Convex Analysis and Convex Optimization

Technical Report · July 2014

1 author:

Debabrata Ghosh
Indian Institute of Technology Kanpur

Convex Analysis and Convex Optimization

Debabrata Ghosh

July 3, 2014

ABSTRACT:

Optimization is a central theme of applied mathematics that involves minimizing or maximizing various quantities.This is an important application of the derivative tests in calculus.In addition to the first and second derivative tests of one-variable calculus, there is the powerful technique of Lagrange multipliers in several variables.In this reading project we are mainly concerned with analogues of these tests that are applicable to functions that are not differentiable.Some different hypothesis must replace differentiability and this is the notion of convexity.It turns out that many applications in economics, business, and related areas involve convex functions.Convex optimization is at the core of many of todays analysis tools for large datasets, and in particular machine learning methods.In this project,we concentrate on the theoretical underpinnings of the subject rather than it's vast applications.But it will surely help to understand the important aspect of constructing algorithms to carry out the programs involved in various applications.

In this project we will concentrate on convex analysis and convex optimization techniques in finite dimensional spaces to reach such condition so that we can easily understand the related problems in infinite dimensional space and get a better insight into it.So in this reading project we studied the duality approach to solving convex optimization problems in detail using tools in convex analysis and the theory of conjugate functions. Conditions for the duality formalism to hold were studied which require that the optimal value of the original problem vary continuously with respect to perturbations in the constraints only along feasible directions; this is sufficient to imply existence for the dual problem and no duality gap. These conditions are also posed as certain local compactness requirements on the dual feasibility set, based on a characterization of locally compact convex sets in locally convex spaces in terms of nonempty relative interiors of the corresponding polar sets.

The general approach followed during studying the project contents is as follows: 1.Identify techniques from algebra, elementary single-variable calculus, or elementary multivariable calculus that can be used to solve optimization problems. 2.Reformulate the solution geometrically. 3.Using geometry for inspiration, generalize the solution, typically to innite-dimensional vector spaces and non-Euclidean norms, and prove (algebraically) that it is still valid.

Contents

Chapter 1

Introduction

Optimization is a rich and thriving mathematical discipline.Properties of minimizers and maximizers of functions rely intimately on a wealth of techniques from mathematical analysis, including tools from calculus and its generalizations, topological notions, and more geometric ideas.The theory underlying current computational optimization techniques grows ever more sophisticated duality-based algorithms,interior point methods,and control-theoretic applications are typical examples.The powerful and elegant language of convex analysis unifies much of this theory.Hence our aim of this projest is going through a concise,accessible account of convex analysis and its applications and extensions to get the essence of optimization.

The very first question comes to our mind while starting this project is :

*WHY IS CONVEXITY SO SPECIAL IN OPTIMIZATION?

What I found as an answer to this is:

1.A convex function has no local minima that are not global.

2.A convex set has a nonempty relative interior.

3.A convex set is connected and has feasible directions at any point.

4.A nonconvex function can be convexified while maintaining the optimality of its global minima.

5.The existence of a global minimum of a convex function over a convex set is conveniently characterized in terms of directions of recession.

6.A polyhedral convex set is characterized in terms of a finite set of extreme points and extreme directions.

7.A real-valued convex function is continuous and has nice differentiability properties.

8.Closed convex cones are self-dual with respect to polarity.

9.Convex,lower semicontinuous functions are self-dual with respect to conjugacy.

Chapter 2

Preliminaries

2.1 Basic concepts

Line segment,convex set and convex hull: Let x and y be two points of R^n. The set [x,y]:={z$\in$ R^n | z=αx+(1-α)y,α $\in$R and 0$\leq$ α $\leq$1} is called closed line segment joining x and y. Also [x,y):=[x,y] {y} = {z $\in$ R^n | z=αx+(1-α)y, α $\in$ R and 0< α $\leq$1},(x,y]:=[x,y] {x}={z$\in$ R^n | z=αx+(1-α)y,α $\in$R and 0$\leq$ α <1}.The set (x,y):=]x,y[={z $\in$ R^n | z=αx+(1-α)y, α $\in$ R and 0< α <1} is called open line segment joining x and y.
Definition 1.1.1. Let u,v $\in$ V . Then the set of all convex combinations of u and v is the set of points {w_λ $\in$ V :w_λ = (1-λ)u+λv, 0$\leq$ λ $\leq$1}.
Definition 1.1.2. Let K$\subset$V.Then the set K is said to be convex provided that given two points u,v $\in$ K the set in 1.1.1. is a subset of K i.e. a subset K of V is said to be convex if it contains every closed line segment joining two points of K. e.g. A disk with center (0,0) and radius c is a convex subset of R^2.
1.1.3.Proposition: If C$\in$ R^n is convex,the cl(C), the closure of C, is also convex.
Proof: Suppose x,y $\in$ cl(C). Then there exist sequences $\{x_n\}_{n=1}^{\infty}$ $\{y_n\}_{n=1}^{\infty}$ in C such that x_n $\rightarrow$ x and y_n $\rightarrow$ y as n$\rightarrow$ ∞. For some λ; 0$\leq$ λ $\leq$1, define z_n :=(1-λ)x_n+λ y_n.Then, by convexity of C,z_n $\in$C. Moreover z_n $\rightarrow$ (1-λ)x+λy as n$\rightarrow$ ∞.Hence this latter point lies in cl(C).
1.1.4.Proposition: The intersection of any number of convex sets is convex.
Proof: Let $\{K_\alpha\}_{\alpha\in A}$ be a family of convex sets, and let K:=$\cap_{\alpha\in A}$ K_α.Then for any x,y$\in$K by definition of the intersection of a family of sets,x,y$\in$ K_α $\forall$

$\alpha \in$A and each of these sets is convex. Hence for any $\alpha \in$A and$\lambda \in$[0,1],(1-λ)x+λy$\in K_\alpha$.Hence,(1-λ)x+λy$\in$K.
1.1.5.Proposition: Let K be a convex set and let λ_1,λ_2,...,$\lambda_p \geq$0 and $sum\_i = 1^p$ $\lambda_i = 1$.If x_1,x_2,...,$x_p \in$ K then, $sum\_i = 1^p$ λ_i $x_i \in$ K.
Proof: We prove the result by induction. Since K is convex,the result is true, trivially, for p = 1 and by definition for p = 2. Suppose that the proposition is true for p = r (induction hypothesis!) and consider the convex combination λ_1 x_1+λ_2 x_2+....+λ_{r+1} x_{r+1}.Define $\Lambda := sum\_i = 1^r$ λ_i.Then since 1-Λ=$sum\_i = 1^{r+1}$ λ_i-$sum\_i = 1^r$ λ_i=λ_{r+1},we have ($sum\_i = 1^r$ λ_i x_i)+λ_{r+1} x_{r+1}=Λ($sum\_i = 1^r$ $\frac{\lambda_i}{\Lambda}$ x_i)+(1-Λ)x_{r+1}.Note that $sum\_i = 1^r$ $\frac{\lambda_i}{\Lambda}$=1 and so, by the induction hypothesis,$sum\_i = 1^r$ $\frac{\lambda_i}{\Lambda}$ $x_i \in$K.Since $x_{r+1} \in$K it follows that the right hand side is a convex combination of two points of K and hence lies in K.
Definition 1.1.6.The convex hull of a set C is the intersection of all convex sets which contain the set C. We denote the convex hull by conv(C).
•Let S $\subset$ V . Then the set of all convex combinations of points of the set S is exactly conv(S).
1.1.7.Caratheodory's theorem: Let S be a subset of R_n. Then every element of conv(S) can be represented as a convex combination of no more than (n+1) elements of S.
Proof: Let x$\in$conv(S).Then we can represent x as $\sum\limits_{i=1}^{m}$ α_i x_i for some vectors $x_i \in$S and scalars $\alpha_i \geq$0 with $\sum\limits_{i=1}^{m}$ α_i=1.Let us suppose that m is the minimal number of vectors for which such a representation of x is possible. In particular, this means that for all i = 1,2,...,m, α_i >0.If we were to have m>n+1,then the vectors x_i-x,i = 1,2,...,m, must be linearly dependent since there are more vectors in this set than the dimension of the space. It follows that there exist scalars λ_2,....,λ_m at least one of which is positive such that $\sum\limits_{i=2}^{m}$ λ_i(x_i-x)=0.Let μ_1 :=-$\sum\limits_{i=2}^{m}$ λ_i, and μ_1 :=λ_i,i=2,3,...,m.Then $\sum\limits_{i=1}^{m}$ μ_i x_i=0,and $\sum\limits_{i=1}^{m}$ μ_i=0,while at least one of the scalars μ_2,μ_3,...,μ_m is positive. The strategy for completing the proof is to produce a convex combination of vectors that represents x and which has fewer than m summands which would then contradict our choice of m as the minimal number of non-zero elements in the representation. To this end,$\hat{\alpha}_i$:=α_i-$\hat{\gamma}$ μ_i,i=1,...,m, where $\hat{\gamma}$ >0 is the largest γ such that α_i-γ $\mu_i \geq$0 $\forall$i. Then, since $\sum\limits_{i=1}^{m}$ μ_i x_i=0 we have $\sum\limits_{i=1}^{m}$ $\hat{\alpha}_i$ x_i=$\sum\limits_{i=1}^{m}$(α_i-$\hat{\gamma}$

$\mu_i)x_i$= $\sum\limits_{i=1}^{m}$ α_i x_i=x.Now,the $\hat{\alpha}_i$ $\geq$0, at least one is zero, and $\sum\limits_{i=1}^{m}$ $\hat{\alpha}_i$=$\sum\limits_{i=1}^{m}$ α_i-$\hat{\gamma}$ $\sum\limits_{i=1}^{m}$ μ_i=$\sum\limits_{i=1}^{m}$ α_i=1. This gives a convex representation of x in terms of fewer than m points in S which contradicts the assumption of minimality of m.

2.2 Hyperplane

1.2.1.Definition: A hyperplane is an affine set of codimension 1. Thus a hyperplane in R^n has dimension n-1.
Hyperplanes are rather special affine sets, and they serve to split the whole space into two pieces. 1.2.2.Proposition: A subset H of R_n is a hyperplane if and only if there is a nonzero vector h$\in$ R^n and a scalar α $\in$R so that H=$\{$x$\in$ R^n : $<$x,h$>$=$\alpha\}$.

