Regularized Least-Squares: Stability Properties and the Maximum Entropy on the Mean Method

Tim Hoheisel

McGill

Autumn School on Constrained Optimization and Machine Learning

Trier, October 9 –11, 2024.

Outline

1. Fundamentals from Convex Analysis

'What's dead may never die!'

Convex sets and cones

"The great watershed in optimization is not between linearity and nonlinearity, but convexity and nonconvexity." (R.T. Rockafellar, *1935)

$S \subset \mathbb{E}$ is said to be

- *convex* if $\lambda S + (1-\lambda)S \subset S \quad (\lambda \in (0,1))$;
- a *cone* if $\lambda S \subset S \quad (\lambda \geq 0)$.

Note that $K \subset \mathbb{E}$ is a *convex cone* iff $K + K \subset K$.

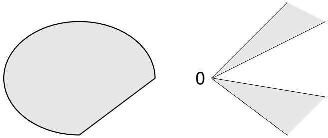


Figure: Convex set/non-convex cone

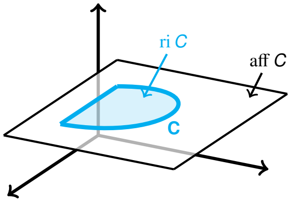
Relative topology and paralell subspaces

Affine set: A set $S = U + x$ with $x \in \mathbb{E}$ and a subspace $U \subset$ is called *affine*. The subspace U is uniquely determined by $U = \operatorname{aff}(S - x) = S - S$.

Affine hull: $\operatorname{aff} M := \bigcap \{S \in \mathbb{E} \mid M \subset S,\ S \text{ affine}\}$.

Relative interior/boundary: $C \subset \mathbb{E}$ convex.

$$\begin{aligned} \operatorname{ri} C &:= \left\{x \in C \,\middle|\, \exists \varepsilon > 0 : \ B_\varepsilon(x) \cap \operatorname{aff} C \subset C\right\} && \text{(relative interior)} \\ x \in \operatorname{ri} C &\Leftrightarrow \operatorname{aff}(C - x) = \mathbb{R}_+(C - x) =: \operatorname{par} C && \text{(parallel subspace)} \end{aligned}$$



C	$\operatorname{aff} C$	$\operatorname{ri} C$
$\{x\}$	$\{x\}$	$\{x\}$
$[x, x']$	$\{\lambda x + (1-\lambda)x' \mid \lambda \in \mathbb{R}\}$	(x, x')
$\overline{B}_\varepsilon(x)$	$\mathbb{E}$	$B_\varepsilon(x)$

Table: Examples of relative interiors

Extended real-valued functions: an epigraphical perspective

Let $f : \mathbb{E} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

- $\mathrm{epi}\, f := \left\{(x, \alpha) \in \mathbb{E} \times \mathbb{R} \;\middle|\; f(x) \le \alpha\right\}$ (epigraph)
- $\mathrm{epi}_{<} f := \left\{(x, \alpha) \in \mathbb{E} \times \mathbb{R} \;\middle|\; f(x) < \alpha\right\}$ (strict epigraph)
- $\mathrm{dom}\, f := \left\{x \in \mathbb{E} \;\middle|\; f(x) < \infty\right\}$ (domain).
- $\mathrm{lev}_r f := \left\{x \;\middle|\; f(x) \le r\right\}$ (level set)

→ f is uniquely determined through $\mathrm{epi}\, f$!

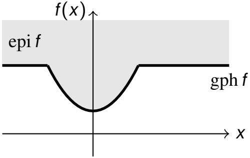


Figure: Epigraph of $f : \mathbb{R} \to \mathbb{R}$

f proper $:\Leftrightarrow$ $-\infty < f \not\equiv +\infty$ $\Leftrightarrow$[1] $\mathrm{dom}\, f \neq \emptyset$

f convex $:\Leftrightarrow$ $\mathrm{epi}\, f / \mathrm{epi}_{<} f$ convex $\Leftrightarrow$[1] $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \;\forall x, y \in \mathbb{E},\; \lambda \in [0,1]$

$\Rightarrow$ $\mathrm{lev}_r f$ convex $\forall r \in \mathbb{R}$.

$$\Gamma := \left\{f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\} \;\middle|\; f \text{ proper, convex}\right\}$$

[1] Only for $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$

Lower semicontinuity

Let $f : \mathbb{E} \to \mathbb{R}$ and $\bar{x} \in \mathbb{E}$.

Lower limit:
$\liminf_{x \to \bar{x}} f(x) := \inf \left\{ \alpha \,\middle|\, \exists\, x_k \to \bar{x} : \; f(x_k) \to \alpha \right\}$

Lower semicontinuity: f is said to be *lsc* (or *closed*) at $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \ge f(\bar{x}).$$

$\Gamma_0 := \{ f \in \Gamma \mid f \text{ closed} \}$

Closure: $\mathrm{cl}\, f : \mathbb{E} \to \overline{\mathbb{R}}, \quad (\mathrm{cl}\, f)(\bar{x}) := \liminf_{x \to \bar{x}} f(x).$

Facts:

- f lsc $\iff$ $\mathrm{epi}\, f$ closed $\iff$ $f = \mathrm{cl}\, f$ $\iff$ $\mathrm{lev}_r f$ closed $(r \in \mathbb{R})$
- $\mathrm{cl}\, f \le f$

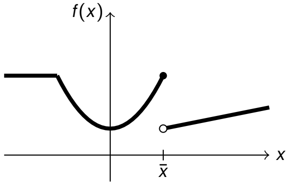


Figure: f <u>*not*</u> lsc at $\bar{x}$

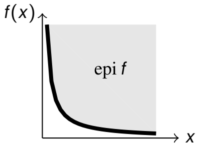


Figure: $f : x \mapsto \begin{cases} \frac{1}{x} & x > 0, \\ +\infty, & \text{else.} \end{cases}$

Convexity preserving operations (*New from old*)

1. Set Operations

   For $C, C_i \ (i \in I) \subset \mathbb{E},\ D \subset \mathbb{E}'$ convex, $F : \mathbb{E} \to \mathbb{E}'$ affine the following sets are convex:

   - $F(C)$ (affine image)
   - $F^{-1}(D)$ (affine pre-image)
   - $C \times D$ (Cartesian product)
   - $C_1 + C_2$ (Minkowski sum)
   - $\bigcap_{i \in I} C_i$ (Intersection)

2. Functional operations

   For $f_i, g : \mathbb{E} \to \overline{\mathbb{R}}$ convex and $F : \mathbb{E}' \to \mathbb{E}$ affine the following functions are convex:

   - (Affine pre-composition) $f := g \circ F$: $\operatorname{epi} f = T^{-1}(\operatorname{epi} g),\ T : (x, \alpha) \mapsto (F(x), \alpha)$
   - (Epi-multiplication) $f := \lambda \star g := \lambda g\left(\frac{\cdot}{\lambda}\right)$: $\operatorname{epi} f = \lambda \operatorname{epi} g$
   - (Pointwise supremum) $f := \sup_{i \in I} f_i$: $\operatorname{epi} f = \bigcap_{i \in I} \operatorname{epi} f_i$
   - (Moreau envelope) $f : x \mapsto \inf_u \left\{ g(u) + \frac{1}{2}\|x - u\|^2 \right\}$: $\operatorname{epi} f = \operatorname{epi} g + \operatorname{epi} \frac{1}{2}\|\cdot\|^2$.

Coercivity notions and existence of minimizers

Let $f:\mathbb{E}\to\overline{\mathbb{R}}$. Then f is called

i) *coercive* if $\lim_{\|x\|\to+\infty} f(x)=+\infty$;

ii) *supercoercive* if $\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|}=+\infty$.

Lemma 1 (Level-boundedness = coercivity).

$f:\mathbb{E}\to\overline{\mathbb{R}}$ is coercive if and only if it is level-bounded, *i.e.,* $\mathrm{lev}_\alpha f$ *is bounded for all* $\alpha\in\mathbb{R}$.

Theorem 2 (Existence of minima).

Let $f:\mathbb{E}\to\mathbb{R}\cup\{+\infty\}$ be proper, lsc and level-bounded. Then $\mathrm{argmin}_{\mathbb{E}} f\neq\emptyset$.

Proof.

Pick $\{x_k\}$ such that $f(x_k)\to f^*:=\inf_{\mathbb{E}} f<\infty$; choose $\alpha\in(f^*,+\infty)$; set $L_\alpha:=\left\{x \;\middle|\; f(x)\leq\alpha\right\}$.

$$L_\alpha \text{compact},\quad x_k\in L_\alpha\ (k\text{ suff. large}) \overset{\text{Bolzano–Weierstrass}}{\Longrightarrow} \exists\bar{x}\in L_\alpha,\ \{x_k\}\to_K \bar{x}$$

$$\Longrightarrow f(\bar{x}) \overset{f\ \mathrm{lsc}}{\leq} \liminf_{x\to\bar{x}} f(x)\leq \liminf_{k\in K} f(x_k)=f^*$$

$$\Longrightarrow \bar{x}\in\underset{\mathbb{E}}{\mathrm{argmin}}\, f.$$

□

Stronger notions of convexity

Let $f \in \Gamma$ and $C \subset \mathrm{dom}\, f$ convex. Then f is said to be

a) *strictly convex* on C if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \quad (x, y, \in C, x \neq y,\ \lambda \in (0,1)).$$

b) *strongly convex* on C if there exists $\sigma > 0$ such that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2 \quad (x, y, \in C,\ \lambda \in (0,1))$$

The scalar $\sigma > 0$ is called *modulus of strong convexity* of f (on C).

For $C = \mathrm{dom}\, f$ we simply call f strictly and strongly convex, respectively.

Proposition 3.

Let $f \in \Gamma$. Then:

a) *f σ-strongly convex $\iff$ $f - \frac{\sigma}{2}\|\cdot\|^2$ convex.*

b) *f σ-strongly convex $\implies$ f supercoercive and strictly convex.*

Guide.

a) Elementary computation.

b) Use the (nontrivial) fact that $f - \frac{\sigma}{2}\|\cdot\|^2$ has an affine minorant $g(x) = \langle v, x\rangle + \beta$ to verify supercoercivity. Strict convexity is straightforward. □

The basic results in convex optimization

Proposition 4.

Let $f \in \Gamma$. Then every local minimizer of f (over $\mathbb{E}$) is a global minimizer and $\operatorname{argmin} f$ is convex (possibly empty).

Proposition 5 (Uniqueness of minimizers).

Let $f \in \Gamma$ be strictly convex. Then f has at most one minimizer.

Corollary 6 (Minimizing the sum of convex functions).

Let $f, g \in \Gamma_0$ such that $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. Suppose that one of the following holds:

- i) *f is supercoercive;*
- ii) *f is coercive and g is bounded from below.*

Then $f + g$ is coercive and has a minimizer (over $\mathbb{E}$). If f or g is strictly convex, $f + g$ has exactly one minimizer.

Guide.

Observe $f + g \in \Gamma_0$. Now show in either case that $f + g$ is coercive, and apply Theorem 2.
The uniqueness result follows immediately from Proposition 5, realizing that $f + g \in \Gamma_0$ is strictly convex if one of the summands is. □

Parametric minimization aka infimal projection

Theorem 7 (Infimal projection).

Let $h : \mathbb{E}_1 \times \mathbb{E}_2 \to \mathbb{R} \cup \{+\infty\}$ *be convex. Then the* optimal value function

$$\varphi : \mathbb{E}_1 \to \overline{\mathbb{R}}, \ \varphi(x) := \inf_{y \in \mathbb{E}_2} h(x, y)$$

is convex. Moreover, the set-valued mapping

$$x \mapsto \operatorname{argmin}_{y \in \mathbb{E}_2} h(x, y) \subset \mathbb{E}_2.$$

is convex-valued.

Proof.

It can easily be shown that $\operatorname{epi}_< \varphi = L(\operatorname{epi}_< h)$ under the linear mapping $L : (x, y, \alpha) \mapsto (x, \alpha)$.
The remaining assertion follows immediately from Proposition 4 , since $y \mapsto h(x, y)$ is convex for all $x \in \mathbb{E}_1$. □

The convex subdifferential

Definition 8.

