Optimizing condition numbers

Pierre Maréchal

Toulouse (France)

Joint work with Jane Ye

Victoria (Canada)

The problem

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

Ω compact convex subset of S_n^+
(not containing the null matrix)

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

Ω compact convex subset of S_n^+
(not containing the null matrix)

$\kappa(A)$ condition number of A

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

Ω compact convex subset of S_n^+
(not containing the null matrix)

$\kappa(A)$ condition number of A

$$\kappa(A) = \begin{cases} \lambda_1(A)/\lambda_n(A) & \text{if } \lambda_n(A) > 0 \\ \infty & \text{if } \lambda_n(A) = 0 \text{ and } \lambda_1(A) > 0 \\ 0 & \text{if } A = 0 \end{cases}$$

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

Ω compact convex subset of S_n^+
(not containing the null matrix)

$\kappa(A)$ condition number of A

$$\kappa(A) = \begin{cases} \lambda_1(A)/\lambda_n(A) & \text{if } \lambda_n(A) > 0 \\ \infty & \text{if } \lambda_n(A) = 0 \text{ and } \lambda_1(A) > 0 \\ 0 & \text{if } A = 0 \end{cases}$$

$\lambda_1(A) \geq \ldots \geq \lambda_n(A)$ eigenvalues of A

The problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

Ω compact convex subset of S_n^+
(not containing the null matrix)

$\kappa(A)$ condition number of A

$$\kappa(A) = \begin{cases} \lambda_1(A)/\lambda_n(A) & \text{if } \lambda_n(A) > 0 \\ \infty & \text{if } \lambda_n(A) = 0 \text{ and } \lambda_1(A) > 0 \\ 0 & \text{if } A = 0 \end{cases}$$

$\lambda_1(A) \geq \ldots \geq \lambda_n(A)$ eigenvalues of A

Properties of κ

Properties of κ

- κ is nonconvex

Properties of κ

- κ is nonconvex
- κ is not everywhere differentiable

Properties of κ

- κ is nonconvex
- κ is not everywhere differentiable
- κ is quasiconvex (i.e. has convex level sets)

Properties of κ

- κ is nonconvex
- κ is not everywhere differentiable
- κ is quasiconvex (i.e. has convex level sets)
- κ is difference convex

Properties of κ

- κ is nonconvex
- κ is not everywhere differentiable
- κ is quasiconvex (i.e. has convex level sets)
- κ is difference convex
- κ is pseudoconvex

κ as a difference-convex function (from JBHU)

κ as a difference-convex function (from JBHU)

$$\kappa = \frac{\lambda_1}{\lambda_n} = \frac{1}{2}\left[\left(\lambda_1 + \frac{1}{\lambda_n}\right)^2 - \left(\lambda_1^2 + \frac{1}{\lambda_n^2}\right)\right]$$

κ as a difference-convex function (from JBHU)

$$\kappa = \frac{\lambda_1}{\lambda_n} = \frac{1}{2}\left[\left(\lambda_1 + \frac{1}{\lambda_n}\right)^2 - \left(\lambda_1^2 + \frac{1}{\lambda_n^2}\right)\right]$$

$1/\lambda_n$ is log-convex (thus convex):

$$\ln \frac{1}{\lambda_n} = -\ln \lambda_n$$

κ as a difference-convex function (from JBHU)

$$\kappa = \frac{\lambda_1}{\lambda_n} = \frac{1}{2}\left[\left(\lambda_1 + \frac{1}{\lambda_n}\right)^2 - \left(\lambda_1^2 + \frac{1}{\lambda_n^2}\right)\right]$$

$1/\lambda_n$ is log-convex (thus convex):

$$\ln \frac{1}{\lambda_n} = -\ln \lambda_n$$

λ_n is concave

κ as a difference-convex function (from JBHU)

$$\kappa = \frac{\lambda_1}{\lambda_n} = \frac{1}{2}\left[\left(\lambda_1 + \frac{1}{\lambda_n}\right)^2 - \left(\lambda_1^2 + \frac{1}{\lambda_n^2}\right)\right]$$

$1/\lambda_n$ is log-convex (thus convex):

$$\ln \frac{1}{\lambda_n} = -\ln \lambda_n$$

λ_n is concave

$-\ln$ is convex decreasing

Example: Markovitz model for portfolio selection

Example: Markovitz model for portfolio selection

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{Minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n, \ \langle c, x \rangle \geq b \end{array} \right.$$

Example: Markovitz model for portfolio selection

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{Minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n,\ \langle c, x \rangle \geq b \end{array} \right.$$

Q is a covariance matrix (to be inferred)

Example: Markovitz model for portfolio selection

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{Minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n, \ \langle c, x \rangle \geq b \end{array} \right.$$

Q is a covariance matrix (to be inferred)

$$\Delta_n = \left\{ x \in \mathbb{R}_+^n \,|\, \textstyle\sum_j x_j \leq 1 \right\}, \ c \in \mathbb{R}^n \text{ and } b \in \mathbb{R}$$

Example: Markovitz model for portfolio selection

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{Minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n,\ \langle c, x \rangle \geq b \end{array} \right.$$

Q is a covariance matrix (to be inferred)

$$\Delta_n = \left\{ x \in \mathbb{R}_+^n \,|\, \textstyle\sum_j x_j \leq 1 \right\},\ c \in \mathbb{R}^n \text{ and } b \in \mathbb{R}$$

Q is constrained to belong to some polytope P

Example: Markovitz model for portfolio selection

$$(\mathscr{M}) \quad \left| \begin{array}{rl} \text{Minimize} & \langle x, Qx \rangle \\ \text{s.t.} & x \in \Delta_n,\ \langle c, x \rangle \geq b \end{array} \right.$$

