Conjugate Functions

Pontus Giselsson

Today's lecture

- conjugates and biconjugates
- Fenchel's inequality
- Fenchel-Young's equality
- conjugation and optimization
- subdifferentials using the conjugate
- conjugates of
  - image functions
  - functions precomposed with linear mappings
- subdifferential calculus rules

Conjugate functions

- standing assumption:

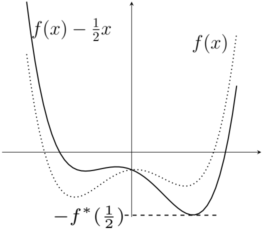
  > we assume that f is proper and has an affine minorizer

- the conjugate function is defined as

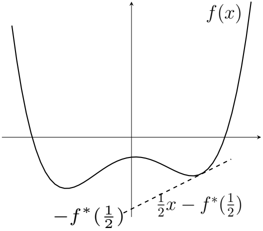
$$f^*(s) \triangleq \sup_x \{\langle s, x\rangle - f(x)\}$$

Graphical interpretation

- consider $f^*(s) = \sup_x \{\langle s, x\rangle - f(x)\} = -\inf_x \{f(x) - \langle s, x\rangle\}$
- "(-) smallest value of f when tilted by $\langle s, x\rangle$"
- example: $f^*(\frac{1}{2})$



interpretation 1



interpretation 2

Conjugate properties

recall from lecture on convex functions:

- the conjugate is convex, since supremum of affine functions
- it is closed since epigraph intersection of closed half-spaces

Further properties

- assume affine minorizer to $f(x)$ on form $\langle s_0, x\rangle - b$
- the conjugate function $f^* \not\equiv \infty$:

$$f^*(s_0) = \sup_x \{\langle s_0, x\rangle - f(x)\} \leq \sup_x \{\langle s_0, x\rangle - \langle s_0, x\rangle + b\} \leq b$$

- the conjugate $f^*(s) > -\infty$ for all s and has affine minorizer:

$$f^*(s) = \sup_x \{\langle s, x\rangle - f(x)\} \geq \langle s, \bar{x}\rangle - f(\bar{x})$$

  where $\bar{x}$ is a points with $f(\bar{x}) < \infty$ (exists by assumption)
  (use same $\bar{x}$ for all s to get affine minorizer)

- conjugate satisfies assumptions for taking conjugate!

Biconjugate

- the biconjugate f^{**} is obtained by conjugating twice, i.e.

$$f^{**}(x) = (f^*)^*(x)$$

- biconjugate can be written as

$$\begin{aligned}
f^{**}(x) &= \sup_s \{\langle x, s\rangle - f^*(s)\} \\
&= \sup_s \left\{ \langle x, s\rangle - \sup_z \{\langle s, z\rangle - f(z)\} \right\} \\
&= \sup_{s,r} \left\{ \langle x, s\rangle - r \mid r = \sup_z \{\langle s, z\rangle - f(z)\} \right\} \\
&= \sup_{s,r} \left\{ \langle x, s\rangle - r \mid r \geq \sup_z \{\langle s, z\rangle - f(z)\} \right\} \\
&= \sup_{s,r} \{\langle x, s\rangle - r \mid r \geq \langle s, z\rangle - f(z) \text{ for all } z\} \\
&= \sup_{s,r} \{\langle s, x\rangle - r \mid \langle s, z\rangle - r \leq f(z) \text{ for all } z\}
\end{aligned}$$

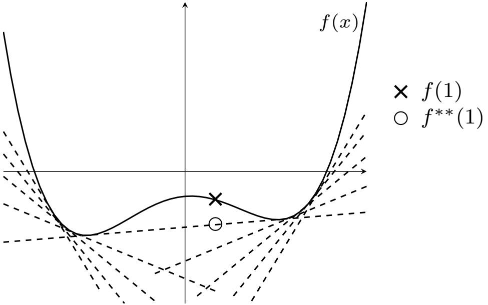
- do you recall this expression?