Let $f:\mathbb{E}\to\overline{\mathbb{R}}$. A vector $v\in\mathbb{E}$ is called a *subgradient* of v at $\bar{x}$ if

$$f(x)\geq f(\bar{x})+\langle v,x-\bar{x}\rangle \quad (x\in\mathbb{E}). \tag{1}$$

We denote by $\partial f(\bar{x})$ the set of all subgradients of f at $\bar{x}$ and call it the *(convex) subdifferential* of f at $\bar{x}$.

The inequality (1) is referred to as *subgradient inequality*.

Slogan: *"The subgradients of f at $\bar{x}$ are the slopes of affine minorants of f that coincide with f at $\bar{x}$".*

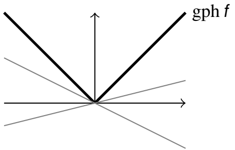


Figure: Affine minorants at a point of nondifferentiability

- $0\in\partial f(x) \iff x\in\operatorname{argmin}_{\mathbb{E}} f$ (Fermat's rule)
- $\partial f(x)$ closed and convex $(x\in\mathbb{E})$
- $\partial f(x)$ compact and nonempty $\iff x\in\operatorname{int}(\operatorname{dom} f) \iff$ locally Lipschitz at x
- $\partial f(x)$ is a singleton $\iff$ f differentiable at x $\iff$ f continuously differentiable at x

Examples of subdifferentiation

- (Indicator function/Normal cone) Let $S \subset \mathbb{E}$.

Indicator function of S:

$$\delta_S : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}, \quad \delta_S(x) := \begin{cases} 0, & x \in S, \\ +\infty, & \text{else.} \end{cases}$$

$$\begin{aligned} \partial \delta_S(\bar{x}) &= \left\{ v \,\middle|\, \delta_S(x) \geq \delta_S(\bar{x}) + \langle v, x - \bar{x} \rangle \ (x \in \mathbb{E}) \right\} \\ &= \left\{ v \in \mathbb{E} \,\middle|\, \langle v, x - \bar{x} \rangle \leq 0 \ (x \in S) \right\} \\ &=: N_S(\bar{x}) \quad (\bar{x} \in S) \end{aligned}$$

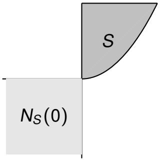


Figure: Normal cone

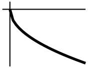
- (Euclidean norm) $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$. Then

$$\partial \|\cdot\|(\bar{x}) = \begin{cases} \left\{ \frac{\bar{x}}{\|\bar{x}\|} \right\} & \text{if} \quad \bar{x} \neq 0, \\ \mathbb{B} & \text{if} \quad \bar{x} = 0. \end{cases}$$

- (Empty subdifferential)

$$f : x \in \mathbb{R} \mapsto \begin{cases} -\sqrt{x} & \text{if} \quad x \geq 0, \\ +\infty & \text{else.} \end{cases}$$

$$\partial f(x) = \begin{cases} \left\{ -\frac{1}{2\sqrt{x}} \right\}, & x > 0, \\ \emptyset, & \text{else.} \end{cases}$$



The Fenchel conjugate

For $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ let $f^* : \mathbb{E} \to \overline{\mathbb{R}}$ be the function whose epigraph encodes the affine minorants of f:

$$\begin{aligned} & \operatorname{epi} f^* \overset{!}{=} \left\{(v,\beta) \,\middle|\, \langle v, x\rangle - \beta \le f(x) \quad (x \in \mathbb{E})\right\} \\ \Longrightarrow \quad & f^*(v) \le \beta \iff \sup_{x\in\mathbb{E}} \{\langle v, x\rangle - f(x)\} \le \beta \quad ((v,\beta) \in \mathbb{E}\times\mathbb{R}) \\ \Longrightarrow \quad & f^*(v) = \sup_{x\in\mathbb{E}} \{\langle v, x\rangle - f(x)\} \quad (v \in \mathbb{E}). \end{aligned} \tag{2}$$

Definition 9 (Fenchel conjugate).

Let $f : \mathbb{E} \to \overline{\mathbb{R}}$ proper. The function $f^* : \mathbb{E} \to \overline{\mathbb{R}}$ defined through (2) is called the (Fenchel) conjugate of f. The function $(f^{**}) := (f^*)^*$ is called the biconjugate of f.

Recall: $\Gamma := \left\{f : \mathbb{E} \to \overline{\mathbb{R}} \mid f \text{ convex and proper}\right\}$ and $\Gamma_0 := \{f \in \Gamma \mid f \text{ closed}\}$.

- f^* closed and convex - proper if $f \not\equiv +\infty$ with an affine minorant
- $f = f^{**}$ *proper* $\iff$ $f \in \Gamma_0$ (Fenchel-Moreau)
- $f^* = (\operatorname{cl} f)^*$ $(f \in \Gamma)$
- $f(x) + f^*(y) \ge \langle x, y\rangle$ $(x, y \in \mathbb{E})$ (Fenchel-Young Inequality)

Support functions: A special case of conjugacy

The *support function* σ_S of $S \subset \mathbb{E}$ (nonempty) is defined by

$$\sigma_S : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}, \quad \sigma_S(z) := \delta_S^*(z) = \sup_{x \in S} \langle x, z \rangle .$$

- σ_S is finite-valued if and only if S is bounded (and nonempty)
- $\sigma_S = \sigma_{\mathrm{conv}\, S} = \sigma_{\overline{\mathrm{conv}}\, S} = \sigma_{\mathrm{cl}\, S}$
- $\sigma_S^* = \delta_{\overline{\mathrm{conv}}\, S}$
- $\partial \sigma_S(x) = \left\{ z \in \overline{\mathrm{conv}}\, S \;\middle|\; x \in N_{\overline{\mathrm{conv}}\, S}(z) \right\}$
- σ_S is a norm if and only if S is symmetric, bounded and $0 \in \mathrm{int}\, S$.

Example: Let $\mathbb{B}_\infty$ be the unit ball in the maximum norm. Then

$$\sigma_{\mathbb{B}_\infty} = \|\cdot\|_1 .$$

Dual correspondences

'Every property of the primal object ($f \in \Gamma_0$) corresponds to a property of the dual object ($f^* \in \Gamma_0$).'

Proposition 10 (Dual correspondences).

Let $f \in \Gamma_0(\mathbb{E})$. Then:

(a) $\inf f = -f^*(0)$ *and* $\operatorname{argmin} f = -\partial f^*(0)$.

(b) *f level-bounded* $\iff 0 \in \operatorname{int}(\operatorname{dom} f^*)$.

(c) *f supercoercive* $\iff \operatorname{dom} f^* = \mathbb{E}$.

(d) *The following are equivalent:*

  (i) *f is essentially strictly convex, i.e. strictly convex on every convex subset of $\operatorname{dom} \partial f$;*

  (ii) *f^* is essentially smooth, i.e. ∂f^* is single-valued. In particular, $\partial f^*(x) = \nabla f^*(x)$ for all $x \in \operatorname{dom} \partial f^* = \operatorname{int}(\operatorname{dom} f^*)$.*

Guide.

These are all 'straightforward' except (d)! □

Interplay of conjugation and subdifferentiation

Theorem 11 (Subdifferential and conjugate function).

Let f be lsc, proper, convex. TFAE:

i) $y \in \partial f(x)$;

ii) $f(x) + f^*(y) = \langle x, y \rangle$;

iii) $x \in \partial f^*(y)$.

In particular, $\partial f^* = (\partial f)^{-1}$.

Proof.

Notice that

$$\begin{aligned} y \in \partial f(x) \quad & \Longleftrightarrow \quad f(z) \geq f(x) + \langle y, z - x \rangle \quad (z \in \mathbb{E}) \\ & \Longleftrightarrow \quad \langle y, x \rangle - f(x) \geq \sup_z \{\langle y, z \rangle - f(z)\} \\ & \Longleftrightarrow \quad f(x) + f^*(y) \leq \langle x, y \rangle \\ & \overset{\text{Fenchel–Young}}{\Longleftrightarrow} \quad f(x) + f^*(y) = \langle x, y \rangle, \end{aligned}$$

Applying the same reasoning to f^* and noticing that $f^{**} = f$ if $f \in \Gamma_0$, gives the missing equivalence. □

Infimal convolution

Definition 12 (Infimal convolution).

Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$. Then the function

$$f \# g : \mathbb{E} \to \overline{\mathbb{R}}, \quad (f \# g)(x) := \inf_{u \in \mathbb{E}} \{f(u) + g(x - u)\}$$

is called the *infimal convolution* of f and g. We call the infimal convolution $f \# g$ *exact at* $x \in \mathbb{E}$ if

$$\operatorname*{argmin}_{u \in \mathbb{E}} \{f(u) + g(x - u)\} \neq \emptyset.$$

We simply call $f \# g$ exact if it is exact at every $x \in \operatorname{dom} f \# g$.

We always have:

- $\operatorname{dom} f \# g = \operatorname{dom} f + \operatorname{dom} g$;
- $f \# g = g \# f$;
- f, g convex, then $f \# g$ convex (as $(f \# g)(x) = \inf_y h(x, y)$ with $h : (x, y) \mapsto f(y) + g(x - y)$ convex).

Example 13 (Distance functions).

Let $C \subset \mathbb{E}$. Then $d_C := \delta_C \# \|\cdot\|$, i.e.

$$d_C(x) = \inf_{u \in C} \|x - u\|$$

is the distance function of C, which is hence convex if C is a convex.

Conjugacy of infimal convolution

Proposition 14 (Conjugacy of inf-convolution).

Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$. Then the following hold:

a) $(f\#g)^* = f^* + g^*$;

b) *If $f, g \in \Gamma_0$ such that $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$, then $(f+g)^* = \mathrm{cl}\,(f^* \# g^*)$.*

Proof.

a) For all $y \in \mathbb{E}$, we have

$$\begin{aligned}(f\#g)^*(y) &= \sup_x \left\{\langle x, y\rangle - \inf_u \{f(u) + g(x-u)\}\right\}\\ &= \sup_{x,u} \{\langle x, y\rangle - f(u) - g(x-u)\}\\ &= \sup_{x,u} \{(\langle u, y\rangle - f(u)) + (\langle x-u, y\rangle - g(x-u))\}\\ &= f^*(y) + g^*(y).\end{aligned}$$

b) $(f^*\#g^*)^* \overset{a)}{=} f^{**} + g^{**} \overset{f,g\in\Gamma_0}{=} f + g \overset{clear?}{\in} \Gamma_0$, hence $(f^*\#g^*) \in \Gamma$.

$$\Longrightarrow \quad \mathrm{cl}\,(f^*\#g^*) = (f^*\#g^*)^{**} = (f+g)^*.$$

□

Drop the closure!

Theorem 15.

Let $f, g \in \Gamma_0$ such that

$$\mathrm{ri}\,(\mathrm{dom}\, f) \cap \mathrm{ri}\,(\mathrm{dom}\, g) \neq 0 \quad (CQ).$$

Then the following hold

a) *('Attouch-Brézis') $(f+g)^* = f^* \# g^*$, and the infimal convolution is exact, i.e. the infimum in the infimal convolution is attained on $\mathrm{dom}\, f^* \# g^*$.*

b) *(Sum rule) $\partial(f+g) = \partial f + \partial g$.*

Proof.

a) Hard work! See, e.g., Rockafellar (1970) or Bauschke/Combettes (2017).

b) Only show "$\subset$":[1] Let $v \in \partial(f+g)(x)$. By a), $\exists \bar{u} : (f+g)^*(v) = f^*(\bar{u}) + g(v - \bar{u})$. Thus,

$$\begin{aligned} v \in \partial(f+g)(x) \quad & \overset{\text{Th. 11}}{\iff} \quad (f+g)(x) + (f+g)^*(v) = \langle v, x\rangle \\ & \iff \quad f(x) + g(x) + f^*(\bar{u}) + g(v-\bar{u}) = \langle \bar{u}, x\rangle + \langle v - \bar{u}, x\rangle \\ & \overset{\text{Fenchel-Young}}{\iff} \quad \bar{u} \in \partial f(x), v - \bar{u} \in \partial g(x) \\ & \implies \quad v \in \partial f(x) + \partial g(x). \end{aligned}$$

□

[1] The converse direction always holds!