Q is a covariance matrix (to be inferred)

$$\Delta_n = \Big\{ x \in \mathbb{R}^n_+ \,|\, \textstyle\sum_j x_j \leq 1 \Big\},\ c \in \mathbb{R}^n \text{ and } b \in \mathbb{R}$$

Q is constrained to belong to some polytope P

$$\left| \begin{array}{rl} \text{Minimize} & \kappa(Q) \\ \text{s.t.} & Q \in S_n^+ \cap P \end{array} \right.$$

An auxiliary function

An auxiliary function

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$$

An auxiliary function

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$$

$$\kappa_p(A) = \frac{\lambda_1^{\frac{p+1}{p}}(A)}{\lambda_n(A)}$$

An auxiliary function

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$$

$$\kappa_p(A) = \frac{\lambda_1^{\frac{p+1}{p}}(A)}{\lambda_n(A)}$$

$$\kappa_p(A) \to \kappa(A) \text{ pointwise as } p \to \infty$$

An auxiliary function

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$$

$$\kappa_p(A) = \frac{\lambda_1^{\frac{p+1}{p}}(A)}{\lambda_n(A)}$$

$\kappa_p(A) \to \kappa(A)$ pointwise as $p \to \infty$

$$\kappa_p^p(A) = \frac{\lambda_1^{p+1}(A)}{\lambda_n^p(A)}$$ is convex !

An auxiliary function

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$$

$$\kappa_p(A) = \frac{\lambda_1^{\frac{p+1}{p}}(A)}{\lambda_n(A)}$$

$\kappa_p(A) \to \kappa(A)$ pointwise as $p \to \infty$

$$\kappa_p^p(A) = \frac{\lambda_1^{p+1}(A)}{\lambda_n^p(A)} \text{ is convex !}$$

κ_p and κ_p^p have the same minimizers

An optimization strategy

An optimization strategy

Replace the original quasiconvex problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

by the surrogate convex problem

$$(\mathscr{P}_p) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa_p^p(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

and let $p \longrightarrow \infty$.

An optimization strategy

Replace the original quasiconvex problem

$$(\mathscr{P}) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

by the surrogate convex problem

$$(\mathscr{P}_p) \quad \left| \begin{array}{rl} \text{Minimize} & \kappa_p^p(A) \\ \text{s.t.} & A \in \Omega \end{array} \right.$$

and let $p \to \infty$.

Question: can we approach a solution to Problem $(\mathscr{P})$ with solutions to the surrogate problems $(\mathscr{P}_p)$?

Main results

Main results

Theorem (Exact approximation)

Let $p_k \uparrow \infty$, and let $\bar{A}_{p_k}$ be a solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_{p_k})$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

Main results

Theorem (Exact approximation)

Let $p_k \uparrow \infty$, and let $\bar{A}_{p_k}$ be a solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_{p_k})$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

Theorem (Inexact approximation)

Let $p_k \uparrow \infty$, and let $\varepsilon_k \downarrow 0$. Let $\bar{A}_k := \bar{A}_{p_k}^{\varepsilon_k}$ an ε_k-solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_k)$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

Main results

Theorem (Exact approximation)

Let $p_k \uparrow \infty$, and let $\bar{A}_{p_k}$ be a solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_{p_k})$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

Theorem (Inexact approximation)

Let $p_k \uparrow \infty$, and let $\varepsilon_k \downarrow 0$. Let $\bar{A}_k := \bar{A}_{p_k}^{\varepsilon_k}$ an ε_k-solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_k)$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

$$\forall A \in \Omega, \quad \kappa(A) \geq \kappa(\bar{A}_k) - \varepsilon_k$$

Quasi-convexity

Quasi-convexity

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *quasi-convex* if it has convex level sets.

Quasi-convexity

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *quasi-convex* if it has convex level sets.

$$\mathrm{lev}_\alpha(f) := \{x \,|\, f(x) \leq \alpha\}$$

Quasi-convexity

Definition

A function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be *quasi-convex* if it has convex level sets.

$$\mathrm{lev}_\alpha(f) := \{x \,|\, f(x) \leq \alpha\}$$

Proposition

κ, κ_p and κ_p^p are quasi-convex.

Proof

Proof

$$\begin{aligned}
\mathrm{lev}_\alpha(\kappa) &:= \left\{A \in S_n^{++} \,|\, \kappa(A) \le \alpha\right\} \\
&= \{0\} \cup \left\{A \in S_n^{++} \,|\, \lambda_1(A) - \alpha\lambda_n(A) \le 0\right\} \\
&= \mathrm{lev}_0(\lambda_1 - \alpha\lambda_n)
\end{aligned}$$

Proof

$$
\begin{aligned}
\mathrm{lev}_\alpha(\kappa) &:= \left\{A \in S_n^{++} \,|\, \kappa(A) \leq \alpha\right\} \\
&= \{0\} \cup \left\{A \in S_n^{++} \,|\, \lambda_1(A) - \alpha\lambda_n(A) \leq 0\right\} \\
&= \mathrm{lev}_0\big(\lambda_1 - \alpha\lambda_n\big)
\end{aligned}
$$

$$
\left.\begin{array}{r} \lambda_1 \ \text{ convex} \\ \lambda_n \ \text{ concave} \end{array}\right\} \Longrightarrow \lambda_1 - \alpha\lambda_n \ \text{ convex}
$$

