Equivalent def^n of Continuous F^n 2

$x_n \to x$ in $C \Rightarrow f(x_n) \to f(x)$

Not continuous f is not continuous at x if $\exists \varepsilon_0$ s.t. $\forall \delta > 0$, we have for $y \in C$ $\|y-x\| < \delta$ and $|f(y) - f(x)| \geq \varepsilon_0$

Ch1 Continuity and existence of optima.

Norm on $\mathbb{R}^d$ a map $x \in \mathbb{R}^d \mapsto \|x\| \in [0,\infty)$ is said to be a norm if

i) $\|x\| \geq 0$ $\forall x \in \mathbb{R}^d$ and $=0$ iff $x = \theta :=$ vector of all zeros

ii) $\|\alpha x\| = |\alpha| \|x\|$ for $\alpha \in \mathbb{R}, x \in \mathbb{R}^d$ and

iii) Δ inequality holds i.e. for $x, y \in \mathbb{R}^d$ $\|x+y\| \leq \|x\| + \|y\|$

$B_r(x) := \{y : \|y-x\| < r\}$

Open Set If for every $x \in A \subset \mathbb{R}^d$, $B_r(x) \subset A$ for some $r > 0$ that can depend on x

If so, we say that A is an open set.

$x_n \to x^*$ For any open set $A \subset \mathbb{R}^d$ such that $x^* \in A$,

$\exists n_0 \geq 1$ such that $n \geq n_0 \Rightarrow x_n \in A$

$B_r(x)$ $r = \infty$ Whole space $r = 0$ empty set.

Two properties of open sets

i) Finite intersection of open sets are open

ii) Arbitrary unions of open set is open.

Suppose $\|\cdot\|'$ and $\|\cdot\|''$ on $\mathbb{R}^d$ lead to the same family of open sets. Then they are said to be compatible or equivalent. In this case $\exists b, c > 0$ s.t. $b\|x\|' \leq \|x\|'' \leq c\|x\|'$

Denseness Set A is dense in $\mathbb{R}^d$ if any open set in $\mathbb{R}^d$ contains an element of A. More generally, a set A is dense in a set B if $A \subset B$ and for any $x \in B$ & any $\varepsilon > 0$, we can find $y \in A$ s.t. $\|x - y\| < \varepsilon$. Equivalently, for any $x \in B \setminus A$, $\exists$ a sequence $\{x_n\} \subset A$ s.t. $x_n \to x$.

Closed Set Set B is closed if, whenever a sequence $\{x_n\} \subset B$ satisfies $x_n \to x^*$ (say), then $x^* \in B$ / B closed if B^c open

for any set $D \subset \mathbb{R}^d$

Interior int(D) Largest open set contained in D / union of open sets contained in D.

Closure $\bar{D}$ smallest closed set containing D / intersection of all closed sets containing D.

$\text{int}(D) \subseteq D \subseteq \bar{D}$, $\text{int}(D)$ can be ϕ even if $D \neq \phi$

Boundary $\partial D := \bar{D} \setminus \text{int}(D)$ $(= \bar{D} \cap (\text{int} D)^c)$

∂D is always a closed set, a closed set contains its boundary & an open set is disjoint from its boundary.

"Relative" For $C \subset \mathbb{R}^d$, $A \subset C$ is relatively open in C if it is the intersection of C with an open set in $\mathbb{R}^d$.

Hyperplane H in $\mathbb{R}^d$ i.e. set of type $\{x \in \mathbb{R}^d : \langle x - x_0, \vec{n} \rangle = 0\}$

Continuous Function a f^n $f : C \subset \mathbb{R}^d \mapsto \mathbb{R}^m$ is continuous at $x \in C$ if given any $\varepsilon > 0$, we can find $\delta > 0$ such that $y \in C$, $\|y - x\| < \delta \Rightarrow \|f(x) - f(y)\| < \varepsilon$.

It is a continuous f^n if it is continuous at all x in C.

An equivalent definition 1

$f^{-1}(A) := \{x \in C : f(x) \in A\}$ is relatively open in C for any open set $A \subset \mathbb{R}^m$.

Uniform Continuity a f^n $f : C \subset \mathbb{R}^d \mapsto \mathbb{R}$ is uniformly continuous if for any $\varepsilon > 0$, we can find a $\delta > 0$ s.t. if $x, y \in C$, $\|x - y\| < \delta$, then $|f(x) - f(y)| < \varepsilon$

Here δ is now independent of x

For $A \subset \mathbb{R}$

$\text{l.u.b.}(A) := \min\{z : z \geq x \ \forall x \in A\}$ supremum of A

$\text{g.l.b.}(A) := \max\{z : z \leq x \ \forall x \in A\}$ infimum of A

If $\exists$ an $x^* \in C$ with $f(x^*) = \inf_C f$ we call the infimum as the minimum & x^* as a minimizer of f in C. The set of such minimizers will be denoted by $\text{argmin}_C f$.

A sequence x_n is monotone if either $x_n \leq x_{n+1} \ \forall n$ or $x_n \geq x_{n+1} \ \forall n$

Say that a set A is bounded if $\exists K$ $0 < K < \infty$ such that $\|x\| \leq K \ \forall x \in A$

Monotone bounded sequences in $\mathbb{R}$ converge.

Bolzano-Weierstrass theorem. Every bounded sequence in $\mathbb{R}^d$ has a convergent subsequence.

Limit Point x^* is a limit point of $\{x_n\}$ if $\exists$ a subsequence $\{x_{n(k)}\}$ of $\{x_n\}$ s.t. $x_{n(k)} \to x^*$

• Sequence $\{x_n\}$ converges to set A if $\inf_{y \in A} \|x_n - y\| \to 0$ as $n \uparrow \infty$

Cauchy Sequence $\{x_n\} \subset \mathbb{R}^d$ is Cauchy if $\lim_{m,n \uparrow \infty} \|x_m - x_n\| = 0$

Theorem 1.2 If $\{x_n\} \subset \mathbb{R}^d$ is Cauchy, then $x_n \to x^*$ for some x^*

Weierstrass Theorem If $A \subset \mathbb{R}^d$ is closed & bounded, and $f: A \mapsto \mathbb{R}$ is continuous, then f attains its maximum & minimum on A i.e. $\exists\, x, x' \in A$ satisfying $f(x) \le f(y) \le f(x')\ \forall y \in A$

Theorem 1.4 Suppose $C \subset \mathbb{R}^d$ is closed & $f: C \mapsto \mathbb{R}$ is continuous and satisfies $\lim_{\|x\| \uparrow \infty} f(x) = \infty$. Then f attains its minimum on C.