Graphical interpretation

- expression:

$$f^{**}(x) = \sup_{y,r} \{\langle y, x\rangle - r \mid \langle y, z\rangle - r \leq f(z) \text{ for all } z\}$$

"search for affine minorizers to f with largest value at x"



- biconjugate is closed convex hull
- $f^{**} \leq f$
- $f = f^{**} \Leftrightarrow$ cl conv$f = f \Leftrightarrow f$ proper closed convex

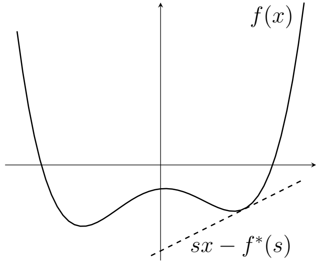
Fenchel's inequality

- from definition of conjugate function

$$f^*(s) = \sup_x \{\langle s, x\rangle - f(x)\}$$

we get for any $x, s \in \mathbf{R}^n$

$$f^*(s) + f(x) \geq \langle s, x\rangle \quad \text{or} \quad f(x) \geq \langle s, x\rangle - f^*(s)$$



- affine function $x \mapsto \langle s, x\rangle - f^*(s)$ minorizes $f(x)$

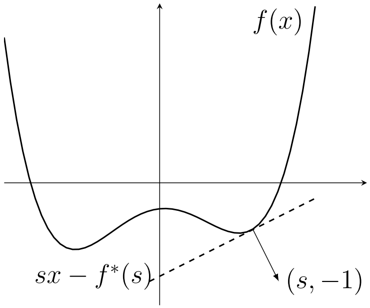
Fenchel-Young's equality

- how do x and s relate when we have equality in

$$f(x) \geq \langle s, x \rangle - f^*(s)$$

i.e., when

$$f(x) = \langle s, x \rangle - f^*(s)$$



- we have equality iff $(s, -1) \in N_{\text{epi } f}(x, f(x))$ or $s \in \partial f(x)$

Proof

$$\boxed{f(x) = \langle s, x\rangle - f^*(s) \Leftrightarrow s \in \partial f(x)}$$

- $s \in \partial f(x)$ iff (defintion of subgradient)

$$\begin{aligned}
& f(y) \geq f(x) + \langle s, y - x\rangle \text{ for all } y \\
\Leftrightarrow \quad & \langle s, y\rangle - f(y) \leq \langle s, x\rangle - f(x) \text{ for all } y \\
\Leftrightarrow \quad & \sup_y \{\langle s, y\rangle - f(y)\} \leq \langle s, x\rangle - f(x) \\
\Leftrightarrow \quad & f^*(s) \leq \langle s, x\rangle - f(x)
\end{aligned}$$

- Fenchel's inequality always holds:

$$f^*(s) \geq \langle s, x\rangle - f(x)$$

  inequality reversed $\Rightarrow$ equality holds

- simple yet powerful result!

Consequence of Fenchel-Young

for general f we have $\boxed{s \in \partial f(x) \Rightarrow x \in \partial f^*(s)}$

- proof: since $s \in \partial f(x)$, Fenchel-Young and $f \geq f^{**}$ gives

$$0 = f^*(s) + f(x) - \langle s, x \rangle \geq f^*(s) + f^{**}(x) - \langle s, x \rangle$$

- Fenchel's inequality says that other direction holds:

$$0 \leq f^*(s) + f^{**}(x) - \langle s, x \rangle$$

  i.e., this implies equality,

$$0 = f^*(s) + (f^*)^*(x) - \langle s, x \rangle$$

  which is equivalent to $x \in \partial f^*(s)$

Consequence of Fenchel-Young

for general f we have $\boxed{x \in \partial f^*(x) \Rightarrow s \in \partial f^{**}(s)}$

- apply $x \in \partial g(s) \Rightarrow s \in \partial g^*(x)$ to $g(s) = f^*(s)$:

$$x \in \partial g(s) = \partial f^*(s) \qquad \Rightarrow \qquad s \in \partial g^*(x) = \partial f^{**}(x)$$