Proof

$$\begin{aligned}
\mathrm{lev}_\alpha(\kappa) &:= \left\{A \in S_n^{++} \,|\, \kappa(A) \le \alpha\right\} \\
&= \{0\} \cup \left\{A \in S_n^{++} \,|\, \lambda_1(A) - \alpha\lambda_n(A) \le 0\right\} \\
&= \mathrm{lev}_0(\lambda_1 - \alpha\lambda_n)
\end{aligned}$$

$$\left.\begin{array}{r} \lambda_1 \ \text{ convex} \\ \lambda_n \ \text{ concave} \end{array}\right\} \Longrightarrow \lambda_1 - \alpha\lambda_n \ \text{ convex}$$

$$\mathrm{lev}_\alpha(\kappa_p) = \mathrm{lev}_0\left(\lambda_1^{\frac{p+1}{p}} - \alpha\lambda_n\right)$$

Proof

$$\begin{aligned}
\mathrm{lev}_\alpha(\kappa) &:= \left\{A \in S_n^{++} \,|\, \kappa(A) \le \alpha\right\} \\
&= \{0\} \cup \left\{A \in S_n^{++} \,|\, \lambda_1(A) - \alpha\lambda_n(A) \le 0\right\} \\
&= \mathrm{lev}_0(\lambda_1 - \alpha\lambda_n)
\end{aligned}$$

$$\left.\begin{array}{r} \lambda_1 \text{ convex} \\ \lambda_n \text{ concave} \end{array}\right\} \Longrightarrow \lambda_1 - \alpha\lambda_n \text{ convex}$$

$$\mathrm{lev}_\alpha(\kappa_p) = \mathrm{lev}_0\left(\lambda_1^{\frac{p+1}{p}} - \alpha\lambda_n\right)$$

$$\kappa_p^p(A) \le \alpha \quad \Leftrightarrow \quad \kappa_p(A) \le \alpha^{1/p}$$

$$\mathrm{lev}_\alpha(\kappa_p^p) = \mathrm{lev}_{\alpha^{1/p}}(\kappa_p)$$

Convexity of κ_p^p

Proposition

Suppose $f : \mathbb{R}^n \to [0, \infty]$ is

- quasiconvex,
- lower semi-continuous,
- positively homogeneous of degree $p \geq 1$ $(\forall t > 0,\ \forall x \in \mathbb{R}^n,\ f(tx) = t^p f(x))$.

Then f is convex.

Building a convex set

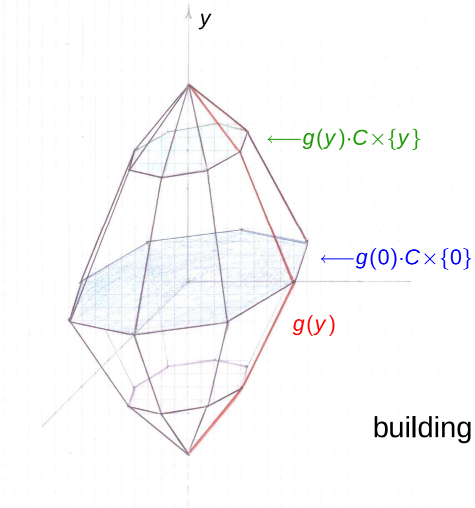
Lemma

The set-valued mapping $r \mapsto r \cdot C$ is increasing on $\mathbb{R}_+$ if and only if $C \subset \mathbb{R}^n$ is a convex set containing the origin.
Consequently, if $g\colon \mathbb{R}^m \to \bar{\mathbb{R}}$ is concave and nonnegative on its domain $\mathrm{dom}\, g$, then the set

$$\bigcup_{y \in \mathrm{dom}\, g} \left(g(y) \cdot C \times \{y\} \right)$$

is a convex subset of $\mathbb{R}^n \times \mathbb{R}^m$.

Illustration



building $\displaystyle\bigcup_{y\in\mathrm{dom}\, g}\big(g(y)\cdot C\times\{y\}\big)$

Proof of the proposition

Proof of the proposition

We can assume WLOG that $\exists x_0 : f(x_0) < \infty$.

Proof of the proposition

We can assume WLOG that $\exists x_0 : f(x_0) < \infty$.

- $f(0) = 0$: Since f is lower semi-continuous, one has:

$$f(0) = f(\lim_{t\downarrow 0} tx_0) \leq \lim_{t\downarrow 0} f(tx_0) = \lim_{t\downarrow 0} t^p f(x_0) = 0.$$

Since f takes its values in $[0,\infty]$, one must have $f(0) = 0$.

Proof of the proposition

We can assume WLOG that $\exists x_0 : f(x_0) < \infty$.

- $f(0) = 0$: Since f is lower semi-continuous, one has:

$$f(0) = f(\lim_{t \downarrow 0} tx_0) \leq \lim_{t \downarrow 0} f(tx_0) = \lim_{t \downarrow 0} t^p f(x_0) = 0.$$

  Since f takes its values in $[0, \infty]$, one must have $f(0) = 0$.

- For all $r > 0$, $\mathrm{lev}_r(f) = r^{1/p} \cdot \mathrm{lev}_1(f)$:

$$\begin{aligned} \mathrm{lev}_r(f) = \{x \,|\, f(x) \leq r\} &= \left\{x \,|\, r^{-1} f(x) \leq 1\right\} \\ &= \left\{x \,|\, f(r^{-1/p} x) \leq 1\right\} \\ &= \left\{r^{1/p} x' \,|\, f(x') \leq 1\right\} = r^{1/p} \mathrm{lev}_1(f) \end{aligned}$$

Proof of the proposition

- $0 \cdot \mathrm{lev}_1(f) := 0^+(\mathrm{lev}_1(f)) = \mathrm{lev}_0(f)$:

$$
\begin{aligned}
0^+(\mathrm{lev}_1(f)) &= 0^+\{x \in \mathbb{R}^n | f(x) \leq 1\} \\
&= \bigcap_{\beta>0} \{\beta x \in \mathbb{R}^n | f(x) \leq 1\} \\
&= \bigcap_{\beta>0} \{x' \in \mathbb{R}^n | f(x'/\beta) \leq 1\} \\
&= \bigcap_{\beta>0} \mathrm{lev}_{\beta^p}(f) \; = \; \mathrm{lev}_0(f)
\end{aligned}
$$