Coercive F^n $f: \lim_{\|x\| \uparrow \infty} f(x) = \infty$

$\liminf_{n\uparrow\infty} x_n = \lim_{n\uparrow\infty} \inf_{m\ge n} x_m = \sup_n \inf_{m\ge n} x_n$

$\limsup_{n\uparrow\infty} x_n := \lim_{n\uparrow\infty} \sup_{m\ge n} x_m = \inf_n \sup_{m\ge n} x_n$

$\inf_{m\ge n} x_m$ $n\ge 1$ is monatone increasing in n.

If $\liminf_{n\uparrow\infty} x_n = \limsup_{n\uparrow\infty} x_n = x^*$, then $x^* = \lim_{n\uparrow\infty} x_n$

In general $\limsup_{n\uparrow\infty} f(x_n) \ge \liminf_{n\uparrow\infty} f(x_n)$

Lower Semi Continuous $f: \mathbb{R}^d \to \mathbb{R}$ is l.s.c. if $\liminf_{n\uparrow\infty} f(x_n) \ge f(x)$ whenever $x_n \to x$ in $\mathbb{R}^d$.

Semi U.S.C. - $\lim_{n\uparrow\infty} \sup f(x_n) \le f(x)$ whenever $x_n \to x$ in $\mathbb{R}^d$

A f^n that is both l.s.c. & u.s.c. will be continuous.

Theorem 1.5 A lower semicontinuous function on a closed bounded set C attains its minimum. An upper semicontinuous f^n on a closed bounded set C attains its minimum.

Theorem 1.6 If f is a pointwise supremum of continuous (more generally l.s.c.) functions, it is l.s.c. Likewise, ~~Theorem 1.7~~ if it is the pointwise infimum of continuous (more generally u.s.c.) functions, it is u.s.c.

Theorem 1.7 A lower semicontinuous f^n f on a closed set C satisfying $\lim_{\|x\|\uparrow\infty} f(x) = \infty$ attains its minimum.

Pointwise & Uniform Convergence Given $f_n, f: C \subset \mathbb{R}^d \mapsto \mathbb{R}$, $f_n \mapsto f$ pointwise if as $n \uparrow \infty$ $f_n(x) \to f(x)\ \forall x \in C$, and uniformly if $\sup_{x\in C} |f_n(x) - f(x)| \to 0$

Theorem 1.8: If continuous functions $f_n: C \subset \mathbb{R}^d \mapsto \mathbb{R}$, $n \ge 1$ converge uniformly to an $f: C \mapsto \mathbb{R}$ as $n \uparrow \infty$, then f is continuous.

Theorem 1.9 Any continuous $f: \mathbb{R}^d \mapsto \mathbb{R}$ is approximated uniformly on any closed bounded set $C \subset \mathbb{R}^d$ by polynomials. If it is k times continuously differentiable (i.e. diff k times with continuous derivatives upto k^{th} order), both the function & its derivatives upto order k or less can be approximated uniformly by a polynomial & its corresponding derivatives. (Weierstrass Approximation Theorem)

Special case of Urysohn's Lemma: Let $A, B \subset \mathbb{R}^d$ be disjoint closed sets and $a, b \in \mathbb{R}$ with $b > a$.

Lemma 1.1. There exists a continuous $f: \mathbb{R}^d \mapsto [a, b]$ s.t. $f(x) = a$ for $x \in A$ and $f(x) = b$ for $x \in B$

Theorem 1.10 (Tietze extension theorem) Let $C \subset \mathbb{R}^d$ be closed and $f: C \mapsto [a, b]$, $b > a$, be continuous. Then f extends continuously to a function $f^*: \mathbb{R}^d \mapsto [a, b]$

Theorem 1.11 (Banach Contraction Mapping Theorem) If $f: \mathbb{R}^d \mapsto \mathbb{R}^d$ satisfies $\|f(x) - f(y)\| \le \alpha \|x - y\|\ \forall x, y \in \mathbb{R}^d$ for some $0 < \alpha < 1$, then f has a unique fixed point x^* and the iteration $x_{n+1} = f(x_n)$ with any choice of $x_0 \in \mathbb{R}^d$ converges to x^*

Theorem 1.12 Arzela-Ascoli Theorem Let $C \subset \mathbb{R}^d$ be closed & bounded. If a family $\mathcal{A}$ of maps $C \mapsto \mathbb{R}$ is equicontinuous & satisfies $\sup_{f\in\mathcal{A}} f(x_0) < \infty$ for some $x_0 \in C$, then every sequence $\{f_n\} \in \mathcal{A}$ has a further subsequence that converges uniformly to some continuous $f: C \mapsto \mathbb{R}$. The converse also holds.

Ch2 Differentiability and local optima

Gateaux Derivative

If the limit $f'(x;h) := \lim_{\varepsilon\to 0} \frac{f(x+\varepsilon h)-f(x)}{h}$ exists

it is called Gateaux derivative along line defined by h.

Equivalently $f(x+\varepsilon h) = f(x) + f'(x;h)\,\varepsilon + o_h(\varepsilon)$

where $\frac{o_h(\varepsilon)}{\varepsilon} \to 0$ as $\varepsilon\to 0$, but at a rate that can depend on h.

Fréchet derivative

Say $f: \mathbb{R}^d \mapsto \mathbb{R}^m$ is Frechet differentiable if $\exists$ a linear map $D_x f: \mathbb{R}^d \mapsto \mathbb{R}^m$ (called Frechet derivative) such that

$$\sup_{\|h\|=1} \left\| \frac{f(x+\varepsilon h)-f(x)}{\varepsilon} - D_x f(h) \right\| \to 0 \text{ as } \varepsilon \to 0$$

$\Leftrightarrow$ $f(x+\varepsilon h) = f(x) + \varepsilon D_x f(h) + o(\varepsilon)$.