Consequence of Fenchel-Young

proper closed convex f:

- we have

$$f(x) + f^*(s) - \langle s, x \rangle = 0 \Leftrightarrow s \in \partial f(x) \Leftrightarrow x \in \partial f^*(s)$$

- proof:
  - First equivalence: Fenchel-Young's equality
  - Second equivalence $\Rightarrow$: as above
  - Second equivalence $\Leftarrow$: follows from $f^{**} = f$:

$$x \in \partial f^*(s) \Rightarrow s \in \partial f^{**}(x) = \partial f(x)$$

Conjugation and optimization

- we have

$$\inf_x f(x) = -\sup_x \{\langle 0, x \rangle - f(x)\} = -f^*(0)$$

- Fermat's rule says:

$$x \text{ minimizes } f(x) \qquad \Leftrightarrow \qquad 0 \in \partial f(x)$$

- can you characterize $\operatorname{Argmin}_x f(x)$ if f proper closed convex?

$$\operatorname*{Argmin}_x f(x) = \partial f^*(0)$$

(since $x \in \partial f^*(0) \Leftrightarrow 0 \in \partial f(x)$)

Subdifferential of conjugate

- the subdifferential to the conjugate function satisfies

$$\partial f^*(s) \supseteq \operatorname*{Argmax}_x \{\langle s, x\rangle - f(x)\}$$

proof:

$$\begin{aligned} x^\star \in \operatorname*{Argmax}_x \{\langle s, x\rangle - f(x)\} &\Leftrightarrow x^\star \in \operatorname*{Argmin}_x \{f(x) - \langle s, x\rangle\} \\ &\Leftrightarrow 0 \in \partial(f(x^\star) - \langle s, x^\star\rangle) \\ \text{(assume)} &\Leftrightarrow 0 \in \partial f(x^\star) - s \\ &\Leftrightarrow s \in \partial f(x^\star) \\ &\Rightarrow x^\star \in \partial f^*(s) \end{aligned}$$

- if in addition f is closed convex, then

$$\partial f^*(s) = \operatorname*{Argmax}_x \{\langle s, x\rangle - f(x)\}$$

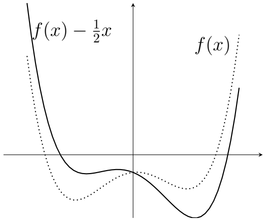
proof: last implication is equivalence in above proof

Proof of assumption

- for any proper f, we have $\partial(f(x) - \langle s, x\rangle) = \partial f(x) - s$
- $u \in \partial(f(x) - \langle s, x\rangle)$ iff

$$\begin{aligned} f(y) - \langle s, y\rangle &\geq f(x) - \langle s, x\rangle + \langle u, y - x\rangle \\ \Leftrightarrow \qquad f(y) &\geq f(x) + \langle u + s, y - x\rangle \end{aligned}$$

  i.e., iff $u + s \in \partial f(x)$ or $\partial(f(x) - \langle s, x\rangle) + s = \partial f(x)$
- example:



Subdifferential of function

- from previous slide: if f is closed convex, then

$$\partial f^*(s) = \operatorname*{Argmax}_x \{\langle s, x\rangle - f(x)\}$$

- apply to f^* (since closed convex):

$$\partial f^{**}(x) = \operatorname*{Argmax}_s \{\langle x, s\rangle - f^*(s)\}$$

- if f closed convex, then $f = f^{**}$ and

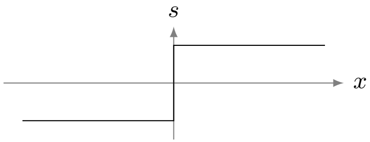
$$\partial f(x) = \operatorname*{Argmax}_s \{\langle x, s\rangle - f^*(s)\}$$

Relation between subdifferentials

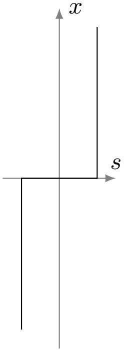
- we know that for proper closed convex f

$$s \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(s)$$

- ∂f and ∂f^* are each others images under mapping $(x, s) \mapsto (s, x)$
- example: $f(x) = |x|$, draw ∂f^*



