyentin tanımından

$$\alpha s^1 + (1-\alpha) s^2 \in \partial f(x)$$

elde edilir ve ispat tamamlanır.

3.3. Subgradyent Metotlar

Subgradyent metotlar türevlenemeyen bir $f(x)$ konveks (konkav) fonksiyonunun minimum (maksimum) yapma problemini çözmek için kullanılabilen iteratif bir kuraldır. Yani

$$\min f(x) \tag{3.3.1}$$

ile aşağıdaki üretken prosedür kullanılır;

Bir x^0 başlangıç noktası seçilir.

$x^{n+1} = x^n - \lambda_n v^n$ kuralı kullanılarak optimal bir çözüme yakınsayan x^n noktalarının dizisi oluşturulur. Burada $\lambda_n > 0$ adım uzunluğu denilen pozitif bir skalerdir ve v^n vektörü ise her bir iterasyon noktasında belirlenen adım yönüdür.

Eğer istenilen optimal çözüm bir Ω kapalı konveks kümesinde aranıyorsa kural biraz değişerek

$$\min f(x) : x \in \Omega \tag{3.3.2}$$

olup üretken prosedür ise şu şekilde olur;

Bir $x^0 \in \Omega$ başlangıç noktası seçilir.

$x^{n+1} = P_{\Omega} (x^n - \lambda_n v^n)$ kuralı kullanılarak optimal bir çözüme yakınsayan x^n noktalarının dizisi oluşturulur. Burada P_{Ω} ifadesi, Ω kümesi üzerine bir izdüşüm fonksiyonudur.

Kuralda her iterasyonda belirlenen adım yönü istenilen sonucu elde etmek için çok önemli bir rol oynar. Hareketin yönünü belirlemedeki stratejilere göre subgradyent optimizasyon metotlarını; pür subgradyent, deflected subgradyent ve koşullu subgradyent metotlar şeklinde ana kategorilere ayırabiliriz.

Pür subgradyent metot, iterasyonların bir dizisini oluştururken adım yönü olarak her bir iterasyon noktasında amaç fonksiyonunun bir subgradyentini kullanır.

Verilen noktadaki subgradyent vektörü hareketin önceki yönü ile geniş açılı ise o zaman zikzaklı bir yol çizmiş olur. Bu duruma birinci tür zikzak durumu denir. Subgradyent algoritmasının herhangi bir aşamasında ortaya çıkabilen böyle bir zikzak durumu kuralın yavaş yakınsamasına sebep olabilir. Bu zorluğun üstesinden gelmek için önceki adım yönü ile güncel adım yönünü birleştirerek daha hızlı yakınsamanın sağlandığı bu metoda deflected subgradyent metot denilir (Guta 2003).

Pür yada deflected subgradyent kurallarıyla iterasyon noktalarını oluştururken iterasyon noktalarının uygun küme üzerine düşmesi için izdüşüm operatörüne ihtiyaç duyulabilir. Uygun kümenin dışına düşecek olan iterasyon noktalarının izdüşüm operatörü yardımıyla uygun kümeye izdüşümü alınır. Bu iz düşüm önceki noktaya çok yakın bir noktaya düşerse yani uygun bölgenin bir yüzünün normal vektörüne yaklaşık olarak paralel ise bir diğer zikzak türü oluşur. Bu zikzak türüne ikinci tür zikzak denir. İkinci tür zikzak durumunu ortadan kaldırmak için verilen noktanın normal konisinden bir vektör ile yine aynı noktanın bir subgradyentini birleştiren metoda koşullu subgradyent metot diyeceğiz. Ayrıca bu ikinci tür zikzak durumunun sadece iterasyon noktalarının uygun kümenin göreceli sınırı boyunca taşınması durumunda elde edileceği gösterilecektir.

3.3.1. Pür Subgradyent metot

Bu bölümde (Guta 2003), (Boyd and Vandenberg 2004) çalışmasındaki sonuçlar incelenmiştir.

Türevlenebilen konveks fonksiyonların minimum değeri genellikle gradyent metotlar tarafından belirlenir. Bir gradyent metot olan steepest descent metoduyla (3.3.1) probleminin optimal çözümünü bulmak için x^0 başlangıç noktasını kullanarak ve

$$x^{n+1} = x^n - \lambda_n \nabla f(x^n)$$

iterasyon ilişkisi ile optimal çözüme yakınsayan bir x^n dizisi oluşturulur. Burada $\lambda_n \geq 0$ uygun adım uzunluğudur ve $\nabla f(x^n)$, x^n noktasında f fonksiyonunun gradyent vektörüdür.

Ancak ele alınan problemlerde f fonksiyonu türevlenemez olacaktır. Böylece f fonksiyonunun tanım kümesindeki bazı noktalarda ∇f gradyent vektörü mevcut olmayacağından gradyent metot kullanılamaz. Bu durumda gradyentler yerine subgradyentlerin konulduğu gradyent metotla uyumlu subgradyent metot kullanılır.

Teorem 3.3: R^n üzerinde tanımlı bir konveks fonksiyonun $x^* \in R^n$ gibi bir noktada minimumunun olması için gerekli ve yeter şart $0 \in \partial f(x^*)$ olmasıdır.

İspat: Subgradyentin tanımından $x^* \in R^n$ için $0 \in \partial f(x^*)$ ise $\forall x \in R^n$ için

$$f(x) - f(x^*) \geq 0, \quad x - x^*$$

olur. Buradan $f(x) \geq f(x^*)$ olup ilk kısım ispat edilmiş olur. Eğer $x^* \in \mathbb{R}^n$ fonksiyonun minimum noktası ise o zaman $\forall x \in \mathbb{R}^n$ için $f(x) \geq f(x^*)$ yazılabilir. Bu eşitsizliğin her iki tarafından $f(x) - f(x^*)$ ifadesi çıkarılırsa

$$f(x) - f(x^*) \geq 0$$

olur. Eşitsizliğin sağ tarafı sıfır olduğundan eşitsizlik

$$f(x) - f(x^*) \geq 0, \quad x - x^*$$

şeklinde yazılabilir. Bu da $0 \in \partial f(x^*)$ demektir. Böylece ispat tamamlanmıştır.

Şimdi subgradyent metodun bazı adım uzunlukları ile yakınsama sonuçlarını vereceğiz. x^* , f nin bir minimumu olsun ve aynı zamanda f nin subgradyentlerinin normu sınırlı olsun. Yani

$$\|s^k\|_2 \leq G$$

olsun.

Standart gradyent descent metot için yakınsamanın ispatının her bir adımda fonksiyonun değerinin azalmasına dayalı olduğunu hatırlayalım. Subgradyent metotda anahtar nitelik fonksiyonun değeri değildir. Çünkü her adımda azalmayabilir. Bunun yerine optimal kümeye Öklid uzaklığı kullanılır.

x^* in f fonksiyonunu minimum yapan keyfi bir optimal nokta olduğunu hatırlayalım.

$x^{k+1} = P_{\Omega} (x^k - \lambda_k s^k)$ eşitliği kullanılıp

$$\begin{aligned} \|x^{k+1} - x^*\|_2^2 &= \|P_{\Omega} (x^k - \lambda_k s^k) - x^*\|_2^2 \\ \|x^{k+1} - x^*\|_2^2 &\leq \|x^k - \lambda_k s^k - x^*\|_2^2 \\ \|x^{k+1} - x^*\|_2^2 &\leq \|x^k - x^*\|_2^2 - 2\lambda_k s^{k^T} (x^k - x^*) + \lambda_k^2 \|s^k\|_2^2 \end{aligned} \quad (3.3.1.1)$$

yazılabilir. Burada subgradyentin tanımı hatırlanacak olursa

$$f(x^*) \geq f(x^k) + s^{k^T} (x^* - x^k)$$

eşitsizliği vardır. Buradan da $s^{k^T} (x^k - x^*) \geq f(x^k) - f(x^*)$ olup bu eşitsizlik (3.3.1.1) ifadesine uygulanacak olursa

$$\|x^{k+1} - x^*\|_2^2 \leq \|x^k - x^*\|_2^2 - 2\lambda_k (f(x^k) - f^*) + \lambda_k^2 \|s^k\|_2^2$$

olur. Bu ifadede $f(x^*) = f^*$ dır.

Bu eşitsizlik toplam sembolü ile

$$\|x^{k+1} - x^*\|_2^2 \leq \|x^1 - x^*\|_2^2 - 2 \sum_{i=1}^k \lambda_i (f(x^i) - f^*) + \sum_{i=1}^k \lambda_i^2 \|s^i\|_2^2$$

şeklinde ifade edilir. Bu ifadede $\|x^{k+1} - x^*\|_2^2 \geq 0$ olduğu göz önüne alınsın. Böylece

$$2 \sum_{i=1}^k \lambda_i \left[f(x^i) - f^* \right] \leq \|x^1 - x^*\|_2^2 + \sum_{i=1}^k \lambda_i^2 \|s^i\|_2^2 \quad (3.3.1.2)$$

olur. Ayrıca

$$\sum_{i=1}^k \lambda_i \left[f(x^i) - f^* \right] \geq \left(\sum_{i=1}^k \lambda_i \right) \min_{i=1, \dots, k} \left[f(x^i) - f^* \right]$$

eşitsizliği yazılabilir olup (3.3.1.2) eşitsizliği ile birleştirirsek

$$f_{best}^k - f^* = \min_{i=1, \dots, k} \left[f(x^i) - f^* \right] \leq \frac{\|x^1 - x^*\|_2^2 + \sum_{i=1}^k \lambda_i^2 \|s^i\|_2^2}{2 \sum_{i=1}^k \lambda_i}$$

elde edilir. Burada f_{best}^k , iterasyon noktaları boyunca fonksiyonun almış olduğu en iyi minimum değeridir. Son olarak $\|s^k\|_2 \leq G$ varsayımını kullanarak

$$f_{best}^k - f^* = \min_{i=1, \dots, k} \left[f(x^i) - f^* \right] \leq \frac{\|x^1 - x^*\|_2^2 + G^2 \sum_{i=1}^k \lambda_i^2}{2 \sum_{i=1}^k \lambda_i}$$

temel eşitsizlik elde edilir.

Bu eşitsizlikten çeşitli yakınsama sonuçları çıkar.

x^* , f nin herhangi bir minimumu olduğu için

$$f_{best}^k - f^* \leq \frac{\text{dist } x^1, X^{*2} + G^2 \sum_{i=1}^k \lambda_i^2}{2 \sum_{i=1}^k \lambda_i} \quad (3.3.1.3)$$

yazılır. Burada X^* ifadesi optimal kümeyi göstermektedir ve $\text{dist } x^1, X^*$ ifadesi x^1 noktasının optimal kümeye olan uzaklığıdır.

Şimdi bazı adım boyları için bu sonucu inceleyelim (Boyd and Vandenberghe 2004).

$\lambda_k = h$ olsun. O zaman

$$f_{best}^k - f^* \leq \frac{\text{dist } x^1, X^{*2} + G^2 h^2 k}{2hk}$$

olur. Sağ taraf $k \rightarrow \infty$ için $\frac{G^2 h}{2}$ ye yakınsar. Böylece $\lambda_k = h$ adım boyu ile subgradyent

metot için $f_{best}^k - f^* \leq \frac{G^2 h}{2}$ olur. Şimdi de $\lambda_k = \frac{h}{\|s^k\|_2}$ adım boyu için bakalım.

$\lambda_k = \frac{h}{\|s^k\|_2}$ için

$$f_{best}^k - f^* \leq \frac{\text{dist } x^1, X^{*2} + h^2 k}{2hk/G}$$

olur. Sağ taraf $k \rightarrow \infty$ için $\frac{Gh}{2}$ ye yakınsar. Böylece $\lambda_k = \frac{h}{\|s^k\|_2}$ adım boyu ile

subgradyent metod için $f_{best}^k - f^* \leq \frac{Gh}{2}$ olur.

Bir başka adım boyu ise $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$, $\sum_{k=1}^{\infty} \lambda_k = \infty$ olması durumudur. O zaman

$$f_{best}^k - f^* \leq \frac{\text{dist } x^1, X^*{}^2 + G^2 \sum_{i=1}^k \lambda_i^2}{2 \sum_{i=1}^k \lambda_i}$$

ifadesi $k \rightarrow \infty$ için sıfıra yakınsar. Bir başka deyişle subgradyent metot yakınsar. Bu adım boyu için bir örnek verecek olursak $\lambda_k = \frac{1}{k}$ verilebilir.