Proof of the proposition

- $0 \cdot \mathrm{lev}_1(f) := 0^+(\mathrm{lev}_1(f)) = \mathrm{lev}_0(f)$:

$$\begin{aligned}
0^+(\mathrm{lev}_1(f)) &= 0^+\{x \in \mathbb{R}^n | f(x) \leq 1\} \\
&= \bigcap_{\beta>0} \{\beta x \in \mathbb{R}^n | f(x) \leq 1\} \\
&= \bigcap_{\beta>0} \left\{x' \in \mathbb{R}^n | f(x'/\beta) \leq 1\right\} \\
&= \bigcap_{\beta>0} \mathrm{lev}_{\beta^p}(f) = \mathrm{lev}_0(f)
\end{aligned}$$

Hence, the formula $\mathrm{lev}_r(f) = r^{1/p} \cdot \mathrm{lev}_1(f)$ holds for every $r \geq 0$.

Proof of the proposition

- $\operatorname{epi} f$ is convex

$$
\begin{aligned}
\operatorname{epi} f &= \{(x,r) \in \mathbb{R}^n \times \mathbb{R}_+ \,|\, f(x) \le r\} \\
&= \bigcup_{r \in \mathbb{R}_+} (\operatorname{lev}_r(f) \times \{r\}) \\
&= \bigcup_{r \in \mathbb{R}_+} \left(r^{1/p} \cdot \operatorname{lev}_1(f) \times \{r\} \right).
\end{aligned}
$$

Proof of the proposition

- $\mathrm{epi}\, f$ is convex

$$\begin{aligned}
\mathrm{epi}\, f &= \{(x,r) \in \mathbb{R}^n \times \mathbb{R}_+ \,|\, f(x) \leq r\} \\
&= \bigcup_{r \in \mathbb{R}_+} (\mathrm{lev}_r(f) \times \{r\}) \\
&= \bigcup_{r \in \mathbb{R}_+} \left(r^{1/p} \cdot \mathrm{lev}_1(f) \times \{r\} \right).
\end{aligned}$$

The conclusion then follows by the Lemma. ■

The Clarke directional derivative

The Clarke directional derivative

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz near x_0. Given v, the ratio

$$\frac{f(x+tv)-f(x)}{t} \quad (t>0)$$

is bounded for (x,t) sufficiently close to $(x_0,0)$. The Clarke directional derivative is then defined as

$$f^\circ(x_0;v) = \limsup_{\substack{x\to x_0\\ t\downarrow 0}} \frac{f(x+tv)-f(x)}{t}.$$

The Clarke subdifferential

The Clarke subdifferential

The Clarke subdifferential of f at x_0 is the subset of $\mathbb{R}^n$ defined by

$$\partial f(x_0) = \Big\{ \xi \in \mathbb{R}^n \;\Big|\; \forall v \in \mathbb{R}^n,\ \langle \xi, v \rangle \leq f^\circ(x_0; v) \Big\}.$$

The Clarke subdifferential

The Clarke subdifferential of f at x_0 is the subset of $\mathbb{R}^n$ defined by

$$\partial f(x_0) = \left\{\xi \in \mathbb{R}^n \;\middle|\; \forall v \in \mathbb{R}^n,\ \langle \xi, v\rangle \leq f^\circ(x_0; v)\right\}.$$

Theorem

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and $x_0 \in \mathbb{R}^n$. Then

- *$\partial f(x_0)$ is convex compact;*
- *for every $v \in \mathbb{R}^n$, $f^\circ(x_0; v) = \max\{\langle \xi, v\rangle | \xi \in \partial f(x_0)\}$.*

A fundamental result

A fundamental result

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Rademacher's theorem then says that f is differentiable almost everywhere.

A fundamental result

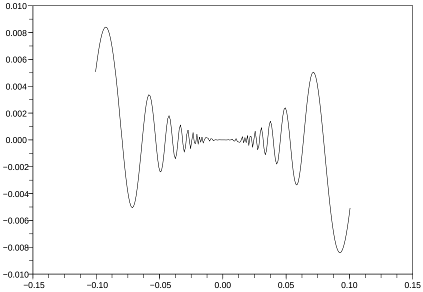
Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Rademacher's theorem then says that f is differentiable almost everywhere.

Theorem

Let $\Omega_f := \{x \in \mathbb{R}^n \,|\, f \text{ is not differentiable at } x\}$. Then $\partial f(x)$ is the convex hull of the set

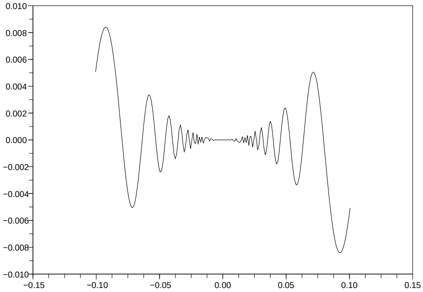
$$\left\{ \lim_{k\to\infty} \nabla f(x_k) \,\middle|\, x_k \in \Omega_f,\ x_k \to x,\ \lim_{k\to\infty} \nabla f(x_k) \text{ exists} \right\}.$$

Example



$$f(x) = x^2 \sin \frac{1}{x}$$

Example



$$f(x) = x^2 \sin \frac{1}{x}$$

- f is differentiable on $\mathbb{R}$;
- $f'(0) = 0$ and $\partial f(0) = [-1, 1]$.