2.3 Relative interior

1.3.1.Definition: If A is a convex subset of R^n, then the relative interior of A, denoted ri(A), is the interior of A relative to aff(A). That is, a$\in$ ri(A) if and only if there is an ϵ >0 so that B_ϵ(a)$\cap$aff(A)$\subset$A.
1.3.2.Lemma: Suppose that A and B are convex subsets of R^n with A$\subset$B.If aff(A) = aff(B), then ri(A)$\subset$ri(B).
Proof:If a$\in$ri(A),then there is an ϵ >0 with B_ϵ(a)$\cap$aff(A)$\subset$A.Because aff(B)=aff(A), we have B_ϵ(a)$\cap$aff(B)$\subset$A $\subset$B, so a$\in$ri(B).
1.3.3.Lemma: Let S=$\{s_1$,....,$s_r\}$ be a finite subset of R^n. Then ri(conv(S))=$\{\sum\limits_{i=1}^{r}$ λ_i s_i : λ_i $\in$(0,1), $\sum\limits_{i=1}^{r}$ λ_i=1$\}$.
1.3.4.Accessibility lemma: Suppose that A is a convex subset of R^n.If a$\in$ ri(A) and b$\in$cl(A), then [a,b) is contained in ri(A).
Proof: Let ϵ >0 be given so that B_ϵ(a)$\cap$aff(A)$\subset$A.We need to show that λa+(1-λ)b$\in$ri(A) for λ $\in$ (0,1).Since b$\in$cl(A), pick c$\in$A so that x=b-c$\in$L(A) satisfies ||x|| $<$ ϵ λ/(2-2λ).Suppose that z$\in$L(A) with ||z|| $<$ ϵ/2.Then λa+(1-λ)b+λz=λa+(1-λ)(c+x)+λz=λ(a+z+(1-λ)x/λ)+(1-λ)c.Since||z+(1-λ)x/λ || $\leq$ ||z||+(1-λ)||x||/λ $<$ ϵ/2+ϵ/2=ϵ,this is a vector in B_ϵ(0)$\cap$L(A).Hence d=a+z+(1-λ)x/λ $\in$A.Thus λa+(1-λ)b+λz=λd+(1-λ)c also lies in A.So λa+(1-λ)b belongs to ri(A).Hence the result.

2.4 Separation theorems

1.4.1.Convex projection theorem: Let C be a nonempty closed convex subset of R^n.For each point x$\in$ R^n, there is a unique point P_C(x) in C which is closest to x.The point P_C(x) is characterized by $<$x-P_C(x),a-P_C(x)$>$ $\leq$0 $\forall$ a$\in$C. Moreover, || P_C(x)-P_C(y)|| $\leq$ ||x-y|| $\forall$ x,y$\in$ R^n.
Proof:Pick any vector v in C,and let R=||x-v||.Then 0$\leq$ inf{||x-a|| : a$\in$C} $\leq$R.The closest point to x in C must belong to C$\cap$ B_R(x).This is the intersection of two closed convex sets, and so is closed and convex. Moreover,C$\cap$ B_R(x) is bounded and thus compact by the HeineBorel Theorem.Then we consider the continuous function on C$\cap$ B_R(x) given by f(a)=||x-a||.By the Extreme Value Theorem there is a point u $\in$ C at which f takes its minimum a closest point.To see that u is unique, we need to use convexity. Suppose that there is another vector t$\in$C with ||x-t||=||x-u||.We will show that b = $\frac{1}{2}$(u+t) is closer, contradicting the choice of u as a closest point.
Now we consider the triangle with vertices x,u and t.As ||x-t||=||x-u||,this triangle is isosceles. As ||b-t||=||b-u||,the point b is the foot of the perpendicular dropped from x to [u,t] and therefore ||x-b|| $<$ ||x-a_i ||,where a_i=u/t.Indeed $\angle$xba_i is a right angle, and the Pythagorean Theorem shows that ||x-a_i || ^{2}=||x-b|| ^{2}+$\frac{1}{4}$ ||u-t|| ^{2}.
It now makes sense to define the function P_C by setting it to be this unique closest point to x.Let a $\neq$ P_C(x) be any other point in C.Then for 0$<$ λ $\leq$1, ||x-u|| ^{2} $<$ ||x-(λa+(1-λu)|| ^{2}=||(x-u)-λ(a-u)|| ^{2}=||x-u|| ^{2}-2λ $<$x-u,a-u$>$+λ^2 ||a-u|| ^{2} so that $<$x-u,a-u$>$ $\leq$ $\frac{\lambda}{2}$ ||a-u|| ^{2}.Let λ tend to 0 to obtain $<$x-u,a-u$>$ $\leq$0.This argument is reversible. If u is any point such that $<$x-u,a-u$>$ $\leq$0 $\forall$a$\in$C,then ||x-a|| ^{2}=||x-u|| ^{2}-2$<$x-u,a-u$>$+||a-u|| ^{2} $\geq$ ||x-u|| ^{2}+||a-u|| ^{2}.Hence u=P_C(x) is the unique closest point.
Let x and y be points in R^n. Apply the inequality once for each of x and y: $<$x-P_C(x),P_C(y)-P_C(x)$>$ $\leq$0 and $<$ P_C(y)-y,P_C(y)-P_C(x)$>$=$<$y-P_C(y),P_C(x)-P_C(y)$>$ $\leq$0.Adding yields $<$(x-y)+(P_C(y)-P_C(x)),P_C(y)-P_C(x)$>$ $\leq$0.Hence || P_C(y)-P_C(x)|| ^{2} $\leq$ $<$y-x,P_C(y)-P_C(x)$>$ $\leq$ ||y-x|| || P_C(y)-P_C(x)|| by the Schwarz inequality.Therefore, || P_C(y)-P_C(x)|| $\leq$ ||x-y||. Hence proved.
1.4.2.Definition: Let C be a closed convex subset of R^n and b $\in$ rbd(C). A supporting hyperplane to C at b is a hyperplane H so that b $\in$ H and C is contained in one of the closed half-spaces determined by H.If aff(C) is a proper subspace of R^n, then it is contained in a hyperplane that supports C. We regard this as a pathological situation and call H a nontrivial supporting hyperplane if it does not contain aff(C).

1.4.3.Support theorem: Let C be a convex subset of R^n and b $\in$ rbd(C). Then there is a nontrivial supporting hyperplane to C at b.

Chapter 3

Inequality Constraints for Optimization

3.1 Minimality Conditions

Initially we are considering the problem of minimizing a function f:C $\rightarrow$ R on a set C in E where E is an arbitrary Euclidean space.We say a point x* in C is a local minimizer of f on C if f(x)$\geq$ f(x*)for all points x close to x* in C. The directional derivative of the function f at x* in a direction d$\in$E is f'(x*;d)=$limit_{t\rightarrow 0}$ $\frac{f(x*+td)-f(x*)}{t}$, if limit exists.Now if f'(x*;d)=<a,d> for some element a of E, then we say f is (G$\hat{a}$teaux) differentiable at x*,with (G$\hat{a}$teaux)derivative $\bigtriangledown f(x*) = a$. And a convex cone used frequently in optimization, normal cone to a convex set C at a point x*$\in$C, N_c(x*) is the set of vectors d in E such that $< d, x - x* >$ $\leq$0 $\forall$ x$\in$C.
3.1.1 First order necessary condition : Suppose that C is a convex set in E, and a point x* in C is a local minimizer of the function f:C$\rightarrow$R. Then for every point x in C, the directional derivative, if exists satisfies f'(x*;x-x*)$\geq$0 i.e. if f is differentiable at x* then the condition -$\bigtriangledown$f(x*)$\in$ N_C(x*) holds.
Proof: If any point x $\in$ C satisfies f'(x*;x-x*)<0, then for all real t>0 f(x*+t(x-x*))<f(x*), but this contradicts the fact that x* is the local minimum of the function.Now suppose -$\bigtriangledown$f(x*)$\notin$ N_C(x*).Now for a vector d and scalar α we have sup { <s+n,d>: s $\in$ $\bigtriangledown$ f(x*),n$\in$ N_C(x*)} $\leq$ α <<0,d>=0. It must be the case that <n,d> $\leq$0 for n$\in$ N_C(x*),for if <n,d>> 0,then <s+λn,d>=<s,d>+λ <n,d>> 0 for very large λ.Therefore,d belongs to N_C(x*)^o=T_C(x*). Now take n=0 and by simple exercise get f'(x*;d)=sup{

$<s,d> : s \in \bigtriangledown f(x^*)\} \leq \alpha < 0$. Which is a contradiction and hence the result.
3.1.2.First order sufficient condition : We suppose that the set C⊂E is convex and the function f:C→R is convex. Then for any points x* and x in C, the directional derivative f'(x*;x-x*)exists in $(-\infty,+\infty]$ and if f'(x*;x-x*)≥0 holds ∀x∈C (i.e. -▽f(x*)∈ N_C(x*) holds), then x* is a global minimizer of f on C.
3.1.3.First order conditions for linear constraints : For a covex set C⊂E and a function f:C→R, a liner map A : E→Y(a Euclidean space) and a point b in Y, if we consider the optimization problem inf{f(x)|x∈C,Ax=b}.
Assuming the point x*∈int C satisfies Ax*=b, we get (a) If x* is a local minimizer for this problem and f is differentiable at x* then ▽f(x*)∈ A*Y and (b) Conversely, if ▽f(x*)∈ A*Y and f is convex then x* is a global minimizer for the problem.
Here the element y∈Y satisfying ▽f(x*)∈ A*Y is called a Lagrange multiplier.
3.1.4.Basic separation theorem : If the set C⊂E is closed and convex and the point y∉C, then ∃ a real b and a nonzero element a of E such that $<a,y> > b \geq <a,x>$ ∀x∈C.
Proof: Let h=P_A(y).As y∉C,a=y-h is a nonzero vector. We defined b=$<a.h>$.Now by Convex Projection theorem(2..),if z∈C, then $<z-h,y-h> \leq 0$. Rewriting this, we obtain $<a,z> \leq <a,h>=b$.On the other hand, $<y-h,y-h>=||a||^2$, which implies that $<a,y> > <a,h>=b$.