Difference with Gateaux differentiability is that $\frac{f(x+\varepsilon h)-f(x)}{\varepsilon}$ converges uniformly in h, $\|h\|=1$

Frechet Differentiability $\Rightarrow$ Gateaux differentiability.

If $m=1$, $D_x f: \mathbb{R}^d \mapsto \mathbb{R}$ is linear $D_x f(y) = \langle \nabla f(y), y\rangle$

For $m\geq 2$, and $f=(f_1,\cdots,f_m)$ $D_x f(y) = A_x y$ $\forall y\in\mathbb{R}^d$ is a linear map for each fixed x for some $m\times d$ matrix A_x

$(i,j)^{th}$ element of A_x is $\partial f_i(x)/\partial x_j$ $1\leq i\leq m$, $1\leq j\leq d$

A_x = Jacobian Matrix of f $(D_x f)$

- $x\mapsto \nabla f(x)$ $\mathbb{R}^d \mapsto \mathbb{R}^d$ Jacobian of this map is called Hessian $\nabla^2 f(x)$ (Symmetric)

First order & second order Taylor Formulas

$$f(x) = f(x_0) + \langle \nabla f(x_0), x-x_0\rangle + o(\|x-x_0\|)$$

$$f(x) = f(x_0) + \langle \nabla f(x_0), x-x_0\rangle + \frac{1}{2}(x-x_0)^T \nabla^2 f(x_0)(x-x_0) + o(\|x-x_0\|^2)$$

If the first (resp. the first two) derivative(s) exist at all points in domain of interest & are continuous, the f^n is said to be continuously (twice continuously) differentiable.

$$f(x) = f(x_0) + \int_0^1 \langle \nabla f((1-t)x_0 + tx), x-x_0\rangle\, dt$$

$$= f(x_0) + \langle \nabla f(x_0), x-x_0\rangle + \frac{1}{2}\int_0^1 (x-x_0)^T \nabla^2 f((1-t)x_0+tx)(x-x_0)\, dt.$$

Theorem 2.1 Fermat

Let f be either Gateaux or Frechet differentiable & let x^* be a local minimum. Then $D_{x^*}f(h)$, resp. $\nabla f(x^*) = \theta$ ($\forall h\in\mathbb{R}^d$ in the former case)

The converse does not hold: a point of zero gradient, also called Critical Point, can correspond to a local maximum, a saddle point (a pt at which for some lines passing through it, the f^n attains its local maximum while along other lines passing through it, attains local minimum), a point of inflection (point on curve where derivative of its slope changes sign).

Theorem 2.2 If x_0 satisfies: $\nabla f(x_0) = \theta$ and $\nabla^2 f(x_0)$ is positive definite, then x_0 is a strict local minimum.

If x_0 is a strict local minimum of f, $\nabla^2 f(x_0)$ is positive semidefinite.

Minimization on subset of $\mathbb{R}^d$.

Karush-Kuhn-Tucker conditions:

Let k, s be nonnegative integers and let $f, g_1,\cdots,g_k; h_1,\cdots,h_s: C\mapsto\mathbb{R}$ for some open set $C\subset\mathbb{R}^d$ be continuous & continuously differentiable.

Theorem 2.3 Let $x_0\in C$ satisfy the equality & inequality constraints: $g_i(x)=0$, $h_r(x)\leq 0$ $\forall i,r$. for $x=x_0$ and $f(x_0)\leq f(x)$ $\forall x\in C$. Then $\exists$ scalars $\lambda_0,\lambda_1,\cdots,\lambda_k$ & $\mu_1,\cdots,\mu_s$

s.t. $\lambda_0 \frac{\partial f}{\partial x_j}(x_0) + \sum_{i=1}^k \lambda_i \frac{\partial g_i}{\partial x_j}(x_0) + \sum_{i=1}^s \mu_i \frac{\partial h_i}{\partial x_j}(x_0) = 0$ $\forall j$

Furthermore (i) $\lambda_0 \geq 0$ and $\mu_r \geq 0$ $\forall r$;

ii) Complementary Slackness Conditions

$h_r(x_0) < 0 \Rightarrow \mu_r = 0$ for $1\leq r\leq s$;

iii) if $\nabla_i g(x_0)$, $1\leq i\leq k$, and those $\nabla h_r(x_0)$, r, $1\leq r\leq s$, for which $h_r(x_0)=0$ are linearly independent, then $\lambda_0 = 1$ w.l.o.g.

The f^n $f: C\times D \mapsto \mathbb{R}$, where $C\subset\mathbb{R}^d$ is open and $D\subset\mathbb{R}^m$ is closed bounded, is continuous and its partial gradient wrt x alonge $\nabla^x f(x) := \left[\frac{\partial f}{\partial x_1}(x,y),\cdots,\frac{\partial f}{\partial x_d}(x,y)\right] \in \mathbb{R}^d$ is assumed to be continuous. Let $g(x) = \max_{y\in D} f(x,y)$

Theorem 2.4 Danskin

The map $g:\mathbb{R}^d\mapsto\mathbb{R}$ has a directional derivative in every direction, given by

$$g'(x;\vec{n}) = \max_{y\in M(x)} \langle \nabla^x f(x,y), \vec{n}\rangle$$

for every unit vector $\vec{n}\in\mathbb{R}^d$

2.5 **Paramatric monotonicity of optimizers.**

$x\wedge y :=$ componentwise min. of x & y | $x\vee y =$ componentwise max. of x & y.

Submodular Function: $f(x)+f(y) \geq f(x\wedge y) + f(x\vee y)$

Supermodular if inequality reversed.

$f:\mathbb{R}^d\times\mathbb{R}^m\mapsto\mathbb{R}$ satisfies increasing (decreasing) differences property if $x\geq x'$, $y\geq y' \Rightarrow$

$$f(x,y)-f(x',y) \geq (\leq)\ f(x,y') - f(x',y')$$

Lemma 2.1: If $f:\mathbb{R}^d\times\mathbb{R}^m\mapsto\mathbb{R}$ is submodular (supermodular) then it has decreasing (increasing) differences.