Clarke regularity

Clarke regularity

Recall that the standard directional derivative is

$$f'(x_0; v) = \lim_{t \downarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

whenever the limit exists.

Clarke regularity

Recall that the standard directional derivative is

$$f'(x_0;v) = \lim_{t\downarrow 0} \frac{f(x_0+tv)-f(x_0)}{t}$$

whenever the limit exists.

The function f is said to be Clarke regular at x_0 (or merely regular at x_0) if, for every $v \in \mathbb{R}^n$, $f'(x_0;v)$ exists and

$$f'(x_0;v) = f^\circ(x_0;v).$$

Clarke regularity

Recall that the standard directional derivative is

$$f'(x_0; v) = \lim_{t \downarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

whenever the limit exists.

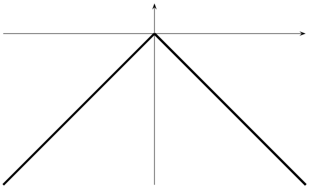
The function f is said to be Clarke regular at x_0 (or merely regular at x_0) if, for every $v \in \mathbb{R}^n$, $f'(x_0; v)$ exists and

$$f'(x_0; v) = f^\circ(x_0; v).$$

Theorem

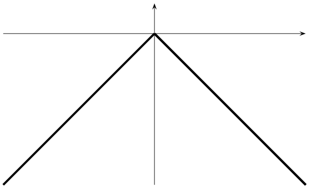
Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then f is Clarke regular and, for every x, $\partial f(x) = \{\nabla f(x)\}$.

Example



$$f(x) = -|x|$$

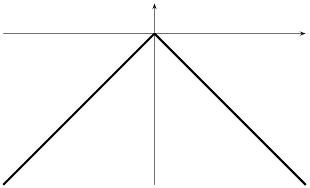
Example



$f(x) = -|x|$

- $\partial f(0) = [-1, 1]$, thus $f^\circ(0, 1) = \sup\{\xi | \xi \in [-1, 1]\} = 1$;
- $f'(0, 1) = -1$.

Example



$$f(x) = -|x|$$

- $\partial f(0) = [-1, 1]$, thus $f^\circ(0, 1) = \sup\{\xi \,|\, \xi \in [-1, 1]\} = 1$;
- $f'(0, 1) = -1$.

f is not regular at 0

Quotient rule

Proposition

Let $f_1, f_2 \colon \mathbb{R}^n \to \bar{\mathbb{R}}$ be Lipschitz near x_0. Assume that

- $f_1(x_0) \geq 0$ *and* $f_2(x_0) > 0$;
- f_1 *and* $-f_2$ *are Clarke regular at* x_0.

Then, f_1/f_2 is Clarke regular at x_0 and

$$\partial \left(\frac{f_1}{f_2} \right) (x_0) = \frac{f_2(x_0)\partial f_1(x_0) - f_1(x_0)\partial f_2(x_0)}{f_2^2(x_0)}.$$

Composition rule

Proposition

Let S be a subset of $\mathbb{R}^n$ and $x_0 \in \operatorname{int} S$. Let $f = g \circ h$, where $h\colon S \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$. Assume that

- *g is continuously differentiable at $h(x_0)$;*
- *h is Lipschitz near x_0.*

Then

$$\partial f(x_0) = g'(h(x_0))\partial h(x_0).$$

Moreover, if g is continuously differentiable in a neighborhood of $h(x_0)$ and h is Clarke regular at x_0, then f is also Clarke regular at x_0.

A regularity trick

Lemma

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz, $x_0 \in \mathbb{R}^n$ and $\varphi\colon \mathbb{R} \to \mathbb{R}$. Assume that

- $-f$ is Clarke regular at x_0;
- φ is continuously differentiable and non-decreasing at $f(x_0)$.

Then $-\varphi \circ f$ is Clarke regular at x_0 and

$$\partial(-\varphi \circ f)(x_0) = -\varphi'(f(x_0))\partial f(x_0) = \varphi'(f(x_0))\partial(-f)(x_0).$$

Proof

The formula is an immediate consequence of the chain rule (since f is Lipschitz near x_0).

Proof

The formula is an immediate consequence of the chain rule (since f is Lipschitz near x_0). By the Mean Value Theorem applied to φ,

$$-\varphi(f(x_0 + tv)) + \varphi(f(x_0)) = \varphi'(u)(-f(x_0 + tv) + f(x_0))$$

for some $u \in [f(x_0), f(x_0 + tv)]$.

Proof

The formula is an immediate consequence of the chain rule (since f is Lipschitz near x_0). By the Mean Value Theorem applied to φ,

$$-\varphi(f(x_0+tv))+\varphi(f(x_0))=\varphi'(u)(-f(x_0+tv)+f(x_0))$$

for some $u\in[f(x_0),f(x_0+tv)]$. Now,

$$\begin{aligned}&\frac{-\varphi(f(x_0+tv))+\varphi(f(x_0))}{t}\\ =\ &\varphi'(u)\frac{-f(x_0+tv)+f(x_0)}{t}\xrightarrow{t\downarrow 0}\varphi'(f(x_0))(-f)'(x_0;v)\end{aligned}$$

(the regularity of $-f$ ensures existence of the limit).