Bir başka adım boyu azalan adım boyu da diyebileceğimiz bir adım boyudur. Eğer λ_k dizisi sıfıra yakınsar ve toplanamaz ise o zaman (3.3.1.3) eşitliğinin sağ tarafı sıfıra yakınsar yani subgradyent metot yakınsar. Bunu göstermek için bir $\varepsilon > 0$ sayısı alalım. O zaman $\lambda_i \leq \frac{\varepsilon}{G^2}$, bütün $i > N_1$ için olacak şekilde bir N_1 tamsayısı vardır. Ayrıca bütün $k > N_2$ için

$$\sum_{i=1}^k \lambda_i \geq \frac{1}{\varepsilon} \left(\|x^1 - x^*\|_2^2 + G^2 \sum_{i=1}^{N_1} \lambda_i^2 \right)$$

olacak şekilde bir N_2 tamsayısı vardır. Çünkü $\sum_{i=1}^{\infty} \lambda_i = \infty$ şeklindedir. Şimdi

$N = \max N_1, N_2$ olsun. O zaman bütün $k > N$ için

$$\min_{i=1, \dots, k} f(x^i) - f^* \leq \frac{\|x^1 - x^*\|_2^2 + G^2 \sum_{i=1}^{N_1} \lambda_i^2}{2 \sum_{i=1}^k \lambda_i} + \frac{G^2 \sum_{i=N_1+1}^k \lambda_i^2}{2 \sum_{i=1}^{N_1} \lambda_i + 2 \sum_{i=N_1+1}^k \lambda_i}$$

$$\begin{aligned}
&\leq \frac{\|x^1 - x^*\|_2^2 + G^2 \sum_{i=1}^{N_1} \lambda_i^2}{\varepsilon \left(\|x^1 - x^*\|_2^2 + G^2 \sum_{i=1}^{N_1} \lambda_i^2 \right)} + \frac{G^2 \sum_{i=N_1+1}^k \frac{\varepsilon}{G^2} \lambda_i}{2 \sum_{i=N_1+1}^k \lambda_i} \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

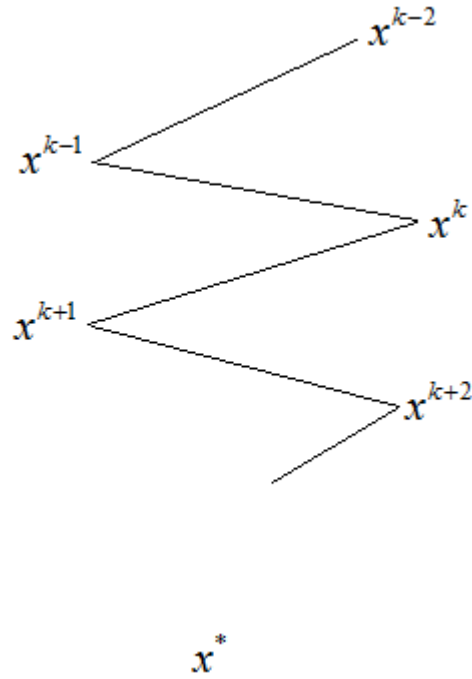
olur.

Bir başka adım boyu ise Polyak's adım boyudur. Eğer verilen fonksiyonun optimal çözümde aldığı değer biliniyorsa $\lambda_k = \frac{f(x^k) - f^*}{\|s^k\|^2}$ adım boyu kullanılarak optimal çözüme gidilebilir.

3.3.2. Deflected Subgradyent metot

Bu bölümde (Guta 2003) çalışmasındaki sonuçlar incelenmiştir.

Subgradyent metodun en önemli davranışlarından birisi her bir x^k iterasyonunda oluşan s^k subgradyentlerinin, x^k noktasından x^* optimal çözümüne gidilirken aralarında açı oluşturmalarıdır. Bu şekilde iterasyon ilerlerken s^k subgradyent yönü ile bir önceki s^{k-1} hareket yönü arasındaki açı geniş açı olabilir. Bu durum gelecek noktanın bir önceki noktaya yakın bir yere düşmesine sebep olabilir. Bu ise yakınsamayı yavaşlatır.



Şekil 3.9. İterasyon noktalarının oluşturduğu zikzaklar

λ_n pozitif bir skaler ve $s^n \in R^m$ olsun. Bir iterasyon kuralı olarak $n = 0, 1, 2, \dots$ için

$$x^{n+1} = P_{\Omega}(x^n - \lambda_n s^n)$$

alalım. Eğer herhangi iki (yada daha fazla) ardışık $x^k, x^{k+1} \in \Omega$ noktalarına uyumlu s^k ve s^{k+1} arasındaki açı geniş açı ise yani;

$$s^k \cdot s^{k+1} < 0$$

ise kural birinci tür zikzak durumundadır.

Böyle bir zikzak oluşumu subgradyent metodu her bir aşamada yavaşlatabilir. Böyle bir durumun olmasını yani güncel adım yönü ile önceki adım yönü arasındaki açının geniş açılı olması durumunu ortadan kaldırmak için subgradyent yönünü saptırmaya ihtiyaç

duyulur. Bu son dediğimizi gerçekleştirmek ve yakınsamanın hızını artırmak için x^k iterasyonunda s^k subgradyent yönü bir deflected (saptırılmış) subgradyent yön olan d^k ile yer değiştirilip pür subgradyent metot modifiye edilir. d^k aşağıdaki prosedürle bulunur:

$$d^k = s^k + \delta_k d^{k-1}$$

burada $s^k \in \partial f(x^k)$ dır ve $\delta_k \geq 0$ saptırma parametresi diye adlandırılan uygun bir skalerdir. $k = 0$ için $d^{k-1} = 0$ dır. Yani deflected subgradyent metot, önceki adımda kullanılan yön ve güncel subgradyent yönün bir lineer kombinasyonu olan d^k araştırma yönünde taşınır.

Bu bölümde deflected subgradyent metodunu göz önüne alacağız. Bu metot, pür subgradyent metoda kıyasla optimal çözüme götüren hareket yönleri arasındaki açılar dar açılı olacaktır. Ayrıca bu bölümde subgradyent vektörün her bir olumlu özelliğinin deflected subgradyent vektöre genişletilebileceği de gösterilecektir.

Deflected subgradyent algoritma

Adım 0: (başlangıç durum)

$x^0 \in \Omega$ bir başlangıç noktasını seç ve $k = 0$ için $d^{k-1} = 0$ olarak al.

Adım 1: bir $s^k \in \partial f(x^k)$ subgradyenti al

$$d^k = s^k + \delta_k d^{k-1} \tag{3.3.2.1}$$

$$x^{k+1} = P_{\Omega}(x^k - \lambda_k d^k) \tag{3.3.2.2}$$

(δ_k ve λ_k karar vermek için kurallar verilecek)

$$k = k + 1$$

olarak al.

Adım 2: İstenilen optimallik elde edilinceye kadar Adım 1 e dön.

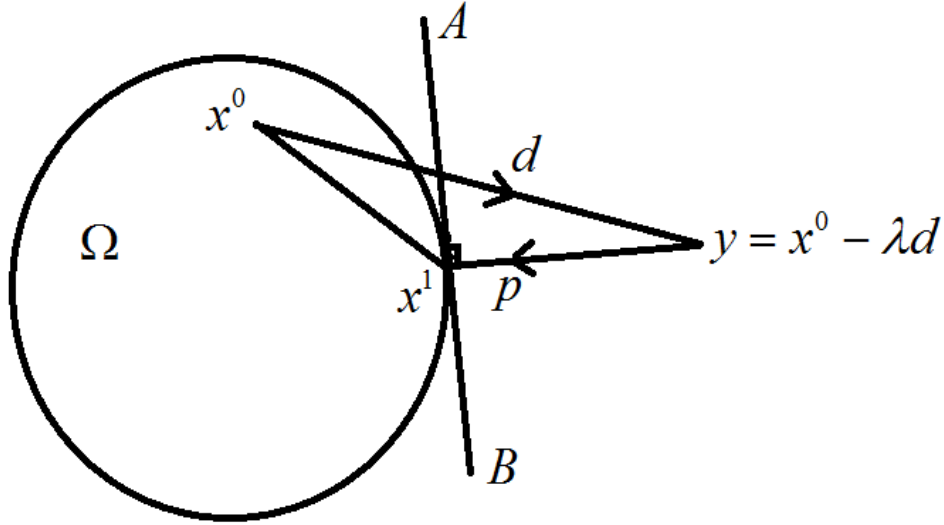
Deflected subgradyent yönlerin bazı özellikleri göz önüne alınacak ve sırasıyla (3.2.2.1) ve (3.2.2.2) eşitliklerinde verilen δ_k saptırma parametresi ve λ_k adım uzunluğuna karar vermek için bir kural verilecektir. Bunun için sonucu gelecek teoremde kullanılacak aşağıdaki lemma göz önüne alınacaktır.

Lemma 3.1: Ω , R^m nin kapalı konveks bir alt kümesi, $x^0 \in \Omega$ ve $y = x^0 - \lambda d$ olsun. Burada d , R^m de bir vektör ve λ bir pozitif skalerdir. Eğer $x^1 = P_{\Omega}(y)$ ve $p = y - x^1$ ise o zaman

$$p \cdot d \leq 0 \text{ ve } \|x^1 - x^0\| \leq \|y - x^0\|$$

şeklindedir.

İspat: Bu lemmanın sonuçları doğrudan konveks kümenin özelliklerinden görülür. Özellikle p vektörü x^1 de Ω nin destek hiperdüzlemine diktir ve böylece $\Delta(x^0, x^1, y)$ üçgen sonucu olarak x^1 de açı geniştir.



Şekil 3.10. Lemmanın ispatının şekil üzerinde gösterimi

Teorem 3.4: $s^k \in \partial f(x^k)$ olsun ve d^k , (3.3.2.1) tarafından verilsin. Eğer $\forall k = 0, 1, 2, \dots$ için

$$0 < \lambda_k < \frac{f(x^k) - f^*}{\|d^k\|^2} \quad (3.3.2.3)$$

ise o zaman

$$d^k(x^k - x^*) \geq s^k(x^k - x^*) \quad (3.3.2.4)$$

olur.

İspat: Tümevarım uygulayarak ispat edelim. (3.3.2.4) $k = 0$ için geçerlidir. Teoremin iddiasının $m = k$ için doğru olduğunu kabul edip $m = k + 1$ için doğru olduğunu göstereyim. (3.3.2.1) denklemini kullanılarak

$$d^{k+1}(x^{k+1} - x^*) = s^{k+1}(x^{k+1} - x^*) + \delta_{k+1} d^k(x^{k+1} - x^*) \quad (3.3.2.5)$$

yazılabilir. Böylece sadece $\delta_{k+1}d^k(x^{k+1} - x^*) \geq 0$ olduğu gösterilmelidir. Şimdi

$$\begin{aligned}
p^k &= -P_\Omega(x^k - \lambda_k d^k) + (x^k - \lambda_k d^k) \text{ olup} \\
d^k(x^{k+1} - x^*) &= d^k(P_\Omega(x^k - \lambda_k d^k) - x^*) \\
&= d^k(x^k - \lambda_k d^k - p^k - x^*) \\
&= d^k(x^k - x^*) - \lambda_k \|d^k\|^2 - d^k p^k \\
&\geq d^k(x^k - x^*) - \lambda_k \|d^k\|^2
\end{aligned} \tag{3.3.2.6}$$

son ifadedeki bu eşitsizlik Lemma 3.1 deki $-p^k d^k \geq 0$ olmasındandır. Diğer yandan teoremden verilen şart kullanılarak $s^k \in \partial f(x^k)$ için aşağıdaki eşitsizlik yazılabilir;

$$0 < \lambda_k \|d^k\|^2 \leq f(x^k) - f^* \leq s^k(x^k - x^*) \leq d^k(x^k - x^*)$$

son eşitsizlik tümevarım hipotezinden yazılmıştır. Böylece

$$d^k(x^k - x^*) - \lambda_k \|d^k\|^2 \geq 0$$

elde edilir.

Bu (3.3.2.6) ile birlikte düşünülürse

$$\delta_{k+1}d^k(x^{k+1} - x^*) \geq 0$$

olduğu görülür. Burada $\delta_{k+1} \geq 0$ dır. Bu son durumla (3.3.2.5) ilişkisi değerlendirilirse

$$d^{k+1}(x^{k+1} - x^*) \geq s^{k+1}(x^{k+1} - x^*)$$

olur ve ispat tamamlanır.

Teorem 3.5: x^k deflected subgradyent kuralı tarafından oluşturulan iterasyonların bir dizisi olsun. Önceki teoremin şartları altında bütün k lar için;

i- $d^k(x^k - x^*) > 0$

ii- $\|x^{k+1} - x^*\| < \|x^k - x^*\|$

şeklindedir. Burada x^k optimal olmayan nokta ve x^* bir optimal çözümdür.

İspat: i- $s^k(x^k - x^*) \geq f^k - f^* > 0$ olduğu için sonuç Teorem 3.4 dan direkt görülür.

ii-

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|P_{\Omega}(x^k - \lambda_k d^k) - x^*\|^2 \\ &\leq \|x^k - \lambda_k d^k - x^*\|^2 \\ &= \|x^k - x^*\|^2 + \lambda_k \left[\lambda_k \|d^k\|^2 - 2d^k(x^k - x^*) \right] \end{aligned} \quad (3.3.2.7)$$

ifadesinde Teorem 3.4 deki şart uygulanırsa

$$\lambda_k \|d^k\|^2 \leq f(x^k) - f^* < 2(f(x^k) - f^*)$$

yazılır. f fonksiyonunun konveksliği ve Teorem 3.4 kullanılırsa aşağıdaki ilişki elde edilir;

$$\lambda_k \|d^k\|^2 \leq 2[f(x^k) - f^*] \leq 2s^k \|x^k - x^*\| \leq 2d^k \|x^k - x^*\|$$

Buradan

$$\lambda_k \|d^k\|^2 - 2d^k \|x^k - x^*\| < 0$$

sonucu çıkarılır. Bu ifade (3.3.2.7) ile birlikte düşünülürse

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|$$

ifadesi elde edilir. Böylece ispat tamamlanmıştır.