Conjugacy for convex-linear composites

Let $f \in \Gamma$ and $L \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$. Then

$$Lf : \mathbb{E}' \to \overline{\mathbb{R}}, \quad (Lf)(y) := \inf \left\{ f(x) \;\middle|\; L(x) = y \right\}$$

is convex[2].

Proposition 16.

Let $g : \mathbb{E} \to \overline{\mathbb{R}}$ be proper and $L \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$ and $T \in \mathcal{L}(\mathbb{E}', \mathbb{E})$. Then the following hold:

a) $(Lg)^* = g^* \circ L^*$.

b) $(g \circ T)^* = \mathrm{cl}\,(T^* g^*)$ if $g \in \Gamma$.

c) *The closure in b) can be dropped and the infimum is attained when finite if $g \in \Gamma_0$ and*

$$\mathrm{rge}\, T \cap \mathrm{ri}\,(\mathrm{dom}\, g) \neq \emptyset. \tag{3}$$

Guide.

a) Straightforward. b) From a) and Fenchel-Moreau.

c) Observe that $(g \circ T)^*(z) = (\delta_{\mathrm{gph}\, T} + \phi)^*(z, 0)$ for $\phi(x, y) \mapsto g(y)$. Apply Attouch-Brézis to the latter realizing that the (CQ) is equivalent to (3). □

[2] Show that $\mathrm{epi}_< Lf = T(\mathrm{epi}_< f)$ for $T : (x, y) \mapsto (Tx, y)$.

Fenchel-Rockafellar duality

Theorem 17 (Fenchel-Rockafellar duality).

Let $\gamma \in \Gamma(\mathbb{E}_1)$, $\phi \in \Gamma(\mathbb{E}_2)$ *and* $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$. *Define*

$$\min_{x \in \mathbb{E}_1} \phi(Lx) + \gamma(x) \quad \textbf{(primal problem)}$$

and

$$\max_{y \in \mathbb{E}_2} -\gamma^*(L^* y) - \phi^*(-y) \quad \textbf{(dual problem)}.$$

Set

$$p := \inf_{x \in \mathbb{E}_1} \{\phi(Lx) + \gamma(x)\} \quad \text{and} \quad d := \sup_{y \in \mathbb{E}_2} \{-\gamma^*(L^* y) - \phi^*(-y)\}.$$

The following hold:

a) *(Weak duality)* $p \geq d$.

b) *(Strong duality)* $p = d$ *if* $\mathrm{ri}\,(\mathrm{dom}\,\phi) \cap \mathrm{ri}\,L(\mathrm{dom}\,\gamma) \neq \emptyset$ *(CQ).*

c) *(Primal-dual recovery) If* $\gamma \in \Gamma_0$ *and* $g \in \Gamma_0$ *the following are equivalent:*

   i) $\bar{x} \in \partial\gamma^*(L^*\bar{y}), \quad L\bar{x} \in \partial\phi^*(-\bar{y})$;

   ii) $p = d, \bar{x} \in \mathrm{argmin}\,\gamma(x) + \phi(Lx), \bar{y} \in \mathrm{argmax} -\gamma^*(L^* y) - \phi^*(-y)$.

Fenchel-Rockafellar duality for regularized least-squares

For $A \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, $b \in \mathbb{E}_2$, $\lambda > 0$ and $g \in \Gamma_0(\mathbb{E}_1)$ consider

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda g(x). \tag{4}$$

To apply the Fenchel-Rockafellar duality scheme (Theorem 17) set

$$\phi := \frac{1}{2}\|(\cdot) - b\|^2, \quad \gamma := \lambda g, \quad L := A.$$

Since $\operatorname{dom}\phi = \mathbb{E}_2$, the qualification condition (CQ) is vacuously satisfied. Moreover

$$\phi^* = \frac{1}{2}\|\cdot\|^2 + \langle b, \cdot\rangle, \quad \gamma^* = \lambda \star g.$$

Consequently, the dual problem of (4) reads

$$\max_y \langle b, y\rangle - \frac{1}{2}\|y\|^2 - \lambda \star g^*(A^* y). \tag{5}$$

Primal-dual recovery: Assume that $\bar{y}$ is the unique (clear?) solution for the dual problem. Then

$$\bar{x} : \bar{x} \in \partial g^*(A^*\bar{y}) \quad \text{and} \quad b - A\bar{x} = \bar{y} \quad \text{solves} \quad (4).$$

Note that, by Proposition 10, $\partial g^*(A^*) = \nabla g^*(A^*\bar{y})$ if g is essentially strictly convex.

The proximal operator

Let $f \in \Gamma_0$ and $\lambda > 0$. Define the proximal operator of f by

$$\mathrm{prox}_f(x) := \operatorname*{argmin}_u \left\{ f(u) + \frac{1}{2}\|x-u\|^2 \right\}.$$

Proposition 18 (Proximal operator).

Let $f \in \Gamma_0$, $\lambda > 0$. Then:

$$\text{a)}\ \ \mathrm{prox}_f = (I + \partial f)^{-1}; \quad \text{b)}\ \ \mathrm{prox}_f \text{ is 1-Lipschitz.}$$

Proof.

a) Optimality conditions.

b) Set $u = \mathrm{prox}_f(x)$, $v := \mathrm{prox}_f(y)$. Then (via a))

$$\begin{aligned}
x-u \in \partial f(u),\ \ y-v \in \partial f(v) \quad &\overset{\text{subgrad. ineq.}}{\Longrightarrow} \quad \begin{cases} f(v) \geq f(u) + \langle x-u,\, v-u\rangle, \\ f(u) \geq f(v) + \langle y-v,\, u-v\rangle \end{cases} \\
&\overset{\text{summ.}}{\Longrightarrow} \quad 0 \geq \langle y-x+u-v,\, u-v\rangle \\
&\Longleftrightarrow \quad \langle x-y,\, u-v\rangle \geq \|u-v\|^2 \\
&\overset{\text{CSI}}{\Longrightarrow} \quad \|u-v\| \leq \|x-y\|.
\end{aligned}$$

□

2. Stability Analysis of regularized least-squares problems

The general setting

Consider the optimization problem

$$\min_{x\in\mathbb{R}^n} h(p,x) + \varphi(x) \tag{6}$$

where

- $h : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ (locally) smooth and convex in x;
- $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ closed, proper, convex.

We are interested in the solution map

$$S(p) \quad := \quad \operatorname*{argmin}_{x\in\mathbb{R}^n} \{h(p,x) + \varphi(x)\}$$
$$\overset{\text{convexity}}{=} \quad \left\{x \in \mathbb{R}^n \;\middle|\; 0 \in \nabla_x h(p,x) + \partial\varphi(x)\right\}.$$

(Smooth case) If $\varphi \in C^2$ then the classical implicit function theorem yields:

$$\bar{x} = S(\bar{p}),\ \nabla^2_{xx} h(\bar{p},\bar{x}) + \nabla^2\varphi(\bar{x}) \succ 0 \quad\Longrightarrow\quad \exists U \in \mathcal{N}(\bar{p}) : S \in C^1(U).$$

Question: What to do when φ is not smooth?

Set-convergence *(by Painlevé-Kuratowski)*

Let $\{C^k\}$ with $C^k \subset \mathbb{R}^n$ for all $k \in \mathbb{N}$. We define

- **(outer limit)**

$$\operatorname*{Lim\,sup}_{k\to\infty} C^k := \left\{x \mid \exists K \subset \mathbb{N}(\text{infinite}), \{x^k\} \to_K x : x^k \in C^k \quad \forall k \in K\right\}$$

- **(inner limit)**

$$\operatorname*{Lim\,inf}_{k\to\infty} C^k := \left\{x \mid \exists k_0 \in \mathbb{N}, \{x^k\} \to x : x^k \in C^k \quad \forall k \geq k_0\right\}.$$

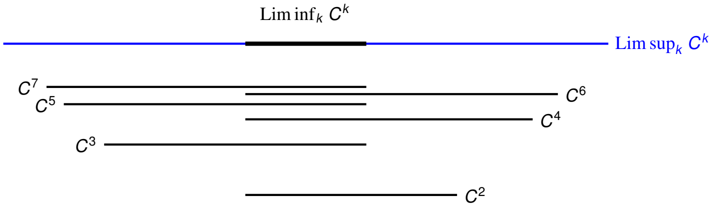


Figure: Example of $\{C^k\}$ non-convergent

Set-valued maps

For a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, we define:

- $\operatorname{dom} S := \left\{ x \in \mathbb{R}^n \;\middle|\; S(x) \neq \emptyset \right\}$ (domain);
- $\operatorname{gph} S := \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \;\middle|\; y \in S(x) \right\}$ (graph);
- $S^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n,\ S^{-1}(y) = \left\{ x \in \mathbb{R}^n \;\middle|\; y \in S(x) \right\}$ (inverse map).

We define the *outer limit* of S at $\bar{x}$.

$$\operatorname*{Lim\,sup}_{x \to \bar{x}} S(x) := \bigcup_{x^k \to \bar{x}} \operatorname*{Lim\,sup}_{k \to \infty} S(x^k) = \left\{ \bar{v} \;\middle|\; \exists : x^k \to \bar{x}, v^k \to \bar{v} : v^k \in S(x^k) \forall_k \right\}$$

We call S *outer semicontinuous (osc)* at $\bar{x} \in \mathbb{R}^n$ if $\operatorname{Lim\,sup}_{x \to \bar{x}} S(x) \subset S(\bar{x})$. Clearly,

$$S \text{ is osc (everywhere)} \iff \operatorname{gph} S \text{ is closed} \iff S^{-1} \text{ is osc (everywhere)}.$$

Example: The subdifferential operator as a set-valued map

The subdifferential operator ∂f for $f \in \Gamma$ is a set-valued mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$.

Proposition 19 (∂f).

Let $f \in \Gamma_0$. Then:

a) $\mathrm{ri}\,(\mathrm{dom}\, f) \subset \mathrm{dom}\,\partial f (\subset \mathrm{dom}\, f)$.

b) *For any $\lambda > 0$, we have*

$$\mathrm{gph}\,\partial f = F_\lambda(\mathbb{R}^n) \quad \text{where} \quad F_\lambda(r) = \left(\mathrm{prox}_{\lambda f}(r), \frac{r - \mathrm{prox}_{\lambda f}(r)}{\lambda}\right) \quad \text{is Lipschitz.}$$

In particular, $\mathrm{gph}\,\partial f$ *is closed.*

c) $(\partial f)^{-1} = \partial f^*$.

d) *(Monotonicity)* $\langle y - y', x - x' \rangle \geq 0 \quad \forall (x, y), (x', y') \in \mathrm{gph}\,\partial f$.

Proof.

a) (Sketch) Prove that $f'(x; \cdot) = \sigma_{\partial f(x)}$ is proper which yields $\partial f(x) \neq \emptyset$ for $x \in \mathrm{ri}\,(\mathrm{dom}\, f)$.

b) Use Proposition 18.

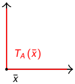
c) Theorem 11.

d) Simple application of the subgradient inequality. □

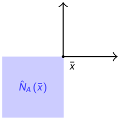
Variational Geometry

Let $A \subset \mathbb{R}^n$ and $\bar{x} \in A$. We define

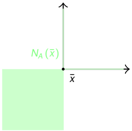
- the *tangent cone* $T_A(\bar{x}) := \operatorname{Lim\,sup}_{t \downarrow 0} \frac{A - \bar{x}}{t}$. The following hold:
  - We have
    $$d \in T_A(\bar{x}) \iff \exists \{t_k\} \downarrow 0, \{x_k \in A\} : \frac{x_k - \bar{x}}{t_k} \to d$$
    $$\iff \exists \{t_k\} \downarrow 0, \{d_k\} \to d : \bar{x} + t_k d_k \in A \ \forall k$$
  - $T_A(\bar{x})$ is a closed cone; convex if A is convex.



- the *regular normal cone* $\hat{N}_A(\bar{x}) = \left\{ v \;\middle|\; \limsup_{x \to_A \bar{x}} \frac{\langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \le 0 \right\}$. The following hold:
  - $\hat{N}_A(\bar{x}) =$ [3] $T_A(\bar{x})^\circ$.
  - $\hat{N}_A(\bar{x})$ is closed and convex.



- the *limiting normal cone* $N_A(\bar{x}) := \operatorname{Lim\,sup}_{x \to_A \bar{x}} \hat{N}_A(x)$. The following hold:
  - $N_A(\bar{x})$ is closed.
  - $N_A(\bar{x}) = \hat{N}_A(\bar{x}) = (A - \bar{x})^\circ$ (hence convex) if A is convex.