Proof

The formula is an immediate consequence of the chain rule (since f is Lipschitz near x_0). By the Mean Value Theorem applied to φ,

$$-\varphi(f(x_0+tv))+\varphi(f(x_0))=\varphi'(u)(-f(x_0+tv)+f(x_0))$$

for some $u\in[f(x_0),f(x_0+tv)]$. Now,

$$\frac{-\varphi(f(x_0+tv))+\varphi(f(x_0))}{t} = \varphi'(u)\frac{-f(x_0+tv)+f(x_0)}{t} \xrightarrow{t\downarrow 0} \varphi'(f(x_0))(-f)'(x_0;v)$$

(the regularity of $-f$ ensures existence of the limit). Thus,

$$(-\varphi\circ f)'(x_0;v)=\varphi'(f(x_0))(-f)'(x_0;v).$$

Proof (end)

Since $\quad \partial(-\varphi \circ f)(x_0) = \varphi'(f(x_0))\partial(-f)(x_0)$,

$$
\begin{aligned}
(-\varphi \circ f)^\circ(x_0; v) &= \max_{s \in \partial(-\varphi \circ f)(x_0)} \langle s, v \rangle \\
&= \max_{s' \in \partial(-f)(x_0)} \varphi'(f(x_0))\langle s', v \rangle \\
&= \varphi'(f(x_0)) \max_{s' \in \partial(-f)(x_0)} \langle s', v \rangle \\
&= \varphi'(f(x_0))(-f)^\circ(x_0; v) \\
&= \varphi'(f(x_0))(-f)'(x_0; v) \qquad [\,\text{regularity of } -f\,] \\
&= (-\varphi \circ f)'(x_0; v) \qquad [\,\text{previous step}\,]
\end{aligned}
$$

Subgradients

Proposition

Let $A \in S_n^{++}$. Then:

$$\partial\kappa(A) = \lambda_1(A)^{-1}\kappa(A)\big[\partial\lambda_1(A) - \kappa(A)\partial\lambda_n(A)\big]$$

$$\partial\kappa_p^p(A) = \kappa(A)^p\big[(p+1)\partial\lambda_1(A) - p\kappa(A)\partial\lambda_n(A)\big]$$

$$\partial\kappa_p(A) = \lambda_1(A)^{\frac{1-p}{p}}\kappa(A)^p\left[\frac{p+1}{p}\partial\lambda_1(A) - \kappa(A)\partial\lambda_n(A)\right]$$

Pseudoconvexity

Pseudoconvexity

Let $\Omega \subset \mathbb{R}^n$, $\bar{x} \in \Omega$ and $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$, lsc and Lipschitz near $\bar{x}$.

Pseudoconvexity

Let $\Omega \subset \mathbb{R}^n$, $\bar{x} \in \Omega$ and $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, lsc and Lipschitz near $\bar{x}$.

(1) We say that f is pseudoconvex at $\bar{x}$ on Ω if

$$\forall x \in \Omega, \quad f^\circ(\bar{x}; x - \bar{x}) \geq 0 \implies f(x) \geq f(\bar{x}).$$

We say that f is pseudoconvex on Ω if f is pseudoconvex at every $\bar{x} \in \Omega$ on Ω.

Pseudoconvexity

Let $\Omega \subset \mathbb{R}^n$, $\bar{x} \in \Omega$ and $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$, lsc and Lipschitz near $\bar{x}$.

(1) We say that f is pseudoconvex at $\bar{x}$ on Ω if

$$\forall x \in \Omega, \quad f^\circ(\bar{x}; x - \bar{x}) \geq 0 \implies f(x) \geq f(\bar{x}).$$

We say that f is pseudoconvex on Ω if f is pseudoconvex at every $\bar{x} \in \Omega$ on Ω.

(2) We say that f is strongly pseudoconvex at $\bar{x}$ on Ω if

$$\forall \xi \in \partial f(\bar{x}), \quad \forall x \in \Omega, \quad \langle \xi, x - \bar{x} \rangle \geq 0 \implies f(x) \geq f(\bar{x}).$$

We say that f is strongly pseudoconvex on Ω if f is strongly pseudoconvex at every $\bar{x} \in \Omega$ on Ω.

Optimality condition for pseudoconvex functions

Theorem

Let $\bar{x} \in \Omega \subset \mathbb{R}^n$, where Ω is closed convex. Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be lower semicontinuous and Lipschitz near $\bar{x}$. If f is pseudoconvex at $\bar{x}$ on Ω, then the following are equivalent:

(a) *$\bar{x}$ is a global minimizer of f on Ω;*

(b) $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$.

Proof of $(b) \Rightarrow (a)$

Proof of $(b) \Rightarrow (a)$

Assume that $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$:

Proof of $(b) \Rightarrow (a)$

Assume that $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$:

There exists $\xi_0 \in \partial f(\bar{x})$ such that $-\xi_0 \in N_\Omega(\bar{x})$

Proof of $(b) \Rightarrow (a)$

Assume that $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$:

There exists $\xi_0 \in \partial f(\bar{x})$ such that $-\xi_0 \in N_\Omega(\bar{x})$

$$\forall x \in \Omega, \quad \langle -\xi_0, x - \bar{x} \rangle \leq 0$$

Proof of $(b) \Rightarrow (a)$

Assume that $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$:

There exists $\xi_0 \in \partial f(\bar{x})$ such that $-\xi_0 \in N_\Omega(\bar{x})$

$$\forall x \in \Omega, \quad \langle -\xi_0, x - \bar{x} \rangle \leq 0$$

$$\forall x \in \Omega, \quad \underbrace{\max\{\langle \xi, x - \bar{x} \rangle \,|\, \xi \in \partial f(\bar{x})\}}_{f^\circ(\bar{x};x-\bar{x})} \geq 0$$

Proof of $(b) \Rightarrow (a)$

Assume that $0 \in \partial f(\bar{x}) + N_\Omega(\bar{x})$:

There exists $\xi_0 \in \partial f(\bar{x})$ such that $-\xi_0 \in N_\Omega(\bar{x})$

$$\forall x \in \Omega, \quad \langle -\xi_0, x - \bar{x} \rangle \leq 0$$

$$\forall x \in \Omega, \quad \underbrace{\max\{\langle \xi, x - \bar{x} \rangle \,|\, \xi \in \partial f(\bar{x})\}}_{f^\circ(\bar{x}; x - \bar{x})} \geq 0$$

$$\forall x \in \Omega, \quad f(x) \geq f(\bar{x})$$ [pseudoconvexity]

Optimality condition for κ

Optimality condition for κ

Proposition

The function κ is strongly pseudoconvex. Consequently, $\bar{A}$ minimizes κ over Ω is and only if $0 \in \partial\kappa(\bar{A}) + N_\Omega(\bar{x})$.

