I. Introduction

1 Optimization problems

find $\hat{u} \in V$ (→ space) such that $\qquad u \in \mathbb{R}^N$

(P) $\qquad \underbrace{\hat{u} \in U}_{\text{constraints}}$ and $\qquad F(\hat{u}) = \inf_{u \in U} \underbrace{F(u)}_{\text{cost-function}}$

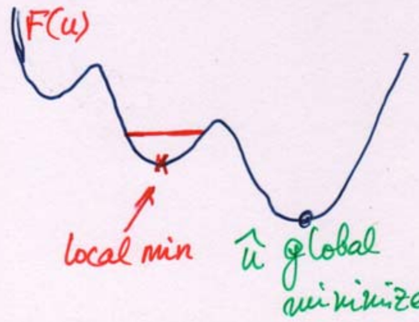
equality constraints $\quad U = \{u \in \mathbb{R}^N : g_i(u) = 0 \; 1 \le i \le p\}$

inequality " $\quad U = \{ \dots \quad h_i(u) \le 0 \; 1 \le i \le q\}$

$\mathcal{O}(U) \equiv$ an open subset containing U

$\hat{u}$ is a local minimum w.r.t. U
if $\exists r > 0$ such that $B(\hat{u}, r) \subset U$
ball of radius r centered at $\hat{u}$

and $\quad F(\hat{u}) \le F(u) \quad \forall u \in B(\hat{u}, r)$



$\hat{u}$ is a global minimizer wrt U
if $\quad F(\hat{u}) \le F(u) \; \forall u \in U$

2 Examples of opt. problems.

- Biochemistry – find the geometry of a molecula = find the positions of atoms in space that minimize the potential energy.
- Meteorology
  minimize cost: actual trend of weather according to previous experience
- Optimal control
  e.g. optimize the trajectory of a vehicle
  optimize the production of power plants.
- Image reconstruction (by minimizing an erro

3 Iterative algorithms

finding the solution at one step is exceptional.
usually iterative algorithms =
$\{u_k\}_{k \ge 0}$ sequence that converges to $\hat{u}$ = solution to (P)

u_0 – initialization

$u_k = \underbrace{G}_{\text{iterative scheme (may be given implicitely)}}(u_{k-1})$

G is constructed using informations on P

e.g. $F(u_k)$, $\nabla F(u_k)$, the constraints ...

Def. $\hat{u}$ is a fixed point for $G \iff G(\hat{u}) = \hat{u}$.

Thm Fixed point theorem.

$G : V \to V$ (complete metric space)

G is said to be a contraction: If $\exists\, c \in (0,1)$ such that

$$d(Gu_1, Gu_2) \le c\, \underset{\text{distance}}{d(u_1,u_2)} \quad \forall u_1, u_2 \in V$$

Then G is Lipschitzian and has a unique fixed point $\hat{u} = G(\hat{u})$.

G is lip. if $\exists$ const. γ s.t. $|G(u) - G(v)| \le \gamma |u-v| \quad \forall\, u,v \in V$

(locally lip. if ---- holds $u, v \in$ neighborhood)

Proof. Uniqueness: let $\hat{u}_1, \hat{u}_2$ 2 fixed point

$$0 < d(G\hat{u}_1, G\hat{u}_2) \le c\, d(\hat{u}_1, \hat{u}_2) < d(\hat{u}_1, \hat{u}_2)$$

$\| \; d(\hat{u}_1, \hat{u}_2)$ contradiction. It remains that $\hat{u}_1 = \hat{u}_2$

Existence – proof using the method of successive projections.

$u_0 \in V$ arbitrarily

Define: $u_1 = G(u_0), \dots, u_k = G(u_{k-1}), \dots$

we will show that $\{u_k\}$ is a Cauchy sequence

$d(u_2, u_1) \le c\, d(u_1, u_0)$

$d(u_3, u_2) \le c\, d(u_2, u_1) \le c^2 d(u_1, u_0)$

...

$d(u_{k+1}, u_k) \le c^k d(u_1, u_0)$ not enough to prove convergence

$d(u_{k+l}, u_k) \le d(u_{k+l}, u_{k+l-1}) + \dots + d(u_{k+1}, u_k)$

(property of distance: $d(u,v) \le d(u,a) + d(v,a) \ \forall a$.)