3.2 Theorems of the alternative

The results discussed below, the first of which appeared in 1902, are concerned with the existence of non-negative solutions of the linear system (I)Ax=b,(II)x≥0, where A is an m×n matrix with real entries, x∈ R^n,b∈ R^m.
3.2.1.The Farkas-Minkowski theorem : A necessary and sufficient condition that (I) and (II) has a solution is that, ∀ y∈ R^m with the property that A^Ty≥0,we have $<b,y> \geq 0$. Or alternatively, the system (I) and (II) has a solution if and only if the system A^Ty≥0,$<b,y> < 0$, has no soluion.
Proof(of the first statement): Firstly it is easy to see that the condition is necessary. Indeed, if the system (I) and (II) has a non-negative solution x*≥0,then,∀y∈ R^m such that A^Ty≥0,we have $<y,b>=<y,Ax^*>=<A^Ty,x^*> \geq 0$,since all ters in the inner product are products of non-negative real num-

bers.
To see that the condition is sufficient we assume that the system (I) and (II) has no solution and show that there is some vector y such that A^Ty$\geq$0 and <b,y> <0.In order to do this, we will apply the basic separation theorem.Consider the set S :={z$\in$ R^m |z=Ax,x$\geq$0}.
Clearly this set is convex and closed.To say that the system (I) and (II) has no solution says that b$\notin$S.Observe that the set {b} is closed, bounded and convex.Hence, by the strict separation theorem, $\exists$ a vector a$\in$ R^m,a$\neq$0 and a scalar α,such that <a,y> < α $\leq$ <a,b> $\forall$y $\in$S.Since 0$\in$S we must have α >0.Hence <a,b> >0.Likewise, <a,Ax> $\leq$ α $\forall$x$\geq$0.From this it follows that A^Ta$\leq$0.Indeed,if the vector w=A^Ta were to have a positive component,say w_j then we can take $\hat{x}$=(0,0,...,0,M,0,...,0)^T where M>0 appears in the jth position.Then certainly $\hat{x}$ $\geq$0 and < A^Ta,$\hat{x}$ >=w_jM, which can be made as large as desired by choosing M sufficiently large.In particular,if we choose M> α / w_j then the bound <a,Ax> $\leq$ α is violated.This shows that A^Ta$\leq$0 and completes the proof.Indeed, we simply set y= -a to get the required result.
3.2.2.Gordan's theorem : Let A be an m$\times$n real matrix, x$\in$ R^n and y$\in$ R^m. Then one and only one of the following conditions holds: 1.There exists an x$\in$ R^n such that Ax<0; 2.There exists a y$\in$ R^m,y$\neq$0 such that A^Ty=0 and y$\geq$0.
Proof : Let $\hat{e}$=$(1, 1, ..., 1)^T$ $\in$ R^m. Then the first condition is equivalent to saying that Ax$\leq$-$\hat{e}$ has a solution.By (3.2.1) this is equivalent to the statement that if y$\geq$0 and A^Ty=0 then <-y,$\hat{e}$ > $\geq$0.Hence there is no y$\neq$0 such that A^Ty=0 and y$\geq$0.
Conversely, if there is a y$\neq$0 such that A^Ty=0 and y$\geq$0,then the condition of the Farkas-Minkowski theorem does not hold and hence Ax<0 has no solution.

3.3 Max-function

We define max-function as g(x)=$max_{i=0,1,....,m}$ {g_i(x)}.It illustrates a pervasive analytic idea in optimization : nonsmoothness.Even if the functions g_0,g_1,....,g_m are smooth, g may not be, and hence the gradient may no longer be a useful notion. 3.3.1.Directional derivatives of max-functions : Let x* be a point in the interior of a set C$\subset$E.We assume that continuous functions g_0,g_1,....,g_m :C$\to$R are differentiable at x*, that g is the max-function and define the index set K={i| g_i(x*)=g(x*)}.Then for all directions d in E, the directional derivative of g is given by g'(x*;d)= $max_{i\in K}$ { $<\bigtriangledown$ g_i (x*),d $>$ }.

Proof : By continuity we can assume, without loss of generality, K={0,1,...,m}: those g_i not attaining the maximum will not affect g'(x*;d).Now for each i, we have the inequality $lim_{t\to 0}$inf$\frac{g(x*+td)-g(x*)}{t}$ $\geq$ $lim_{t\to 0}$ $\frac{g_i(x*+td)-g_i(x*)}{t}$=$<\bigtriangledown$ g_i(x*),d$>$. Suppose $lim_{t\to 0}$sup$\frac{g(x*+td)-g(x*)}{t}$ $>$ max_i {$<\bigtriangledown$ g_i(x*),d$>$}.Then some real sequence t_k $\to$0 and real ϵ $>$0 satisfy $\frac{g(x*+t_kd-g(x*)}{t_k}$ $\geq$ max_i{$<$ $\bigtriangledown$ g_i(x*),d$>$+ϵ $\forall$k$\in$N(where N is the sequence of natural numbers).We can now select a subsequence R of N and a fixed index j so that all integers k in R satisfy g(x*+t_kd)=g_j(x*+t_kd).In the limit we get the contradiction $<$ $\bigtriangledown$ g_j(x*),d$>$ $\geq$ max_i {$<\bigtriangledown$ g_i(x*),d$>$}+ϵ. Hence $lim_{t\to 0}$sup$\frac{g(x*+td)-g(x*)}{t}$ $\leq$ max_i {$<\bigtriangledown$ g_i(x*),d$>$}, and the result follows.

3.3.2.Here in this project we mostly consider optimization problems of the form : inf f(x); subject to g_i(x)$\leq$0,for i$\in$I, h_j(x)=0,for j$\in$J,x$\in$C, where C$\subset$E,I and J are finite index sets, and all the functions concerned are continuous from C to R.A point x in C is feasible if it satisfies the constraints mentioned above, and the set of all such points is called the feasible region.

3.3.3.Here we deal with the following inequality constrained problem : inf f(x); subject to g_i(x) $\leq$0 for i=1,2,....,m; x$\in$C.For a feasible point x* we define the active set I(x*)={i| g_i(x*)=0}.Here, we assume x*$\in$intC and call a vector $\lambda\in R^m_+$ a Lagrange multiplier vector for x* if x* is a critical point of the Lagrangian L(x;λ)=f(x)+$\sum\limits_{i=1}^{m}$ λ_i g_i(x) and complementary slackness holds: λ_i=0 for indices i$\notin$I(x*).

Fritz John conditions : Suppose the previous problem has a local minimizer x*$\in$intC.If the functions f,g_i (i$\in$I(x*)) are differentiable at x* then $\exists$ λ_0,$\lambda_i\in R_+$, (i$\in$I(x*)), not all zero, satisfying λ_0 $\bigtriangledown$f(x*)+$\sum_{i\in I(x*)}$ λ_i $\bigtriangledown$ g_i(x*)=0.

Proof : Here we consider the function g(x)=max{f(x)-f(x*), g_i(x)(i$\in$I(x*))}.Since x* is a local minimizer for the previous problem, it is a local minimizer of the function g, so all directions d$\in$E satisfy the inequality g'(x*;d)=max{$<$ $\bigtriangledown$f(x*),d$>$,$<\bigtriangledown$ g_i(x*),d$>$ (i$\in$I(x*))} $\geq$0, by the First order necessary condition and 3.3.1. Thus the system $<\bigtriangledown$f(x*),d$>$ $<$0, $<\bigtriangledown$ g_i(x*),d$>$ $<$0 (i$\in$I(x*)) has no solution, and the result follows by Gordan's theorem.

The Mangasarian-Fromovitz constrait qualification : There is a direction d in E satisfying $<\bigtriangledown$ g_i(x*),d$>$ $<$0 for all indices i in the active set I(x*).

3.3.5.Karush-Kuhn-Tucker conditions : Suppose the previously mentioned

problem has a local minimizer x* in intC.If the functions f,g_i (i∈I(x*)) are differentiable at x*, and if the Mangasarian-Fromovitz constraint qualification holds, then ∃ a Lagrange multiplier vector for x*.
Proof : By the trivial implication in Gordan's theorem, the constraint qualification ensures $\lambda_0 \neq 0$ in the Fritz John conditions and hence the result.

Chapter 4

Fenchel Duality

4.1 Subgradients

4.1.1.Sublinearity : A function f:E$\rightarrow(-\infty,+\infty]$ is sublinear if and only if it is positively homogeneous and subadditive. 4.1.2.Sublinearity of directional derivative : If the function f:E$\rightarrow(-\infty,+\infty]$ is convex then for any point x* in core(dom f) the directional derivative f'(x*;·) is everywhere finite and sublinear.
4.1.3.Subgradients : Suppose that C is a convex subset of E,a$\in$C, and f is a convex function on C. A subgradient of f at a is a vector v$\in$E such that f(x)$\geq$f(a)+$<$x-a,v$>$ $\forall$x$\in$C.
The set of all subgradients of f at a is the subdifferential and is denoted by δf(a).
The gradient of f at a, denoted $\bigtriangledown$f(a), is the n-tuple of partial derivatives: ($\frac{\delta}{\delta x_1}$f(a),.....,$\frac{\delta}{\delta x_n}$f(a)).The differential of f at a is the hyperplane given by those points (x,t)$\in$E$\times$R so that t=f(a)+$<$x-a,$\bigtriangledown$f(a)$>$.
4.1.4.Subgradients at optimality : For any proper function f:E$\rightarrow(-\infty,+\infty]$, the point x* is a (global) minimizer of f if and only if the condition 0$\in$ δf(x*) holds. Proof : By definition,0$\in$ δf(x*) if and only if f(x)$\geq$f(x*)+$<$x-x*,0$>$=f(x*) $\forall$x$\in$C. 4.1.5.Subgradients and directional derivatives: If the function f:E$\rightarrow(-\infty,+\infty]$ is convex and the point x* lies in dom f, then an element ϕ of E is a subgradient of f at x* if and only if it satisfies $<\phi,\cdot>$ $\leq$f'(x*;·). Proof : Directly follows from the definition of subgradients. 4.1.6. Now we will use the following lemma to prove the max formula:
Suppose that the function s:E$\rightarrow(-\infty,+\infty]$ is sublinear, and that the point x*

lies in core(dom s). Then the function t($\cdot$)=s'(x*;$\cdot$) satisfies the conditions 1.t(λx*)=λs(x*) $\forall$real λ, 2.t$\leq$s, and 3.lin t $\supset$lin s+span {x*}. 4.1.7.Max formula : If the function f:E$\rightarrow(-\infty,+\infty]$ is convex then any point x* in core(dom f) and any direction d$\in$E satisfy f'(x;d)=max{$<\phi$,d> | $\phi \in \delta$f(x*)}.Particularly, the subdifferential δf(x*) is nonempty.

Proof : According to 4.1.5., we simply have to show that for any fixed d in E there is a subgradient ϕ satisfying $<\phi$,d>=f'(x*;d). We select a basis {e_1,e_2,...e_n} for E with e_1=d if d is nonzero. Now we define a sequence of functions p_0,p_1,...,p_n recursively by p_0(*)=f'(x*;*), and p_k=p'_{k-1}(e_k;*), for k=1,2,....,n. We will show that p_n(*) is the required subgradient.

First note that,each p_k is everywhere finite and sublinear.Now by 4.1.6. we know lin p_k $\subset$lin p_{k-1}+span{e_k}, for k=1,2,...,n, so p_n is linear. Thus there is an element ϕ of E satisfying $<\phi$,*>=p_n(*).Also by 4.1.6. $p_n \leq p_{n-1} \leq$.....p_0, so certainly, by 4.1.5., any point x in E satisfies p_n(x-x*) $\leq$ p_0(x-x*)=f'(x*;x-x*)$\leq$f(x)-f(x*).Thus ϕ is a subgradient.If d is 0 then we have p_n(0)=0=f'(x*;0).Finally, if d is nonzero then by 4.1.6 again we see p_n(d) $\leq$ p_0(d)=p_0(e_1)=-p'_0(e_1;-e_1)=-p_1(-e_1)=-p_1(-d) $\leq$-p_n(-d)=p_n(d), whence p_n(d)=p_0(d)=f'(x*;d).

4.1.8.Differentiability of convex functions : Suppose that the function f:E$\rightarrow(-\infty,+\infty]$ is convex, and that the point x* lies in core(dom f). Then f is Gâteaux differentiable at x* exactly when f has a unique subgradient at x*.

4.2 Value function

4.2.1.Lagrangian sufficient conditions : If the point x* is feasible for the convex program

• inf f(x); subject to g_i(x) $\leq$ 0, for i=1,2,....,m; x$\in$E

and there is a Lagrange multiplier vector, then x* is optimal.