Teorem 3.5'in i- şikkında görüldüğü gibi deflected subgradyent vektörün subgradyent vektöre benzediği görülür. Fakat deflected subgradyent vektör x^* optimal noktasına öncülük eden her bir noktadaki hareket yönü ile dar açıdır. Üstelik Teorem 3.5 in ikinci kısmı, $\|x^k - x^*\|$ dizisinin azalan olduğunu garantiler ve bunun bir sonucu olarak deflected subgradyent metot her bir iterasyon sonucunda optimal çözüme yaklaşmış olur. Aşağıdaki teorem δ_k saptırma parametresinin seçiminde yardımcı olacaktır.

Teorem 3.6: x^k deflected subgradyent prosedür tarafından oluşturulan bir dizi olsun.

Teorem 3.4 ün şartları altında eğer $s^k d^{k-1} < 0$ ise

$$\delta_k = -\tau_k \frac{s^k d^{k-1}}{\|d^{k-1}\|^2} \quad (3.3.2.8)$$

olur. Diğer durumlarda ise $\delta_k = 0$ olur. Burada $0 \leq \tau_k < 2$ şeklindedir. Böylece

$$\mathbf{i-} \quad \frac{d^k \cdot x^k - x^*}{\|d^k\|} \geq \frac{s^k \cdot x^k - x^*}{\|s^k\|} \quad (3.3.2.9)$$

ii- Eğer d^k ve s^k vektörleri $x^k - x^*$ vektörü ile sırasıyla ϕ_d^k ve ϕ_s^k açılarını oluşturuyorsa o zaman

$$0 \leq \phi_d^k \leq \phi_s^k \leq 90$$

şeklindedir.

İspat: i- Eğer $s^k d^{k-1} \geq 0$ ise o zaman $\delta_k = 0$ dır ve böylece $d^k = s^k$ olup (3.3.2.9) görülür. Şimdi de $s^k d^{k-1} < 0$ olması durumunu göz önüne alalım. Bu durumda teoremin kabulünden

$$\delta_k = -\tau_k \frac{s^k d^{k-1}}{\|d^{k-1}\|^2}$$

yazılır. O zaman

$$\begin{aligned} \|d^k\|^2 - \|s^k\|^2 &= \|s^k + \delta_k d^{k-1}\|^2 - \|s^k\|^2 \\ &= \delta_k \cdot 2s^k d^{k-1} + \delta_k \|d^{k-1}\|^2 \\ &= \delta_k \cdot 2s^k d^{k-1} - \tau_k s^k d^{k-1} \\ &= \delta_k (2 - \tau_k) s^k d^{k-1} \\ &< 0 \end{aligned}$$

olarak elde edilir. Bu son eşitsizlik τ_k için verilen şart ve $s^k d^{k-1} < 0$ olmasındandır.

Böylece

$$\|d^k\|^2 < \|s^k\|^2 \quad (3.3.2.10)$$

olur. (3.3.2.10) ifadesi, Teorem 3.4 ile beraber düşünülürse

$$\frac{d^k \cdot (x^k - x^*)}{\|d^k\|} \geq \frac{s^k \cdot (x^k - x^*)}{\|s^k\|}$$

olarak elde edilir.

ii- ϕ_d^k ve ϕ_s^k nin dar açılı olduğu açıktır.

$$\cos \phi_d^k = \frac{d^k \cdot (x^k - x^*)}{\|d^k\| \|x^k - x^*\|}, \quad \cos \phi_s^k = \frac{s^k \cdot (x^k - x^*)}{\|s^k\| \|x^k - x^*\|}$$

olup teoremin birinci kısmından

$$\cos \phi_d^k \geq \cos \phi_s^k$$

yazılabilir. Böylece açılar dar açılı olduğundan ve kosinüs fonksiyonu $[0, 90]$ aralığında monoton azalan olduğundan

$$\phi_d^k \leq \phi_s^k$$

olur. Subgradyent yönü s^k hareketin önceki yönü ile geniş açılı olduğu zaman Teorem 3.6'nın (3.3.2.8) kuralına göre seçilen bir saptırma parametresi iterasyon dizisinin birinci tür zikzak yapmasının önüne geçer.

Teorem 3.6'nin ikinci şikkından çıkaracağımız sonuçlardan biri de deflected subgradyent yönü olan d^k daima en az subgradyent yönü kadar iyidir. Çünkü deflected subgradyent adım yönü olan d^k pür subgradyent vektör yönüne göre optimal çözüme giden en iyi yol ile daha dar açılıdır. Böylece yakınsamanın hızı artar.

Teorem 3.7: d^k , k .iterasyonda deflected subgradyent yön olsun. Teorem 3.6'nin şartları altında eğer $\tau_k \geq 1$ ise o zaman

$$d^k d^{k-1} \geq 0$$

olur.

İspat: $s^k d^{k-1} \geq 0$ durumunda $d^k = s^k$ dir. Böylece iddia doğrulanır. Böylece biz $s^k d^{k-1} < 0$ durumu göz önüne alınacaktır. Bu durumda $\tau_k \geq 1$ olduğundan

$$\begin{aligned} d^k d^{k-1} &= s^k + \delta_k d^{k-1} d^{k-1} \\ &= s^k d^{k-1} - \tau_k s^k d^{k-1} \left(\delta_k = -\tau_k s^k d^{k-1} / \|d^{k-1}\|^2 \right) \\ &= 1 - \tau_k s^k d^{k-1} \\ &\geq 0 \end{aligned}$$

elde edilir. Böylece ispat tamamlanmıştır.

Teorem 3.7'nin önemi $1 < \tau_k < 2$ aralığından alınan τ_k ile seçilen saptırma parametresi önceki adım yönü ile güncel deflected subgradyent yönü arasındaki açının dar açı olmasını sağlamasıdır. Böylece bu metot pür subgradyent metotta oluşabilen birinci tür zikzak durumunu engellemiş olur.

$\tau_k = 1$ seçimi önceki yöne ortogonal bir yön seçimi için kullanılır. Bu yukarıdaki teoremin ispatında da açıkça görülmektedir. Ama τ_k için önerilen kullanım $\tau_k = 1,5$ kullanımındır. Deflected subgradyent algoritmanın performansı pratikte pür subgradyent algoritmanın performansından üstündür. Aslında literatürde saptırma parametresinin seçiminde çeşitlilik vardır. (3.3.2.8) ifadesindeki gibi verilen saptırma parametresi ile deflected subgradyent prosedür tarafından oluşturulan hareketin yönü, pür subgradyent prosedürün birinci tür zikzak yaptığı durumda onun subgradyent yönünü sadece döndürebilir. Ayrıcı saptırma, önceki yön ile güncel yön arasındaki açının geniş aç olması durumunda yapılır.

Bir başka saptırma parametresi seçimi, güncel subgradyent olan s^k ile d^{k-1} arasındaki açının tipini hesaba katmadan, önceki hareketin yönü ile güncel hareketin yönü arasındaki açığı ikiye bölen bir hareket yönü seçimi için her bir iterasyon noktasında saptırma yapmayı önermiştir. Bu yönü elde etmek için saptırma parametresi

$$\delta_k = \frac{\|s^k\|}{\|d^{k-1}\|}$$

eşitliği ile bulunur. Saptırma parametresinin bu seçimi ile $d^k = s^k + \delta_k d^{k-1}$ yönü s^k ile d^{k-1} arasındaki açığı ikiye böler ve bundan dolayı bu stratejiye ortalama yön stratejisi denir. Bundan başka subgradyent s^k ve hareketin önceki yönü d^{k-1} in konveks kombinasyonu bazı zamanlarda güncel adım yönü d^k yı belirlemek için kullanılır. Yani, $\alpha_k \in 0,1$ için

$$d^k = \alpha_k s^k + 1 - \alpha_k d^{k-1}$$

şeklindedir. Bu durumda eğer $s^k d^{k-1} \geq 0$ ise o zaman $d^k d^{k-1} \geq 0$ olur. (herhangi bir $\alpha_k > 0$ için) Ancak eğer $s^k d^{k-1} < 0$ ise o zaman birinci tür zikzağı ortadan kaldırmak için α_k nın kısıtlanmasına ihtiyaç vardır. Eğer $s^k d^{k-1} < 0$ ise α_k için

$$d^k d^{k-1} \geq 0 \Leftrightarrow 0 \leq \alpha_k \leq \frac{\|d^{k-1}\|}{\|d^{k-1}\|^2 - s^k d^{k-1}}$$

dir. Böylece eğer $s^k d^{k-1} < 0$ ise bu durumda α_k nın seçimi $0 < \alpha_k < \alpha_k^*$ dir. Burada

$$\alpha_k^* = \min \left\{ 1, \frac{\|d^{k-1}\|}{\|d^{k-1}\|^2 - s^k d^{k-1}} \right\}$$

olursa birinci tür zikzak ortadan kalkar.

3.3.3. Koşullu Subgradyent metot

Bu bölümde (Guta 2003) çalışmasındaki sonuçlar incelenmiştir.

Koşullu subgradyent metot, deflected subgradyent metot tarafından ortadan kaldırılamayan pür subgradyent metodun ikinci tür zikzak durumunu ortadan kaldırmada kullanılır. Deflected subgradyent metodun s^k subgradyent yönü ile hareketin önceki yönü arasındaki açının, geniş olması durumunda kullanıldığını hatırlayalım. Ancak bazı durumlarda s^k subgradyent yönü ve hareketin önceki yönü arasındaki açı, dar olmasına rağmen istenilen bölgeye $x^k - \lambda_k s^k$ nın projeksiyonu, x^k ya çok yakın bir şekilde bölgenin sınırına düşebilir. Böyle bir durumda iterasyon noktaları hemen hemen hiç değiştirilmemiş gibi olur ya da prosedür çok yavaş gelişip anlamlı bir değişiklik olmaz. Aşağıdaki problem pür (ve deflected) subgradyent metotlarının muhtemel yavaş yakınsamasını gösterir.

Örnek 3.6: $\min x_1 - 2x_2$

$$x_1, x_2 \in \Omega$$

Burada $\Omega = \{x_1, x_2 \in \mathbb{R}^2 : x_1 - x_2 \geq 0, 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$ şeklindedir.

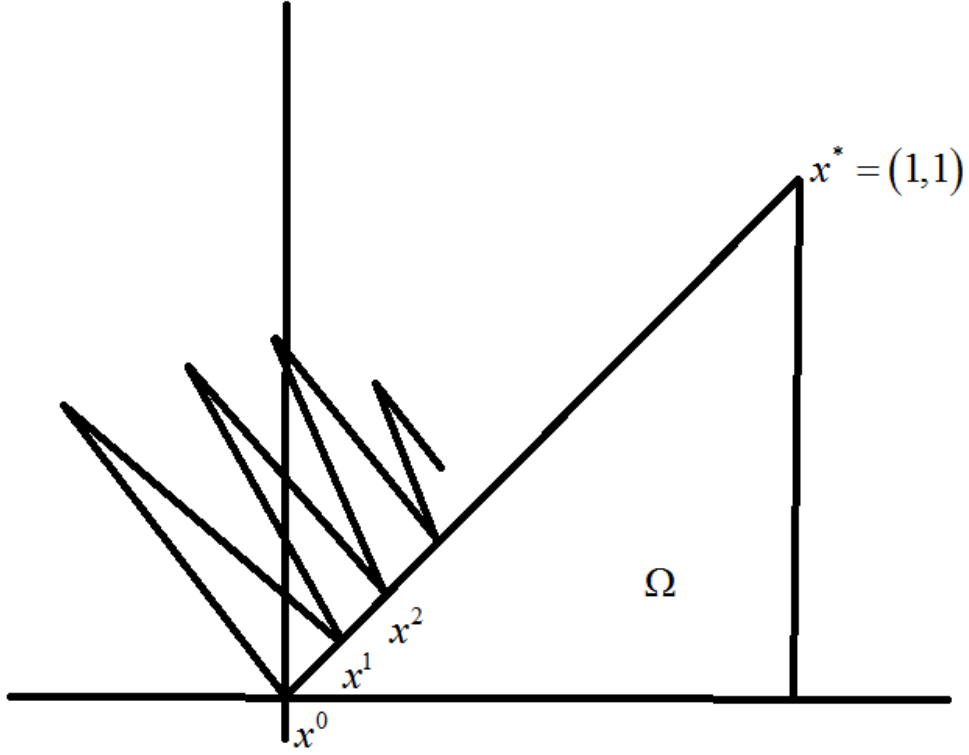
Optimal amaç değeri -1 ve optimal çözüm $x^* = (1, 1)$ dir. Herhangi bir $x^k \in \Omega$ iterasyon noktasında subgradyent $s^k = (1, -2)$ olup hareket yönü olan $-s^k = (-1, 2)$, x^* optimal çözüme götüren yön ile dar açıdır ve önceki adım yönü ile asla geniş açı olamaz. Yani adım yönü daima $(-1, 2)$ subgradyent vektörünün yönüdür. (3.3.2.8) saptırma kuralı ile veya ortalama yön stratejisi ile ya da konveks kombinasyon adım yönü ile deflected subgradyent metotlarının adım yönü daima pür subgradyent metot ile çakışır. Bundan dolayı başlangıç noktasını $(0, 0)$ seçip

$$\lambda_k = \frac{f(x^k) - f^*}{\|s^k\|^2} = \frac{1 + f(x^k)}{5}$$

adım kuralını kullanarak pür subgradyent metotla çözelim. O zaman iterasyonlar

$$x^k = P_{\Omega} \left(x^{k-1} - \frac{1 + f(x^{k-1})}{5} (1, -2) \right) = \left(1 - \frac{9^k}{10^k}, 1 - \frac{9^k}{10^k} \right)$$

şeklinde elde edilir. Amaç değeri $f(x^k) = -1 + \frac{9^k}{10^k}$ dir. Bu örnek için kuralın optimal noktaya yakınsaması çok yavaştır.