$\le c^k (c^{l-1} + \dots + c^0)\, d(u_1, u_0) \le \frac{c^k}{1-c} d(u_1, u_0)$

$\Rightarrow d(u_{k+l}, u_k) \xrightarrow[k\to\infty]{} 0 \Rightarrow u_k$ Cauchy sequence

V complete $\Rightarrow \exists$ limit point $\hat{u} \in V$

$u_k \to \hat{u}$ and G continuous $\Rightarrow G(u_k) \to G(\hat{u})$

$$0 = \lim_{k\to\infty} (u_{k+1} - G(u_k)) = \hat{u} - G(\hat{u}) = 0$$

4. Speed of Convergence

(to characterize asymptotically the speed of $u_k \to \hat{u}$)

Two classical ways to measure (cf. book ORTEG

– Q-convergence factor (Q = quotient)

$$Q((u_k)_{k\in\mathbb{N}}) = \limsup_{k\to\infty} \frac{\|u_{k+1} - \hat{u}\|}{\|u_k - \hat{u}\|}$$

– R-convergence factor (ROOT)

$$R((u_k)_{k\in N}) = \limsup_{k\to\infty} \|u_k - \hat{u}\|^{1/k}$$

Q

- the worst speed is $\sup_{u_0} Q((u_k))$
- Q depends on the norm
- $Q(G_1,\hat{u}) < Q(G_2,\hat{u})$ means G_1 faster than G_2
- The convergence said
  - super-linear: $Q(G,\hat{u})=0$
  - Q-linear if $0<Q<\infty$
  - sublinear if $Q(G,\hat{u})=\infty$ very bad

R

- independent of the norm $\|\cdot\|$
- $u_k \to \hat{u}$ then $0 \leq R \leq 1$
- the worst speed $R(G,\hat{u}) = \sup_{u_0} R((u_k)_{k\in N})$
- $R(G_1,\hat{u}) < R(G_2,\hat{u})$ means G_1 is faster than G_2
  - super-linear $R(G,\hat{u})=0$
  - R linear if $0<R<1$
  - sublinear $R(G,\hat{u})=1$

Theorem (Ostrowski)

matrix $A \in \mathbb{R}^{N\times N}$

$G: R^N \to R^N$ differentiable at $\hat{u}$ and $G(\hat{u})=\hat{u}$ fixed point

If $\max_{1\leq i\leq N} |\lambda_i \nabla G(\hat{u})| < 1 \Rightarrow \exists\, \mathcal{O}(\hat{u})$, $u_k \in \mathcal{O}(\hat{u})$ and $u_k \to \hat{u}$.

λ_i eigenvalue, $\mathcal{O}(\hat{u})$ neighborhood

The linear convergence theorem

same conditions $\max_{1\leq i\leq N} |\underbrace{\lambda_i}_{\text{eigenvalue}}(\nabla G(\hat{u}))| = R(G,\hat{u})$ root conv. factor

5. Existence, uniqueness

Def. $F:\mathbb{R}^N \to \mathbb{R}$ is said to be coercive if $\lim_{\|u\|\to\infty} F(u) = \infty$

Thm. (Existence) $F: V \to \mathbb{R}$ continuous, V finite dim.

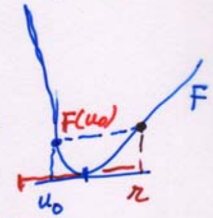
$U \subset V$. If U is unbounded we suppose that F is coercive

$\Rightarrow \exists\, \hat{u} \in U$ such that $F(\hat{u}) = \inf_{u\in U} F(u)$

Proof. (compacity)

- If U bounded, closed $\Rightarrow$ U compact, F continuous $\xRightarrow{\text{Weierstrass thm}}$ $\exists\, \hat{u}$.
- choose $u_0 \in U$ arbitrarily

since F coercive, $\exists\, r>0$ such that $\|u\| > r \Rightarrow F(u_0) < F(u)$



$\Rightarrow \hat{u}$ satisfies $F(\hat{u}) = \inf \{ F(u) : u \in \underbrace{\overline{B(u_0,r)} \cap U}_{\text{compact set}} \}$ closed ball

Apply the Weierstrass thm again $\Rightarrow \exists\, \hat{u}$.

Second approach for the proof. (SKETCH)

minimizing sequence $(u_k)_{k\geq 0} \Leftrightarrow u_k \in U$ and $\lim_{k\to\infty} F(u_k) = \inf_{u\in U} F(u)$

$(u_k)_{k\geq 0}$ is bounded (because either U bounded or F coercive)

$\Rightarrow \exists$ convergent subsequence $u_{k'} \in U$

$u_{k'} \to \hat{u}$

F continuous $\Rightarrow F(\hat{u}) = \lim_{k'\to\infty} F(u_{k'}) = \inf_{u\in U} F(u)$

because $(u_{k'})_{k'}$ minimizing sequence

- possible extension to Hilbert spaces.
- only sufficient condition!!!!