4.2.2.We perturb the previous convex problem and analyze the resulting value function v:r^m $\rightarrow[-\infty,+\infty]$, defined by the equation

v(b)=inf{f(x)|g(x)$\leq$b}.

Also epigraph of a function h:E$\rightarrow[-\infty,+\infty]$ is defined by epi(h)={(y,r)$\in$E$\times$R |h(y)$\leq$r}.

Lemma:If the function h:E$\rightarrow[-\infty,+\infty]$ is convex and some point $\hat{y}$ in core(dom h) satisfies h($\hat{y}$)>-∞,then h never takes the value -∞.

Proof : Suppose some point y in E satisfies h(y)=-∞.Since $\hat{y}$ lies in core(dom h) there is a real t>0 with $\hat{y}$+t($\hat{y}$-y) in dom(h), and hence a real r with

($\hat{y}$+t($\hat{y}$-y),r) in epi(h). Now for any real s,(y,s) lies in epi(h), so we get ($\hat{y}$,$\frac{r+ts}{1+t}$)=$\frac{1}{1+t}$($\hat{y}$+t($\hat{y}$-y),r)+$\frac{t}{1+t}$(y,s)∈epi(h), and now letting s→-∞ gives a contradiction.Hence the result.
4.2.3.Now the Slater constraint qualification for the previous convex problem is as follows: There exists $\hat{x}$ in dom f with $g_i(\hat{x})$ <0 for i=1,2,...,m.
Lagrangian necessary conditions : Suppose that the point x* in dom (f) is optimal for the convex program mentioned in the beginning of this section , and that the Slater condition holds. Then there is a Lagrange multiplier vector for x*.
Proof : According to the definition of value function v, certainly v(0)>-∞, and the Slater condition shows 0∈core(dom v),so from the previous lemma v never takes the value -∞.We now deduce the existence of a subgradient -λ of v at 0,by the Max formula.Any vector b in R_+^m obviously satisfies g(x*)≤b,whence the inequality f(x*)=v(0)≤v(b)+λ^Tb.Hence λ lies in R_+^m. Furthermore, any point x in dom f clearly satisfies f(x)≥v(g(x))≥v(0)-λ^Tg(x)=f(x*)-λ^Tg(x).The case x=x*, using the inequalities λ ≥0 and g(x*)≤0,shows λ^Tg(x)=0, which yields the complementary slackness conditions. Finally, all points x in dom f must satisfy f(x)+λ^Tg(x)≥f(x*)=f(x*)+λ^Tg(x*) and hence the result.

4.3 Fenchel conjugate

The Fenchel conjugate of a function h : E→[-∞,+∞] is the function h* : E→[-∞,+∞] defined by h*(ϕ)=$sup_{x\in E}$ {< ϕ,x>-h(x)}.
The function h* is convex and if the domain of h is nonempty then h* never takes the value -∞.The conjugacy operation is order-reversing: for functions f, g : E→[-∞,+∞],the inequality f≥g implies f*≤g*.
4.3.1.Fenchel-Young inequality: Any points ϕ in E and x in the domain of a function h : E→[-∞,+∞] satisfy the inequality h(x)+h*(ϕ)≥ < ϕ,x>. Equality holds if and only if $\phi \in \partial$h(x).
Proof : < ϕ,x>-h(x) ≤ h*(ϕ)(from the definition of Fenchel conjugate) ⟶ h(x)+h*(ϕ)≥ < ϕ,x>. If ϕ is a subgradient of h at x then: h(x)+< ϕ,y-x> ≤ h(y) ⟶ h(x)-< ϕ,x> ≤h(y)-< ϕ,y> ⟶ h(x)-< ϕ,x> ≤ -h*(ϕ) ⟶ h(x)+h*(ϕ) ≤ < ϕ,x> Therefore h(x)+h*(ϕ)= < ϕ,x>.
The Fenchel-Young Inequality has a geometric interpretation if we view the graph of h(x) as a surfaceover the x hyperplane in the z-direction. We can rearrange the inequality to become: < ϕ,x>-h*(ϕ) ≤ h(x).The inequality then

states that out of all hyperplanes with slope vector ϕ the greatest z-intercept attainable while still remaining below the surface is -h*(ϕ).

4.3.2.Fenchel duality and convex calculus: For given functions f : E→[-∞,+∞] and g : Y→[-∞,+∞] and a linear map A : E → Y, let p,d ∈ [-∞,+∞] be primal and dual values defined respectively by the optimization problems 1. p = $inf_{x \in E}$ {f(x)+g(Ax)} and 2. d = $sup_{\phi \in Y}$ {-f*(A*ϕ)-g*(-ϕ)}.These values satisfy the weak duality inequality p≥d.

If furthermore f and g are convex and satisfy the condition 0∈core (dom g Adom f), or the stronger condition Adom f∩cont g ≠ ∅, then the values are equal i.e. p=d, and the supremum in the dual problem is attained if finite.

Proof : The weak duality inequality follows immediately from the Fenchel-Young inequality. To prove equality we define an optimal value function h : Y→[-∞,+∞] by h(u)=$inf_{x \in E}${f(x)+g(Ax+u)}.It can be easly checked that h is convex, and domh = dom gAdom f. If p is -∞ there is nothing to prove, while if first condition holds and p is finite then by 4.2.2. and the Max formula we can say there is a subgradient $\phi \in \partial$h(0). Hence we deduce h(0) ≤ h(u)+< ϕ,u> ∀ u∈Y,

≤f(x)+g(Ax+u)+< ϕ,u>, ∀ u∈Y,x∈E,

={f(x)-<A*ϕ,x>}+{g(Ax+u)-<-ϕ,Ax+u>}.

Taking the infimum over all points u, and then over all points x gives the inequalities h(0)≤-f*(A*ϕ)-g*(-ϕ)≤d≤p=h(0).Thus ϕ attains the supremum in the given problem, and hence p=d.

4.3.3.Fenchel duality for linear constraints: Given any function f : E→(-∞,+∞], any linear map A : E → Y, and any element b of Y, the weak duality inequality $inf_{x \in E}$ {f(x)|Ax=b}≥ $sup_{\phi \in Y}$ {<b,ϕ >-f*(A*ϕ)} holds.If f is convex and b belongs to core (Adom f) then equality holds, and the supremum is attained when finite.

4.3.4.Cones: A wonderful application of the Fenchel duality circle of ideas is the calculation of polar cones. The (negative) polar cone of the set K ⊂ E is the convex cone K^{-}={ϕ ∈ E| < ϕ,x> ≤0, ∀ x∈K}, and the cone K^{--} is called the bipolar.

•Self dual cones : $(R^n_+)^-$ = -R^n_+, and $(S^n_+)^-$ = -S^n_+.

•Bipolar cone: The bipolar cone of any nonempty set K⊂E is given by K^{--}= cl(conv(R_+K)).

Pointed cones: A closed convex cone K⊂E is pointed if and only if there is an element y of E for which the set C={x∈K| <x,y>=1} is compact and generates K i.e. K=R+C.

Chapter 5

Convex Analysis

5.1 Continuity of convex functions

For a real K$\geq$0,we say that a function f : E$\rightarrow$(-∞,+∞] is Lipschitz (with constant K) on a subset C of dom f if |f(x)-f(y)| $\leq$ L||x-y|| for any points x and y in C.If f is Lipschitz on a neighbourhood of a point z then we say that f is locally Lipschitz around z. 5.1.1.Local boundedness : Let f : E$\rightarrow$(-∞,+∞] be a convex function. Then f is locally Lipschitz around a point z in its domain if and only if it is bounded above on a neighbourhood of z.
Proof : Without loss of generality we take z=0,f(0)=0, and assume f$\leq$1 on 2B and we shall deduce f is Lipschitz on B.Notice first the bound f$\leq$-1 on 2B, since convexity implies f(-x)$\leq$-f(x) on 2B.Now for any distinct points x and y in B,define α=||y-x|| and fix a point w=y+α^{-1}(y-x), which lies in 2B.By convexity we obtain f(y)-f(x)$\leq$ $\frac{1}{1+\alpha}$f(x)+$\frac{\alpha}{1+\alpha}$f(w)-f(x)$\leq$ $\frac{2\alpha}{1+\alpha}$ $\leq$ 2||y-x||, and the result follows,since x and y can be interchanged according to our wish.
5.1.2.Lemma : Let $\triangle$ be the simplex $\{$x$\in$ R^n_+ $\mid$ $\sum x_i$ $\leq$1$\}$.If the function g:$\triangle$ $\rightarrow$R is convex then it is continuous on int$\triangle$.
Proof : From 5.1.1. we can say that we just need to show g is bounded above on $\triangle$.But any point in $\triangle$ satisfies g(x)=g($\sum\limits_{i=1}^{n}$ x_i e_i+(1-$\sum x_i$)0)$\leq$ $\sum\limits_{i=1}^{n}$ x_ig(e_i)+(1-$\sum x_i$)g(0)$\leq$ max$\{$g(e_1),g(e_2),....,g(e_n),g(0)$\}$, where$\{e_1,e_2,....,e_n\}$ is the standard basis in R_n.Hence the result follows.
5.1.3.Convexity and continuity : Let f : E$\rightarrow$(-∞,+∞] be a convex function. Then f is continuous(in fact locally Lipschitz) on the interior of its domain.
Proof : Assume 0$\in$int(dom f).We can choose α >0 sufficiently small such that the convex hull V=conv$\{$0,α e_1),....,α e_n)$\}$, where the e_i are the vectors of the

canonical basis of R_n, is contained in E. We can recall that V has a nonempty interior and thus contains an open ball B.Any x in V may be expressed as: x=λ_0f(0)+$\sum_{i=1}^{n}$ λ_if(α e_i) $\leq$ max{f(α e_1),f(α e_2),....,f(α e_n),f(0)}.Therefore f is bounded above on B and the result follows from 5.1.1.

•N.B. : The core and the interior of any convex set in E are identical and convex.

5.1.4.The conjugate of the gauge function γ_C is the indicator function of a set $C^o \subset$ E defined by C^o={ϕ $\in$E| $<\phi$,x$>$ $\leq$1 $\forall$x$\in$C} and call it the polar set for C.

Bipolar set : The bipolar set of any subset C of E is given by C^{oo}=cl(conv(C$\cup${0})).

5.1.5.Supporting hyperplane : Suppose that the convex set C$\subset$E has nonempty interior, and that the point x* lies on the boundary of C. Then there is a supporting hyperplane to C at x*: there is a nonzero element a of E satisfying <a,x> $\geq$ <a,x*> $\forall$x$\in$C.The set {x$\in$E| <a,x-x*>=0} is the supporting hyperplane. Lemma:Given a supporting hyperplane H of a convex set C$\subset$E, any extreme point of C$\cap$H is also an extreme point of C. Or,Let C be a convex set,and H a supporting hyperplane of C.Define the (convex)set T:=C$\cap$H.Then every extreme point of T is also an extreme point of C.