Şekil 3.11. Pür subgradyent metotla oluşan ikinci tür zikzak

Bu örnekte oluşan zikzak subgradyentler uygun kümenin $x_1, x_2 : x_1 = x_2$ yüzeyine hemen hemen dik olmasından dolayıdır. Subgradyent metodun böyle bir zikzak durumuna ikinci tür zikzak diyeceğiz. Bunun için önce aşağıdaki tanımları vermeliyiz.

Bazı $x \in \Omega$ noktalarında Ω nın bir normal konisi

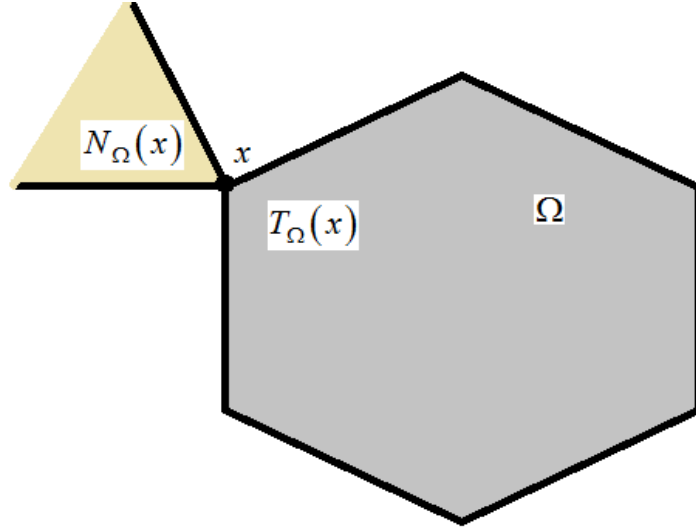
$$N_{\Omega} x = \{y \in R^m : y \cdot z - x \leq 0, \forall z \in \Omega\}$$

kümesidir.

Bazı $x \in \Omega$ de Ω kümesinin bir tanjant konisi

$$T_{\Omega} x = \{z \in R^m : z \cdot y \leq 0, \forall y \in N_{\Omega} x\}$$

kümesidir. $N_{\Omega}(x)$ ve $T_{\Omega}(x)$ her ikisi sıfırı içeren boştan farklı kapalı konveks alt kümelerdir.



Şekil 3.12. $x \in bd \Omega$ noktasının tanjant ve normal konisi

Eğer $x \in \text{int } \Omega$ ise burada $\text{int } \Omega$, Ω nın içini ifade eder. O zaman $N_{\Omega}(x) = 0$ ve $T_{\Omega}(x) = R^m$ olur. $N_{\Omega}(x)$ ve $T_{\Omega}(x)$ in elemanları x de sırasıyla normal vektörler ve tanjant vektörler diye adlandırılır. Herhangi bir $z \in \Omega$ için $y \in N_{\Omega}(x)$ olmak üzere $y \cdot (z - x) \leq 0$ olduğu için $z - x \in T_{\Omega}(x)$ olduğuna dikkat edelim yani $\Omega - x \subseteq T_{\Omega}(x)$ şeklindedir. Yani tanjant koninin tanımı aynı zamanda $\Omega - x$ tarafından oluşturulan koninin kapanışı olarak da ifade edilebilir.

$$T_{\Omega}(x) = \{ z \in R^m : z = \alpha(y - x), y \in \Omega, \alpha > 0 \}$$

Ya da buna eşit olarak

$$T_{\Omega}(x) = \{ z \in R^m : x + \lambda z \in \Omega, \lambda > 0 \}$$

yazılabilir. Eğer Ω bir polihedral küme ise o zaman

$$T_{\Omega} x = \{z \in R^m : x + \lambda z \in \Omega, \text{bazı } \lambda > 0 \text{ için}\}$$

olduğuna dikkat edelim.

Eğer $x \in bd \Omega$ de $d \notin T_{\Omega} x$ ise $d \in R^m$ ye x noktasında uygun olmayan yöndür denir. Burada $bd \Omega$, Ω nin sınırındaki noktaların kümesidir. Yani eğer

$$\text{bazı } v \in N_{\Omega} x \text{ için } dv > 0$$

ise d vektörü $x \in bd \Omega$ noktasında uygun olmayan bir yöndür. Aşağıda ikinci tür zikzağın formal bir tanımı verilecek.

λ_n pozitif bir skaler ve $d^n \in R^m$ olsun. Bir iterasyon prosedürü

$$x^{n+1} = P_{\Omega} (x^n + \lambda_n d^n)$$

olsun. Eğer herhangi iki (yada daha fazla) ardışık iterasyon noktası $x^k, x^{k+1} \in \Omega$ da

$$d^k v > 0 \text{ ve } d^{k+1} w > 0 \tag{3.3.3.1}$$

olacak şekilde $v \in N_{\Omega} x^k$ ve $w \in N_{\Omega} x^{k+1}$ vektörleri oluşuyorsa ikinci tür zikzak formu oluşur. Minimumu bulunacak fonksiyonlar için d^n , negatif subgradyenttir. Tanıma dikkat edilirse ikinci tür zikzak sadece iterasyon noktaları Ω nın sınırında olduğu zaman meydana çıkar. Yoksa (3.3.3.1) şartını sağlayan v ve w oluşmaz. Bu bölümde adım yönüne karar verirken ve koşullu subgradyent metot yakınsamasını

kurarken, uygun küme göz önüne alınarak pür subgradyent metod genellenecek. Koşullu subgradyent metot ikinci tür zikzak oluşumunu önlediği için pür subgradyentten daha iyi performansa sahiptir.

Bir problemi göz önüne alalım. $f : R^m \rightarrow R$ konveks bir fonksiyon olsun. Böylece f süreklidir ama her yerde türevlenebilir olması gerekmez. Ayrıca $\Omega \in R^m$ boştan farklı kapalı ve konveks küme olsun ve bazı $x^* \in \Omega$ için $f^* = f(x^*) > -\infty$ olduğunu varsayalım. Problem

$$f^* = \min_{x \in \Omega} f(x) \quad (3.3.3.2)$$

olarak göz önüne alınır. Bununla birlikte

$$\Omega^* = \{ \bar{x} \in \Omega : f(\bar{x}) = f^* \}$$

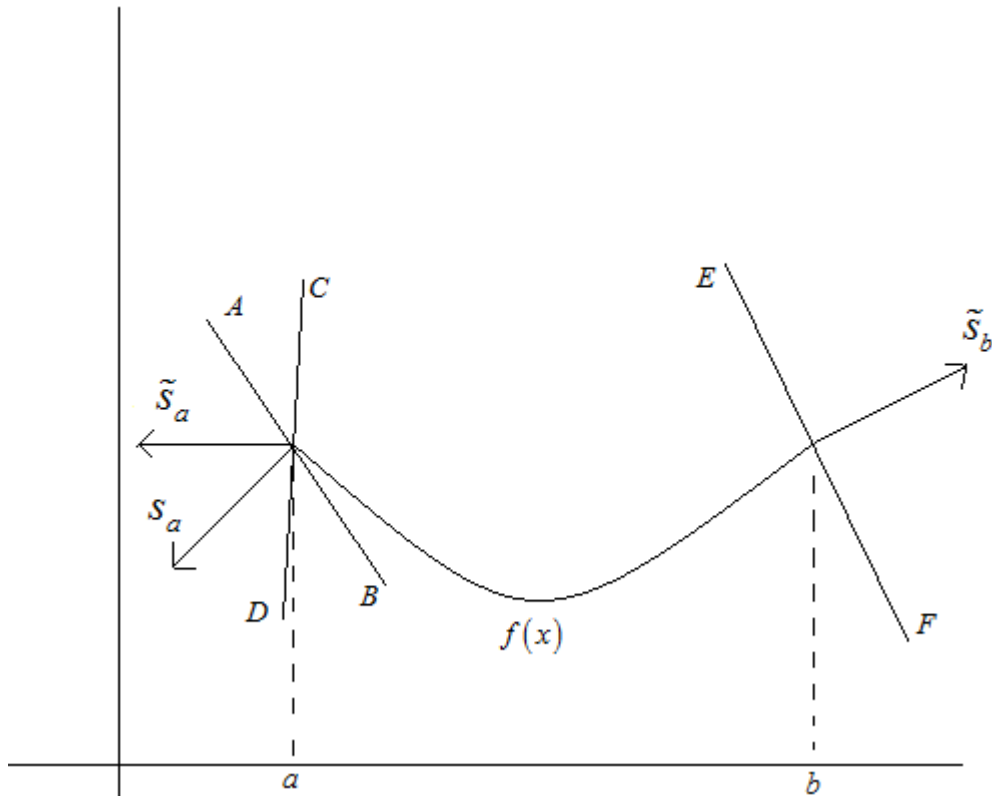
optimal çözümlerin konveks kümesi ve boştan farklıdır. (3.3.3.2) probleminin özellikleri aksi söylenmedikçe bütün tartışma boyunca geçerlidir. Subgradyent ve subdiferansiyelin tanımında Ω uygun bölgesinin bir etkisi yoktu. Şimdi Ω uygun bölgesini de hesaba katarak koşullu subgradyent ve koşullu subdiferansiyelin tanımlarını verelim.

Tanım 3.3: $f : R^m \rightarrow R$ konveks bir fonksiyon ve $\Omega \subseteq R^m$ olsun. $x \in \Omega$ de f nin koşullu subdiferansiyeli

$$\partial^\Omega f(x) = \{ \tilde{s} \in R^m : f(x) + \tilde{s} \cdot (z - x) \leq f(z), \forall z \in \Omega \}$$

kümesidir. $\tilde{s} \in \partial^\Omega f(x)$ elemanı x de f nin koşullu subgradyenti diye adlandırılır.

Ω nın sınırında olmayan yani içinde olan noktalarda subdiferansiyel ile koşullu subdiferansiyel aynıdır. Aşağıdaki Şekil 3.13 de AB çizgisinin normal vektörü \tilde{s}_a , R de düzgün konveks fonksiyonun bir koşullu subgradyentidir. İlgili kümemiz $\Omega = a, b$ kapalı aralıktır. Dikkat edilirse grafik s_a yı da içeren bir çok koşullu subgradyente sahip iken CD tanjant çizgisine a noktasında normal olan sadece s_a subgradyent vektörüdür. Benzer olarak EF doğrusunun \tilde{s}_b normal vektörü b noktasındaki f nin koşullu subgradyentlerinden biridir. Koşullu subgradyent sadece fonksiyona değil aynı zamanda ilgili Ω kümesine de bağlıdır. Verilen bir konveks f fonksiyonu ve onun tanım kümesindeki bir x noktası için $\partial f(x)$ boştan farklı bir kümedir. $\partial f(x) \subseteq \partial^\Omega f(x)$ olduğu açıktır. Böylece $\partial^\Omega f(x)$ boştan farklıdır.



Şekil 3.13. $\Omega = a, b$ kümesinin uç noktaları için subgradyent ve koşullu subgradyentler

Teorem 3.15: (optimallik şartları) Aşağıdaki ifadeler geçerlidir.

a) $\bar{x} \in \Omega^*$ olması için gerek ve yeter şart $0 \in \partial^\Omega f \bar{x}$ olmasıdır.

b) $\bar{x} \in \Omega^*$ olması için gerek ve yeter şart $-\partial f \bar{x} \cap N_\Omega \bar{x} \neq \emptyset$ olmasıdır.

İspat: a) Verilen ifade f nin konveksliği ve koşullu subgradyent tanımından kolayca görülür. b) $\bar{x} \in \Omega^*$ olsun ve $-\partial f \bar{x} \cap N_\Omega \bar{x} = \emptyset$ olduğunu varsayalım. Bu bize bütün $s \in -\partial f \bar{x}$ ve bazı $x \in \Omega$ için $s \cdot (x - \bar{x}) > 0$ olduğunu ima eder. $0 \in -\partial f \bar{x}$ olduğu için $0 \cdot (x - \bar{x}) > 0$ çelişkisi elde edilir. O halde $\bar{x} \in \Omega^* \Leftrightarrow -\partial f \bar{x} \cap N_\Omega \bar{x} \neq \emptyset$ olur.

Diğer taraftan $-\partial f \bar{x} \cap N_\Omega \bar{x} \neq \emptyset$ olduğunu kabul edelim. Bu $s \in \partial f \bar{x}$ ve bütün $x \in \Omega$ için $s \cdot (x - \bar{x}) \geq 0$ olduğu anlamındadır. O zaman koşullu subgradyentin tanımından $s \in \partial f \bar{x} \subseteq \partial^\Omega f \bar{x}$ olduğu için

$$f(x) \geq f(\bar{x}) + s \cdot (x - \bar{x})$$

olup $s \cdot (x - \bar{x}) \geq 0$ olduğuda düşünülürse $f(x) \geq f(\bar{x})$ olup ispat tamamlanır.

Teorem 3.16: (koşullu subdiferansiyelin karakterizasyonu) Her bir $x \in \Omega$ için

$$\partial^\Omega f(x) = \partial f(x) + N_\Omega(x)$$

şeklindedir.