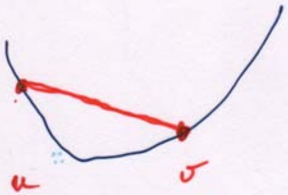
Def. (Convexity)

$F: U \to \mathbb{R}$ convex $\iff \forall\, u, v \in U$ and $\theta \in (0,1)$

$F(\theta u + (1-\theta) v) \le \theta F(u) + (1-\theta) F(v)$

F strictly convex if equality $\iff u = v$

$U \subset V$ convex $\iff \forall\, u, v \in U$ and $\theta \in (0,1)$
$\theta u + (1-\theta) v \in U$



non convex

F can be convex but not coercive

Thm. $F: U \to \mathbb{R}$ convex, continuous

(a) if F has a min at $\hat{u}$ $\Rightarrow$ this minimum is global (with respect to U)

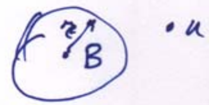
(b) $\hat{U} = \{ \hat{u} \in U \text{ such that } F(\hat{u}) = \inf_{u \in U} F(u) \}$ is closed and convex

(c) if F is strictly convex, then F has at most one minimum and the latter is strict.

Proof of the thm.

(a) $\exists\, r > 0$ such that $F(\hat{u}) \le F(u)$ $\forall u \in \overline{B(\hat{u}, r)} \cap U$

$u \in U \setminus \overline{B(\hat{u}, r)}$ (arbitrarily)



$\theta \overset{\text{def}}{=} \dfrac{r}{\|u - \hat{u}\|} \quad 0 < \theta < 1$

$u_\theta = (1-\theta)\hat{u} + \theta u \Rightarrow u_\theta \in \overline{B(\hat{u}, r)} \cap U$

because
$u_\theta - \hat{u} = \theta(u - \hat{u}) = r \dfrac{u - \hat{u}}{\|u - \hat{u}\|}$

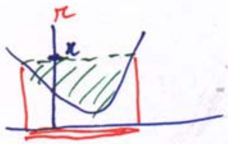
$F(\hat{u}) \le F(u_\theta) = F((1-\theta)\hat{u} + \theta u)$ F is convex

$\le (1-\theta) F(\hat{u}) + \theta F(u)$

$\Rightarrow \cancel{\theta} F(\hat{u}) \le \cancel{\theta} F(u) \Rightarrow \hat{u}$ is a global minimizer

Remark: if F strictly convex then strict minimum

Definition Epigraphe epi $F = \{(u, r) \in \mathbb{R}^N \times \mathbb{R} : F(u) \le r\}$



- F convex $\iff$ epi F convex
- F continuous $\Rightarrow$ epi F closed in $\mathbb{R}^N \times \mathbb{R}$.

(b) $\hat{U} = \{ u \in \mathbb{R}^N : F(u) \le F(\hat{u}) \} \cap U \Rightarrow$ convex closed

proj. of epigraph. (convex) closed.

convex closed

Reminders

- $f: \mathbb{R}^N \to \mathbb{R}$ differentiable at u if
  $\exists$ $Df(u)$ linear operator ($\mathbb{R}^N \to \mathbb{R}$) such that

$$\boxed{f(u+v) = f(u) + Df(u)\, v + \|v\|\, \varepsilon(v) \quad \text{where } \varepsilon(v) \xrightarrow[\|v\| \to 0]{} 0}$$

- Differential and gradient: (canonical isomorphism)

$$\langle \nabla f(u), v \rangle = Df(u)\, v, \quad \forall v$$

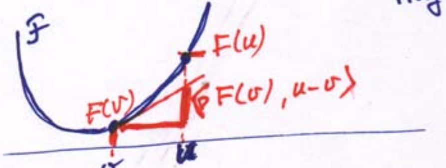
scalar product

Several properties:

an open set containing U

① • $F: U \to \mathbb{R}$ differentiable on $\mathcal{O}(U)$ and U convex

F is convex on U
$\iff F(u) \geq F(v) + \langle \nabla F(v), u-v \rangle, \ \forall u, v \in U.$



Proof ($\Rightarrow$) F convex, $v, u \in U$, $\theta \in (0,1)$

Using the definition of convexity!

$$\underbrace{F(\theta u + (1-\theta) v)}_{v + \theta(u-v)} - F(v) \leq \theta(F(u) - F($$

divide both sides by θ and let $\theta \searrow 0$

$$\langle \nabla F(v), u-v \rangle = \lim_{\theta \searrow 0} \frac{F(v + \theta(u-v)) - F(v)}{\theta} \leq F(u) - F(v)$$