$s \in \partial f(x)$



$x \in \partial f^*(s)$

Conjugate of image function and precomposition

- next we will compute the conjugates of image functions (Lg):

$$(Lg)(x) = \inf_{Ly=x} g(y)$$

- and the functions with precomposition $(g \circ L)$:

$$(g \circ L)(x) = g(Lx)$$

Conjugate of image function

- let g be proper with affine minorizer and L be a linear mapping
- assume:

$$\mathrm{Im} L \cap \mathsf{dom} g \neq \emptyset$$

- then Lg is proper and has an affine minorizer and its conjugate is

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

- proof: (first show that Lg is proper and has affine minorizer)

$$\begin{aligned}
(Lg)^*(s) &= \sup_x \left\{ \langle s, x \rangle - \inf_{Ly=x} g(y) \right\} \\
&= \sup_{x, Ly=x} \{ \langle s, x \rangle - g(y) \} \\
&= \sup_y \{ \langle s, Ly \rangle - g(y) \} \\
&= \sup_y \{ \langle L^* s, y \rangle - g(y) \} \\
&= g^*(L^* s) = (g^* \circ L^*)(s)
\end{aligned}$$

Conjugate of precomposition function

- let g be proper closed convex and L be a linear operator
- assume:

$$\mathrm{Im}L \cap \text{ri dom}g \neq \emptyset$$

- then $(g \circ L)^* = L^* g^*$ and for every $s \in \text{dom}(g \circ L)^*$,

$$(g \circ L)^*(s) = L^* g^*(s) = \min_p \{g^*(p) \mid L^* p = s\}$$

  i.e., the minimum is attained
- proof: apply previous result:

$$(L^* g^*)^* = g^{**} \circ L^{**} = g \circ L$$

  taking again the conjugate:

$$(g \circ L)^* = (L^* g^*)^{**} \stackrel{?}{=} L^* g^*$$

  where the last equality holds if $(L^* g^*)$ is proper closed convex
- proper and convex shown before, closedness can be shown if $\mathrm{Im}L \cap \text{ri dom}g \neq \emptyset$

Key result 1

- let's summarize the results from the previous slides:

Assume that g is proper closed and convex, that L is a linear operator, and that $\mathrm{Im} L \cap \mathrm{ri}\ \mathrm{dom} g \neq \emptyset$ then for $s \in \mathrm{dom}\ (g \circ L)^*$:

$$(g \circ L)^*(s) = (L^* g^*)(s) = \min_p \{g^*(p) \mid L^* p = s\}$$

i.e., the the conjugate of the precomposition function $g \circ L$ is the image function $(L^* g^*)$, and the minimum in the image function definition is attained.

- this result will be the main result from which we can:
  - prove subdifferential calculus rules
  - derive strong duality
  - show necessary and sufficient optimality conditions

Key result 2

- let f, g be proper closed convex and L be linear and assume

ri dom$g \cap$ ri L(dom$f$) $\neq \emptyset \quad \iff \quad$ ri dom$(g \circ L) \cap$ ri dom $f \neq \emptyset$

- then

$$\min_{\mu} \{f^*(s - L^*\mu) + g^*(\mu)\} = (f + g \circ L)^*(s)$$

- (note that minimum attained)
- (actually a Corollary of Key result 1)

Proof sketch of "Key result 2"

- let

$$h(x, y) := f(x) + g(y)$$
$$Kx := (x, Lx)$$

- then: $(f + g \circ L)^* = (h \circ K)^*$
- properties of h and K: assumption that

$$\begin{aligned}
& \text{ri dom } f \cap \text{ri dom } (g \circ L) \neq \emptyset \\
\iff \quad & \exists x | x \in \text{ri dom } f \cap \text{ri dom } (g \circ L) \\
\iff \quad & \exists x | (x, x) \in \text{ri dom } f \times \text{ri dom } (g \circ L) \\
\iff \quad & \exists x | (x, Lx) \in \text{ri dom } f \times \text{ri dom } g \\
\iff \quad & \exists x | (x, Lx) \in \text{ri } (\text{dom } f \times \text{dom } g) \\
\iff \quad & \text{Im } K \cap \text{ri dom } h \neq \emptyset
\end{aligned}$$

- $\Rightarrow$ can apply "Key result 1"!