C: closed & bounded, C&D closed under $\vee,\wedge$. $M(y) := \{x\in C: f(x,y) = \min_{x'} f(x',y)\}$

Theorem 2.5: Let f be continuous. If f satisfies decreasing differences and $f(\cdot,y)$ is submodular for each $y\in D$, then

i) $\forall y$, $M(y)$ is nonempty closed & bounded

ii) $M(y)$ satisfies: $x,x'\in M(y) \Rightarrow x\wedge x', x\vee x' \in M(y)$

iii) $\forall y$, $M(y)$ contains a minimal element $x_*(y)$, resp. a maximal $x^*(y)$ $\{$i.e. $\nexists x\in M(y)$ s.t. $x\neq x_*(y)$ & $x\leq x_*(y)\}$

iv) $y\mapsto x_*(y)$ is non-increasing,

v) If f satisfies strict decreasing differences, then $y\geq y', y\neq y' \Rightarrow \forall x\in M(y), x'\in M(y'), x\geq x'$

Theorem 2.6 Ekeland Variational Principle.

Let $f: C\subset\mathbb{R}^d\mapsto\mathbb{R}$, C open, be lower semicontinuous and bounded from below, and $\varepsilon>0$. Let $x^*\in C$ satisfy

$$f(x^*) \leq \inf_{x\in C} f(x) + \varepsilon$$

Then for any $\lambda>0$, $\exists x_\lambda\in C$ s.t.

a) $f(x_\lambda)\leq f(x^*)$ b) $\|x_\lambda - x^*\|\leq\lambda$

c) $f(x_\lambda) < f(x) + \frac{\varepsilon}{\lambda}\|x-x_\lambda\|$, $\forall x\neq x_\lambda$.

Theorem 2.7 The Mountain Pass Theorem

Let $f:\mathbb{R}^d\mapsto\mathbb{R}$ be a continuously differentiable f^n satisfying $\lim_{\|x\|\uparrow\infty} f(x) = \infty$, with at least two strict local minima x_1, x_2. Then it also has a third critical point x^* characterized by

$$f(x^*) = \inf_{C\in\mathcal{C}} \max_C f(x)$$

where $\mathcal{C} := \{C\subset\mathbb{R}^d : C$ is closed, bounded and connected, and $x_1,x_2\in C\}$.

Date ______
Page ______

Ch3 Convex Sets

Introduction A convex set $C \subset \mathbb{R}^d$ is a set such that any line segment joining two distinct points in C lies entirely in C.
Equivalently for any $n \geq 2$, $x_i \in C$, $1 \leq i \leq n$, and $\alpha_i \in [0,1]$, $1 \leq i \leq n$, with $\sum \alpha_i = 1$, we have $\sum_i \alpha_i x_i \in C$.

x is Strict convex combination if $\alpha_i \in (0,1)\ \forall i$

If C is closed, a simpler requirement suffices:
C is convex if $x, y \in C \Rightarrow \frac{1}{2}(x+y) \in C$.

Simple properties of convex sets

P1 A convex set is connected i.e. it cannot be written as the union of two sets that have disjoint open neighborhoods

P2 Intersection of an arbitrary collection of convex sets is convex when nonempty.

P3 Union of two convex sets need not be convex.

P4 Interior and closure of a convex set are convex

P5 Image of a convex set under a linear (affine) transformation is convex.

Theorem 3.1 Let $d \geq 2$ and $f: \mathbb{R}^d \mapsto \mathbb{R}^m$ be a continuous injective function such that f maps convex sets to sets. Then f is affine.

Convex Hull $co(A)$ smallest convex set containing arbitrary $A \subset \mathbb{R}^d$, alternatively as set of convex combinations of points in A.

Closed Convex Hull $\overline{co}(A)$: smallest closed convex set containing A / intersection of closed convex sets containing / closure of its convex hull.

The minimum distance problem

Let $C \subset \mathbb{R}^d$ be a closed convex set C & let $x \notin C$.

Theorem 3.2 There is a unique $x^* \in C$, such that
$$\|x - x^*\| = \min_{y \in C} \|x - y\|$$

If C is a set such that this theorem holds true, then the set C is called a Chebyshev set. Every Chebyshev set is convex in finite dimensions.

The above x^* is called the projection of x on C.

Theorem 3.3
Let C, D be disjoint closed convex sets in $\mathbb{R}^d$ with C bounded. Then $\exists\, x^* \in C$, $y^* \in D$ such that
$$0 < \|x^* - y^*\| = \min_{x \in C, y \in D} \|x - y\|$$
No uniqueness can be claimed.

Theorem 3.4
For a closed convex $C \subset \mathbb{R}^d$ and $x \notin C$, the point $x^* := \arg\min_{y \in C} \|x - y\|$ is characterized by $\forall y \in C$, $\langle y - x^*, x - x^* \rangle \leq 0$

Separation Theorems
Consider hyperplane $H := \{x \in \mathbb{R}^d : \langle x - x_0, \vec{n} \rangle = 0\}$ in $\mathbb{R}^d$
H defines two closed half spaces $L_H := \{x \in \mathbb{R}^d : \langle x - x_0, \vec{n} \rangle \leq 0\}$ and $U_H := \{x \in \mathbb{R}^d : \langle x - x_0, \vec{n} \rangle \geq 0\}$ that intersect in H.

Theorem 3.5
i) If C is closed convex and $x \notin C$, then $\exists$ a hyperplane H separating the two.
ii) If C, D are disjoint closed convex sets, there exists a hyperplane separating the two.

Corollary 3.1
The minimum distance between C and D equals the maximum separation between pairs of parallel hyperplanes separating the two.

Theorem 3.6
If C is closed convex & $x^* \in \partial C$, then $\exists$ a hyperplane H passing through x^* such that $C \subset L_H$ and $int(C) \subset int(L_H)$
Such a hyperplane is said to be a support hyperplane of C at x^*.

Theorem 3·7 ~~Theorem~~ Let C, D be closed convex sets such that $\text{int}(C) \neq \phi$, $D \cap \text{int}(C) = \phi$ and $D \cap C \neq \phi$
($\Rightarrow \exists r \in \partial C \cap D$)

There exists a hyperplane H such that $C \subset L_H$ and $D \subset U_H$

The condition that one of the two convex sets have a non-empty interior cannot be relaxed.