Proof

Proof

$$\partial\kappa(\bar{A}) = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\big)$$

Proof

$$\partial\kappa(\bar{A}) = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\big)$$

$$V \in \partial\kappa(\bar{A}) \iff V = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\underbrace{V_1}_{\in\partial\lambda_1(\bar{A})} - \kappa(\bar{A})\underbrace{V_n}_{\partial\lambda_n(\bar{A})}\big)$$

Proof

$$\partial\kappa(\bar{A}) = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\big)$$

$$V \in \partial\kappa(\bar{A}) \;\Leftrightarrow\; V = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\Big(\underbrace{V_1}_{\in\partial\lambda_1(\bar{A})} - \kappa(\bar{A})\underbrace{V_n}_{\partial\lambda_n(\bar{A})}\Big)$$

$$\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) = \underbrace{\lambda_1(A) - \lambda_1(\bar{A})}_{\substack{\geq\langle V_1, A-\bar{A}\rangle \\ [\lambda_1 \text{ convex}]}} + \kappa(\bar{A})\Big(\underbrace{-\lambda_n(A) + \lambda_n(\bar{A})}_{\substack{\geq\langle -V_n, A-\bar{A}\rangle \\ [-\lambda_n \text{ convex}]}}\Big)$$

Proof

$$\partial\kappa(\bar{A}) = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\big)$$

$$V \in \partial\kappa(\bar{A}) \;\Leftrightarrow\; V = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\underbrace{V_1}_{\in\partial\lambda_1(\bar{A})} - \kappa(\bar{A})\underbrace{V_n}_{\partial\lambda_n(\bar{A})}\big)$$

$$\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) = \underbrace{\lambda_1(A) - \lambda_1(\bar{A})}_{\substack{\geq\langle V_1, A-\bar{A}\rangle \\ [\lambda_1 \text{ convex}]}} + \kappa(\bar{A})\big(\underbrace{-\lambda_n(A) + \lambda_n(\bar{A})}_{\substack{\geq\langle -V_n, A-\bar{A}\rangle \\ [-\lambda_n \text{ convex}]}}\big)$$

$$\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) \geq \langle \underbrace{V_1 - \kappa(\bar{A})V_n}_{\lambda_1(\bar{A})\kappa(\bar{A})^{-1}V}, A - \bar{A}\rangle = \lambda_1(\bar{A})\kappa(\bar{A})^{-1}\langle V, A - \bar{A}\rangle$$

Proof

$$\partial\kappa(\bar{A}) = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\partial\lambda_1(\bar{A}) - \kappa(\bar{A})\partial\lambda_n(\bar{A})\big)$$

$$V \in \partial\kappa(\bar{A}) \;\Leftrightarrow\; V = \lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\underbrace{V_1}_{\in\partial\lambda_1(\bar{A})} - \kappa(\bar{A})\underbrace{V_n}_{\partial\lambda_n(\bar{A})}\big)$$

$$\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) = \underbrace{\lambda_1(A) - \lambda_1(\bar{A})}_{\substack{\geq\langle V_1, A-\bar{A}\rangle \\ [\lambda_1 \text{ convex}]}} + \kappa(\bar{A})\big(\underbrace{-\lambda_n(A) + \lambda_n(\bar{A})}_{\substack{\geq\langle -V_n, A-\bar{A}\rangle \\ [-\lambda_n \text{ convex}]}}\big)$$

$$\lambda_1(A) - \kappa(\bar{A})\lambda_n(A) \geq \langle \underbrace{V_1 - \kappa(\bar{A})V_n}_{\lambda_1(\bar{A})\kappa(\bar{A})^{-1}V}, A - \bar{A}\rangle = \lambda_1(\bar{A})\kappa(\bar{A})^{-1}\langle V, A - \bar{A}\rangle$$

$$\langle V, A - \bar{A}\rangle \geq 0 \implies \lambda_1(A) - \kappa(\bar{A})\lambda_n(A) \geq 0 \iff \kappa(A) \geq \kappa(\bar{A})$$

Our main convergence result

Our main convergence result

Theorem (Exact approximation)

Let $p_k \uparrow \infty$, and let $\bar{A}_{p_k}$ be a solution to Problem $(\mathscr{P}_{p_k})$. Then the sequence $(\bar{A}_{p_k})$ has a subsequence which converges to a global solution $\bar{A}$ of Problem $(\mathscr{P})$.

Proof

Proof

Since Ω is compact, taking a subsequence, we can assume that $(\bar{A}_{p_k})$ converges to some $\bar{A} \in \Omega$. We shall prove that $0 \in \partial\kappa(\bar{A})$. The conclusion will follow by the pseudoconvexity argument.

Proof

Since Ω is compact, taking a subsequence, we can assume that $(\bar{A}_{p_k})$ converges to some $\bar{A} \in \Omega$. We shall prove that $0 \in \partial\kappa(\bar{A})$. The conclusion will follow by the pseudoconvexity argument.