İspat: $v \in \partial^\Omega f(x)$ olduğunu varsayalım. Burada $x \in \Omega$ keyfi sabittir. Böylece

$$f(z) - f(x) \geq v(z-x) \quad \forall z \in \Omega \text{ için}$$

şeklindedir. $h: \Omega \rightarrow \mathbb{R}$, $h(z) = f(z) - f(x) - v(z-x)$ yardımcı fonksiyonunu tanımlayalım. h fonksiyonunun, Ω da konveks olduğu açıktır. Bütün $z \in \Omega$ için $h(z) \geq 0$ ve $h(x) = 0$ dır. Böylece x , Ω kümesinde h için bir minimum noktadır. Böylece optimallik koşullarından

$$-\partial h(x) \cap N_{\Omega}(x) \neq \emptyset$$

olur. Üstelik h nın tanımından $\partial h(x) = \partial f(x) - v$ yazılır. Böylece $-s_x + v \in N_{\Omega}(x)$ olacak şekilde bir $s_x \in \partial f(x)$ oluşur. Yani bazı $n_x \in N_{\Omega}(x)$ için $n_x = -s_x + v$ ya da $v = s_x + n_x \in \partial f(x) + N_{\Omega}(x)$ olur. Böylece $\partial^{\Omega} f(x) \subset \partial f(x) + N_{\Omega}(x)$ olur. Diğer taraftan $v_1 \in \partial f(x)$ ve $v_2 \in N_{\Omega}(x)$ olmak üzere $v = v_1 + v_2 \in \partial f(x) + N_{\Omega}(x)$ olduğunu varsayalım. Böylece; bütün $z \in \mathbb{R}^m$ ve $v_1 \in \partial f(x)$ için

$$f(z) - f(x) \geq v_1(z-x) \tag{3.3.3.3}$$

olur. $\forall z \in \Omega$ ve $v_2 \in N_{\Omega}(x)$ için

$$v_2(z-x) \leq 0 \tag{3.3.3.4}$$

olur. (3.3.3.3) ve (3.3.3.4) ifadeleri taraf tarafa toplanırsa

$$f(z) - f(x) \geq v_1(z-x) + v_2(z-x) = v(z-x)$$

olur. Bu da $v = v_1 + v_2 \in \partial^\Omega f(x)$ anlamındadır. Yani $\partial f(x) + N_\Omega(x) \subset \partial^\Omega f(x)$ olup ispat tamamlanır.

$\tilde{s} \in \partial^\Omega f(x^k)$, $x^k \in \Omega$ da f nin bir koşullu subgradyenti olsun. Bir koşullu subgradyent optimizasyon verilen bir $x^0 \in \Omega$ noktasından başlar. (3.3.3.2) problemi $k = 0, 1, 2, \dots$ için

$$x^{k+1} = P_\Omega(x^k - \lambda_k \tilde{s}^k) \quad (3.3.3.5)$$

kuralı ile bir x^k iterasyon dizisi oluşturulur. Burada $\tilde{s}^k = s^k + v^k$ şeklindedir. $s^k \in \partial f(x^k)$, $v^k \in N_\Omega(x^k)$ ve $\lambda_k > 0$ yakınsamayı garanti edecek bir kurala göre seçilen adım uzunluğudur. Koşullu subgradyent metot, pür ve deflected subgradyent metotların bazı eksikliklerini hafiflettiği gibi onların önemli iki özelliğini de korur. Bu özellikler şunlardır: $x^k \subseteq \Omega$, $\tilde{s} \in \partial^\Omega f(x^k)$ ile (3.3.3.5) koşullu subgradyent kuralına göre oluşturulan bir dizi ise

i-optimal olmayan bütün x^k lar için $\tilde{s}^k \cdot (x^k - x^*) > 0$ şeklindedir. Çünkü koşullu subgradyentin tanımından

$$\tilde{s}^k \cdot (x^k - x^*) \geq f(x^k) - f(x^*) > 0 \quad (3.3.3.6)$$

olduğunu biliyoruz. Bu yüzden koşullu subgradyentin optimal çözüme götüren bir yön ile dar açılı olması subgradyent vektör ile benzer özelliğidir.

$$\mathbf{ii-} \quad 0 < \lambda_k < \frac{2(f(x^k) - f(x^*))}{\|\tilde{s}^k\|^2} \Rightarrow \|x^{k+1} - x^*\| < \|x^k - x^*\| \quad (3.3.3.7)$$

şeklindedir. Yani $\|x^k - x^*\|$ dizisi kesin azalandır.

Aşağıdaki teorem λ_k adım uzunluğunun seçimine bir şart koyarak koşullu subgradyent prosedür tarafından oluşturulan x^k iterasyon dizisinin yakınsamasını sağlar.

Teorem 3.17: bütün $k=0,1,2,\dots$ için $x^k \subseteq \Omega$, $\lambda_k > 0$ adım uzunluğu ile (3.3.3.2) problemine uygulanan (3.3.3.5) koşullu subgradyent prosedür tarafından oluşturulan iterasyonların bir dizisi olsun. Burada $\lambda_k > 0$ ve

$$\lim_{k \rightarrow \infty} \lambda_k = 0, \quad \sum_{k=0}^{\infty} \lambda_k = \infty \quad \text{ve} \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty \quad (3.3.3.8)$$

olsun. Eğer $\sup_k \|\tilde{s}^k\| < \infty$ ise o zaman x^k , Ω^* nin bir elemanına yakınsar.

İspat: $x^* \in \Omega^*$ ve $k \geq 1$ olsun. Her bir k iterasyonunda

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|P_{\Omega} (x^k - \lambda_k \tilde{s}^k) - x^*\|^2 \\ &\leq \|x^k - \lambda_k \tilde{s}^k - x^*\|^2 \\ &= \|x^k - x^*\|^2 - 2\lambda_k \tilde{s}^k \cdot (x^k - x^*) + \lambda_k^2 \|\tilde{s}^k\|^2 \end{aligned} \quad (3.3.3.9)$$

yazılır. (3.3.3.9) ifadesinin tekrarlı uygulaması

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 - 2 \sum_{j=0}^{k-1} \lambda_j \tilde{s}^j \cdot (x^j - x^*) + \sum_{j=0}^{k-1} \lambda_j^2 \|\tilde{s}^j\|^2 \quad (3.3.3.10)$$

olur. O zaman (3.3.3.6)ifadesinden $\forall j \geq 0$ için

$$\tilde{s}^j \|x^j - x^*\| > 0 \quad (3.3.3.11)$$

yazılır. Böylece (3.3.3.10) ve (3.3.3.11) ifadelerinden

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 + \sum_{j=0}^{k-1} \lambda_j^2 \|\tilde{s}^j\|^2 \quad (3.3.3.12)$$

yazılır. $c = \sup_j \|\tilde{s}^j\|$ ve $p = \sum_{j=0}^{\infty} \lambda_j^2$ olarak tanımlanırsa herhangi bir $j \geq 0$ için

$$\|\tilde{s}^j\| \leq c \text{ ve } \sum_{j=0}^{k-1} \lambda_j^2 < p$$

elde edilir. (3.3.3.12) ifadesinden herhangi bir $k \geq 1$ için

$$\|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2 + pc^2$$

yazılır. Bu $\|x^k - x^*\|$ dizisinin sınırlı olduğu anlamına gelir. Bu yüzden $\|x^k - x^*\|$ da sınırlıdır. Şimdi $\|\tilde{s}^t\| \|x^t - x^*\| \rightarrow 0$ olacak şekilde $\|x^k - x^*\|$ nın $\|x^t - x^*\|$ alt dizisi olmadığını varsayalım. O zaman bütün $k \geq K$ için $\|\tilde{s}^k\| \|x^k - x^*\| \geq \delta$ ile $\delta > 0$ olmalıdır. Burada K yeterince büyük doğal sayıdır. (3.3.3.11) dan dolayı dizi negatif değildir. Bu $\sum_{j=0}^{\infty} \lambda_j = \infty$ şartıyla birlikte düşünülürse

$$\lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \lambda_j \tilde{s}^j \|x^j - x^*\| = \infty$$

olur. Üstelik

$$\lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} \lambda_j^2 \|\tilde{s}^j\|^2 \leq c^2 \text{ ve } \lim_{k \rightarrow \infty} \sum_{j=0}^{\infty} \lambda_j^2 < \infty$$

olup (3.3.3.10) dan $\|x^k - x^*\| \rightarrow -\infty$ olur. Bu mümkün değildir. x^k dizisi $\tilde{s}^t x^t - x^* \rightarrow 0$ olacak şekilde x^t alt dizisi içermelidir. (3.3.3.6) dan $f x^t \rightarrow f^*$ olur. Üstelik x^k sınırlı olduğu için x^t dizisinin bir \bar{x} gibi bir yığılma noktası oluşur. f nin sürekliliğinden $f \bar{x} = f^*$ dir. Böylece $\bar{x} \in \Omega^*$ dir. Şimdi \bar{x} e yakınsayan bütün x^k dizilerini göstermek için

$$\|x^N - \bar{x}\| < \varepsilon \text{ ve } \sum_{t=N}^{\infty} \lambda_t^2 < \frac{\varepsilon}{2c^2}$$

olacak şekilde $\varepsilon > 0$ ve $N = N(\varepsilon)$ bulunsun. O zaman herhangi bir $k > N$ için benzer olarak (3.3.3.12) ifadesindeki gibi

$$\|x^k - \bar{x}\|^2 \leq \|x^N - \bar{x}\|^2 + \sum_{j=N}^{k-1} \lambda_j^2 \|\tilde{s}^j\|^2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2c^2} c^2 = \varepsilon$$

elde edilir. Bu keyfi herhangi bir küçük $\varepsilon > 0$ için olduğundan teoremin ispatı tamamlanır.

Koşullu subgradyentlerin \tilde{s}^k dizisinin sınırlılığı (3.3.3.5) metoduunda v^k nın uygun seçimiyle garantilenebilir. Mümkün yollardan biri

$$v^k = P_{N_{\Omega} x^k} -s^k \quad (3.3.3.13)$$

şeklindedir. Burada $s^k \in \partial f x^k$ şeklindedir. $\partial f x^k$ sınırlı bir konveks küme ve (3.3.3.13) ifadesine göre seçilen bir v^k da sınırlı olduğu için

$$\tilde{s}^k = v^k + s^k$$

dizisi de sınırlıdır.

Doğal sonuç: (uyarlamalı adım uzunluğu seçim kuralı)

(3.3.3.5) metodundaki λ_k nın yerine adım uzunluklarının keyfi bir dizisi η_k olsun. Her ikisi de (3.3.3.8) ifadesini sağlayan $\underline{\lambda}_k$ ve $\overline{\lambda}_k$ dizileri olup bütün k lar için $\underline{\lambda}_k \leq \eta_k \leq \overline{\lambda}_k$ olursa teorem 3.17 nin iddiası gerçekleşir.

$\underline{\lambda}_k$ ve $\overline{\lambda}_k$ dizisinin elemanları keyfi olarak küçük ve büyük yapılabilir. Örnek olarak sırasıyla (3.3.3.8) şartları sağlanırken

$$\underline{\lambda}_k = \frac{\alpha}{\beta + k} \text{ ve } \overline{\lambda}_k = \frac{M}{\beta + k}$$

yazılsın. Burada $\alpha > 0$ yeterince küçük, $M > 0$ gerektiği kadar büyük ve $\beta > 0$ dir. Doğal sonuç 3.18 in şartı bize adım boyu seçiminde esneklik sağlar. Aşağıdaki örneği daha önce pür subgradyentle çözmüştük. Şimdi koşullu subgradyentle çözüp koşullu subgradyentin etkisini göreceğiz.

Örnek 3.7: Örnek 3.6 durumuna Polyak's adım uzunluğunu kullanarak (3.3.3.5) metodunu uygulayacağız. $x^0 = 0,0$ ile başlayalım. x^0 da normal koni

$$N_{\Omega} x^0 = \{y \in \mathbb{R}^2 : y_1 + y_2 \leq 0, y_1 \leq 0\}$$

şeklindedir. $s^0 = 1, -2$ olup hareket yönü $-s^0 = -1, 2$ olur. $-s^0 = -1, 2$ nin normal koni üzerine iz düşümü $v^0 = -\frac{3}{2}, \frac{3}{2}$ dir. Böylece x^0 da bir koşullu subgradyent

$$\tilde{s}^0 = -\frac{3}{2}, \frac{3}{2} + 1, -2 = -\frac{1}{2}, -\frac{1}{2}$$

olur. Buradan $\lambda_0 = \frac{1}{\|\tilde{s}^0\|^2} = 2$ olur. Böylece

$$x^1 = P_{\Omega} x^0 - \lambda_0 \tilde{s}^0 = P_{\Omega} 1, 1 = 1, 1$$

bulunur. Yani optimal çözüme sadece bir iterasyonda ulaşıldı. Dikkat edilirse koşullu subgradyent metodun etkinliği seçilen koşullu subgradyent yöne bağlıdır ve keyfi seçilen koşullu subgradyent 2.tür zigzağı ortadan kaldırmayabilir. Örneğin yukarıdaki örnekte $-1, 1$ de $N_{\Omega} x^0$ in bir elemanıdır ve böylece $\tilde{s}^0 = 1, -2 + -1, 1 = 0, -1$ x^0 da bir koşullu subgradyenttir. Ama uygun bir yön değildir.

3.3.4. Zikzak-Free Subgradyent metot

Bu bölümde (Guta 2003) çalışmasındaki sonuçlar incelenmiştir.

Subgradyent metotların adım yönlerinden meydana çıkan zikzak davranışı prosedürün yavaş yakınsamasına sebep olup, metodun performansını azaltmaktadır. Daha önce bahsedildiği gibi zikzak olgusu, birinci tür yada ikinci tür zikzak şeklinde ortaya çıkabilir. Bütün prosedürler her iki zikzak türüne maruz kalabilir. Bu durum özellikle geniş boyutlu problemleri oldukça verimsizleştirir.