Proof : Suppose x*$\in$T is not an extreme point of C.Then we can find a λ $\in$(0,1) such that x*=(1λ)x_1+λ x_2 for some x_1,x_2 $\in$ C, $x_1 \neq x_2$. Assume, without loss of generality, that H is described by <x,a>=c and that the convex set C lies in the positive half-space determined by H. Then $<x_1$,a$>$ $\geq$ c and $<x_2$,a$>$ $\geq$c.But since x*$\in$H, c=<x*,a> =(1λ)$<x_1$,a$>$+λ $<x_2$,a$>$, and thus x_1 and x_2 lie in H. Hence x_1,x_2 $\in$T and hence x* is not an extreme point of T. Contrapositively, the result is proved.

5.1.6.Minkowski's theorem : Any closed compact convex set C$\subset$E is the closure of convex hull of its extreme points.

Proof: We begin by observing that since the set C is bounded, it can contain no line. Moreover, the smallest convex set that contains the non-empty set ext (C) is just the convex hull of this latter set. So we certainly have that C$\supset$cl{conv(ext(C))}$\neq \emptyset$.Denote the closed convex hull of the extreme points by K. We remark that, since C is bounded, K is necessarily bounded.We need to show that C$\subset$K.Assume the contrary, namely that there is a y$\in$C with y$\notin$K.Then by the separation theorem, there is a hyperplane H separating y and K.Thus, for some a$\neq$0,<y,a> $< inf_{x\in K}$ <x,a>.Let c_o=$inf_{x\in C}$ <x,a>.The number c_o is

finite and there is an $\hat{x}$ $\in$C such that $<\hat{x}$,a$>$=c_o because, by the theorem of Weierstrass, a continuous function (in this case x$\rightarrow$ <x,a>) takes on its minimum value over any closed bounded set.It follows that the hyperplane H ={x$\in$ R^n | <x,a>=c_o} is a supporting hyperplane to C. It is disjoint from K since c_o < $inf_{x\in K}$ <x,a>.The preceeding two results then show that, the set H$\cap$C has an extreme point which is also a fortiori an extreme point of C and which cannot then be in K. This is a contradiction.

5.2 Fenchel biconjugation

We say the function h:E$\rightarrow$[-∞,+∞] is closed if its epigraph is a closed set and we say h is lower semicontinuous at a point x in E if lim inf h(x_r)(=$lim_{s\rightarrow\infty}$ $inf_{r\geq s}$h(x_r)) $\geq$ h(x) for any sequence x_r $\rightarrow$x.The function h is lower semicontinuous if it is lower semicontinuous at every point in E.Any two functions h and g satisfying h<g (in which case we call h a minorant of g) must satisfy h*>g*, and hence h**<g**.

5.2.1.Fenchel biconjugation : The properties below are equivalent,for any function h:E$\rightarrow$[-∞,+∞]: (i)h is closed and convex; (ii)h=h**; (iii)h(x)=sup{α(x) | α an affine minorant of h} $\forall$x$\in$E.

Proof : We can assume h is proper. Since conjugate functions are always closed and convex we know property (b) implies property (a).Also any affine minorant α of h satisfies α=α**$\leq$h**$\leq$h,and hence property (c)implies (b).It remains to show (a) implies (c).

Fix a point x_0 in E.We assume $x_0 \in$ cl(dom h), and fix any real r<h(x_0).Since h is closed, the set {x|h(x)>r} is open, so there is an open convex neighbourhood U of x_0 with h(x)> on U. Now note that the set dom h$\cap$cont δ_U is nonempty, so we can apply the Fenchel theorem to deduce that some element ϕ of E satisfies r$\leq$ inf_x {h(x)+δ_U(x)}={-h*(ϕ)-δ_U^*(-ϕ)}.Now define an affine function α(*)=< ϕ,*>+δ_U^*(-ϕ)+r. Now we can see that α minorizes h,and by definition we know $\alpha(x_0)$ $\geq$ r.Since r was arbitrary,(c) follows at the point x = x_0.

Suppose on the other hand x_0 does not lie in cl(dom h).By the Basic separation theorem there is a real b and a nonzero element a of E satisfying <a,x_0 > >b$\geq$ <a,x>, $\forall$x$\in$dom h.The argument in the preceding paragraph shows there is an affine minorant α of h.But now the affine function α(*)+k(<a,*>-b) is a minorant of h $\forall$k=1,2,.... .Evaluating these functions at x=x_0 proves property (c) at x_0.Thus finally we get the result.

5.2.2.Support functions : Fenchel conjugacy induces a bijection between

everywhere-finite sublinear functions and nonempty compact convex sets in E: (i)If the set C⊂E is compact, convex and nonempty then the support function δ_C^* is everywhere finite and sublinear. (ii)If the function h:E→R is sublinear then h*=δ_C, where the set C={ϕ ∈E| < ϕ,d> ≤h(d) ∀d∈E} is nonempty, compact and convex.
5.2.3.Moreau-Rockafellar theorem : A closed convex proper function on E has bounded level sets if and only if its conjugate is continuous at 0.
Proof : We know a convex function f:E→(-∞,+∞] has bounded level sets if and only if it satisfies the growth condition $lim_{||x||\to\infty}$inf $\frac{f(x)}{||x||}$ >0.Since f is closed, it is equivalent to the existence of a minorant of the form ϵ ||*||+k≤f(*), for some constants ϵ >0 and k. Taking conjugates, this is in turn equivalent to f* being bounded above near 0, and the result then follows by local boundedness.
5.2.4.•Strict-smooth duality: A proper closed convex function on E is essentially strictly convex if and only if its conjugate is essentially smooth.
•Lower semicontinuity and closure : A convex function f:E→[-∞,+∞] is lower semicontinuous at a point x where it is finite if and only if f(x)=(cl f)(x). In this case f is proper.
5.2.5.Theorem : Suppose the function h:E→[-∞,+∞] is convex. (i) If h** is somewhere finite then h**=cl h. (ii)For any point x where h is finite, h(x) = h**(x) if and only if h is lower semicontinuous at x.
Proof : We can notice that h** is closed and minorizes h and h**≤cl h≤h. If h** is somewhere finite then h** and according to 5.2.4. cl h also can never attain -∞.In either case, the Fenchel biconjugation theorem implies cl h=(cl h)**≤h**≤cl h, so cl h=h**.Part (i) is now immediate and using 5.2.4. (ii) is proved similarly.

5.3 Lagrangian duality

5.3.1.Primal problem : In this section, we consider a possibly non-convex optimization problem p*:=min_x f_0(x) : f_i(x) ≤0. i=1,....,m, where the functions f_0,f_1,...,f_m are convex. We denote by D the domain of the problem(which is intersection of the domains of all the functions involved),and by X⊆D its feasible set.
We will refer to the above as the primal problem, and to the decision variable x in that problem, as the primal variable.
5.3.2.Dual problem : Lagrangian : To the problem we associate the La-

grangian L:$R_n \times R_m \rightarrow$R,with values L(x,$\lambda$):=$f_0$(x)+$\sum_{i=1}^{m}$ λ_i f_i(x). The variables $\lambda \in R_m$, are called Lagrange multipliers.
We observe that, for every feasible x$\in$X,and every $\lambda \geq$0, f_0(x) is bounded below by L(x,λ) : $\forall$x$\in$X,$\forall$ $\lambda \in R_m^+$: f_0(x)$\geq$L(x,λ).
The Lagrangian can be used to express the primal problem as an unconstrained one. Precisely: p*=min_x $max_{\lambda\geq0,v}$L(x,λ),where we have used the fact that, for any vectors f$\in R_m$,we have $max_{\lambda\geq0}$ λ^Tf=0 if f$\leq$0, otherwise $max_{\lambda\geq0}$ λ^Tf=+∞.
5.3.3.Lagrange dual function: We then define the Lagrange dual function (dual function for short) the function g(λ) :=min_xL(x,λ).
Note that, since g is the pointwise minimum of affine functions (L(x,*) is affine for every x), it is concave. Note also that it may take the value -∞.Also $\forall$x$\in$X, $\forall$ $\lambda \geq$0:f_0(x)$\geq$ $min_{x'}$L(x',λ)=g(λ),which, after minimizing over x the left-hand side, leads to the lower bound $\forall$ $\lambda \in R_+^m$,v $\in R_p$: p*$\geq$ g(λ).
Lagrange dual problem: The best lower bound that we can obtain using the above bound is p*$\geq$d*,where d*=$max_{\lambda\geq0,v}$ g(λ).
We refer to the above problem as the dual problem, and to the vector $\lambda \in R^m$ as the dual variable.
5.3.4.Weak duality: For the general (possibly non-convex) problem mentioned at the very beginning of this section,weak duality holds: p*$\geq$d*.In this case we have a duality gap.
We define primal value function v:$R^m \rightarrow$[-∞,+∞] as v(b)=inf{f_0(x)|f(x)$\leq$b}.
Dual optimal value: 1.The primal optimal value p is v(0). 2.The conjugate of the value function satisfies v*(-λ)=-g(λ), if $\lambda \geq$0,otherwise v*(-λ)=+∞.
3.The dual optimal value d is v**(0).
Proof : Statement 1. is just the definition and 2. follows from identities.
Now v*(-λ)=sup{-λ^Tb-v(b)|b$\in R^m$} =sup{-λ^Tb-f_0(x)|f(x)+z=b, x$\in$dom f_0, b$\in R^m$, z$\in R_+^m$} =sup{-λ^T(f(x)+z)-f_0(x)|x$\in$domf_0}+sup{-λ^Tz|z$\in R_+^m$} = -g(λ), if $\lambda \geq$0,and otherwise +∞. Finally we can see d=$sup_{\lambda\in R_+^m}$ g(λ)= -$inf_{\lambda\in R_+^m}$ -{-g(λ)} = -$inf_{\lambda\in R_+^m}$ v*(- λ)=v**(0), so follows statement 3.
5.3.5.Zero duality gap: Suppose the value of the primal problem is finite. Then the primal and dual values are equal if and only if the value function v is lower semicontinuous at 0. In this case the set of optimal dual solutions is -δv(0).
Proof : By the previous result, there is no duality gap exactly when the value function satisfies v(0) = v**(0), so Theorem 5.2.5. proves the first

assertion.By part 2. of the previous result, dual optimal solutions λ are characterized by the property $0 \in \delta$v*(-λ), or equivalently,v*(-λ)+v**(0) = 0. But we know v(0) = v**(0), so this property is equivalent to the condition -$\lambda \in \delta$v(0).
5.3.6.Dual attainment: If the Slater condition holds for the primal problem then the primal and dual values are equal, and the dual value is attained if finite.
Proof : If p is -∞ there is nothing to prove, since we know p$\geq$d.If on the other hand p is finite then, as in the proof of the Lagrangian necessary conditions,the Slater condition forces δv(0)$\neq \emptyset$.Hence v is finite and lower semicontinuous at 0 and the result is followed by the conditions of zero duality gap.
5.3.7.Primal attainment: Suppose that the functions f,g_1,g_2,....,g_m:E$\rightarrow$(-∞,+∞] are closed, and that for some real $\lambda_0 \geq 0$ and some vector λ in R^m_+,the function λ_0f+λ^Tg has compact level sets. Then the value function v is closed, and the infimum in this equation is attained when finite. Consequently, if the functions f,g_1,g_2,....,g_m are in addition convex and the dual value for the problem mentioned in the beginning is not -∞, then the primal and dual values, p and d, are equal, and the primal value is attained when finite.