Extreme Points

A point $x \in C$ (a convex set) $\subseteq \mathbb{R}^d$ is said to be its extreme point if it cannot be expressed as a strict convex combination of two distinct points in C.

That is $x = \alpha y + (1-\alpha) z$, $y, z \in C$ $\alpha \in (0,1) \Rightarrow x = y = z$.

Clearly $x \in \partial C$

Lemma 3·1 A closed bounded convex set $C \subset \mathbb{R}^d$ has at least one extreme point

Theorem 3·8 A closed and convex set has an extreme point iff it contains no lines.

Let H be supporting hyperplane of C at $\bar{x}$. Then, $H \cap C$ lies in $(d-1)$ space & has no lines.

Consider $K := \{x \in \mathbb{R}^d : Ax \leq b\}$ $A_{m\times d}$, $b_{m\times 1}$

Theorem 3·9 Let $x \in K$. Then x is an extreme point of K iff some d inequalities corresponding to d linearly independent rows of the system $Ax \leq b$ are equalities i.e. $\langle a_i, x\rangle = b_i$ for i corresponding to those linearly independent rows.

Corollary 3·2 Set of extreme points of K is finite.

Simple facts about extreme points

Q1 Every boundary point may not be an extreme point (pts on boundary of rectangle other than its corners are not extreme points.)

Q2 A convex set may not have an extreme point if it is not closed (e.g. open ball in $\mathbb{R}^d$)

Q3 If $x \in \partial C$ is not an extreme point and $y, z \in C$ are such that $x = \alpha y + (1-\alpha) z$ for some $\alpha \in (0,1)$, then $y, z \in \partial C$.

Q4 An extreme point may have a unique supporting hyperplane that contains other non-extreme points.

$e(C)$ = the set of extreme points of C.

Theorem 3·10 (Krein-Milman theorem in finite dimension)

A closed bounded convex set C is the closed convex hull of $e(C)$.

Theorem 3·11

The set of extreme points of a closed bounded convex set $C \subset \mathbb{R}^d$ can be written as a countable intersection of relatively open sets in C.

Lemma 3·2

Let $C \subset \mathbb{R}^d$ be a closed bounded convex set and H a support hyperplane of C such that $C \subset L_H$ (say). Then $G := C \cap H$ is closed bounded and convex, and $e(G) \subset e(C)$.

Theorem 3·12 Caratheodory's Theorem

Every $x \in$ a closed bounded convex set $C \subset \mathbb{R}^d$ can be written as a convex combination of m points in $e(C)$ for some $m \leq d+1$

If C is bounded and $e(C)$ is finite, C is said to be a convex polytope & pts in $e(C)$ its 'corners'.

Define $e_i = x_{i+1} - x_i$ $1 \leq i \leq m-1$, then at most d vectors $\{e_i\}$ can be linearly independent if $m > d+1$ they are not.

Then for each $x \in C$, $x - x_1$ can be written as convex combination of e_i in more than one way.

If $\{e_i\}$ are linearly independent, these representations are unique. In such a case C is said to be a simplex. A simplex in $\mathbb{R}^d$ with non-empty interior is said to be a d-dimensional simplex or simply d-simplex.

Let S be a d-simplex with $e(C)=\{x_1,\cdots,x_{d+1}\}$

S1. For any m with $1\le m\le d$, and distinct $i_1,\cdots,i_m$ in $\{1,\cdots,d+1\}$, $co(\{x_{i_1},\cdots,x_{i_m}\})=\overline{co}(\{x_{i_1},\cdots,x_{i_m}\})\subset\partial S$ and is an $(m-1)$-simplex. These are called $(m-1)$-dimensional faces of S.

S2 ∂S is the union of $d+1$ distinct $(d-1)$ ~~indices~~ simplices whose intersections are $d-2$ simplices in their respective boundaries.

S3 Intersection of S with a hyperplane H not intersecting its interior is one of its faces. If H intersects $int(S)$, then $S\cap H$ is a polytope whose relative interior is $Int(S)\cap H$.

Let $C\subset\mathbb{R}^d$ be closed, bounded and convex let $H_i:=\{x\in\mathbb{R}^d:\langle r_i,x\rangle\le c_i\}$, $1\le i\le m$, be closed half spaces in $\mathbb{R}^d$ for some m, $1\le m\le d$. Let $H:=\bigcap_{i=1}^m H_i$ and $C^*:=C\cap H$. Then C^* is also closed bounded and convex. $\tilde{H}_i:=\{x\in\mathbb{R}^d:\langle r_i,x\rangle=c_i\}$, $1\le i\le m$, $\tilde{H}:=\partial C^*\cap(\bigcup_i\tilde{H}_i)$

Theorem 3.13 Every $x\in e(C^*)$ can be written as a convex combination of at most $m+1$ elements of $e(C)$.

Theorem 3.14 Shapley Folkman

Let $S_i\subset\mathbb{R}^d$, $1\le i\le n$, and $S:=$ the Minkowski sum $\sum_{i=1}^n S_i=\{\sum_{i=1}^n x_i : x_i\in S_i, 1\le i\le n\}$

Then each $x\in Co(S)$ can be written as $x=\sum_{i=1}^n x_i$ where $x_i\in Co(S_i)$ $\forall i$ and $x_i\in S_i$ for at least $n-d$ indices i.

Lemma 3.3 (Radon)

Let $A:=\{x_1,\cdots,x_m\}\subset\mathbb{R}^d$ where $m\ge d+2$. Then $\exists$ $A_1,A_2\subset A$ such that $A=A_1\cup A_2$, $A_1\cap A_2=\phi$ and $co(A_1)\cap co(A_2)\neq\phi$

Theorem 3.15 (Helly)

Let $C_i\in\mathbb{R}^d$, $1\le i\le m$, be convex, with $m\ge d+1$. If $\bigcap_{i\in I}C_i\neq\phi$ for any $I\subset\{1,\cdots,m\}$ with $|I|=d+1$, then $\bigcap_i C_i\neq\phi$

Brouwer fixed point theorem.

Take $B:=$ the closed unit ball in $\mathbb{R}^d$ centred at origin, though proofs extend easily to closed bouned convex set B.