[3] For a convex set K its *polar cone* is $K^\circ := \{ v \mid \langle x, v \rangle \le 0 \ \forall x \in K \}$.

Basic tangent and normal cone calculus

Proposition 20 (Change of coordinates).

Let $D \subset \mathbb{R}^m$ and $C = F^{-1}(D)$ for $F : \mathbb{R}^n \to \mathbb{R}^m$ smooth and $\operatorname{rank} F'(\bar{x}) = m$ for $\bar{x} \in C$. Then for $\bar{u} = F(\bar{x})$:

$$\text{a) } T_C(\bar{x}) = F'(\bar{x})^{-1} T_D(\bar{u}); \quad \text{b) } \hat{N}_C(\bar{x}) = F'(\bar{x})^* \hat{N}_D(\bar{u}) \quad \text{c) } N_C(\bar{x}) = F'(\bar{x})^* N_D(\bar{u}).$$

Guide for $m = n$.

a) Apply inverse function theorem to $F(x) = u$ at $(\bar{x}, \bar{u})$.

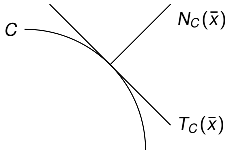
b) Use $\hat{N}_C(\bar{x}) = \left(F'(\bar{x})^{-1} T_D(\bar{u})\right)^\circ$ and invertibility of $F'(\bar{x})^*$.

c) Apply b) locally around $\bar{x}$, and $\operatorname{Lim\,sup}_{u \to_D \bar{u}} F'(\bar{x}) \hat{N}_D(u) = F'(\bar{x})^* \operatorname{Lim\,sup}_{u \to_D \bar{u}} \hat{N}_D(u)$. □

Corollary 21 (Smooth manifolds).

In Proposition 20 let $D := \{0\}$. Then:

$$\text{a) } T_C(\bar{x}) = \ker F'(\bar{x}); \quad \text{b) } \hat{N}_C(\bar{x}) = N_C(\bar{x}) = \operatorname{rge} F'(\bar{x})^*.$$



Graphical differentiation of set-valued maps

Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(\bar{x}, \bar{y}) \in \operatorname{gph} S$.

We define the *graphical derivative* $DS(\bar{x}|\bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ via

$$v \in DS(\bar{x}|\bar{y})(u) \;:\Longleftrightarrow\; (u, v) \in T_{\operatorname{gph} S}(\bar{x}, \bar{y}).$$

We define the *coderivative* $D^*S(\bar{x}|\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ via

$$v \in D^*S(\bar{x}|\bar{y})(u) \;:\Longleftrightarrow\; (v, -u) \in N_{\operatorname{gph} S}(\bar{x}, \bar{y}).$$

- When S is single valued (at $\bar{x}$) we write $D^{(*)}S(\bar{x}) := D^{(*)}S(\bar{x}|S(\bar{x}))$.
- Both $DS(\bar{x}|\bar{y})$ and $D^*S(\bar{x}|\bar{y})$ are positively homogenous maps, i.e.,

$$D^{(*)}S(\bar{x})(\lambda z) = \lambda D^{(*)}S(\bar{x})(z) \quad \forall \lambda > 0 \quad \text{and} \quad 0 \in D^{(*)}S(\bar{x})(0).$$

Example: Coderivative of $\partial\|\cdot\|_1$

Observe that

$$\partial\|\cdot\|_1(x) = \bigtimes_{i=1}^{n} \partial|\cdot|(x_i), \quad \partial|\cdot|(t) = \begin{cases} \{\mathrm{sgn}(t)\}, & t \neq 0, \\ [-1,1], & t = 0. \end{cases} \tag{7}$$

Consequently

$$\mathrm{gph}\,\partial\|\cdot\|_1 = \bigtimes_{i=1}^{n} \mathrm{gph}\,\partial|\cdot|.$$

Thus for $(x,v) \in \mathrm{gph}\,\partial\|\cdot\|_1$:

$$N_{\mathrm{gph}\,\partial\|\cdot\|_1}(x,v) = \bigtimes_{i=1}^{n} \begin{cases} \{0\}\times\mathbb{R}, & x_i \neq 0, \\ \mathbb{R}\times\{0\}, & x_i = 0, |v_i| < 1, \\ \mathbb{R}\times\{0\} \cup \{0\}\times\mathbb{R} \cup \mathbb{R}_+\times\mathbb{R}_-, & x_i = 0, v_i = -1, \\ \mathbb{R}\times\{0\} \cup \{0\}\times\mathbb{R} \cup \mathbb{R}_-\times\mathbb{R}_+, & x_i = 0, v_i = 1. \end{cases}$$

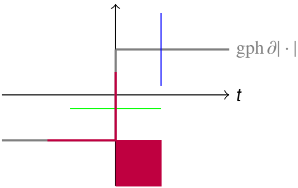


Figure: Normal cones to gph $\partial|\cdot|$

Hence

$$z \in D^*(\partial\|\cdot\|_1)(x|v)(w) \iff (z_i, -w_i) \in \begin{cases} \{0\}\times\mathbb{R}, & x_i \neq 0, \\ \mathbb{R}\times\{0\}, & x_i = 0, |v_i| < 1, \\ \mathbb{R}\times\{0\} \cup \{0\}\times\mathbb{R} \cup \mathbb{R}_+\times\mathbb{R}_-, & x_i = 0, v_i = -1, \\ \mathbb{R}\times\{0\} \cup \{0\}\times\mathbb{R} \cup \mathbb{R}_-\times\mathbb{R}_+, & x_i = 0, v_i = 1. \end{cases} \tag{8}$$

Calculus rules for Co- and Graphical derivatives

Proposition 22.

Let $S:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $(\bar{x},\bar{v})\in \mathrm{gph}\, S$, $F:\mathbb{R}^n\to\mathbb{R}^m$ continuously differentiable (at $\bar{x}$). The following hold:

(a) *(Inversion rule) We have*

$$y\in DS(\bar{x}|\bar{v})(s) \iff s\in D(S^{-1})(\bar{v}|\bar{x})(y) \quad \text{and} \quad z\in D^*S(\bar{x}|\bar{v})(w) \iff -w\in D^*(S^{-1})(\bar{v}|\bar{x})(-z)$$

b) *(Sum rule) We have* $D^{(*)}(S+F)(\bar{x}|\bar{v}+F(\bar{x}))(w) = D^{(*)}S(\bar{x}|\bar{v})(w) + F'(\bar{x})^{(*)}w$.

Proof.

a) $\mathrm{gph}\, S^{-1} = G^{-1}(\mathrm{gph}\, S)$ for $G(x,v)=(v,x)$. Then apply Proposition 20 (coordinate change).

b) (Coderivative statement) With $G(x,v)=(x,v+F(x))$, we have $\mathrm{gph}\,(S+F) = G^{-1}(\mathrm{gph}\, S)$.

$$\begin{aligned} z\in D^*(S+F)(\bar{x}|\bar{v}+F(\bar{z}))(w) \quad &\iff \quad (z,-w)\in N_{G^{-1}(\mathrm{gph}\, S)}(\bar{x},\bar{v}+F(\bar{x}))\\ &\overset{\text{Prop.20}}{\iff} \quad (z,-w)\in \begin{pmatrix} I & -F'(\bar{x})^* \\ 0 & I\end{pmatrix} N_{\mathrm{gph}\, S}(\bar{x},\bar{v})\\ &\iff^{4} \quad (z-F'(\bar{x})^*w,-w)\in N_{\mathrm{gph}\, S}(\bar{x},\bar{v})\\ &\iff \quad z\in D^*S(\bar{x},\bar{v})(w)+F'(\bar{x})^*w. \end{aligned}$$

□

[4] $\begin{pmatrix} I & -B \\ 0 & I\end{pmatrix}^{-1} = \begin{pmatrix} I & B \\ 0 & I\end{pmatrix}$

Locally Lipschitz maps and graphical differentiation

Let $F:\mathbb{R}^n\to\mathbb{R}^m$. We call F *locally Lipschitz*[5] at $\bar{x}$ if

$$\exists L,\varepsilon>0: \|F(x)-F(x')\|\leq L\|x-x'\|\quad \forall x,x'\in B_\varepsilon(\bar{x}).$$

We call

$$\mathrm{Lip}F(\bar{x}):=\limsup_{x,x'\to\bar{x}}\frac{\|F(x)-F(x')\|}{\|x-x'\|}$$

the *Lipschitz modulus of F at* $\bar{x}$. Clearly

$$F \text{ locally Lipschitz at } \bar{x} \iff \mathrm{Lip}F(\bar{x})<\infty.$$

Fact: Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be locally Lipschitz at $\bar{x}$. Then:

- (Scalarization formula) $D^*F(\bar{x})(w)=\partial(\langle w,\ F\rangle)(\bar{x})$[6] is nonempty, compact.
- (Lipschitz modulus) We have

$$\mathrm{Lip}F(\bar{x})=|D^*F(\bar{x})|^+:=\sup_{v\in\mathbb{B}}\ \sup_{v\in D^*F(\bar{x})(z)}\|v\| \tag{9}$$

- (Relation to Clarke Jacobian[7]) $\mathrm{conv}\, D^{(*)}F(\bar{x})(w)=\partial_C F(\bar{x})^{(*)}w$.

[5] In Rockafellar-Wets, this property is called strict continuity.

[6] For $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, we define the *limiting subdifferential* $\partial g(\bar{x}):=\left\{v\;\middle|\;(v,-1)\in N_{\mathrm{epi}\,g}(\bar{x},g(\bar{x})\right\}$

[7] $\partial_C F(\bar{x}):=\mathrm{conv}\left\{V\;\middle|\;\exists\{x^k\}\to\bar{x}: F'(x^k)\to V\right\}$

Definiteness properties of the coderivative

Proposition 23.

Let $f \in \Gamma_0$, and let $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$. Then

$$z \in D^*(\partial f)(\bar{x}|\bar{v})(w) \quad \Rightarrow \quad \langle z,\, w\rangle \geq 0.$$

Proof.

Recall from Proposition 18 that $P_\lambda := \operatorname{prox}_{\lambda f} = (I + \lambda\partial f)^{-1}$ for all $\lambda > 0$. Thus

$$\begin{aligned} z \in D^*(\partial f)(\bar{x}|\bar{v})(w) &\overset{\text{Prop.20}}{\iff} \lambda z \in D^*(\lambda\partial f)(\bar{x}|\lambda\bar{v})(w) \\ &\overset{\text{Prop.22(b)}}{\iff} \lambda z + w \in D^*(I + \lambda\partial f)(\bar{x}|\bar{x} + \lambda\bar{v})(w) \\ &\overset{\text{Prop.22(a)}}{\iff} -w \in D^* P_\lambda(\bar{x} + \lambda\bar{v})(-\lambda z - w) \\ &\overset{\text{pos. hom.}}{\iff} -\frac{w}{\|\lambda z + w\|} \in D^* P_\lambda(\bar{x} + \lambda\bar{v})\left(-\frac{\lambda z + w}{\|\lambda z + w\|}\right) \end{aligned}$$

Therefore

$$\frac{\|w\|}{\|\lambda z + w\|} \leq \sup_{\|r\|=1} \sup_{s \in D^* P_\lambda(\bar{x}+\lambda\bar{v})(r)} \|s\| \overset{\text{Eq.(9)}}{=} \operatorname{Lip} P_\lambda(\bar{x} + \lambda\bar{v}) = 1.$$

Hence

$$\|w\|^2 \leq \|\lambda z + w\|^2 = \lambda^2\|z\| + 2\lambda\langle z,\, w\rangle + \|w\|^2 \quad \overset{:\lambda,\, \lambda\downarrow 0}{\Rightarrow} \quad 0 \leq \langle z,\, w\rangle.$$

□

The Aubin property and the Mordukhovich criterion

Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with closed graph at $(\bar{x}, \bar{y}) \in \operatorname{gph} S$. We say that S has the *Aubin property* at $\bar{x}$ for $\bar{y}$ if there exist neighborhoods V of $\bar{x}$ and W of $\bar{y}$ as well as $\kappa > 0$ such that

$$S(x') \cap W \subset S(x) + \kappa \|x' - x\| \mathbb{B} \quad \forall x, x' \in V.$$

Remark: The Aubin property is a local property in that if S has the Aubin property at $\bar{x}$ for $\bar{y}$ then it has the Aubin property for every point $(x, y) \in \operatorname{gph} S$ sufficiently close to $(\bar{x}, \bar{y})$.