Chapter 6

Duality Revisited

6.1 Polyhedral convex set and function

Polyhedron or polyhedral set is a finite intersection of closed halfspaces in E.We say a function f:E$\rightarrow$[-∞,+∞] is polyhedral if its epigraph is polyhedral. Moreover, a polytope is the convex hull of a finite subset of E, and we call a subset of E finitely generated if it is the sum of a polytope and a finitely generated cone.
6.1.1.Polyhedral functions : Suppose the function f:E$\rightarrow$[-∞,+∞] is polyhedral.Then f is closed and convex, and can be decomposed in the form f=$max_{i\in I}\ g_i$+δ_P, where the index set I is finite (and possibly empty), the functions g_i are affine,and the set P$\subset$E is polyhedral (and possibly empty). Thus the domain of f is polyhedral, and coincides with dom δf if f is proper.
Proof: Since any polyhedron is closed and convex, so is f, and the decomposition follows directly from the definition. If f is proper then both the sets I and P are nonempty in this decomposition. At any point x in P (= dom f) we know 0$\in$ ∂ δ_P(x),and the function $max_i\ g_i$ certainly has a subgradient at x since it is everywhere finite. Hence we deduce the condition ∂f(x)$\neq$ $\emptyset$.
6.1.2.Finitely generated functions : Suppose the function f:E$\rightarrow$[-∞,+∞] is finitely generated. Then f is closed and convex, and dom f is finitely generated. Furthermore, f* is polyhedral.
6.1.3.Lemma : Lemma: Let P be a polyhedron. The following are equivalent.(i)x is a vertex (unique maximizer);(ii)x is an extreme point(not convex combination of other points);(iii)x is a basic feasible solution (BFS)(tight constraints have rank n).

Proof: Proof of (i)$\rightarrow$(ii):x is a vertex$\rightarrow$ $\exists$ c such that x is unique maximizer of c^Tx over P. Suppose x = αy + (1-α)z where y,z$\in$P and α $\in$(0,1). Suppose y$\neq$x. Then c^Tx = α c^Ty + (1-α) c^Tz< α c^Tx + (1-α) c^Tx=c^Tx, which is a contradiction. So y=x. Symmetrically, z=x. So x is an extreme point of P.
Proof Idea of (ii)$\rightarrow$(iii):Assume x not a BFS.Hence rank A_x $\leq$n-1. •Each tight constraint removes one degree of freedom.
•At least one degree of freedom remains.
•So x can wiggle while staying on all the tight constraints.
•Then x is a convex combination of two points obtained by wiggling.
•So x is not an extreme point.
Proof of (ii)$\rightarrow$(iii):Assume x is not a BFS. Hence rank A_x <n. Claim: $\exists$w$\in$ R^n,w$\neq$0, such that a_i^Tw=0 $\forall$ a_i $\in$ A_x.Now let M be matrix whose rows are the a_is in A_x.Hence dim row-space(M) + dim null-space(M)=n. But dim row-space(M)<n. So,$\exists$ w$\neq$0 in the null space.Now let y=x+ϵw and z=x-ϵw, where ϵ >0. Now claim:If ϵ is very small then y,z$\in$P.Now first consider tight constraints at x.a_i^Ty=a_i^Tx+ϵ a_i^Tw=b_i+0.So y satisfies this constraint. Similarly for z. Next consider the loose constraints at x.b_i-a_i^Ty=b_i-a_i^Tx-ϵ a_i^Tw$\geq$0.So y satisfies these constraints. Similarly for z. Then x=αy+(1-α)z, where y,z$\in$P,y,z$\neq$x,and α=1/2.So x is not an extreme point.Similarly we can easily do the (iii) to (i) part.
6.1.4.Lemma : Any polyhedron has at most finitely many extreme points.
Proof: Any polyhedron can be described by m$\in$Z constraints, thus there are atmost m
$n$ways to choose constraints to be satisfied by the basic feasible solution,and thus finitely many such points. Since every extreme point is a basic feasible solution, there are no more extreme points than there are basic feasible solutions. Thus there are finitely many extreme points.
6.1.5.Features of polyhedron: 1.Any compact polyhedron is a polytope. 2.Any polyhedral cone is a finitely generated cone.
6.1.6.Polyhedrality : A set or function is polyhedral if and only if it is finitely generated.
Proof: For finite sets $\{a_i \mid$ i$\in$I$\}$ $\subset$E and $\{b_i \mid$ i$\in$I$\}$ $\subset$R,consider the polyhedron in E defined by P=$\{$x$\in$E $\mid$ < a_i,x> $\leq$ b_i for i$\in$I$\}$.The polyhedral cone in E$\times$R defined by Q=$\{$(x,r)$\in$E$\times$R$\mid$ < a_i,x>-b_ir$\leq$0 for i$\in$I$\}$ is finitely generated,by the previous lemma, so there are finite subsets $\{x_j$ $\mid$j$\in$J$\}$ and $\{y_t$ $\mid$t$\in$T$\}$ of E with Q=$\{$ $\sum\limits_{j\in J}$ λ_j(x_j,1)+$\sum\limits_{t\in T}$ μ_t(y_t,0) $\mid$ λ_j $\in$ R_+ for j$\in$J,μ_t $\in$ R_+ for t$\in$T$\}$.We now deduce P=$\{$(x,1)$\in$Q$\}$ =conv$\{x_j$ $\mid$ j$\in$J$\}$+$\sum\limits_{t\in T}$ μ_ty$\mid$ μ_t $\in$ R_+ for t$\in$T$\}$,so

P is finitely generated. We have thus shown that any polyhedral set (and hence function) is finitely generated.
Conversely, suppose the function f : E→[-∞,+∞] is finitely generated. Consider first the case when f is proper.By 6.1.2. f* is polyhedral, and hence (by the above argument) finitely generated. But f is closed and convex, by 6.1.2. again,so the Fenchel biconjugation theorem implies f=f**.By 6.1.2 once again we see f** (and hence f) is polyhedral.
6.1.7.Polyhedral algebra :Consider a Euclidean space Y and a linear map A : E → Y.
(a) If the set P ⊂ E is polyhedral then so is its image AP.
(b) If the set K ⊂ Y is polyhedral then so is its inverse image A^{-1}K.
(c) The sum and pointwise maximum of finitely many polyhedral functions are polyhedral.
(d) If the function g : Y → [$infty$,+$infty$] is polyhedral then so is the composite function g o A.
(e) If the function q : E × Y → [$infty$,+$infty$] is polyhedral then so is the function h : Y → [$infty$,+$infty$] defined by h(u)= $inf_{x\in E}$ q(x, u).
6.1.8.Polyhedral Fenchel duality: All the conclusions of the Fenchel duality theorem remain valid if the regularity condition is replaced by the assumption that the functions f and g are polyhedral with dom g ∩ Adom f nonempty.
6.1.9.Mixed Fenchel duality: All the conclusions of the Fenchel duality theorem remain valid if the regularity condition is replaced by the assumption that dom g ∩ Acont f is nonempty and the function g is polyhedral.

6.2 Duality for linear and semidefinite programming

Linear programming(LP) is the study of optimization problems involving a linear objective function subject to linear constraints.Linear programs are inherently polyhedral, so our main development follows directly from the polyhedrality section.Given vectors a_1,a_2,....,a_m and c in R^n and a vector b in R^m, cnsider the prima linear program : inf <c,x> subject to $<a_i$,x>-$b_i \leq$0, for i=1,2,...,m, x∈ R^n. Denote the primal optimal value by p ∈ [-∞,+∞].In the Lagrangian duality framework, the dual problem is : sup -b^T μ subject to $\sum\limits_{i=1}^{m}$ μ_i a_i = -c, $\mu \in R^m_+$, with dual optimal value d∈[-∞,+∞].We know the weak duality inequality p≥d. If the primal problem satisfies the Slater con-

dition then the Dual attainment theorem shows p=d with dual attainment when the values are finite.
Suppose the primal value p is finite. Then it is easy to see that the homogenized system of inequalities in R^{n+1}, $< a_i$,x>-b_iz $\leq$0, for i=1,2,...,m; -z $\leq$0, and <-c,x>+pz>0,x$\in$ R^n,z$\in$R,has no solution. Applying the Farkas lemma to this system, we deduce there is a vector $\mu \in R^n_+$ and a scalar $\alpha \in R_+$ satisfying $\sum_{i=1}^{m}$ $\mu_i(a_i$,-$b_i)$+α(0,-1)=(-c,p).Thus μ is a feasible solution for the dual problem, with objective value at least p. The weak duality inequality now implies μ is optimal and p=d.
We can be more systematic using our polyhedral theory. Suppose that Y is a Euclidean space, that the map A : E $\rightarrow$ Y is linear, and consider cones H $\subset$ Y and K $\subset$ E. For given elements c of E and b of Y, consider the primal abstract linear program : inf<c,x> subject to Ax-b $\in$ H, x$\in$ K. As usual, we denote the optimal value by p. We can write this problem in Fenchel form if we define functions f on E and g on Y by f(x) = <c,x>+δ_K(x) and g(y)=δ_h(y-b).Then the Fenchel dual problem will look like : sup <b,ϕ > subject to A*ϕ-c $\in K^-$, $\phi \in H^-$, with dual optimal value d.
6.2.1.Cone programming duality : Suppose the cones H and K in the previous problems are convex. Then (i)If any of the conditions (a) b$\in$int (AK-H), (b) b$\in$ AK-int H, or (c) b$\in$ A(int K)-H, and H is polyhedral or A is surjective, hold then there is no duality gap (p = d) and the dual optimal value d is attained if finite. (ii)Suppose H and K are also closed. If any of the conditions (a)-c$\in$ int(A*H^-+K^-), (b)-c$\in$ A*H^-+int K^-, or (c)-c$\in$ A*(intH^-)+K^-, and K is polyhedral or A* is surjective, hold then there is no duality gap and the primal optimal value p is attained if finite.
6.2.2.Linear programming duality: Suppose the cones H and K in the the dual pair of problems discussed earlier are polyhedral. If either problem has finite optimal value then there is no duality gap and both problems have optimal solutions.
6.2.3.Linear programming has an interesting matrix analogue. If we have matrices A_1,A_2,.....,A_m and C in S^n_+ and a vector v in R^m, we consider the primal semidefinite program : inf tr(CX) subject to tr(A_iX) = v_i, for i=1,2,....,m, X$\in$ S^n_+ and the dual problem : sup v^T ϕ subject to C-$\sum_{i=1}^{m}$ ϕ_i $A_i \in S^n_+$, $\phi \in R^m$.
Semidefinite programming duality: When the primal problem has a positive definite feasible solution, there is no duality gap and the dual optimal value

is attained when finite. On the other hand, if there is a vector ϕ in R^m with C-$\sum_{i=1}^{m} \phi_i A_i$ positive definite then once again there is no duality gap and the primal optimal value is attained when finite.