Lemma 3.4

Let $\phi:\mathbb{R}^d\to\mathbb{R}^d$ is continuously differentiable with $\phi(x)=x$ when $\|x\|\ge 1$. Then ϕ is onto.

Corollary 3.3

If $\phi:B\mapsto\mathbb{R}$ is continuous with $\phi(x)=x$ on ∂B, then $\forall y\in B$, $y=\phi(x)$ for some $x\in B$

Corollary 3.4 (No retract theorem)

There is no continuous map $\phi:B\mapsto\partial B$ for which $\phi(x)=x$ for $x\in\partial B$

Theorem 3.16 (Brouwer Fixed Point Theorem)

Any continuous function $\phi:B\mapsto B$ has a fixed point.

Ch4 Convex Functions

Basic Properties : A f^n $f: C \mapsto R$ for a convex set $C \subset R^d$ is said to be convex if for any $\lambda \in [0,1]$ and $x, y \in C$, we have $f(\lambda x + (1-\lambda)(y)) \le \lambda f(x) + (1-\lambda) f(y)$.

Equivalent Def^n $f(\sum_{i=1}^n \lambda_i x_i) \le \sum_{i=1}^n \lambda_i f(x_i)$ where $\sum_i \lambda_i = 1$, $\lambda_i \in [0,1]$

Strictly Convex If the inequalities above are strict when x_i distinct

When we talk about multiple convex f^n, we talk on the intersection of their domains, which is convex when nonempty.

C1 Positive linear combinations of convex f^n are convex. $a_i \ge 0$ $\sum_{i=1}^n a_i f_i$ is cvx. It is strictly convex if at least one f_i for which $a_i > 0$ is.

C2 Pointwise limits of convex functions are convex. if f_n are cvx & $f(x) = \lim_{n \to \infty} f_n(x)$ exists $\forall x \in C$, then $f: C \to R$ is cvx. Even if all f_n are strictly convex, f need not be.

C3 Pointwise maxima or suprema of convex function are cvx.

C4 Pointwise minima of cvx f^n need not be cvx. However, if $f: C \times D \to R$ is cvx for cvx sets $C \subset R^d, D \subset R^m$, & $g(x) = \inf_{y \in D} f(x,y) > -\infty$ $\forall x$, then g is convex.

C5 If $f_1, \dots, f_m$ are cvx & $g: R^m \to R$ is cvx & componentwise increasing. then $h(\cdot) = g(f_1(\cdot), f_2(\cdot), \dots, f_m(\cdot)) : C \to R$ is cvx

C6 Let $\{f_\alpha\}$ be family of cvx f^n & $\varphi: R^m \mapsto [0, \infty)$. Suppose $f^*(x) := \int f_\alpha(x) \varphi(\alpha) d\alpha$ is well defined as Riemann integral $\forall x$. Then f^* is convex.

C7 If f is componentwise convex, it need not be jointly cvx. if $f(x, \cdot)$ & $f(\cdot, y)$ are cvx, $f(\cdot, \cdot)$ need not be. ex $f(x,y) = xy$

C8 Let $f: R^d \to R$ be cvx, $b \in R^d$ and A a $d \times m$ matrix. Then $x \in R^m \mapsto f(Ax + b) \in R$ is convex.

4.2 Continuity

Epigraph: set of points on or above graph of f. $\{(x,y) : y \ge f(x)\}$

Lemma 4.1 f is a convex f^n iff epi(f) is a convex set.

Application C3 $\text{epi}(\sup_\alpha f_\alpha) = \bigcap_\alpha \text{epi}(f_\alpha)$, C4 epi$(g)$ is the projection of epi(f) on x-space.

Let $f: C \mapsto R$ be convex for a convex $C \subset R^d$ with nonempty interior.

Theorem 4.1

f is continuous at any $x_0 \in \text{int}(C)$

In case $f: C \mapsto R$ is continuous, epi(f) is closed.

The converse is not true.

4.3 Differentiability

Theorem 4.2 A convex f^n $f: R^d \mapsto R$ is locally Lipschitz and is twice differentiable almost everywhere (diff. outside a set of measure zero)

- A celebrated theorem of Rademacher states that a Lipschitz f^n, is differentiable almost everywhere.

Lemma 4.2 1. Let $f: C \mapsto R$ be convex, where $C \subset R^d$ is convex and open. Then for any $x \in C$ where f is diff.

$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x) \quad \forall y \in C.$$

Conversely, if f is continuously differentiable & this inequality holds $\forall x, y \in C$, then f is convex.

2. If f above is twice continuously differentiable at $x \in C$, $\nabla^2 f(x)$ is positive semi-definite. Conversely if $\nabla^2 f$ is positive definite in C, f must be convex.

Theorem 4.3 Let $f: U \to R$ be a cvx f^n where $U \subseteq R^d$ is open. Suppose $z \in U$ is such that partial derivatives $\frac{\partial f}{\partial x_1}(z), \dots, \frac{\partial f}{\partial x_d}(z)$ exist. Then f is Frechet differentiable at z.

If f is not differentiable at x, define $\xi \in R^d$ to be a subgradient of f at x if

$$f(y) \ge f(x) + \langle \xi, y - x \rangle \qquad \xi =: \partial f(x)$$

Set of subgradients, called a subdifferential, is a closed cone.

For $f, g \in C \to R$ with C open

G1) $\partial(\lambda f)(x) = \lambda \partial f(x)$ $\forall x \in C$

G2) $\partial(f + g) \subset \partial f + \partial g$ $\forall x \in C$

G3) If f is Gateaux differentiable at x, $\partial f(x) = \{\text{Gateaux derivatives of } f \text{ at } x\}$

If it is Frechet differentiable, $\partial f(x) = \{\nabla f(x)\}$. The converse also holds.

G4 If $g(t) := f(x+tv)$ for $t \in \mathbb{R}$, then
$\partial g(t) \subset \{\langle \xi, v\rangle : \xi \in \partial f(x+tv)\}$

G5 For $x \neq y$, $f(y) - f(x) \in \{\langle \xi, y-x\rangle : \xi \in \partial f(x)\}$

G6 A point $x^* \in \text{int}(C)$ is a local minimum of f if $\theta \in \partial f(x^*)$. The reverse implication is true.