Theorem 24 (Mordukhovich criterion).

Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with closed graph at $(\bar{x}, \bar{u}) \in \operatorname{gph} S$. Then the following are equivalent:

- *S has the Aubin property at $\bar{x}$ for $\bar{y}$;*
- $D^* S(\bar{x}|\bar{y})(0) = \{0\}$.

Excursion: Monotonicity

We call $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ <u>*monotone*</u> if

$$\langle y - y', y - y' \rangle \quad \forall (x, y), (x', y') \in \operatorname{gph} T.$$

Example:

- $T = \partial f$ for $f \in \Gamma$.
- $T : x \mapsto Ax$ for $A \succeq 0$.

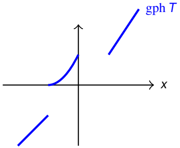


Figure: T monotone

Definition 25 (Maximal montonicity).

A monotone map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called <u>*maximally monotone*</u> if there is no enlargement of gph T possible without destroying monotonicity, i.e.,

$$\forall (\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n \setminus \operatorname{gph} T \; \exists (x, y) \in \operatorname{gph} T : \langle \hat{x} - x, \hat{y} - y \rangle < 0.$$

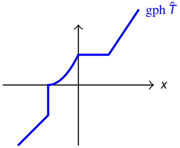


Figure: $\hat{T}$ maximally monotone

Facts:

- T (maximally) monotone $\iff$ T^{-1} (maximally) monotone.
- T maximally monotone $\Rightarrow$ gph T closed.
- T maximally monotone $\Rightarrow$ $T(x)$ closed, convex $\forall x$.

From Aubin property to local Lipschitzness

Proposition 26.

Let $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ have the Aubin property at $\bar{x}$ for $\bar{y} \in G(\bar{x})$ and assume that G is monotone. Then the following hold:

a) *G has a Lipschitz continuous single-valued localization at $\bar{x}$ for $\bar{y}$, i.e., there exist neighborhoods V of $\bar{x}$ and W of $\bar{y}$ such that $\hat{G} : x \in U \mapsto G(x) \cap W$ is single-valued and Lipschitz.*

b) *If G is convex-valued, then G is, in fact, single-valued and (locally) Lipschitz around $\bar{x}$.*

Proof.

a) Blackboard. b) Exercise! □

Corollary 27.

Under the assumptions of Proposition 26 assume that G is maximally monotone. Then G is single-valued and (locally) Lipschitz around $\bar{x}$.

Proof.

Follows from Proposition 26 as G is convex-valued. □

Locally Lipschitz implicit functions

Theorem 28.

Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at $(\bar{p}, \bar{x}) \in \operatorname{gph} S$ such that $f(p, \cdot)$ is monotone near $\bar{p}$, let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone. Define $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ by

$$S(p) = \left\{x \in \mathbb{R}^n \;\middle|\; 0 \in f(p, x) + F(x)\right\}, \quad \forall p \in \mathbb{R}^d.$$

Assume that

$$0 \in D_x f(\bar{p}, \bar{x})^* w + D^* F(\bar{x} | - f(\bar{p}, \bar{x}))(w) \quad \Rightarrow \quad w = 0 \quad \text{(Mordukhovich criterion).} \tag{10}$$

Then S is locally Lipschitz at $\bar{p}$.

High-level guide.

Set $Q := f(\bar{p}, \cdot) + F$. By the coderivative calculus from Proposition 22 find that

$$(10) \iff D^*(Q^{-1})(0|\bar{x})(0) = \{0\} \iff Q^{-1} \text{has Aubin property at } 0 \text{ for } \bar{x}$$

Since Q, thus Q^{-1} is maximally monotone that means that Q^{-1} is locally Lipschitz around 0. This now has to be leveraged to show that S is locally Lipschitz around $\bar{p}$; this hinges on the fact that perturbation (of f) enters smoothly (hence the difference between $f(\bar{p}, \cdot)$ and $f(p, \cdot)$ is controllable). □

The Mordukhovich criterion for regularized linear least-squares

Consider

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \lambda g(x), \quad (g \in \Gamma_0,\ \lambda > 0). \tag{11}$$

Let $\bar{x}$ solve (11), i.e. $\bar{u} := \frac{1}{\lambda}A^T(b - A\bar{x}) \in \partial g(\bar{x})$, i.e.

$$0 \in \underbrace{\frac{1}{\lambda}A^*(A\bar{x}-b)}_{=f(A,b,\lambda,\cdot)(\bar{x})} + \underbrace{\partial g}_{F}(\bar{x}).$$

Let $0 \in D_x f(A,b,\lambda,\bar{x})^* w + D^*F(\bar{x}|\bar{u})(w) = \frac{1}{\lambda}A^*Aw + D^*(\partial g)(\bar{x}|\bar{u})(w)$, i.e.

$$-\frac{1}{\lambda}A^*Aw \in D^*(\partial g)(\bar{x}|\bar{u})(w). \tag{12}$$

By Proposition 23 we have

$$0 \le \langle w, -A^*Aw\rangle$$

Inserting into (12) yields

$$0 \in D^*(\partial g)(\bar{x}|\bar{u})(w) \quad \overset{(\partial g)^{-1}=\partial g^*}{\Longleftrightarrow} \quad -w \in D^*(\partial g^*)(\bar{u}|\bar{x})(0).$$

Hence

$$\ker A \bigcap D^*(\partial g^*)(\bar{u}|\bar{x})(0) = \{0\} \iff \text{Mordukhovich criterion holds} \tag{13}$$

Example: the LASSO problem, i.e., $g = \|\cdot\|_1$

Set $g := \|\cdot\|_1$. Let $\bar{x}$ be a solution of the LASSO problem

$$\min \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1.$$

Thus

$$\bar{u} := \frac{1}{\lambda}A^T(b - A\bar{x}) \in \partial\|\cdot\|_1(\bar{x}) \quad \overset{(7)}{\iff} \quad \bar{u}_i \in \begin{cases} \{\mathrm{sgn}(\bar{x}_i)\}, & \bar{x}_i \neq 0, \\ \in [-1, 1], & \bar{x}_i = 0. \end{cases}$$

We note that

$$\begin{aligned} w \in D^*(\partial g^*)(\bar{u}|\bar{x})(0) \quad &\iff \quad 0 \in D^*(\partial g)(\bar{x}|\bar{u})(w) \\ &\overset{(8)}{\iff} \quad (0, -w_i) \in \begin{cases} \{0\} \times \mathbb{R}, & \bar{x}_i \neq 0, \\ \mathbb{R} \times \{0\}, & \bar{x}_i = 0, |\bar{u}_i| < 1, \\ \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R} \cup \mathbb{R}_+ \times \mathbb{R}_-, & \bar{x}_i = 0, \bar{u}_i = -1, \\ \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R} \cup \mathbb{R}_- \times \mathbb{R}_+, & \bar{x}_i = 0, \bar{u}_i = 1. \end{cases} \\ &\Longrightarrow \quad w_i = 0 \ \forall i \notin J := \{i \mid |\bar{u}_i| = 1\}. \end{aligned}$$

For $A_J = [a_j \ (i \in J)]$, the matrix whose columns are the columns of A corresponding to J we thus find:

$$\ker A \cap D^*(\partial g^*)(\bar{u}|\bar{x})(0) = \{0\} \iff \ker A_J = \{0\}.$$

Towards more general results: PLQ penalties

Let $\mathcal{P} = \left\{z \in \mathbb{R}^n \;\middle|\; \langle p_i, z\rangle \le \beta_i \ (i = 1, \dots, k)\right\} \subset \mathbb{R}^n$ be polyhedron and let $B \in \mathbb{S}^n_+$. We define the *piecewise-linear quadratic (PLQ) penalty*

$$\theta_{\mathcal{P},B}(y) = \sup_{z\in\mathcal{P}} \left\{\langle y, z\rangle - \frac{1}{2}\langle Bz, z\rangle\right\}.$$

Example: $\|\cdot\|_1 = \theta_{\mathcal{P},B}$ for $\mathcal{P} = \mathbb{B}_\infty,\ B = 0$. We note that:

- $\theta_{\mathcal{P},B} = (\delta_{\mathcal{P}} + q_B)^* \in \Gamma_0$ for $q_B(y) = \frac{1}{2}\langle Bz, z\rangle, \mathcal{P} \neq \emptyset$
- $\partial\theta^*_{\mathcal{P},B} = N_{\mathcal{P}} + B$.

Fact: $D^*N_{\mathcal{P}}(u|v)(0) = \operatorname{span}\left\{p_i \;\middle|\; i \in \mathcal{A}(u)\right\}$ where $\mathcal{A}(u) = \{i \in \{1, \dots, k\} \mid \langle p_i, u\rangle = \beta_i\}$. Thus, for $(\bar{x}, \bar{u}) \in \operatorname{gph}\theta_{\mathcal{P},B}$, we have

$$\begin{aligned} D^*(\partial\theta^*_{\mathcal{P},B})(\bar{u}|\bar{x})(0) &= D^*(N_{\mathcal{P}} + B)(\bar{u}|\bar{x})(0) \\ &= D^*N_{\mathcal{P}}(\bar{u}|\bar{x} - B\bar{u})(0) + B\cdot 0 \\ &= \operatorname{span}\left\{p_i \;\middle|\; i \in \mathcal{A}(\bar{u})\right\} \\ &= \operatorname{par}\partial\theta_{\mathcal{P},B}(\bar{u}). \end{aligned}$$

Towards quantitative results

Theorem 29.

Under the assumptions of Theorem 28 define $S:\mathbb{R}^d \rightrightarrows \mathbb{R}^n$ *by*

$$S(p) = \left\{x \in \mathbb{R}^n \;\middle|\; 0 \in f(p,x) + F(x)\right\}, \quad \forall p \in \mathbb{R}^d.$$

Assume that

$$0 \in D_x f(\bar{p},\bar{x})^* w + D^* F(\bar{x}| - f(\bar{p},\bar{x}))(w) \quad \Rightarrow \quad w = 0 \quad \text{(Mordukhovich criterion)}. \tag{14}$$

Then S *is locally Lipschitz at* $\bar{p}$ *with modulus*

$$L \le \limsup_{p\to\bar{p}} \max_{\|q\|\le 1} \inf_{w \in DS(p)(q)} \|w\|.$$

If F *is* proto-differentiable[8] *at* $(\bar{x}, -f(\bar{p},\bar{x}))$, S *is directionally differentiable at* $\bar{p}$ *with locally Lipschitz directional derivative (for* $G(p,x) := f(p,x) + F(x)$*) given by*

$$S'(\bar{p};q) = \left\{w \in \mathbb{R}^n \;\middle|\; 0 \in DG(\bar{p},\bar{x}|0)(q,w)\right\} \quad \forall q \in \mathbb{R}^d.$$

[8] $\partial\phi$ is proto-differentiable at $(\bar{x},\bar{u})$, e.g., if $\phi = g \circ H$ is *fully amenable*, i.e., g PLQ and $H \in C^2$ such that

$$\ker H'(\bar{x})^* \cap N_{\mathrm{dom}\, g}(H(\bar{x})) = \{0\} \quad \text{(basic constraint qualification)}$$

Application: unconstrained LASSO (stability) (Berk, Brugiapaglia, H. '23)

Apply Theorem 29 with $f(b,\lambda,x) := \frac{1}{\lambda}A^T(Ax-b)$, $\quad F := \partial\|\cdot\|_1$ such that

$$S(b,\lambda) = \left\{x \;\middle|\; 0 \in f(b,\lambda,x) + F(x)\right\} = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1\right\} \quad (\lambda > 0).$$

For $(\bar{b},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}_{++}$ let $\bar{x} \in S(\bar{b},\bar{\lambda})$. Assume that

$$\ker A_J = \{0\}.$$

Then S is locally Lipschitz and directionally differentiable at $(\bar{b},\bar{\lambda})$ with Lipschitz modulus

$$L \leq \frac{1}{\sigma_{\min}(A_J)^2}\left(\sigma_{\max}(A_J) + \left\|\frac{A_J^T(A\bar{x}-\bar{b})}{\bar{\lambda}}\right\|\right).$$

Moreover, the directional derivative $S'((\bar{b},\bar{\lambda});(\cdot,\cdot)) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ is locally Lipschitz and given as follows: for $(q,\alpha) \in \mathbb{R}^m \times \mathbb{R}$ there exists an index set $K = K(q,\alpha)$ with $I \subseteq K \subseteq J$ such that

$$S'((\bar{b},\bar{\lambda});(q,\alpha)) = L_K\left((A_K^TA_K)^{-1}A_K^T\left(q + \frac{\alpha}{\bar{\lambda}}(A\bar{x}-\bar{b})\right), 0\right).$$

3. The Maximum Entropy on the Mean Method for Linear Inverse Problems

Reminder: Probability measures and measure transformation

Let Ω be a nonempty set and let $\mathcal{F}$ be a σ-algebra[9] on Ω.