Chapter 7

Nonsmooth Optimization

7.1 Generalized Directional Derivatives

We will consider here three types of directional derivatives for a function f, locally Lipschitz around the point x in E.
1.Dini directional derivative : f^-(x;h)=$lim_{t\to 0}$inf$\frac{f(x+th)-f(x)}{t}$.
• A disadvantage of this idea is that f^-(x;*) is not usually sublinear(e.g. f=-|*| on R), so we could not expect an analogue of the Max formula.
2.Clarke directional derivative : f^o(x;h)=$lim_{y\to x}$ $sup_{t\to 0}$ $\frac{f(y+th)-f(y)}{t}$ =$inf_{\delta>0}$ $sup_{||y-x||\leq\delta,0<t<\delta}$ $\frac{f(y+th)-f(y)}{t}$.
3.Michel-Penot directional derivative : $f^\diamond$(x;h)=$sup_{v\in E}$ $lim_{t\to 0}$sup$\frac{f(x+th+tv)-f(x+tv)}{t}$.
7.1.1.Proposition:If the real function f has Lipschitz constant K around the point x in E then the Clarke and Michel-Penot directional derivatives f^o(x;*) and $f^\diamond$(x;*) are sublinear and satisfy f^-(x;*)$\leq$ $f^\diamond$(x;*)$\leq$ f^o(x;*)$\leq$ K||*||.
•We can now define the Clarke subdifferential as δ_of(x)={ϕ $\in$E| < ϕ,h> $\leq$ f_0(x;h)$\forall$h$\in$E},and similarly the Dini and Michel-Penot subdifferentials δ_-f(x)and $\delta_\diamond$f(x).
7.1.2.Nonsmooth max formulae : If the real f has Lipschitz constant K around the point x in E then the Clarke and Michel-Penot subifferentials δ_of(x) and $\delta_\diamond$f(x) are nonempty, compact and convex, and satisfy δ_-f(x)$\subset$ $\delta_\diamond$f(x)$\subset$ δ_of(x)$\subset$ KB and the Clarke and Michel-Penot directional derivatives are the support functions of the corresponding subdifferentials: f^o(x;h)=max{< ϕ,h> | $\phi\in\delta_o$f(x)}, and $f^\diamond$(x;h)=max{< ϕ,h> | $\phi\in\delta_\diamond$f(x)} for any direction h in E. 7.1.3.Nonsmooth calculus : If the real functions f and g are locally Lipschitz around x$\in$E, then the Clarke subdifferential satisfies (i)δ_0(f+g)(x)$\subset$

δ_0f(x)+δ_0g(x) and (ii)δ_0(f$\vee$g)(x)$\subset$ conv(δ_0f(x)$\cup$ δ_0g(x)), where f$\vee$g denotes the function x$\to$max{f(x),g(x)}. 7.1.4.Nonsmooth necessary condition : We assume the point x* as a local minimizer for the problem : inf{f(x)| g_i(x)$\leq$0 (i$\in$I)}, where the real functions f and g_i (for i in finite index set I) are locally Lipschitz around x*. Let I(x*)={i| g_i(x)=0} be the active set. Then $\exists$ real λ_0,λ_i $\geq$0, for i in I(x*), not all zero, satisfying 0$\in$ λ_0 $\delta_\diamond$f(x*)+$\sum_{i\in I(x*)}$ λ_i $\delta_\diamond$ g_i(x*). If furthermore some direction d in E satisfies $g_i^\diamond$(x*;d)<0$\forall$i$\in$I(x*) then we can assume λ_0=1.
Proof : Firstly, from the conditions given we note that x* is a local minimizer of the function x$\to$max{f(x)-f(x*),g_i(x) (i$\in$I(x*))}. We deduce 0$\in$ $\delta_\diamond$(max{f(x)-f(x*),g_i(x) (i$\in$I(x*))})(x*)$\subset$ conv($\delta_{diamond}$f(x*)$\cup$ $\cup_{i\in I(x*)}$ $\delta_\diamond$ g_i(x*)),by 7.1.3. If condition given in the problem holds and λ_0 is 0 in that case, we obtain the contradiction 0$\leq$max{$<\phi$,d> | $\phi\in\sum_{i\in I(x*)}$ λ_i $\delta_\diamond$ g_i(x*)}= $\sum_{i\in I(x*)}$ $g_i^\diamond$(x*;d) <0. Thus λ_0 is strictly positive, and hence without loss of generality equals 1.Hence proved.

7.2 Nonsmooth regularity

7.2.1.Unique Michel-Penot subgradient : A real function f which is locally Lipschitz around the point x in E has a unique Michel-Penot subgradient ϕ at x if and only if ϕ is the Gâteaux derivative $\bigtriangledown$f(x).
Proof : If f has a unique Michel-Penot subgradient ϕ at x, then all directions d in E satisfy $f^\diamond$(x;d)=$sup_{v\in E}$ $lim_{t\to 0}$sup$\frac{f(x+td+tv)-f(x+tv)}{t}$=$<\phi$,d>. The cases d=w with v=0, and d=-w with v=w show $lim_{t\to 0}$sup$\frac{f(x+tw)-f(x)}{t}$ $\leq$ $<\phi$,w> $\leq$ $lim_{t\to 0}$inf$\frac{f(x+tw)-f(x)}{t}$, so we deduce f'(x;w)=$<\phi$,w> as required.
Conversely, if f has Gâteaux derivative ϕ at x then any directions d and v in E satisfy $lim_{t\to 0}$sup$\frac{f(x+td+tv)-f(x+tv)}{t}$ $\leq$ $lim_{t\to 0}$sup$\frac{f(x+t(d+v))-f(x)}{t}$- $lim_{t\to 0}$inf$\frac{f(x+tv)-f(x)}{t}$=f'(x;d+v)-f'(x;v)=$<\phi$,d+v>-$<\phi$,v>=$<\phi$,d>=f'(x;d)$\leq$ $f^\diamond$(x;d). Now, we take the supremum over v to show $f^\diamond$(x;d)=$<\phi$,d> $\forall$d.
7.2.2.Regularity of convex functions : Suppose the function f:E$\to$(-∞,+∞] is convex. If the point x lies in int(dom f)then f is regular at x, and hence the convex, Dini, Michel-Penot and Clarke subdifferentials all coincide : δ_of(x)=$\delta_\diamond$f(x)=δ_-f(x)=δf(x).
7.2.3.Strict differentiabilty : A real function f has strict derivative ϕ at a point x in E if and only if it is locally Lipschitz around x with $lim_{y\to x,t\to 0}$ $\frac{f(y+th)-f(y)}{t}$=$<\phi$,h>, $\forall$h$\in$E.

7.3 Tangent cones

7.3.1.Distance function : We define the distance function to the nonempty set C⊂E by d_C(x)=inf{||y-x|| |y∈C}. We can easily check that d_C has Lipschitz constant 1 on E.
7.3.2.Exact penalization : For a point x in a set C⊂E,we assume that the real function f is locally Lipschitz around x. If x is a local minimizer of f on C then for real L which is sufficiently large, x is a local minimizer of f + Ld_C.
Proof : Suppose the Lipschitz constant is not larger than L. We fixed a point z close to x. Clearly d_S(z) is the infimum of ||zy|| over points y close to x in C, and such points satisfy f(z)+Ld_S(z)≥f(y)+L(d_S(z)||zy||)≥f(x)+L(d_S(z)||zy||).The result follows by taking the supremum over y.
7.3.3.Tangent cones : From 7.3.2. we come to know that any direction h in E satisfies 0≤(f+Ld_C)^o(x;h)≤ f^o(x;h)+Ld_C^o(x;h),and hence the Clarke directional derivative satisfies f^o(x;h)≥0 wheneverh lies in the set T_C(x)={h| d_C^o(x;h)=0}.This set is a closed convex cone called Clarke tangent cone.Analogously, Dini directional derivative gives rise to the cone K_C(x)={h| d_C^-(x;h)=0},a (nonconvex) closed cone containing T_C(x) called the contingent cone.In case of ordinary directional derivative, it is the closed convex cone T_C(x)={h| d'_C(x;h)=0} called the tangent cone.
7.3.4.Theorem : Suppose the point x lies in a set C in E.Now (i)The contingent cone K_C(x) consists of those vectors h in E for which there are sequences $t_r \to 0$ in R and $h_r \to h$ in E such that x+t_r h_r lies in C∀r. (b) The Clarke tangent cone T_C(x) consists of those vectors h in E such that for any sequences $t_r \to 0$ in R and $x_r \to x$ in S, there is a sequence $h_r \to h$ in E such that x_r+t_r h_r lies in C∀r.
7.3.5.Convex tangent cone : If the point x lies in the convex set C⊂E,then C is tangentially regular at x, with T_C(x)=K_C(x)=clR_+(C-x).
Proof : The regularity follows from the regularity of convex functions discussed earlier.The identity K_C(x)=clR_+(C-x) follows easily from 7.3.4.
7.3.6.Theorem : For a point x in a set C⊂E, the Clarke normal cone, defined by N_C(x)=T_C(x)$^-$, is cl(R_+ δ_o d_C(x)).
Proof : By the Bipolar cone theorem,all we need to show is (δ_o d_C(x))$^-$=T_C(x), and this follows from the Max formula discussed earlier.
7.3.7.Nonsmooth necessary conditions : For a point x in a set C⊂E,suppose the real function f is locally Lipschitz around x. Any local minimizer x of f on C must satisfy the condition 0∈ $\delta_{diamond}$f(x)+N_C(x).
Proof : We know that for large real L, the point x is a local minimizer of

f+Ld_C(x) by the Exact penalization proposition, so it satisfies 0$\in$ $\delta_{diamond}$(f+Ld_C)(x)$\subset$ $\delta_{diamond}$f(x)+L$\delta_{diamond}$ d_C(x)$\subset$ $\delta_{diamond}$f(x)+N_C(x), using the nonsmooth sum rule.

7.3.8.Contingent necessary condition : Suppose a point x is a local minimizer of the real function f on the set C$\subset$E.If f is Frechet differentiable at x, then the condition -$\bigtriangledown$f(x)$\in$ $K_C(x)^-$ must hold.

Proof : If the condition fails then there is a direction h in K_C(x) which satisfies $<\bigtriangledown$f(x),h$>$ $<$0.Then by 7.3.4. $\exists$ sequences $t_r \to 0$ in R and $h_r \to h$ in E such that x+t_r h_r lies in C$\forall$r. But then, since we know $lim_{r\to\infty} \frac{(f(x+t_r h_r)-f(x)-<\bigtriangledown f(x),t_r h_r>}{t_r||h_r||}$=0, we deduce f(x+$t_r$ h_r)<f(x) $\forall$large r, contradicting the local minimality of x.Hence follows the result.