Optimality condition for $\bar{A}_{p_k}$:

$$0 \in \underbrace{\lambda_1(\bar{A}_{p_k})^{\frac{1-p_k}{p_k}} \kappa(\bar{A}_{p_k}) \left(\frac{p_k+1}{p_k} \partial\lambda_1(\bar{A}_{p_k}) - \kappa(\bar{A}_{p_k}) \partial\lambda_n(\bar{A}_{p_k}) \right)}_{\partial\kappa_{p_k}(\bar{A}_{p_k})} + N_\Omega(\bar{A}_{p_k})$$

Proof

Since Ω is compact, taking a subsequence, we can assume that $(\bar{A}_{p_k})$ converges to some $\bar{A} \in \Omega$. We shall prove that $0 \in \partial\kappa(\bar{A})$. The conclusion will follow by the pseudoconvexity argument.

Optimality condition for $\bar{A}_{p_k}$:

$$0 \in \underbrace{\lambda_1(\bar{A}_{p_k})^{\frac{1-p_k}{p_k}} \kappa(\bar{A}_{p_k}) \left(\frac{p_k+1}{p_k} \partial\lambda_1(\bar{A}_{p_k}) - \kappa(\bar{A}_{p_k}) \partial\lambda_n(\bar{A}_{p_k}) \right)}_{\partial\kappa_{p_k}(\bar{A}_{p_k})} + N_\Omega(\bar{A}_{p_k})$$

Thus, there exist $V_1^{(k)} \in \partial\lambda_1(\bar{A}_{p_k})$ and $V_n^{(k)} \in \partial\lambda_n(\bar{A}_{p_k})$ such that

$$0 \in \lambda_1(\bar{A}_{p_k})^{\frac{1-p_k}{p_k}} \kappa(\bar{A}_{p_k}) \left(\frac{p_k+1}{p_k} V_1^{(k)} - \kappa(\bar{A}_{p_k}) V_n^{(k)} \right) + N_\Omega(\bar{A}_{p_k})$$

Proof (end)

Reminder: $\partial\lambda_k(A) = \mathrm{co}\left\{xx^\top \,|\, x \in \mathbb{R}^n,\ \|x\| = 1,\ Ax = \lambda_k(A)\cdot x\right\}$ (Cox, Overton, Lewis).

Proof (end)

Reminder: $\partial\lambda_k(A) = \mathrm{co}\left\{xx^\top \,|\, x \in \mathbb{R}^n,\ \|x\| = 1,\ Ax = \lambda_k(A)\cdot x\right\}$ (Cox, Overton, Lewis). Thus $\partial\lambda_k(A)$ is compact, and it follows that, taking a subsequence, we can assume that

$$V_1^{(k)} \to \bar{V}_1 \in \partial\lambda_1(\bar{A}) \quad \text{and} \quad V_n^{(k)} \to \bar{V}_n \in \partial\lambda_1(\bar{A}),$$

by closedness of the multifunctions $\partial\lambda_1$ and $\partial\lambda_n$.

Proof (end)

Reminder: $\partial\lambda_k(A) = \operatorname{co}\left\{xx^\top \,|\, x \in \mathbb{R}^n,\ \|x\| = 1,\ Ax = \lambda_k(A)\cdot x\right\}$ (Cox, Overton, Lewis). Thus $\partial\lambda_k(A)$ is compact, and it follows that, taking a subsequence, we can assume that

$$V_1^{(k)} \to \bar{V}_1 \in \partial\lambda_1(\bar{A}) \quad \text{and} \quad V_n^{(k)} \to \bar{V}_n \in \partial\lambda_1(\bar{A}),$$

by closedness of the multifunctions $\partial\lambda_1$ and $\partial\lambda_n$.
Since N_Ω is also a closed multifunction, we can pass to the limit in

$$0 \in \lambda_1(\bar{A}_{p_k})^{\frac{1-p_k}{p_k}}\kappa(\bar{A}_{p_k})\left(\frac{p_k+1}{p_k}V_1^{(k)} - \kappa(\bar{A}_{p_k})V_n^{(k)}\right) + N_\Omega(\bar{A}_{p_k}).$$

Proof (end)

Reminder: $\partial\lambda_k(A) = \mathrm{co}\left\{xx^\top \,|\, x \in \mathbb{R}^n,\ \|x\| = 1,\ Ax = \lambda_k(A)\cdot x\right\}$ (Cox, Overton, Lewis). Thus $\partial\lambda_k(A)$ is compact, and it follows that, taking a subsequence, we can assume that

$$V_1^{(k)} \to \bar{V}_1 \in \partial\lambda_1(\bar{A}) \quad \text{and} \quad V_n^{(k)} \to \bar{V}_n \in \partial\lambda_1(\bar{A}),$$

by closedness of the multifunctions $\partial\lambda_1$ and $\partial\lambda_n$.
Since N_Ω is also a closed multifunction, we can pass to the limit in

$$0 \in \lambda_1(\bar{A}_{p_k})^{\frac{1-p_k}{p_k}}\kappa(\bar{A}_{p_k})\left(\frac{p_k+1}{p_k}V_1^{(k)} - \kappa(\bar{A}_{p_k})V_n^{(k)}\right) + N_\Omega(\bar{A}_{p_k}).$$

and obtain

$$0 \in \underbrace{\lambda_1(\bar{A})^{-1}\kappa(\bar{A})\big(\underbrace{\bar{V}_1}_{\in\partial\lambda_1(\bar{A})} - \kappa(\bar{A})\underbrace{\bar{V}_n}_{\in\partial\lambda_n(\bar{A})}\big)}_{\in\partial\kappa(\bar{A})} + N_\Omega(\bar{A}) \subset \partial\kappa(\bar{A}) + N_\Omega(\bar{A}).$$