- $(\Omega, \mathcal{F})$ is called a measure space.
- A function $\mu : \mathcal{F} \to \mathbb{R}_+$ is called a *measure on* $(\Omega, \mathcal{F})$ *if:*
  - $\mu(\emptyset) = 0$;
  - *For* $A_k \in \mathcal{F}$ $(k \in \mathbb{N})$ *with* $A_k \cap A_j = \emptyset$ $(k \neq j)$: $\mu\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \sum_{k\in\mathbb{N}} \mu(A_k)$.

  If, in addition, $\mu(\Omega) = 1$, *we call* μ *a* probability measure, *and* $(\Omega, \mathcal{F}, \mu)$ *a* probability space.

Example: the Lebesgue measure comes with the measure space $(\mathbb{R}^n, \mathbb{B}_n)$, where $\mathbb{B}_n$ is the σ-algebra generated by the open sets in $\mathbb{R}^n$.

Theorem 30 (Measure transformation).

Let $(\Omega, \mathcal{F}, P)$ *be a probability space, and let* $(\Omega', \mathcal{F}')$ *be a measure space. Let* $f : \Omega \to \Omega'$ *be measurable. Moreover, let* $\phi : \Omega' \to \mathbb{R}$ *be measurable. Then:*

a) *For* $\mu := P \circ f^{-1}$ *we find that* $(\Omega', \mathcal{F}', \mu)$ *is a probability space.*

b) *It holds that*

$$\int_{\Omega} \phi \circ f \, dP = \int_{\Omega'} \phi d\mu.$$

---

[9] A collection of sets closed under complements and countable unions containing Ω.

Distiributions and expectations of random vectors

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}^n$ be a random vector (i.e., its components $X_i : \Omega \to \mathbb{R}$ are random variables).

- We call $\mu = P \circ X^{-1}$ the *distribution* or *law* of X, and we write $X \sim \mu$.
- The *expectation* or *mean* of f is

$$E[X] := [E[X_1], \dots, E[X_n]]^T \in \mathbb{R}^n \quad \text{for} \quad E[X_i] = \int_\Omega X_i dP.$$

Proposition 31 (Expectation of a random vector[10]).

Under the assumptions above we have:

$$E[X] = \left[\int_{\mathbb{R}^n} x_1 \mu(dx), \dots, \int_{\mathbb{R}^n} x_n \mu(dx)\right]^T =: E_\mu.$$

Proof.

Define $\pi_i : \mathbb{R}^n \to \mathbb{R}, \pi_i(x) = x_i$. Then we have

$$E[X_i] = E[\pi_i \circ X] = \int_\Omega \pi_i \circ X \, dP \overset{Th.\,30}{=} \int_{\mathbb{R}^d} \pi_i d\mu.$$

□

[10] '$(\Omega, \mathcal{F}, P)$ never mattered'

A Hölder-type inequality

Proposition 32.

Let μ be a probability measure on $\mathbb{R}^n$ and $f, g : \mathbb{R}^n \to \mathbb{R}$ measurable. Then

$$\int \exp(\lambda f + (1-\lambda)g)\, d\mu \leq \left(\int \exp f\, d\mu\right)^{\lambda} \cdot \left(\int \exp g\, d\mu\right)^{1-\lambda} \quad \forall \lambda \in (0,1).$$

When $\int \exp g d\mu$ and $\int \exp f d\mu$ are finite equality holds if and only if $f = g + \gamma$ for some $\gamma \in \mathbb{R}$.

Proof.

Prove the elementary inequality

$$a^{\lambda} b^{\lambda} \leq \lambda a + (1-\lambda) b \quad \forall a, b \geq 0 \quad ('=' \text{ iff } a = b). \tag{15}$$

Now set $a := \frac{\exp f}{\int \exp f\, d\mu}$ and $b := \frac{\exp g}{\int \exp g\, d\mu}$. Then

$$\frac{\exp(\lambda f + (1-\lambda)g)}{\left(\int \exp f\, d\mu\right)^{\lambda} \left(\int \exp g\, d\mu\right)^{1-\lambda}} = \left(\frac{\exp f}{\int \exp f\, d\mu}\right)^{\lambda} \left(\frac{\exp g}{\int \exp g\, d\mu}\right)^{1-\lambda} \overset{(15)}{\leq} \lambda \frac{\exp f}{\int \exp f\, d\mu} + (1-\lambda) \frac{\exp g}{\int \exp g\, d\mu}.$$

Therefore (applying integration on both sides yields)

$$\frac{\int \exp(\lambda f + (1-\lambda)g)\, d\mu}{\left(\int \exp f\, d\mu\right)^{\lambda} \left(\int \exp g\, d\mu\right)^{1-\lambda}} \leq \lambda \frac{\int \exp f\, d\mu}{\int \exp f\, d\mu} + (1-\lambda) \frac{\int \exp g\, d\mu}{\int \exp g\, d\mu} = 1,$$

which gives the desired result. □

Radon-Nikodym theorem - a tour de force

Let μ and ν be measures on the measure space $(\Omega, \mathcal{F})$. Then we call ν *absolutely continuous* with respect to μ (write: $\nu \ll \mu$) if for all $A \in \mathcal{F}$:

$$\mu(A) = 0 \quad \Longrightarrow \quad \nu(A) = 0.$$

Theorem 33 (Radon-Nikodym).

Let $(\Omega, \mathcal{F})$ be a measure space, and let μ and ν be finite[11,12] measures on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$. Then there exists a $(\mathcal{F})$-measurable function $f : \Omega \to \mathbb{R}_+$ such that

$$\forall A \in \mathcal{F} : \quad \nu(A) = \int_A f \, d\mu.$$

Remark: The function f in Theorem 33 is unique (up to changes on μ-null sets). We often write $\frac{d\nu}{d\mu}$ and call it the *Radon-Nikodym derivative* (of ν w.r.t. μ). When ν is probability measure (distribution) then $\frac{d\nu}{d\mu}$ is called a μ-density.

Let $\nu \ll \mu \ll \lambda$ be measures on $(\Omega, \mathcal{F})$. Then:

- $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda}$ λ-a.e.
- If g is measurable then $\int_\Omega g d\nu = \int_\Omega g \frac{d\nu}{d\mu} \, d\mu$.
- If $\mu \ll \mu$ (and $\nu \ll \mu$): $\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$ ν-a.e.

[11] That is $\mu(\Omega), \nu(\Omega) < \infty$.

[12] Or, more generally, σ-finite.

Higher level approach to linear inverse problems

The canonical linear inverse problem $Ax \approx b$ is usually solved via an optimization problem

$$\min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2}\|Ax - b\|^2 + g(x) \right\}$$

- $A \in \mathbb{R}^{m \times n}$: linear (forward) operator
- $b \in \mathbb{R}^m$: measurement vector
- g: (convex) regularizer

Higher level approach: Interpret the ground truth as a random vector with unknown distribution. Solve for a distribution Q that is close to a prior (guess) μ and such that its mean[13] E_Q satisfies $C \cdot E_Q \approx b$. This leads to

$$\min_Q \frac{1}{2}\|AE_Q - b\|^2 + K_\mu(Q)$$

where K_μ measures the compliance with (or distance to) μ.

- *Is this useful?*
- What is our choice of K_μ?

[13] i.e. $E_Q = \int_{\mathbb{R}^n} yQ(dy)$

Measuring compliance: the KL divergence

Let μ be a (prior) distribution, i.e., a probability measure on $\mathcal{X} \subset \mathbb{R}^n$ (i.e. $\mu = P \circ X^{-1}$ where X takes values in $\mathcal{X}$). The measure of compliance of another distribution Q with μ is measured by the **Kullback-Leibler divergence** $\mathrm{KL}(\cdot \mid \cdot) : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$[14] $\to \mathbb{R} \cup \{+\infty\}$,

$$\mathrm{KL}(Q \mid \mu) = \begin{cases} \int_\Omega \log\left(\frac{dQ}{d\mu}\right) dQ, & Q \ll \mu, \\ +\infty, & \text{otherwise}, \end{cases}$$

where $\frac{dQ}{d\mu}$ is the *Radon-Nikodym derivative*.

- $\mathrm{KL}(\cdot \mid \cdot)$ is convex, $\mathrm{KL}(\cdot \mid \mu)$ strictly convex for all $\mu \in \mathcal{P}(\mathcal{X})$.
- $\mathrm{KL}(Q \mid \mu) \geq 0$; equality if and only if $Q = \mu$ a.e.

[14] $\mathcal{P}(\mathcal{X})$: (convex) set of probability measures on $\mathcal{X}$.

KL divergence concretely

Let $\mu \in \mathcal{P}(\mathcal{X})$ be our prior/reference distribution. We are mainly interested in two cases:

1. $\mathcal{X} = \mathbb{R}^n$ and μ is absolutely continuous w.r.t. the Lebesgue measure ν, i.e. has a density $p = \frac{d\mu}{d\nu}$. In this case, if $Q \ll \mu$, Q has a density $\frac{dQ}{d\nu} = q$, and

$$\mathrm{KL}(Q \mid \mu) = \int_{\mathbb{R}^n} \log\left(\frac{q(x)}{p(x)}\right) q(x)dx.$$

   Note that we cover the case where $\mathcal{X} \subset \mathbb{R}^n$ via $\mu(X) = 1$.

2. μ is a discrete probability distribution, i.e., $\mathcal{X}$ is countable, and the probability mass function $p(x) = \mu(\{x\})$ has $\sum_{x\in\mathcal{X}} p(x) = 1$. Then $Q \ll \mu$ implies that μ has a probability mass function q and it holds that

$$\mathrm{KL}(Q \mid \mu) = \sum_{x\in\mathcal{X}} q(x) \log\left(\frac{q(x)}{p(x)}\right).$$

Example: Let μ be the uniform distribution on $\mathcal{X} := \{1, \ldots, N\}$, i.e. $p(i) = 1/N$ for all $i = 1, \ldots, N$. Then for $Q \ll \mu$ with PMF q, we have

$$\mathrm{KL}(Q \mid \mu) = \sum_{i=1}^{N} q(i) \underbrace{\log\left(\frac{q(i)}{1/N}\right)}_{\log(N)+\log(q(i))} = \log(N) + \sum_{i=1}^{N} \log(q(i))q(i).$$

The MEM re-formulation

Given a prior $\mu \in \mathcal{P}(\mathcal{X})$, the *maximum entropy on the mean method (MEMM)* for the linear inverse problem $Ax \approx b$ reads:

Determine $\bar{Q}$ as the solution of

$$\min_{Q \in \mathcal{P}(\mathcal{X})} \left\{ \frac{1}{2}\|A \cdot E_Q - b\|^2 + \alpha \mathrm{KL}(Q \mid \mu) \right\}, \tag{16}$$

and set $\bar{x} := E_{\bar{Q}}$ to be the estimate for the ground truth.