Chapter 8

KKT Conditions

8.1 Weak Metric Regularity

Now we will go through the problems inf[f(x)|h(x)=0] by linearizing the feasible region h^{-1}(0),using the contingent cone.Here we assume S⊂E ia an open set and C⊂S a closed set, h:S→Y a continuous map where Y is a Euclidean space.
8.1.1. If h is Frechet differentiable at x∈S then $K_{h^{-1}(h(x))}$(x)⊂ N(▽h(x)).
8.1.2. Ekeland variational principle: Suppose the functions f:E→($-\infty$,+∞] is closed and the point x∈E satisfies f(x)≤inf f+ϵ, for some real ϵ >0. Then for any real λ >0 ∃ a point a∈E satisfying the conditions (a) $||x-a|| \leq \lambda$, (b) f(a)≤ f(x) and (c) a is the unique minimizer of the function f(*)+$\frac{\epsilon}{\lambda}$ $||*-a||$.
Proof: We can consider f as proper, and so bounded below.Since the function f(*)+$\frac{\epsilon}{\lambda}$ $||*-x||$ then has compact level sets, and its set of minimizers M_{min} ∈E is nonempty and compact. If we choose a minimizer a for f on M_{min}, then for z≠a in M_{min} f(a)≤f(z)< f(z)+$\frac{\epsilon}{\lambda}$ $||z-a||$, while for z not in M_{min} we have f(v)+$\frac{\epsilon}{\lambda}$ $||a-x||$ < f(z)+$\frac{\epsilon}{\lambda}$ $||z-x||$. By triangle inequality we get part (c). Since a lies in M_{min} we get f(z)+$\frac{\epsilon}{\lambda}$ $||z-x||$ ≥f(a)+$\frac{\epsilon}{\lambda}$ $||a-x||$ ∀z∈E.Setting z=x shows the inequalities f(a)+ϵ ≥ inf f+ϵ ≥f(x)≥f(a)+$\frac{\epsilon}{\lambda}$ $||a-x||$.Hence parts (a) and (b) follow.
8.1.3. Now we say h is weakly metrically regular on C at the point x in C if there is a real constant k such that $d_{C\cap h^{-1}(h(x))}(z) \leq$ k$||h(z)-h(x)$ ∀ z∈C close to x.Conversely, if suppose 0∈C and h(0)=0 then if h is not weakly metrically regular on C at 0,∃a seqence $a_r \to 0$ in C such that h(a_r)≠0∀r, and a strictly positive sequence $\delta_r \to 0$ such that the function $||h(*)||$+δ_r

$|| * -a_r||$ is minimized on C at a_r. 8.1.4. Theorem: If h is strictly Frechet differentiable at the point x in C and $\bigtriangledown$h(x)(T_s(x))= Y then h is weakly metrically regular on C at x.
8.1.5. Liusternik Theorem: If h is strictly Frechet differentiable at the point x and $\bigtriangledown$h(x) is surjective, then the set h^{-1}(h(x)) is tangentially regular at x and $K_{h^{-1}(h(x))}$(x)=N($\bigtriangledown$h(x)).
Proof : We assume x=0 and h(0)=0 without loss of generality. By 8.1.1 we can say, it is sufficient if we prove N($\bigtriangledown$h(0))$\subset$ $T_{h^{-1}(0)}(0)$.Fixing an element v of N($\bigtriangledown$h(0)) and considering a sequence $x_r \to$0 in $h^{-1}(0)$ and $t_r \to$0 in R_{++}. From previous theorem h is weakly metricly regular at 0, so $\exists$ a real constant k such that $d_{h^{-1}(0)}(x_r$+t_rp)$\leq$k$||h(x_r + t_r p)||$ holds $\forall$ large r, and hence there are points z_r in $h^{-1}(0)$ satisfying $||x_r + t_r p - z_r|| \leq$ k$||h(x_r + t_r p)||$.
If we define directions p_r=t_r $^{-1}$(z_r-x_r) then the points x_r+t_r p_r lie in $h^{-1}(0)$ for large r, and since $||p - p_r||$=$||(x_r + t_r p) - z_r$ $/t_r \leq$ k$||h(x_r + t_r p) - h(x_r)||$ $\to$ k$||(\bigtriangledown h(0))p||$ =0,and from here we deduce p$\in$ $T_{h^{-1}(0)}(0)$.

8.2 The Karush-Kuhn-Tucker Theorem

The central result of optimization theory describes the first order necessary optimality conditions for the general nonlinear problem inf{f(x)|x$\in$S},where, given an open set S$\subset$E, the objective function f:S$\to$R and the feasible region R is described by equality and inequality constraints: R={x$\in$S| g_i(x)$\leq$0 for i=1,2,.....,m, h(x)=0}. The equality constraint map h:S$\to$Y and the the inequality constraint functions g_i:S$\to$R are all continuous and I(x*)={i| g_i(x)=0} is the set of indices. 8.2.1.Linear Independence constraint qualification : The equality constraint gradients $\bigtriangledown$ h_i(x*),i=1,.....,m and the active inequality constraint gradients $\bigtriangledown$ g_i(x*),i$\in$I(x*) are linearly independent.But many problems satisfy KKT without LICQ, as we can see in the following example with x*=0.
minimize f(x)=x_2; subject to g_1(x)=$-x_1^2$+x_2 $\leq$0, g_2(x)=-x_2 $\leq$0. 8.2.2.The Mangasarian-Fromovitz constraint qualification : The active constraint functions g_i (for i in I(x*)) are Frechet differentiable at the point x*,the equality constraint map h is strictly differentiable at x*, and the set {p$\in$N($\bigtriangledown$h(x*))| $<$ $\bigtriangledown g_i$(x*),p$>$ $<$0 for i in I(x*)} is nonempty.
We say that MFCQ holds at x* when the equality constraint gradients are linearly independent and $\exists$ a vector d$\in$E such that $\bigtriangledown$h(x*)^{T}d=0 and $\bigtriangledown$ g_i(x*)^{T}d<0$\forall$i$\in$I(x*). 8.2.3.Relation between LICQ and MFCQ :

Theorem:If x*∈C satisfies LICQ, then x* satisfies MFCQ.
Proof: Suppose without loss of generality that I(x*)={1,.....,q}. Consider the matrix M=($\bigtriangledown$ h_1(x*) ...$\bigtriangledown$ h_m(x*)$\bigtriangledown$ g_1(x*)....$\bigtriangledown$ g_q(x*)$)^T$ and b∈R$^m + q$ given by b_i=0∀i=1,...,m and b_j=-1∀j∈{m+1,.....,m+q}.Since the rows of M are linearly independent, the system Md=b has a solution.Let d* be a solution.Then $\bigtriangledown$h(x*$)^T$d*=0, and $\bigtriangledown$ g_i(x*$)^T$d*=-1<0∀i∈I(x*),completing the proof.
The next example shows that MFCQ does not imply LICQ. Consider the functions g_i:R^2 →R(i=1,2,3) defined by g_1(x)=(x_1-1$)^2$+(x_2-1$)^2$-2;g_2(x)=(x_1-1$)^2$+(x_2+1$)^2$-2;g_3(x)=-x_1,and the feasible point x*=(0,0$)^T$. Note that {$\bigtriangledown$ g_i(x*),i=1,2,3} is linearly dependent. On the other hand, taking d=(1,0$)^T$,we have $\bigtriangledown$ g_i(x*$)^T$d<0 for i=1,2,3,which means that MFCQ holds.
8.2.4.Theorem:Suppose the Mangasarian-Fromovitz constraint qualification holds.Then the contingent cone to the feasible region R defined earlier is given by K_R(x*)={p∈N($\bigtriangledown$h(x*))| < $\bigtriangledown g_i$(x*),p> ≤0 for i ∈ I(x*)}. Lemma:Any linear maps A:E→ R^q and H:E→Y satisfy {x∈N(H)|Ax≤0$\}^-$=A*R^q_+ + H*Y.
8.2.5.Karush-Kuhn-Tucker conditions : We assume that the point x* is a local minimizer for the problem introduced in the beginning of this section and the objective function f is Frechet differentiable at x*.If the Mangasarian-Fromovitz constraint qualification holds then ∃ multipliers λ_i in R_+ (for i∈I(x*)) and μ ∈Y satisfying $\bigtriangledown$f(x*)+$\sum_{i\in I(x*)}$ λ_i $\bigtriangledown$ g_i(x*)+$\bigtriangledown$h(x*)*μ=0.
Proof : The Contingent necessary condition shows -$\bigtriangledown$f(x*)∈ K_R(x*$)^-$
={p∈N($\bigtriangledown$h(x*))| < $\bigtriangledown$ g_i(x*),p> ≤0 for i∈I(x*)$\}^-$=$\bigtriangledown$f(x*)+$\sum_{i\in I(x*)}$ R_+ $\bigtriangledown$ g_i(x*)+$\bigtriangledown$h(x*)*Y,using the theorem and lemma of 8.2.4. and the result follows.

Chapter 9

Journey from finite to infinite dimensions

Now we are done with all sorts of convex analysis and all important aspects of convex optimization in finite dimensional space.These things make the basis of the convex optimization in infinite dimensional space also.
One way to think of functional analysis is as the branch of mathematics that studies the extent to which the properties possessed by finite dimensional spaces generalize to infinite dimensional spaces. In the finite dimensional case there is only one natural linear topology. In that topology every linear functional is continuous, convex functions are continuous (at least on the interior of their domains), the convex hull of a compact set is compact, and nonempty disjoint closed convex sets can always be separated by hyperplanes. On an infinite dimensional vector space, there is generally more than one interesting topology, and the topological dual, the set of continuous linear functionals, depends on the topology. In infinite dimensional spaces convex functions are not always continuous, the convex hull of a compact set need not be compact, and nonempty disjoint closed convex sets cannot generally be separated by a hyperplane. However, with the right topology and perhaps some additional assumptions, each of these results has an appropriate infinite dimensional version.
Nowadays, innite-dimensional optimization problems appear in a lot of active elds of optimization, such as PDE-constrained optimization [7], with applications to optimal control, shape optimization or topology optimization. Moreover, the generalization of many classical nite optimization problems to a continuous time setting lead to innite-dimensional problems.

So, finally after doing this project we are in such a position that we can motivate us to go through the convex optimization problems in finite dimensional space and undoubtedly we can start thinking about these type of problems in infinite dimensional space also.

Chapter 10

Acknowledgement and References

1.Real Analysis with Real Applications by Kenneth R. Davidson and Allan P. Donsig.
2.Convex Analysis and Nonlinear Optimization by Jonathan M. Borwein and Adrian S. Lewis.
3.Infinite Dimensional Analysis : A Hitchhikers Guide by Charalambos D. Aliprantis and Kim C. Border.
4.Convex Optimization by Stephen Boyd and Lieven Vandenberghe.
Finally I want to thank Dr. P. Shunmugaraj sir, professor of Mathematics and Statistics department,IIT Kanpur for helping me by every means throughout this project and guiding me so adorably.I am really obliged to you.