($\Leftarrow$) $F(u) \geq F(u + \theta(v-u)) - \theta \langle \nabla F(u + \theta(v-u)), v-u \rangle$ multip. $\times\, 1-\theta$ both sides

$F(v) \geq F(u + \theta(v-u)) + (1-\theta) \langle \nabla F(u + \theta(v-u)), v-u \rangle$ $\times\, \theta$

ADD:

$$(1-\theta) F(u) + \theta(v) \geq F(u + \theta(v-u))$$

② F strictly convex on U (convex) $\iff F(u) > F(v) + \langle \nabla F(v), v-u \rangle$ $\forall u, v \in U$ $u \neq v$

Suppose F is twice differentiable

③ F is convex on $U \iff \langle \nabla^2 F(v)(u-v), u-v \rangle \geq 0$ $\forall u, v \in$

Hessian matrix

④ If $\text{- - - - - - - -} > 0$ $\forall u, v \in$ $u \neq v$
then F is strictly convex on U

(the converse statement can be false, e.g. $F(u) = u^4$ for $u = 0$: F is strictly convex and $F''(0) = 0$.)

Definition F is elliptic if $F \in C^1$ and $\exists \mu > 0$ such that
$\langle \nabla F(u) - \nabla F(v), u-v \rangle \geq \mu \|u-v\|^2$ $\forall u, v$

example: the standard form of quad. func. $F(u) = \frac{1}{2}\langle Bu, u \rangle - \langle c, u \rangle$
F elliptic $\iff B$ = symmetric and positive definite

Properties

⑤ $F: \mathbb{R}^N \to \mathbb{R}$ elliptic $\Longrightarrow$ F is strictly convex, coercive and
$F(u) - F(v) \geq \langle \nabla F(v), u-v \rangle + \frac{\mu}{2}\|u-v\|^2$
$\forall u, v \in \mathbb{R}^N$

By Taylor expansion

$$F(u) - F(v) = \int_0^1 \langle \overbrace{\nabla F(v + t(u-v))}^{-\nabla F(v) + \nabla F(v)}, u-v \rangle\, dt = \langle \nabla F(v), u-v \rangle$$

$$+ \int_0^1 \frac{1}{t} \langle \nabla F(\overbrace{v + t(u-v)}^{u}) - \nabla F(v), \overbrace{t(u-v)}^{u-v} \rangle\, dt$$

$$\geq \langle \nabla F(v), u-v \rangle + \int_0^1 \frac{1}{\cancel{t}} \mu t^{\cancel{2}} \|u-v\|^2\, dt \qquad \int_0^1 t\, dt$$

$$= \langle \nabla F(v), u-v \rangle + \frac{1}{2}\mu \|u-v\|^2 \Longrightarrow F \text{ coercive}$$

$F(u) > F(v) + \langle \nabla F(v), u-v \rangle$ so F strictly convex

⑥ Let F be twice differentiable.

F elliptic $\iff \langle \nabla^2 F(v)u, u\rangle \geq \mu \|u\|^2 \quad \forall u, v.$

Proof. $(\Rightarrow)$

def. of second derivative

$$\langle \nabla^2 F(u) w, w\rangle = \lim_{t\to 0} \frac{\langle \nabla F(\overbrace{u+tw}^{v}) - \nabla F(u), tw\rangle}{t^2} \geq$$

def. of ellipticity

$$\geq \lim_{t\to 0} \mu \frac{\cancel{t^2}\|w\|^2}{\cancel{t^2}} = \mu \|w\|^2$$

$(\Leftarrow)$ $G(w) \overset{\text{def}}{=} \langle \nabla F(w), v-u\rangle$

Mean-value-thm: $\theta \in (0,1)$

$$\langle \nabla F(v) - \nabla F(u), v-u\rangle = G(v) - G(u) = \langle \nabla G(u+\theta(v-u)), v-u\rangle$$

$$= \lim_{t\to 0} \frac{G(u+(\theta+t)(v-u)) - G(u+\theta(v-u))}{t}$$

no need.

$$= \lim_{t\to 0} \frac{\langle \nabla F(u+(\theta+t)(v-u)) - \nabla F(u+\theta(v-u)), \cancel{t}(v-u)\rangle}{t^{\cancel{2}}}$$

$$= \langle \nabla^2 F(\underbrace{u+\theta(v-u)}_{v})(\underbrace{v-u}_{u}), v-u\rangle \geq \mu \|v-u\|^2 \quad \blacksquare$$