We observe that the MEM problem can be reformulated as follows:

$$\begin{aligned} \inf_{Q \in \mathcal{P}(\mathcal{X})} \left\{ \frac{1}{2}\|A \cdot E_Q - b\|^2 + \alpha \mathrm{KL}(Q \mid \mu) \right\} &= \inf_{\substack{(Q,x) \in \mathcal{P}(\mathcal{X}) \times \mathbb{R}^d: \\ E_Q = x}} \left\{ \frac{1}{2}\|A \cdot x - b\|^2 + \alpha \mathrm{KL}(Q \mid \mu) \right\} \\ &= \inf_{x \in \mathbb{R}^d} \left\{ \frac{1}{2}\|A \cdot x - b\|^2 + \alpha \underbrace{\inf_{\substack{Q \in \mathcal{P}(\mathcal{X}): \\ E_Q = x}} \mathrm{KL}(Q \mid \mu)}_{:= \kappa_\mu(x)} \right\} \\ &= \inf_{x \in \mathbb{R}^d} \left\{ \frac{1}{2}\|A \cdot x - b\|^2 + \alpha \kappa_\mu(x) \right\}. \end{aligned}$$

The MEM functional and the dual problem

We obtained the *reformulated problem*

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}\|A\cdot x - b\|^2 + \alpha\kappa_\mu(x). \tag{17}$$

with the *MEM functional* $\kappa_\mu : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$,

$$\kappa_\mu(x) = \inf_{Q\in\mathcal{P}(\Omega)} \{\mathrm{KL}(Q\mid\mu) + \delta_{\{0\}}(E_Q - x)\}.$$

- $\kappa_\mu \geq 0$; $\kappa_\mu(y) = 0$ if $y = E_\mu$, in particular, κ_μ proper if E_μ exists.
- κ_μ is convex (infimal projection!).

The million dollar question: *Who is κ_μ really?*

Cramér's function

Given a distribution $\mu \in \mathcal{P}(\mathcal{X})$, its *moment-generating function* is

$$M_\mu : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad M_\mu(z) := \int_{\mathcal{X}} \exp(\langle z, y\rangle)\mu(dy).$$

The *log-moment-generating function* or *cumulant generating function* $L_\mu : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ of $\mu \in \mathcal{P}(\mathcal{X})$ is

$$L_\mu(z) := \log \int_{\mathcal{X}} \exp(\langle z, \cdot\rangle) d\mu = \log(M_\mu(z)).$$

Its conjugate $L_\mu^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$,

$$L_\mu^*(y) := \sup_{z \in \mathbb{R}^d} \{\langle y, z\rangle - L_\mu(z)\}$$

is called *Cramér's function*[15] (fundamental in *large deviations theory*)

The key to computational tractability of the reformulated MEMM problem is to establish conditions (on μ) under which Cramér's function equals the MEM functional, i.e.

$$\kappa_\mu = L_\mu^*.$$

[15] Named after Swedish mathematician and statistician Harald Cramér who is considered as *'one of the giants of statistical theory'*.

Convexity of the log-MGF

Proposition 34 (Convexity of L_μ).

Let μ be a probability measure on $\mathcal{X} \subset \mathbb{R}^n$. Then L_μ is proper and strictly convex. In particular, $L_\mu \in \Gamma$.

Proof.

Note that $L_\mu(0) = \log \int_{\mathcal{X}} 1 d\mu = \log 1 = 0$, so L_μ is proper. Now note that, for $\lambda \in (0,1)$,

$$\begin{aligned} M_\mu(\lambda z + (1-\lambda)v) &= \int_{\mathcal{X}} \exp(\langle \lambda z + (1-\lambda)v, \cdot\rangle)\, d\mu \\ &\overset{\text{Prop. 32}^{16}}{\leq} \left(\int \exp\langle z, \cdot\rangle\ d\mu\right)^{\lambda} \left(\int \exp\langle v, \cdot\rangle\ d\mu\right)^{1-\lambda}. \end{aligned}$$

Therefore

$$L_\mu(\lambda z + (1-\lambda)v) \leq \log\left(\left(\int \exp\langle z, \cdot\rangle\ d\mu\right)^{\lambda}\left(\int \exp\langle v, \cdot\rangle\ d\mu\right)^{1-\lambda}\right) = \lambda L_\mu(z) + (1-\lambda)L_\mu(v).$$

If $z, v \in \operatorname{dom} L_\mu$, by Proposition 32, this can only be an equality if $\langle z, \cdot\rangle = \langle v, \cdot\rangle + \gamma$ for some $\gamma \in \mathbb{R}$, i.e. $z = v$. This shows that L_μ is, in fact, strictly convex. □

[16] With $f := \langle z, \cdot\rangle$ and $g := \langle v, \cdot\rangle$

The compact case

Proposition 35.

Let $\mathcal{X} \subset \mathbb{R}^n$ be compact, and let $\mu \in \mathcal{P}(\mathcal{X})$. Then the following hold:

a) *L_μ is strictly convex and (locally Lipschitz) continuous. In fact, L_μ is continuously differentiable with*

$$\nabla L_\mu(y) = \frac{\int_{\mathcal{X}} x \exp\langle y, \cdot\rangle \, d\mu}{M_\mu(y)}$$

b) *We have $\kappa_\mu = L_\mu^*$. In particular, $\kappa_\mu \in \Gamma_0$ is supercoercive, and essentially strictly convex.*

Guide.

a) By Proposition 34 L_μ is strictly convex. But by compactness of X, for any $z \in \mathcal{X}$, there is $\bar{s} = \mathrm{argmax}_{s\in\mathcal{X}} \exp\langle z, s\rangle$, so that

$$L_\mu(z) = \log \int_X \exp\langle z, \cdot\rangle \, d\mu \le \log \int_X \exp\langle z, \bar{s}\rangle \, d\mu = \langle z, \bar{s}\rangle .$$

Hence, L_μ is finite-valued and convex, hence (locally Lipschitz) continuous. The formula for the gradient follows from 'differentiation under the integral'.

b) The identity $\kappa_\mu = L_\mu^*$. is hard work (more later). □

The dual problem

Recall the (primal) MEM problem

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}\|A\cdot x-b\|^2+\alpha\kappa_\mu(x). \tag{18}$$

Proposition 36.

Under the assumptions of Proposition 35 the following hold:

a) *The dual problem of* (18) *(in the sense of Theorem 17) reads:*

$$\min_z \frac{\alpha}{2}\|z\|^2-\langle b,z\rangle+L_\mu(A^Tz). \tag{19}$$

b) *Let $\bar z$ be the unique solution of* (19). *Then $\bar x := \nabla L_\mu(A^T\bar z)$ solves* (18).

Proof.

a) $\kappa_\mu^* = L_\mu$ by Proposition 35.

b) The dual problem is strongly convex, so has a unique solution $\bar z$ (Prop. 3/6). The primal-dual recovery is given in Theorem 17 using that L_μ is smooth (Prop. 35). □

Applications

To solve the dual problem, one can use standard solvers like e.g. L-BFGS which was successfully done for (blind and non-blind) deblurring of

- *Barcodes/QR-codes*.

  Prior μ: Bernoulli.

  Reference: G. Rioux et al.: *Blind Deblurring of Barcodes via Kullback-Leibler Divergence.* IEEE TPAMI 43(1), 2021, pp.77-88.

- *General images.*

  Prior μ: Uniform on box.

  Reference: G. Rioux et al.: *The Maximum Entropy on the Mean Method for Image Deblurring. Inverse Problems 37, 2021.*

Fig. 11. Out of focus image of a QR code.

Fig. 12. Result of applying our method to a processed version of Fig. 11.

A data-driven approach for the MEM framework: the main idea

Recall the MEM dual problem for the linear inverse problem $Ax \approx b$:

$$\min_{z \in \mathbb{R}^m} \frac{\alpha}{2}\|z\|^2 - \langle b, z\rangle + L_\mu(A^T z), \tag{20}$$

where L_μ is the log-moment generating function $\log \int_{\mathcal{X}} \exp\langle \cdot, s\rangle \, d\mu$.

The obvious question: *'How to choose the prior μ?'*.

Idea for a data-driven approximation scheme: Let $X_1, X_2, \dots$ be a sequence of i.i.d.[17] $\mathcal{X}$-valued random variables on the probability space $(\Omega, \mathcal{F}, P)$ with shared distribution $\mu = P \circ X_1^{-1}$. Let $X_1(\omega), X_2(\omega), \dots,$ be a realization of the sequence.[18] Pick the first n-realizations (data!!). They give rise to the empirical distribution

$$\mu_n^{(\omega)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i(\omega)} \quad \text{for} \quad \mathbb{1}_{X_i(\omega)}(A) = \begin{cases} 1, & X_i(\omega) \in A, \\ 0, & \text{else} \end{cases} \quad \forall A \in \mathbb{B}_n \cap \mathcal{X}.$$

[17] Each $X_i \sim \mu$ and for all $n \in \mathbb{N}$ the RVs $X_1, \dots, X_n$ are independent.

[18] Pick one $\omega \in \Omega$, i.e, 'throw the dice once'.

The empirical dual

Plugging the empirical distribution $\mu_n^{(\omega)} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i(\omega)}$ into the log-moment generating function yields:

$$L_{\mu_n^{(\omega)}}(u) = \log \int_{\mathcal{X}} \exp\langle u, \cdot\rangle \, d\mu_n^{(\omega)} = \log\left(\frac{1}{n}\sum_{i=1}^n \exp\langle u, X_i(\omega)\rangle\right).$$

We now define the *'empirical dual'*

$$\min_{z\in\mathbb{R}^m} \frac{\alpha}{2}\|z\|^2 - \langle b, z\rangle + \log\left(\frac{1}{n}\sum_{i=1}^n \exp\left\langle A^T z, X_i(\omega)\right\rangle\right). \tag{21}$$

This problem has a unique solution $z_n(\omega)$. Define the vector (primal-dual recovery!)

$$x_n(\omega) := \nabla L_{\mu_n^{(\omega)}}(A^T z_n(\omega))$$

The million dollar question: Does $x_n(\omega)$ converge to the solution of the MEM problem as $n \to \infty$?

Excursion: Functional convergence

Let $f_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ $(k \in \mathbb{N})$.

$$
\begin{array}{llll}
f_k \xrightarrow{p} f & :\iff & f_k(x) \to f(x) \quad \forall x \in \mathbb{R}^n & \textbf{(pointwise)} \\
f_k \xrightarrow{e} f & :\iff & \operatorname{epi} f_k \to \operatorname{epi} f & \textbf{(epigraphical)} \\
f_k \xrightarrow{c} f & :\iff & f_k(x^k) \to f(x) \quad \forall x \in \mathbb{R}^n,\ \{x^k\} \to x & \textbf{(continuous)}
\end{array}
$$

Fact: $f_k \xrightarrow{e} f \iff \begin{cases} \liminf_{k\to\infty} f^k(x^k) \geq f(x) & \forall\, x^k \to x, \\ \limsup_{k\to\infty} f^k(x^k) \leq f(x) & \exists\, x^k \to x. \end{cases} \quad \forall x \in \mathbb{R}^n$

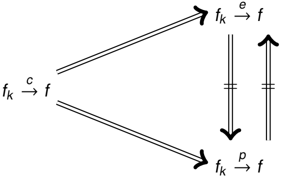


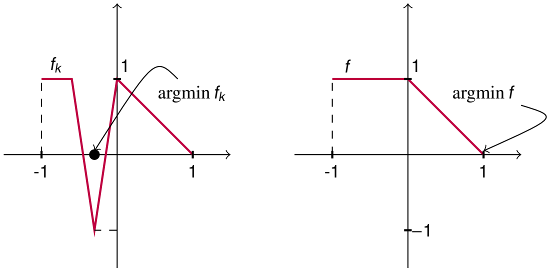
Figure: Connections between the convergence concepts

Pointwise convergence is not enough!

Consider the sequence of functions f^k, defined by

$$f^k(x) = \min\{1 - x, 1, 2k|x + \frac{1}{k}| - 1\} \quad \text{for any} \quad x \in [-1, 1].$$

For any $x \in \mathbb{R}$ we have $f^k(x) \to f(x) := \min\{1 - x, 1\}$ as $k \to \infty$.



The features of epigraphical convergence[19]

Proposition 37 (Poor man's sum rule).

Let $f_k \xrightarrow{e} f$ and let g be continuous and finite-valued. Then $f_k + g \xrightarrow{e} f + g$

Proposition 38.

Let $f_k \xrightarrow{e} f$. Then $\operatorname*{Lim\,sup}_{k\to\infty}(\operatorname{argmin} f_k) \subset \operatorname{argmin} f$.

The convex case allows for even stronger statements.

Proposition 39.