Sırasıyla Bölüm 3.3.2 ve Bölüm 3.3.3 de tartışılan deflected subgradyent ve koşullu subgradyent metotlar, pür subgradyent metodun zikzak oluşumunu engelleme amaçlı dizayn edilmiştir. Ancak deflected subgradyent metot, birinci tür zikzak durumunda işe yarasa da ikinci tür zikzak ile ilgili değildir. Diğer taraftan koşullu subgradyent metot, ikinci tür zikzak durumunda zikzakları kaldırsa da birinci tür zikzak oluşumunu engelleyememektedir.

Bu bölümde seçilen adım yönü ile her iki zikzak türünü de tamamen ortadan kaldıracak yeni bir strateji dizayn edilecektir. Bu strateji, deflected ve koşullu subgradyent prosedürleri birlikte düşünüp birleştirmeyi amaçlar. 3.bölümde ikinci tür zikzak durumunun, iterasyon noktalarının uygun bölgenin göreceli sınırı boyunca değiştiği zaman olacağı söylenmişti. Bu şekilde oluşan zikzak durumunun, uygun bir koşullu subgradyent adım yönü ile kontrol altına alınabileceği ifade edilmişti. Deflected subgradyent metot ise uygun bir saptırma parametresi ile iterasyon noktaları uygun kümenin içinde hareket ederken, birinci tür zikzak durumunun oluşumunu ortadan kaldırmaya yardım ediyordu. Şimdi öyle bir prosedür inşa edilecek ki; iterasyon noktaları uygun bölgenin sınırı boyunca hareket ederken koşullu subgradyenti taklit edecek ve her ne zaman gerekirse deflected subgradyent prosedüre geçecek. Yani dizayn edilen prosedür, her bir noktada sonuç subgradyent yönünü inceleyecek ve ortaya çıkan zikzak türünün eğilimi hakkında bilgi verecek. Bu bilgiye dayalı olarak uygun adım yönü seçilerek zikzak oluşumu engellenecek.

Bu bölümde özel kurallar yardımıyla seçilen subgradyent adım yönü ile zikzak olgusunu tamamen ortadan kaldırabilen yeni bir algoritma sunulacaktır. Bir d^k adım yönü bir $x^k \in bd\Omega$ iterasyon noktasında bazı $v \in N_{\Omega} x^k$ için $d^k v > 0$ ise bu v yönünün uygun olmayacağı söylenebilir.(yani d^k vektörü, x^k noktasında bir normal vektörle dar açılı şekildedir.) Çünkü böyle bir yönle, ikinci tür zikzak olgusu daha da kötüleşebilir. Bu zorluktan kurtulmak için uygun olmayan bu yönü kullanmaktansa $x^k \in bd\Omega$ için uygun bir tanjant vektörünü yön olarak seçip bir sonraki iterasyon noktasına gidilebilir. x^k noktasında bu yön

$$\Delta^k = s^k + v^k \quad (3.3.4.1)$$

kuralı tarafından belirlenebilir. Burada $s^k \in \partial f(x^k)$ ve $v^k = P_{N_\Omega(x^k)}(-s^k)$ dir. (eğer s^k vektörü, x^k noktasında uygun bir yön değilse) $\Delta^k = s^k + P_{N_\Omega(x^k)}(-s^k)$ vektörü, $N_\Omega(x^k)$ normal konisine ortogonal olduğu için Δ^k vektörü, x^k noktasında Ω kümesinin tanjant vektörüdür.

Δ^k vektörü, (3.3.4.1) kuralı ile elde edilen x^k noktasında f fonksiyonunun bir koşullu subgradyentidir. Ancak Δ^k vektörünü daima bir adım yönü olarak kullanan koşullu subgradyent metoda zıt olarak bu prosedürde Δ^k vektörünü, sadece iterasyon noktasında subgradyent uygun bir yön olmadığında kullanılır. Diğer durumlarda yani eğer s^k uygun olmayan yön değilse ya da $x^k \in \text{int } \Omega$ ise birinci tür zikzak oluşumunu engellemek için deflected subgradyent stratejisine geçilir. Böylece Ω de x^n iterasyonlarının dizisini oluşturan üretken prosedür aşağıdaki gibi olur.

Üretken hibrid subgradyent algoritma

Adım 0: $k = 0$, $\Delta^{-1} = 0$ ve $x^0 \in \Omega$ olarak al.

Adım 1: (adım yönüne karar verme)

Eğer x^k da s^k uygun yön değil ve $x^k \in \text{bd}\Omega$ ise

$$\Delta^k = s^k + v^k \quad (3.3.4.2)$$

Diğer durumlarda

$$\Delta^k = s^k + \delta_k \Delta^{k-1}$$

olarak al. Burada $s^k \in \partial f(x^k)$, $v^k = P_{N_{\Omega}(x^k)} -s^k$ ve δ_k uygun seçilen saptırma parametresidir.

Adım 2:

$$x^{k+1} = P_{\Omega}(x^k - \lambda_k \Delta^k) \quad (3.3.4.3)$$

Adım 3: $k = k + 1$ al ve istenilen optimallik gerçekleşene kadar Adım 1 e dönerek tekrar et.

(3.3.4.2) tarafından verilen adım yönü ya koşullu subgradyent ya da bir deflected subgradyent yön olması nedeniyle hibrid (melez) subgradyent yön diye adlandırılır. Bu metodu kullanırken zikzak oluşumunu ortadan kaldırmak için uygun bir λ_k adım uzunluğu ve δ_k saptırma parametresi seçilmelidir. Ayrıca bu prosedürde subgradyentin iki önemli özelliği olan

i- x^k da adım yönü optimal çözüme öncülük eden adım yönü ile dar açılı olmalı yani

$$\Delta^k \cdot (x^k - x^*) > 0$$

ii- $\|x^k - x^*\|$ dizisi kesin azalan olmalı ve 0 a yakınsamalı

özellikleri korunmalıdır. Aşağıdaki Lemma'lar teoremlerin ispatında kullanılacaktır.

Lemma 3.2: Ω , boştan farklı kapalı konveks ve R^m nin bir alt kümesi olsun. Eğer $v \in R^m$ ve $P_v = P_\Omega v$ ise o zaman

i- herhangi bir $x \in \Omega$ için $\|P_v - x\| \leq \|v - x\|$

ii- herhangi bir $\alpha > 0$ skaleri ve herhangi bir $x \in \Omega$ için $\alpha \|v - x - P_v - v\| \leq 0$

ispat: $v - P_v \in N_\Omega P_v$ olduğuna dikkat edelim. Bu yüzden herhangi bir $x \in \Omega$ için

$$x - P_v \cdot v - P_v \leq 0 \quad (3.3.4.4)$$

olur.

i- (3.3.4.4) ifadesindeki eşitsizlikten direkt olarak $xv - xP_v \leq vP_v - \|P_v\|^2$ elde edilir.

Şimdi biz aşağıdaki ilişkide bu son eşitsizlik kullanılırsa

$$\begin{aligned} \|P_v - x\|^2 - \|v - x\|^2 &= \|P_v\|^2 - \|v\|^2 + 2 \cdot xv - xP_v \\ &\leq \|P_v\|^2 - \|v\|^2 + 2vP_v - 2\|P_v\|^2 \\ &= -\|P_v - v\|^2 \\ &\leq 0 \end{aligned}$$

olup birinci eşitsizlik elde edilir.

ii- tekrar (3.3.4.4) ü kullanarak

$$\begin{aligned} x - v + v - P_v \cdot v - P_v &\leq 0 \\ x - v \cdot v - P_v &\leq -\|v - P_v\|^2 \leq 0 \end{aligned}$$

$$v - x \quad P_v - v \leq 0$$

yazılır. Lemma 3.2 için aşağıdakileri söylenebilir. x^k (3.3.4.3) kuralı tarafından oluşturulan iterasyonların bir dizisi olsun. O zaman

$x = x^*$ ve $v = x^k - \lambda_k \Delta^k$ yazılırsa Lemma 3.2 nin i. şikkından

$$\|x^{k+1} - x^*\| \leq \|x^k - \lambda_k \Delta^k - x^*\| \quad (3.3.4.5)$$

yazılır.

$x = x^k$, $v = x^k - \lambda_k \Delta^k$ ve $\alpha = \frac{1}{\lambda_k}$ yazılarak Lemma 3.2 nin ii. şikkından

$$\Delta^k \quad x^{k+1} - x^k + \lambda_k \Delta^k \geq 0$$

$$\Delta^k \quad x^{k+1} - x^k \geq -\lambda_k \|\Delta^k\|^2 \quad (3.3.4.6)$$

yazılır. Bu ilişkiler aşağıdaki lemmanın ispatında kullanılacaktır.

Lemma 3.3: $x^k \subseteq \Omega$ (3.3.4.3) tarafından oluşturulan iterasyonların bir dizisi ve Δ^k , adım yönü (3.3.4.2) tarafından verilen hibrid subgradyent yönü olsun. Eğer

$$0 < \lambda_k \leq \frac{f(x^k) - f^*}{\|\Delta^k\|^2} \quad (3.3.4.7)$$

ise o zaman herhangi bir x^k iterasyon noktası için

$$\Delta^k x^k - x^* \geq f(x^k) - f(x^*)$$

olur.

İspat: Tümevarım yöntemini kullanalım. $k=0$ olsun. O zaman ya $\Delta^0 = s^0 + P_{N_\Omega x^0} -s^0 \in \partial^\Omega f(x^0)$ ya da $\Delta^0 = s^0 \in \partial f(x^0)$ dır. Her iki durum için kural doğrudur. Çünkü bu kural hem koşullu subgradyent hem de pür subgradyent için geçerliydi. $k \in N$ için iddianın doğru olduğunu kabul edelim. O zaman x^{k+1} de s^{k+1} uygun olmayan yön ve $x^{k+1} \in bd\Omega$ durumunda

$$\Delta^{k+1} = s^{k+1} + P_{N_\Omega x^{k+1}} -s^{k+1}$$

elde edilir. Bu yine koşullu subgradyent yön olduğundan bunun için kural doğrudur. Yani

$$\Delta^{k+1} x^{k+1} - x^* \geq f(x^{k+1}) - f(x^*)$$

elde edilir. Diğer durum yani $x^k \in \text{int } \Omega$ durumunda inceleyelim. $k=0$ durumunda $\Delta^0 = s^0 \in \partial f(x^0)$ olup kural sağlanır. $k \in N$ için iddianın doğru olduğunu kabul edelim ve $k+1$ için doğru olduğunu göstereyim. $\Delta^{k+1} = s^{k+1} + \delta_{k+1} \Delta^k$ dir. Böylece

$$\Delta^{k+1} x^{k+1} - x^* = s^{k+1} x^{k+1} - x^* + \delta_{k+1} \Delta^k x^{k+1} - x^* \quad (3.3.4.8)$$

Şimdi $\Delta^k x^{k+1} - x^* \geq 0$ olduğunu göstereceğiz.

$$\Delta^k x^{k+1} - x^* = \Delta^k x^{k+1} - x^k + x^k - x^*$$

$$\begin{aligned}
&= \Delta^k x^k - x^* + \Delta^k x^{k+1} - x^k \\
&\geq f x^k - f x^* + \Delta^k x^{k+1} - x^k \quad (\text{hipotezden}) \\
&\geq f x^k - f x^* - \lambda_k \|\Delta^k\|^2 \quad [(3.3.4.6) \text{ den}] \\
&\geq 0 \quad [\lambda_k \text{ da verilen şartlardan}]
\end{aligned}$$

olur. Bu elde edilen sonuç (3.3.4.8) ifadesinde kullanılırsa

$$\Delta^{k+1} x^{k+1} - x^* \geq s^{k+1} x^{k+1} - x^* \geq f x^{k+1} - f x^*$$

elde edilir.

Teorem 3.18: x^k , (3.3.4.3) tarafından oluşturulan iterasyonların bir dizisi ve Δ^k adım yönü (3.3.4.2) tarafından verilen hibrid subgradyent adım yönü olsun. Eğer

$$0 < \lambda_k \leq \frac{f x^k - f^*}{\|\Delta^k\|^2} \quad (3.3.4.9)$$

ise o zaman bütün $x^k \notin \Omega^*$ için

i- $\Delta^k x^k - x^* > 0$

ii- $\|x^{k+1} - x^*\| < \|x^k - x^*\|$

iii- Eğer bazı x^n iterasyon noktalarında $\Delta^n = 0$ ise o zaman x^n bir optimal çözümdür.