4.4 An approximation theorem

Theorem 4.4 A convex $f: \mathbb{R}^d \mapsto \mathbb{R}$ can be uniformly approximated on a closed bounded convex set $C \subset \mathbb{R}^d$ by convex polynomials on C.

cvx this time.

Weierstrass theorem. $\max_{x \in A} |f(x) - P_\varepsilon(x)| < \varepsilon$

4.5 Convex extensions.

Theorem 4.5 A convex f^n $f: C \subset \mathbb{R}^d \mapsto \mathbb{R}$ for a cvx and bounded C can be extended to a cvx f^n $\mathbb{R}^d \mapsto \mathbb{R}$ iff it is Lipschitz on C.

not necessarily convex

Theorem 4.6 A bounded convex $f: D \subset \mathbb{R}^d \mapsto \mathbb{R}$ for a bounded $D \subset \mathbb{R}^d$ extends to a convex $\tilde{f}: co(D) \mapsto \mathbb{R}$

Further properties of Gradients of Convex Functions

Given a convex f^n $f: \mathbb{R}^d \to \mathbb{R}$

Monotonicity of Gradient $\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq 0$

Lemma 4.4 Let $f: \mathbb{R}^d \to \mathbb{R}$ be a convex & differentiable f^n s.t. $\langle \nabla f(x), x\rangle \geq 0$ for each $x \in \mathbb{R}^d$. Then zero vector θ is minimizer of f.

Theorem 4.7 Maximal monotonicity of Gradient

Let $f: \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex f^n s.t.
$\langle \nabla f(x) - q, x-y\rangle \geq 0 \; \forall x \in \mathbb{R}^d$, then $\nabla f(y) = q$

Theorem 4.8 Let $f: \mathbb{R}^d \to \mathbb{R}$ be convex & differentiable $\forall x \in \mathbb{R}^d$
Then f is continuously differentiable.

Theorem 4.9 Minty's surjectivity theorem Let $f: \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable convex function. Then for each $u \in \mathbb{R}^d$, the equation $x + \nabla f(x) = u$ has a unique solution (i.e. $x \mapsto x + \nabla f(x)$ is invertible)

Moreau Envelope $f_\mu(x) = \inf_{y \in \mathbb{R}^d} f(y) + \frac{1}{2\mu}\|x-y\|^2$

Proximal Operator $\text{prox}_{\mu f}(x) = \arg\min_{y \in \mathbb{R}^d} \left(f(y) + \frac{1}{2\mu}\|x-y\|^2 \right)$

For $\mu = 1$ $\quad x = \text{prox}_f(u)$ iff $x + \nabla f(x) = u$

If f is Lipschitz continuous, $f_\mu \to f$ uniformly as $\mu \to 0$

① $\text{prox}_{\mu f}(x) + \text{prox}_{\mu f^*}(x) = x$ ② $f_\mu(x) + f^*_\mu(x) = \frac{1}{2}\|x\|^2$

Ch5 Convex Optimization

Introduction Consider a convex continuous $f: C \subset \mathbb{R}^d$ where C is closed & cvx

C1 If C is bounded as well, f attains a minimum. Immediate from Weierstrass theorem.

C2 Local minima of f are also its global minima.

C3 The set of (necessarily global) minima of f is a closed convex set. The set of minima of a continuous f^n is in any case closed.

Legendre transform or Conjugate convex function of f:

$y \mapsto f^*(y) := \sup_{x\in C}\{\langle x,y\rangle - f(x)\}$

Here we will consider f^* to be defined on the set

$C^* := \{y\in \mathbb{R}^d : f^*(y) < \infty\}$.

f^* is pointwise supremum of affine f^n hence convex.

Fenchel-Young Inequality $f(x) + f^*(y) \geq \langle x,y\rangle$

Theorem 5.1 $(f^*)^*(x) = f(x)$; $x \in C$

Theorem 5.2 $\partial f := \{y\in C^* : f^*(y) = \langle x,y\rangle - f(x)\}$ i.e. the subdifferential of f at x is precisely the set of y for which equality in the Fenchel-Young inequality is attained.

The non differentiability of f corresponds to a flat patch for the graph of f^* and vice versa.

A f^n $g: D \to \mathbb{R}$, $D \subset \mathbb{R}^d$ closed convex, is concave if for any $\lambda\in[0,1]$ & $x,y\in D$ $\lambda g(x) + (1-\lambda) g(y) \leq g(\lambda x + (1-\lambda) y)$

Line segment joining two points on curve lies below graph of g

Hypograph of g, hypo(g) $\{(x,r)\in D\times\mathbb{R} : r \leq g(x)\}$, is cvx.

Continuity of $g \Rightarrow$ hypo(g) is closed. The concave conjugate f^n for g is: $g_*(y) = \inf_{x\in D} \langle x,y\rangle - g(x)$

viewed as real concave f^n on $D^* := \{y\in\mathbb{R}^d : g_*(y) > -\infty\}$.

Corresponding FY inequality. $g(x) + g_*(y) \leq \langle x,y\rangle$

Theorem 5.3 Fenchel Duality Theorem For f, g, C, D as above, assume $\text{int}(C)\cap\text{int}(D) \neq \phi$

Then $C^*\cap D^* \neq \phi$ and

$$\inf_{x\in C\cap D} f(x) - g(x) = \max_{y\in C^*\cap D^*} g_*(y) - f^*(y)$$

min. vertical separation b/w epigraph of f & hypograph of g. = max. vertical separation b/w pairs of parallel hyperplanes separating

Legendre transform of a continuous but possibly nonconvex f.

f^{**} is the convex minorant of f i.e the largest convex function $\leq f$

An imp result in non convex case:

Theorem 5.4 Let $f: \mathbb{R}^d \to \mathbb{R}$ be differentiable f^n then a point $\bar{x}$ is a global min of f iff $\nabla f(\bar{x}) = 0$ and $f^{**}(\bar{x}) = f(\bar{x})$

Lagrange Multiplier Rule

Consider continuous convex $f: C \mapsto \mathbb{R}$ on a closed cvx $C\subset\mathbb{R}^d$ subject to constraints $g_i(x) \leq 0$, $1\leq i\leq m$ [$g(x)\leq\theta$] where $g_i: C\mapsto\mathbb{R}$ are convex & continuous.