Let $\{f_k \in \Gamma_0\}$. Then the following hold:

a) *(Wijsman) $f_k \xrightarrow{e} f \iff f_k^* \xrightarrow{e} f^*$.*

b) *(Attouch) $f_k \xrightarrow{e} f \implies \operatorname{gph} \partial f_k \to \operatorname{gph} \partial f$.*

c) *If $f_k \xrightarrow{e} f$ f level-bounded and $x_k \in \operatorname{argmin} f_k$ for all $k \in \mathbb{N}$. Then $\{x_k\}$ is bounded and every cluster point belongs to $\operatorname{argmin} f$. If f is, in addition strictly convex and $\bar{x} = \operatorname{argmin} f$, then $x_k \to \bar{x}$.*

[19] See Rockafellar/Wets, Chapter 7 for details.

Epi-convergence of the empirical dual objective

Recall the empirical dual

$$\min_{z\in\mathbb{R}^m} \phi_n^\omega(z) := \frac{\alpha}{2}\|z\|^2 - \langle b,\, z\rangle + L_n^\omega(A^T z),$$

where $L_n^\omega(u) = \log\left(\frac{1}{n}\sum_{i=1}^n \exp\langle u,\, X_i(\omega)\rangle\right)$. We record that:

- ϕ_n^ω is strongly convex.
- $\phi_n^\omega = g + L_n^\omega \circ A^T$ where g is finite-valued and continuous.

In view of Proposition 37 and Proposition 39 for ϕ_n^ω to epigraphically converge to the objective function

$$\phi(z) := \frac{\alpha}{2}\|z\|^2 - \langle b,\, z\rangle + L_\mu(A^T z)$$

of the MEM dual, it suffices to show that $L_n^\omega \circ A^T \xrightarrow{e} L_\mu \circ A^T$. This is a probabilistic statement which reads like this, and leverages the theory of *epi-consistency* by King and Wets.

Proposition 40 (Choksi, King-Roskamp, H. '24).

Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space. Then

$$L_n^\omega \circ A^T \xrightarrow{e} L_\mu \circ A^T \quad (P) - a.e.$$

From empirical dual solutions to primal solutions

As a corollary of Proposition 40, we find that the objective function ϕ_n^ω of the empirical dual converges epigraphically to that of the MEM dual for almost every $\omega \in \Omega$. Smoothness and Attouch's theorem (Proposition 39 b)) now yield the following:

Corollary 41.

Let $\hat{z} \in \mathbb{R}^m$, and let $z_n \to \hat{z}$ be any sequence converging to $\hat{z}$. Then for almost every $\omega \in \Omega$,

$$\nabla L_n^\omega(A^T z_n) \to \nabla L_\mu(A^T \hat{z}).$$

Our derivations suggest the following scheme to solve a data-driven MEM approach for the linear inverse problem $Ax \approx b$.

(S1) Generate realizations $x_1, x_2, \ldots x_n$ (data!) of i.i.d. random vectors $X_i \sim \mu$.

(S2) Determine

$$\bar{z}_n := \operatorname*{argmin}_z \frac{\alpha}{2}\|z\|^2 - \langle b, z\rangle + \log\left(\frac{1}{n}\sum_{i=1}^n \exp\left\langle A^T z, x_i\right\rangle\right).$$

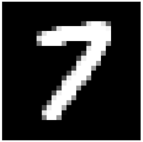
(S3) Set $\bar{x}_n := \nabla L_\mu(A^T \bar{z}_n)$.

A demonstration

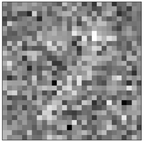
Want to recover a hand drawn digit x from noisy observations $b = x + \eta$. Construct $\mu_n^{(\omega)}$ for the MEM framework by sampling from the MNIST digits dataset.

(S1) For given n, draw sample $x_1, \ldots x_n$ uniformly at random from MNIST.

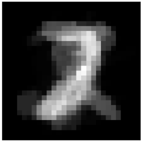
(S2 & S3) Using preferred method (e.g. here L-BFGS) find $\overline{z}_n = \operatorname{argmin}_z \phi_n(z)$. Set $\overline{x}_n = \nabla L_{\mu_n^{(\omega)}}(\overline{z}_n)$.



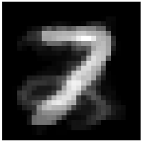
(a) Ground Truth x



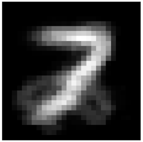
(b) Observed b, $\eta \sim \mathcal{N}(0, 0.1\|x\|_2)$



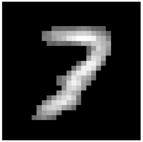
(c) x_n, $n = 100$



(d) x_n, $n = 5000$



(e) x_n, $n = 60000$



(f) Post-processed

The general setting

Given $\mathcal{X} \subset \mathbb{R}^n$, and $\mu \in \mathcal{P}(\mathcal{X})$, recall the MEM functional $\kappa_\mu : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$,

$$\kappa_\mu(x) = \inf_{Q \in \mathcal{P}(\Omega)} \{\mathrm{KL}(Q \mid \mu) + \delta_{\{0\}}(E_Q - y)\},$$

and the log-moment generating function $L_\mu : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$,

$$L_\mu(z) = \log \int_{\mathcal{X}} \exp \langle z, \cdot \rangle \, d\mu.$$

We want to find the crucial identity $\kappa_\mu = L_\mu^*$ for the two essential cases

- $\mathcal{X} = \mathbb{R}^d$ and μ is absolutely continuous w.r.t. to the Lebesgue measure;
- $\mathcal{X}$ is countable ($\mu(\mathcal{X} \cap A) = \sum_{x \in X} P(\{f = x\}) \mathbb{1}_{\{x\}}(A)$ for all $A \in \mathbb{B}_n$).

Key ingredient: Exponential families and Legendre-type functions.

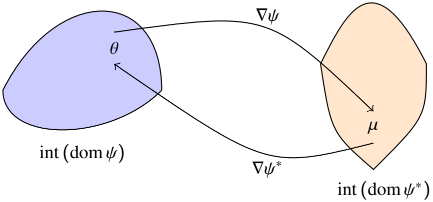
1st Ingredient: Legendre-type functions

Let $\psi \in \Gamma_0$. Say that ψ is of Legendre-type if it is both (cf. Proposition 10)

- essentially strictly convex;
- essentially smooth.

Rockafellar (1970): Let $\psi \in \Gamma_0$. Then

- ψ of Legendre-type $\iff$ ψ^* is of Legendre type.
- In this case: $\nabla\psi : \mathrm{int}\,(\mathrm{dom}\,\psi) \to \mathrm{int}\,(\mathrm{dom}\,\psi^*)$ is a bijection (with $(\nabla\psi)^{-1} = \nabla\psi^*$).



2nd ingredient: Exponential families

Let $\mu \in \mathcal{P}(\mathcal{X})$. The *natural parameter space* for μ is simply the domain of its (log-)MGF, i.e.,

$$\Theta_\mu := \left\{\theta \in \mathbb{R}^d \mid \int_{\mathcal{X}} \exp(\langle\theta, \cdot\rangle) d\mu < +\infty\right\} (= \operatorname{dom} L_\mu) .$$

The *standard exponential family* generated by μ is given by

$$\mathcal{F}_\mu := \left\{f_{\mu_\theta} \;\middle|\; f_{\mu_\theta}(y) := \exp(\langle y, \theta\rangle - \psi_\mu(\theta)), \quad \theta \in \Theta_\mu\right\}.$$

Properties and connections

- $\int_\mu f_{\mu_\theta} d\mu = 1$, thus $\mu_\theta := \mu \circ f_{\mu_\theta}^{-1}$ is a probability measure with $\frac{d\mu_\theta}{d\mu} = f_\theta$ $(\theta \in \Theta_\mu)$.
- For $y \in \operatorname{int}(\Theta_\mu)$ we have: $\bar{Q} \in \operatorname{argmin}_{Q:E_Q=y} \mathrm{KL}(Q \mid \mu) \implies \exists f \in \mathcal{F}_\mu : \; d\bar{Q} = f \cdot d\mu$.

The main result

The (standard) exponential family $\mathcal{F}_\mu$ is called

- *minimal* [20] if $\operatorname{int}\Theta_\nu \neq \emptyset$ and $\operatorname{int}(\operatorname{conv} S_\mu) \neq \emptyset$ [21];
- *steep* if ψ_ν is essentially smooth (automatically satisfied if Θ_ν open).

Theorem 42 (Vaisbourd et al.).

Suppose $\mu \in \mathcal{P}(\mathcal{X})$ generates a minimal and steep exponential family. Moreover, suppose one of the following holds:

- *S_μ is uncountable (absolutely continuous case);*
- *S_μ is countable and $\operatorname{conv} S_\mu$ is closed (which is always the case if S_μ is finite).*

Then $\boxed{\kappa_\mu = L_\mu^*.}$ *In this case, $0 \leq \kappa_P \in \Gamma_0$ is of Legendre type and coercive.*

[20] This can essentially be assumed w.l.o.g. by going to relative topology.

[21] S_μ: support of μ, i.e. the smallest closed set $\mu \subset \Omega$ s.t. $\mu(\mathcal{X} \setminus A) = 0$.

How is $\kappa_\mu = L_\mu^*$ useful?

If $\mu \in \mathcal{P}(\Omega)$ is separable (i.e. $\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_d$), then $M_\mu(\theta) = \prod_{i=1}^d M_{\mu_i}(\theta_i)$. Hence

$$\begin{aligned} L_\mu^*(y) &= \sup_{\theta \in \mathbb{R}^d} \left\{ \langle y, \theta \rangle - \log M_\mu(\theta) \right\} \\ &= \sum_{i=1}^d \sup_{\theta_i \in \mathbb{R}} \left\{ y_i \theta_i - \log M_{\mu_i}(\theta_i) \right\}. \end{aligned}$$

In many cases this yields analytic formulas for L_μ^*, i.e. κ_P (even without separability!).

Example: If μ is the multivariate normal distribution $N(E, \Sigma)$ for $\Sigma \succ 0$, i.e. $M_P(\theta) = \exp\left(\langle E, \theta \rangle + \frac{1}{2}\theta^T \Sigma \theta\right)$, then

$$\begin{aligned} L_\mu^*(y) &= \sup_{\theta \in \mathbb{R}^n} \{ \langle y, \theta \rangle - \log M_\mu(\theta) \} \\ &= \sup_{\theta \in \mathbb{R}^n} \left\{ \langle y - E, \theta \rangle - \frac{1}{2}\theta^T \Sigma \theta \right\} \\ &= \frac{1}{2}(y - E)^T \Sigma^{-1} (y - E). \end{aligned}$$

Examples of Cramér's function

Reference Distribution (μ)	Cramér Rate Function ($L_\mu^*(y)$)	dom L_μ^*
Multivariate Normal $\mu \in \mathbb{R}^d, \Sigma \in \mathbb{S}^d, \Sigma \succ 0$	$\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)$	$\mathbb{R}^d$
Poisson ($\lambda \in \mathbb{R}_{++}$)	$y\log(y/\lambda) - y + \lambda$	$\mathbb{R}_+$
Gamma ($\alpha, \beta \in \mathbb{R}_{++}$)	$\beta y - \alpha + \alpha \log\left(\frac{\alpha}{\beta y}\right)$	$\mathbb{R}_{++}$
Normal-inverse Gaussian $\alpha, \beta, \delta \in \mathbb{R} : \alpha \geq \lvert\beta\rvert$, $\delta > 0, \gamma := \sqrt{\alpha^2 - \beta^2}$	$\alpha\sqrt{\delta^2 + (y-\mu)^2} - \beta(y-\mu) - \delta\gamma$	$\mathbb{R}$
Multinomial ($p \in \Delta_d, n \in \mathbb{N}$)	$\sum_{i=1}^d y_i \log\left(\frac{y_i}{np_i}\right)$	$n\Delta_d \cap I(p)$[22]

In addition: Laplace, (Negative) Multinomial, Continuous/Discrete Uniform, Logistic, Exponential/Chi-Squared/Erlang (via Gamma), Binomial/Bernoulli/Categorical (via Multinomial), Negative Binomial & Shifted Geometric (via Negative Multinomial).

[22] $I(p) := \left\{x \in \mathbb{R}^d \mid x_i = 0 \text{ if } p_i = 0\right\}$