İspat: i-) Lemma 3.3 den bütün $x^k \notin \Omega^*$ için

$$\Delta^k \|x^k - x^*\| \geq f(x^k) - f(x^*) > 0$$

olduğu açıktır.

ii-) Lemma 3.2 kullanılarak

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - \lambda_k \Delta^k - x^*\|^2 \\ &= \|x^k - x^*\|^2 + \lambda_k \left[\lambda_k \|\Delta^k\|^2 - 2\Delta^k \|x^k - x^*\| \right] \end{aligned}$$

yazılır. Şimdi $\left[\lambda_k \|\Delta^k\|^2 - 2\Delta^k \|x^k - x^*\| \right] < 0$ olduğu gösterilmelidir. λ_k daki şartlar ve

Lemma 3.3 kullanılarak

$$\lambda_k \|\Delta^k\|^2 \leq f(x^k) - f^* \leq \Delta^k \|x^k - x^*\| < 2\Delta^k \|x^k - x^*\|$$

olur. Böylece $\lambda_k \|\Delta^k\|^2 - 2\Delta^k \|x^k - x^*\| < 0$ olur ve ispat tamamlanır.

iii-) $f^* = \min \{ f(x^n) : x^n \in \Omega \} \leq f(x^n)$ olup diğer taraftan lemma 3.3 kullanılarak

$$f(x^n) - f^* \leq \Delta^n \|x^n - x^*\| = 0$$

yazılır. Böylece $f(x^n) \leq f^*$ olur. Buda $f(x^n) = f^*$ olduğunu gösterir. Teorem 3.18 ile görüldü ki hibrid subgradyent, subgradyentlerin iki önemli özelliğini korudu. Bunlara değinilecek olursa

i-) herhangi bir iterasyon noktasındaki Δ^k hibrid subgradyent yön x^k noktasından bir optimal çözüm olan x^* noktasına öncülük eden yön ile dar açılı formdadır.

ii-) $\|x^k - x^*\|$ dizisi monoton azalandır.

Teorem 3.19: $x^k \subseteq \Omega$ dizisi $0 < \varepsilon_0 \leq \mu_k \leq 1$ için

$$\lambda_k = \mu_k \frac{f(x^k) - f^*}{\|\Delta^k\|^2} \quad (3.3.4.10)$$

adım uzunluğunu kullanarak hibrid subgradyent prosedür tarafından bir dizi olsun. Eğer $\sup \|\Delta^k\| : k=0,1,2,\dots < \infty$ ve Ω^* boştan farklı ise o zaman bazı $x^* \in \Omega^*$ için $f(x^k) \rightarrow f^*$ ve $x^k \rightarrow x^*$ olur.

İspat: Lemma 3.2 kullanılarak

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - \lambda_k \Delta^k - x^*\|^2 \\ &= \|x^k - x^*\|^2 + \lambda_k^2 \|\Delta^k\|^2 - 2\lambda_k \Delta^k \cdot (x^k - x^*) \\ &= \|x^k - x^*\|^2 + \mu_k^2 \frac{f(x^k) - f^*}{\|\Delta^k\|^2} - 2\mu_k \frac{f(x^k) - f^*}{\|\Delta^k\|^2} \Delta^k \cdot (x^k - x^*) \\ &\leq \|x^k - x^*\|^2 - \mu_k \frac{f(x^k) - f^*}{\|\Delta^k\|^2} \end{aligned} \quad (3.3.4.11)$$

$$\leq \|x^k - x^*\|^2 - \frac{\varepsilon_0}{c^2} (f(x^k) - f^*)^2 \quad (3.3.4.12)$$

yazılabilir. Burada $c = \sup \|\Delta^k\|: k=0,1,2,\dots$ şeklindedir. (3.3.4.11) ifadesi Lemma 3.3 kullanılarak elde edildi ve (3.3.4.12) ifadesi, $2 - \mu_k \geq 1$ ve $\mu_k \geq \varepsilon_0$ olduğundan yazıldı. (3.3.4.12) den $\|f(x^k) - f^*\|^2 \rightarrow 0$ ya da $f(x^k) \rightarrow f^*$ sonucu çıkarılır.

Teorem 3.20: Δ^k , hibrid adım yönü (3.3.4.2) tarafından verilsin. Adım uzunluğu $\lambda_k > 0$ ve saptırma parametresi δ_k ya

$$\text{i-)} \text{ eğer } s^k \Delta^{k-1} < 0 \text{ ise } \delta_k = -\tau_k \frac{s^k \Delta^{k-1}}{\|\Delta^{k-1}\|^2}, \text{ diğer durumlarda } \delta_k = 0 \quad (3.3.4.13)$$

burada $1 < \tau_k < 2$ ve $s^k \in \partial f(x^k)$ dir. Ya da

$$\text{ii-)} \delta_k = \frac{\|s^k\|}{\|\Delta^{k-1}\|} \quad (3.3.4.14)$$

tarafından verilsin. O zaman hibrid subgradyent prosedür, 1.tür ve 2.tür zikzaklardan kurtulur.

İspat: s^k , uygun olmayan yön olsun yani; $x^k \in \text{bd } \Omega$ ve s^k , x^k daki Ω nın bir normal vektörü ile dar açılı olsun. O zaman $\Delta^k = s^k + P_{N_{\Omega}(x^k)}(-s^k)$ bir normal vektöre ortogondur ve böylece $\Delta^k \in T_{\Omega}(x^k)$ olur. Böylece 2.tür zikzak ortaya çıkmaz. Eğer s^k uygun olmayan yön değilse bu durumda sadece 1.tür zikzak ortaya çıkabilir. Bunun içinde $\Delta^k = s^k + \delta_k \Delta^{k-1}$ kullanılır.

i-) şimdi eğer δ_k (3.3.4.13) e göre seçilirse δ_k ya 0 dır ya da pozitif bir skalerdir. Eğer $s^k \Delta^{k-1} \geq 0$ ise o zaman $\delta_k = 0$ olur ve $\Delta^k = s^k$ elde edilir. Böylece $s^k \Delta^{k-1} < 0$ durumunu göz önünde bulunduralım. O zaman

$$\begin{aligned}\Delta^k \Delta^{k-1} &= s^k \Delta^{k-1} + \delta_k \|\Delta^{k-1}\|^2 \\ &= s^k \Delta^{k-1} - \tau_k s^k \Delta^{k-1} \\ &= (1 - \tau_k) s^k \Delta^{k-1} \\ &> 0\end{aligned}$$

olur. Burada $1 - \tau_k < 0$ ve $s^k \Delta^{k-1} < 0$ olduğundan yazıldı. Böylece aynı zamanda 1.tür zikzak da oluşmamış oldu.

ii-) eğer δ_k , (3.3.4.14) tarafından verilirse o zaman $\Delta^k = s^k + \frac{\|s^k\|}{\|\Delta^{k-1}\|} \Delta^{k-1}$ şeklindedir.

Her iki taraf Δ^{k-1} ile çarpılırsa

$$\begin{aligned}\Delta^k \Delta^{k-1} &= s^k \Delta^{k-1} + \|s^k\| \|\Delta^{k-1}\| \\ &\geq \|s^k\| \|\Delta^{k-1}\| - |s^k \Delta^{k-1}| \quad [s^k \Delta^{k-1} \geq -|s^k \Delta^{k-1}|] \\ &\geq 0 \quad [\text{cauchy-schwarz eşitsizliği}]\end{aligned}$$

bulunur.

3.4. Nümerik Uygulamalar

Örnek 3.8: $f(x, y) = \max \{-x, x + y, x - 2y\}$ fonksiyonunun minimum noktasını veya

noktalarından birini bulunuz. Burada $\Omega = R \times R$ ve $u^0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ olsun. ($u^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ dır.)

Çözüm: Azalan adım boyu kuralı ile birlikte pür subgradyent metodunu kullanalım. Daha sonra da deflected subgradyent metodu kullanarak çözüme gidip iki metodu kıyaslayalım. Görüldüğü üzere burada Ω bölgesiyle alakalı herhangi bir kısıtlama olmadığından 2.tür zikzak oluşmaz. Dolayısıyla bu örnek için koşullu subgradyent metottan yararlanmayacağız. Şimdi azalan adım boyu kuralını $\lambda_k = \frac{1}{k+1}$ ile birlikte pür subgradyent metodu uygulayalım. $f(x, y) = \max\{-x, x+y, x-2y\}$ fonksiyonu $u^0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ noktasında $f_2 = x+y$ üzerindedir. Dolayısıyla subgradyent $s_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ olur. Pür subgradyent algoritmasını uygulayacak olursak;

$$u^1 = u^0 - \lambda_0 s_0$$

$$u^1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} - 1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

olur. Şimdi de $u^1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ noktasının $f(x, y) = \max\{-x, x+y, x-2y\}$ fonksiyonunda hangi alt fonksiyon üzerine düştüğüne bakalım. $u^1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ noktası $f_2 = x+y$ üzerindedir. Pür subgradyent yine $s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ dir. O halde

$u^2 = u^1 - \lambda_1 s_1$ ifadesinden

$$u^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0,5 \\ 1,5 \end{pmatrix}$$

olur. Bu şekilde devam ederse;

$$u^0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \rightarrow \|u^0 - u^*\| = 3,605551$$

$$u^1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \rightarrow \|u^1 - u^*\| = 2,236068$$

$$u^2 = \begin{pmatrix} 0,5 \\ 1,5 \end{pmatrix} \rightarrow \|u^2 - u^*\| = 1,581139$$

$$u^3 = \begin{pmatrix} 0,1666 \\ 1,1666 \end{pmatrix} \rightarrow \|u^3 - u^*\| = 1,178511$$

$$u^4 = \begin{pmatrix} -0,0833 \\ 0,9166 \end{pmatrix} \rightarrow \|u^4 - u^*\| = 0,920447$$

$$u^5 = \begin{pmatrix} -0,2833 \\ 0,7166 \end{pmatrix} \rightarrow \|u^5 - u^*\| = 0,770642$$

$$u^6 = \begin{pmatrix} -0,4500 \\ 0,5500 \end{pmatrix} \rightarrow \|u^6 - u^*\| = 0,710634$$

$$u^7 = \begin{pmatrix} -0,3071 \\ 0,5500 \end{pmatrix} \rightarrow \|u^7 - u^*\| = 0,62995$$

$$u^8 = \begin{pmatrix} -0,1821 \\ 0,5500 \end{pmatrix} \rightarrow \|u^8 - u^*\| = 0,579376$$

$$u^9 = \begin{pmatrix} -0,2933 \\ 0,4389 \end{pmatrix} \rightarrow \|u^9 - u^*\| = 0,527881$$

$$u^{10} = \begin{pmatrix} -0,1933 \\ 0,4389 \end{pmatrix} \rightarrow \|u^{10} - u^*\| = 0,479581$$

$$u^{11} = \begin{pmatrix} -0,2842 \\ 0,4380 \end{pmatrix} \rightarrow \|u^{11} - u^*\| = 0,449304$$

$$u^{12} = \begin{pmatrix} -0,2009 \\ 0,3480 \end{pmatrix} \rightarrow \|u^{12} - u^*\| = 0,401827$$

$$u^{13} = \begin{pmatrix} -0,1240 \\ 0,3480 \end{pmatrix} \rightarrow \|u^{13} - u^*\| = 0,369432$$

$$u^{14} = \begin{pmatrix} -0,1954 \\ 0,2766 \end{pmatrix} \rightarrow \|u^{14} - u^*\| = 0,338658$$

$$u^{15} = \begin{pmatrix} -0,1287 \\ 0,2766 \end{pmatrix} \rightarrow \|u^{15} - u^*\| = 0,305075$$

$$u^{16} = \begin{pmatrix} -0,1912 \\ 0,2141 \end{pmatrix} \rightarrow \|u^{16} - u^*\| = 0,287047$$

$$u^{17} = \begin{pmatrix} -0,1324 \\ 0,2141 \end{pmatrix} \rightarrow \|u^{17} - u^*\| = 0,251732$$

$$u^{18} = \begin{pmatrix} -0,0794 \\ 0,2141 \end{pmatrix} \rightarrow \|u^{18} - u^*\| = 0,227458$$

$$u^{19} = \begin{pmatrix} -0,1294 \\ 0,1615 \end{pmatrix} \rightarrow \|u^{19} - u^*\| = 0,206947$$

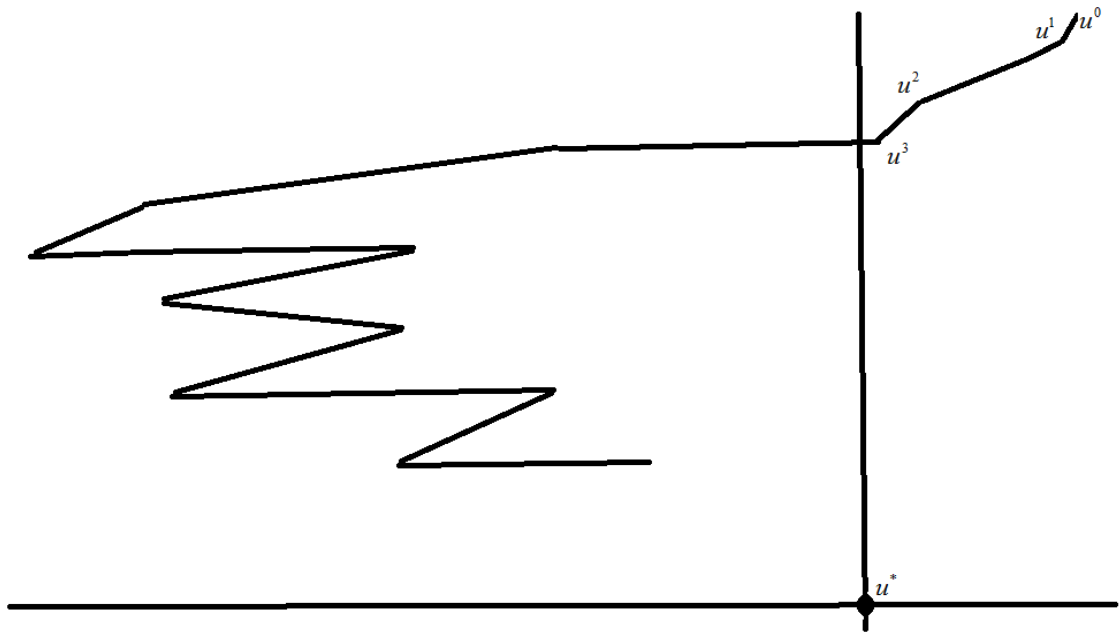
$$u^{20} = \begin{pmatrix} -0,0794 \\ 0,1615 \end{pmatrix} \rightarrow \|u^{20} - u^*\| = 0,179964$$

$$u^{21} = \begin{pmatrix} -0,1270 \\ 0,1139 \end{pmatrix} \rightarrow \|u^{21} - u^*\| = 0,170593$$

$$u^{22} = \begin{pmatrix} -0,0815 \\ 0,1139 \end{pmatrix} \rightarrow \|u^{22} - u^*\| = 0,140054$$

$$u^{23} = \begin{pmatrix} -0,0380 \\ 0,1139 \end{pmatrix} \rightarrow \|u^{23} - u^*\| = 0,120071$$

şeklinde devam eder. Dikkat edilirse her bir iterasyonda minimum noktaya biraz daha yaklaştık. Şimdi bunu grafik üzerinde görelim.