$\tilde{C} := C\cap(\cap_{i=1}^m \{x : g_i(x)\leq 0\})$ is closed convex. Assume $\mu_0 := \inf_{x\in\tilde{C}} f(x)$ is finite.

Theorem 5.5 Lagrange Multiplier Rule

Suppose $\exists x_1$ satisfying $g(x_1) < \theta$ (Slater Rule). Then $\exists$ (Lagrange Multiplier) $\Lambda \geq \theta$ such that

$$\mu_0 = \inf_{x\in C} f(x) + \Lambda^T g(x)$$

Furthermore, if this infimum is attained at some $x^*\in\tilde{C}$, then x^* minimizes f on $\tilde{C}$ and $\langle \Lambda, g(x^*)\rangle = 0$ (Complementarity Cond^n)

Suppose the infimum is attained at $x_0\in C$. Define 'Lagrangian' $L(x,y) = f(x,y) + y^T g(x)$

Theorem 5.6 Saddle point property. The Lagrangian $L(x,y) = f(x) + y^T g(x)$ has saddle point at (x_0, Λ) i.e. $L(x_0, y) \leq L(x_0,\Lambda) \leq L(x,\Lambda)$ $\forall x\in C$, $\theta\leq y\in\mathbb{R}^m$

The minimization of $x\mapsto L(x,\Lambda)$ over $x\in C$ & the maximization of $y\mapsto L(x_0,y)$ over $y\geq\theta$ are viewed as being dual to each other. This is Lagrangian Duality

Let C be a closed convex set in $(\mathbb{R}^d)^+ = \{x = [x_1, \cdots, x_d] \in \mathbb{R}^d : x_i > 0 \ \forall i\}$

A point $x = [x_1, \cdots, x_d]$ in C is said to be a Pareto point if one has: $y = [y_1, \cdots, y_d] \in C$ satisfies $y_i \le x_i \ \forall i$ iff $x = y$.

Theorem 5.7 Arrow-Barankin-Blackwell Theorem. Let $Q(C) := \{x \in C : \exists\, l \in (\mathbb{R}^d)^+ \text{ s.t. } \langle l, x\rangle = \min_{y \in C} \langle l, y\rangle\}$.

Then $Q(C)$ is dense in $P(C)$.

Corollary 5.1 If C above is a polytope, then $Q(C) = P(C)$

Linear Programming

$$\inf_{x \in \mathbb{R}^d} \langle c, x\rangle \quad \text{s.t. } A_{m \times d}\, x_{d \times 1} = b,\ x \ge \theta$$

Let $F = \{x \in \mathbb{R}^d_+ : Ax = b\}$ be the feasible set.

Theorem 5.8 If F has an extreme point and the LP has an optimal solution, then the LP has an optimal solⁿ which is an extreme point of F.

Corresponding to LP, which we call the Primal problem, we have a Dual problem given by

$$\begin{cases} \sup_{y \in \mathbb{R}^m} \langle y, b\rangle \\ \text{s.t. } y^T A \le c^T \end{cases}$$

Let $\alpha = \inf_{x \in F} \langle c, x\rangle$ & $\beta = \sup_{\{y : y^T A \le c^T\}} \langle y, b\rangle$

Theorem 5.9 Weak Duality. Let x, y be feasible solⁿ for Primal & Dual. Then $\langle y, b\rangle \le \langle c, x\rangle$. In particular $\beta \le \alpha$.

Theorem 5.10 Strong Duality. If one of the Primal or Dual is feasible and bounded, then $\alpha = \beta$.

Applications to game theory.

Let $f : C \times D \mapsto \mathbb{R}$ denote payoff function. A pair (x^*, y^*) is said to be a saddle point for f if

$$f(x^*, y) \le f(x^*, y^*) \le f(x, y^*) \quad \forall x \in C, y \in D$$

If saddle point exists $\max_y f(x^*, y) \le f(x^*, y^*) \le \min_x f(x, y^*)$

$$\Rightarrow \min_x \max_y f(x,y) \le \max_y \min_x f(x,y).$$

Since $\min_x \max_y f(x,y) \ge \max_y \min_x f(x,y)$

Thus $\min_x \max_y f(x,y) = \max_y \min_x f(x,y)$

Key step in establishing equality is existence of saddle point.

Theorem 5.11 Min-max Theorem

Let C, D closed bounded convex subsets of $\mathbb{R}^d, \mathbb{R}^n$, $f(x,y) : (x,y) \in C \times D \mapsto \mathbb{R}$ continuous, $f(\cdot, y)$ convex $\forall y \in D$, $f(x, \cdot)$ concave $\forall x \in C$. Then

$$\min_x \max_y f(x,y) = \max_y \min_x f(x,y)$$

Existence of Nash Equilibrium

N-person noncooperative game with payoff f^n

$f = [f_1, \cdots, f_N]^T : C = \prod_{i=1}^N C_i \subset \mathbb{R}^{\sum_{i=1}^N d_i} \mapsto \mathbb{R}^N$

where $C_i \subset \mathbb{R}^{d_i}$, $d_i \ge 1$, is closed bounded & convex.

For $x_i \in C_i$, $1 \le i \le N$, $x_{-i} := [x_1 : \cdots : x_{i-1} : x_{i+1} : \cdots : x_N]$ write $x = [x_i : x_{-i}]$ for any i. Assume f is componentwise convex i.e. $x_i \mapsto f_i([x_i, x_{-i}])$ is convex and continuous for each i & each fixed value of x_{-i}.

We call $x^* = [x_1^* : \cdots : x_N^*]$, $x_i^* \in \mathbb{R}^{d_i}$, a Nash Eq if for each i, x_i^* minimizes the map $x \in \mathbb{R}^{d_i} \mapsto f_i([x, x^*_{-i}])$

Assume f_1 is cvx in x for each fixed y & f_2 is cvx in y for each fixed x. Also assume both f_1 & f_2 are jointly continuous.

Theorem 5.12 Existence of Nash Equilibrium.

There exists a Nash Equilibrium x^*.