Şekil 3.14. Pür subgradyentle optimal noktaya yaklaşma

Şekle bakıldığında bir çok kez zikzak yapıldığı görülür. Bu zikzaklar Bölüm 3.3.2 de tanımlanan 1.tür zikzaklardır. Bunlardan kurtulmak için deflected subgradyent metot kullanılacak ve sonunda da deflected subgradyent metodun minimuma yaklaşma hızının pür subgradyent metottan çok daha iyi olduğunu görülecektir.

Deflected subgradyent algoritmada da $u^0 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ başlangıç noktasıyla başlayalım.

Başlangıç için $d^0 = s^0$ durumu vardır. O halde

$$u^1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

olur. Şimdi hareketin önceki yönü $d^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ oldu. Güncel subgradyent yönü ise $s^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

dir. Hatırlanacağı üzere hareketin yeni yönünü belirlemek için hareketin önceki yönü ile güncel subgradyent yönün çarpımına bakıyorduk. Eğer çarpım sıfırdan küçükse yani hareketin önceki yönü ile güncel subgradyent arasındaki açı geniş ise o zaman hareketin yeni yönü

$$d^k = s^k + \delta_k d^{k-1}$$

formülü ile belirlenir. Eğer çarpım sıfıra eşit ya da sıfırdan büyükse o zaman hareketin yönü güncel subgradyenttir. O halde çarpıma bakalım

$$d^0 s^1 = 1,1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2 > 0$$

olup yeni adım yönü güncel subgradyent yönü olacaktır. O halde

$$u^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0,5 \\ 1,5 \end{pmatrix}$$

olur. Bu şekilde her bir adımda hareketin yönü belirlenerek devam edilirse

$$u^3 = \begin{pmatrix} 0,1666 \\ 1,1666 \end{pmatrix}, u^4 = \begin{pmatrix} -0,0833 \\ 0,9166 \end{pmatrix}, u^5 = \begin{pmatrix} -0,2833 \\ 0,7166 \end{pmatrix}, u^6 = \begin{pmatrix} -0,4500 \\ 0,5500 \end{pmatrix}$$

olup buraya kadar hareketin yeni yönü hep güncel subgradyent yön olur. u^6 noktası için hareketin önceki yönü $d^5 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ dir. Güncel subgradyent yönünü bulmak için u^6 noktasının hangi alt fonksiyon üzerine düştüğünü bulmalıyız. u^6 noktası $f_1 = -x$ alt fonksiyonu üzerine düşmektedir. O halde güncel subgradyent $s^6 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ dir. Şimdi hareketin yeni yönünü tespit etmek için hareketin önceki yönü ile güncel subgradyent yönün çarpımına bakacağız.

$$d^5 s^6 = 1,1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = -1 < 0$$

olup hareketin yeni yönünü belirlemek için $d^k = s^k + \delta_k d^{k-1}$ kuralını kullanacağız. Bu kurala göre hareketin yeni yönü

$$d^6 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \frac{3}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0,25 \\ 0,75 \end{pmatrix}$$

olur. Buna göre

$$u^7 = \begin{pmatrix} -0,4500 \\ 0,5500 \end{pmatrix} - \frac{1}{7} \begin{pmatrix} -0,25 \\ 0,75 \end{pmatrix} = \begin{pmatrix} -0,414286 \\ 0,442857 \end{pmatrix}$$

olur. Bu şekilde devam edilirse

$$u^0 = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \rightarrow \|u^0 - u^*\| = 3,605551$$

$$u^2 = \begin{pmatrix} 0,5 \\ 1,5 \end{pmatrix} \rightarrow \|u^2 - u^*\| = 1,581139$$

$$u^3 = \begin{pmatrix} 0,1666 \\ 1,1666 \end{pmatrix} \rightarrow \|u^3 - u^*\| = 1,178511$$

$$u^4 = \begin{pmatrix} -0,0833 \\ 0,9166 \end{pmatrix} \rightarrow \|u^4 - u^*\| = 0,920447$$

$$u^5 = \begin{pmatrix} -0,2833 \\ 0,7166 \end{pmatrix} \rightarrow \|u^5 - u^*\| = 0,770642$$

$$u^6 = \begin{pmatrix} -0,4500 \\ 0,5500 \end{pmatrix} \rightarrow \|u^6 - u^*\| = 0,710634$$

$$u^7 = \begin{pmatrix} -0,414286 \\ 0,442857 \end{pmatrix} \rightarrow \|u^7 - u^*\| = 0,606428$$

$$u^8 = \begin{pmatrix} -0,289286 \\ 0,442857 \end{pmatrix} \rightarrow \|u^8 - u^*\| = 0,52897$$

$$u^9 = \begin{pmatrix} -0,178175 \\ 0,442857 \end{pmatrix} \rightarrow \|u^9 - u^*\| = 0,477356$$

$$u^{10} = \begin{pmatrix} -0,128175 \\ 0,342857 \end{pmatrix} \rightarrow \|u^{10} - u^*\| = 0,366033$$

$$u^{11} = \begin{pmatrix} -0,219084 \\ 0,251948 \end{pmatrix} \rightarrow \|u^{11} - u^*\| = 0,33388$$

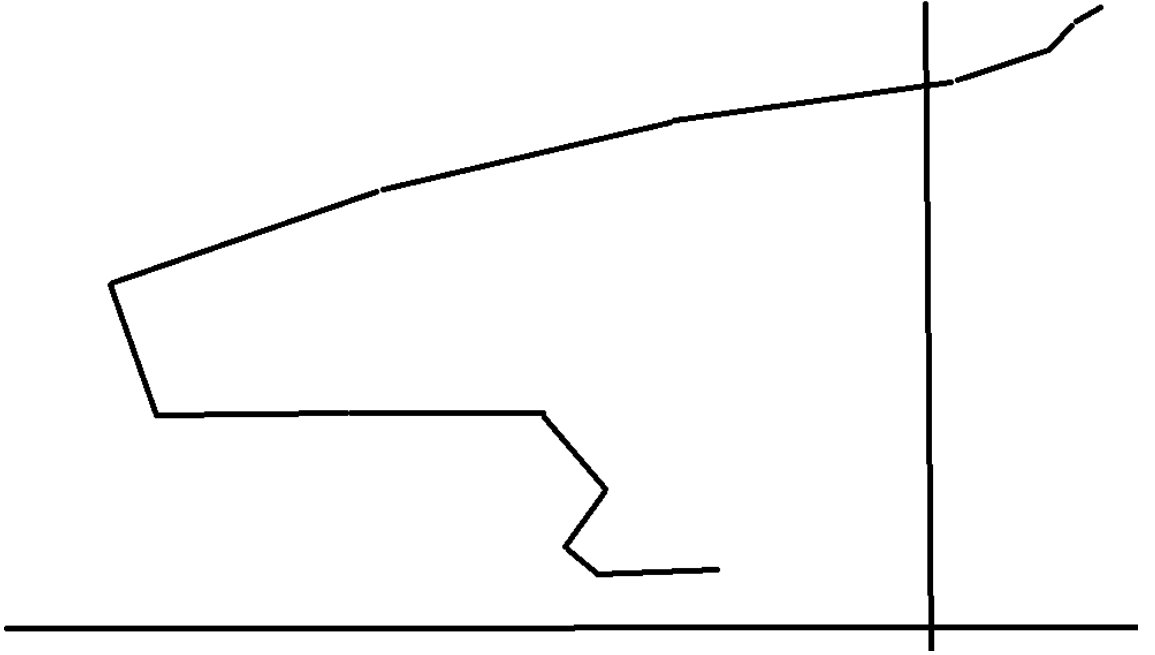
$$u^{12} = \begin{pmatrix} -0,198251 \\ 0,189448 \end{pmatrix} \rightarrow \|u^{12} - u^*\| = 0,274215$$

$$u^{13} = \begin{pmatrix} -0,121328 \\ 0,189448 \end{pmatrix} \rightarrow \|u^{13} - u^*\| = 0,224969$$

$$u^{14} = \begin{pmatrix} -0,049899 \\ 0,189448 \end{pmatrix} \rightarrow \|u^{14} - u^*\| = 0,195908$$

$$u^{15} = \begin{pmatrix} -0,016566 \\ 0,122781 \end{pmatrix} \rightarrow \|u^{15} - u^*\| = 0,123895$$

şeklinde devam eder. Her bir iterasyonda minimuma yaklaşımları noktasında her iki metoda baktığımızda pür subgradyent metot ile deflected subgradyent metot arasında gözle görülür bir fark vardır. Pür subgradyent metotla 23.adımda elde ettiğimiz yakınlığı deflected subgradyent metot ile 15.adımda elde ediyoruz. Bunun böyle olmasının sebebi deflected subgradyent metodun pür subgradyentte oluşan zikzakları ortadan kaldırmasıdır. Bunu deflected subgradyentle elde ettiğimiz grafiğe bakınca da anlayabiliriz.



Şekil 3.15. Deflected subgradyentle optimale yaklaşma

Görüldüğü üzere 1.tür zikzak davranışı deflected subgradyent sayesinde ortadan kalkmıştır. Bu da algoritmanın minimum noktaya olan yaklaşmasını hızlandırmıştır.

4. ARAŐTIRMA BULGULARI

Bu tezde subgradyent ve subdiferansiyel kavramları ifade edildikten sonra subgradyent metot verilmiŐtir. Subgradyent metodun modifiye edilmiŐ halleri olan pür subgradyent metot, deflected subgradyent metot, koŐullu subgradyent metot ve zikzak-free subgradyent metotları verilip, bu metotların uygun adım uzunluđuyla optimal noktaya yaklaŐıkları ispat edilmiŐtir. Pür subgradyent metodun bazı iterasyonlarda maruz kaldıđı birinci tür zikzak durumunun deflected subgradyent metot tarafından ortadan kaldırıldıđı görölmüŐtür. Bunun bir sonucu olarak deflected subgradyent metodun, optimum noktaya ulaŐma hızı aŐısından pür subgradyent metottan çok daha iyi olduđu, nümerik uygulamalar kısmındaki örnek yardımıyla görölmüŐtür. Yine pür subgradyent metodun bazı iterasyonlarda maruz kaldıđı ikinci tür zikzak durumunun koŐullu subgradyent metot tarafından ortadan kaldırıldıđı görölmüŐtür. Bölüm 3.3.3 deki örnek yardımıyla koŐullu subgradyent metodun bu özelliđinden dolayı optimum noktaya ulaŐma hızı pür subgradyentten çok daha iyi olduđu görölmüŐtür.

5. SONUÇ

Tezde türevlenemeyen fonksiyonların minimum noktalarının subgradyent metot yardımıyla nasıl bulunacağı verilmiştir. Metodu daha etkili kılmak için subgradyent metot üzerinde yapılmış yenilikler incelenmiştir. Yapılan bu değişikliklerle metodun aktivitesinin arttığı ve minimum noktaya daha hızlı ulaşıldığı görülmüştür. Verilen nümerik örnekler yardımıyla da bu durum açık bir şekilde ortaya konmuştur.

KAYNAKLAR

- Akgül, M., Topics in RelaxionandEllipsoidalMethods, 1984, Volume 97 of ResearchNotes in Mathematics, Pitman.
- Bayraktar, M., 2010, Analiz, TOBB Ekonomi ve Teknoloji Üniversitesi, Matematik Bölümü
- Bertsekas, D. P., 1999, Nonlinear Programming, Athena Scientific, second edition.
- Boyd, S. and Vandenberghe, L., 2004, Convex Optimization, Cambridge University Press.
- Dedetürk, M., 2007, Konveks Fonksiyonların ve Sonlu Konveks Fonksiyonların Subdiferansiyellenebilmeleri, Yüksek Lisans Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü.
- Guta, B., 2003, Subgradient Optimization Methods in Integer Programming with an Application to a Radiation, Dissertation, Vom Fachbereich Mathematik der Universität Kaiserslautern.
- Shor, N. Z., 1985, Minimization of Methods for Non-Differentiable Functions, Springer-Verlag, Berlin.

ÖZGEÇMİŞ

1987 yılında Adana ilinde doğdu. İlk, orta ve lise öğrenimini Adana'da tamamladı. Atatürk üniversitesi Eğitim Fakültesi Matematik Öğretmenliği Bölümü'nden 2010 yılında mezun oldu. Aynı yıl Atatürk Üniversitesi Fen Bilimleri Enstitüsü'nde Yüksek Lisans programına başladı.