Proof continued

- to use "Key result" $\big((h \circ K)^*(s) = (K^* h^*)(s)\big)$ compute K^*, h^*:
- adjoint K^* to $Kx = (x, Lx)$ is given by:

$$\begin{aligned} \langle Kx, (y,z) \rangle &= \langle x, y \rangle + \langle Lx, z \rangle = \langle x, y \rangle + \langle x, L^* z \rangle \\ &= \langle x, y + L^* z \rangle = \langle x, K^*(y,z) \rangle \end{aligned}$$

  i.e., $K^*(y,z) = y + L^* z$ and
- conjugate h^* is given by:

$$\begin{aligned} h^*(\lambda, \mu) &= \sup_{x,y} \{ \langle (\lambda, \mu), (x,y) \rangle - f(x) - g(y) \} \\ &= \sup_{x,y} \{ \langle \lambda, x \rangle + \langle \mu, y \rangle - f(x) - g(y) \} \\ &= \sup_{x} \{ \langle \lambda, x \rangle - f(x) \} + \sup_{y} \{ \langle \mu, y \rangle - g(y) \} \\ &= f^*(\lambda) + g^*(\mu) \end{aligned}$$

Proof continued

- apply "Key result 1":

$$
\begin{aligned}
(f + g \circ L)^*(s) &= (h \circ K)^*(s) \\
&= (K^* h^*)(s) \\
&= \min_{K^*(\lambda,\mu)=s} h^*(\lambda, \mu) \\
&= \min_{\mu,\lambda} \{f^*(\lambda) + g^*(\mu) : \lambda + L^*\mu = s\} \\
&= \min_{\mu} \{f^*(s - L^*\mu) + g^*(\mu)\}
\end{aligned}
$$

- where we get existence of μ and λ due to "Key result 1"

Notes on Key results

- Key result 2 is Corollary of Key result 1
- Key result 1 and 2 will be used to show when

$$\partial(f + g \circ L) = \partial f + L^* \circ \partial g \circ L$$

  (which will be used to show optimality conditions)
- Key result 2 will also be used to show when strong duality holds

Subdifferential sum

- for differentiable f and g, the chain-rule gives

$$\nabla(f + g \circ L) = \nabla f + L^* \circ \nabla g \circ L$$

- for subdifferentiable functions f and g, when do we have

$$\partial(f + g \circ L) = \partial f + L^* \circ \partial g \circ L?$$

Subdifferential sum

- we start with the case where $L = \mathrm{Id}$
- assume that f, g are proper closed and convex and that

$$\text{ri dom} g \cap \text{ri dom} f \neq \emptyset$$

then

$$\partial(f+g)(x) = \partial f(x) + \partial g(x)$$

for every $x \in$ dom $(f+g) =$ dom $f \cap$ dom g

Proof of subdifferential sum $\Leftarrow$

- assume that $s_1 \in \partial f(x)$ and $s_2 \in \partial g(x)$
- add definitions of subdifferential operator

$$
\begin{aligned}
& f(y) + g(y) \geq f(x) + g(x) + \langle s_1 + s_2, y - x \rangle \\
\iff \quad & (f+g)(y) \geq (f+g)(x) + \langle s_1 + s_2, y - x \rangle
\end{aligned}
$$

- therefore $s_1 + s_2 \in \partial (f+g)(x)$

Proof of subdifferential sum $\Rightarrow$

- assume that $s \in \partial(f+g)(x)$
- Fenchel-Young's equality gives

$$(f+g)^*(s) + (f+g)(x) - \langle s, x\rangle = 0$$

- since $s \in \mathrm{dom}(f+g)^*$ we apply "Key result 2", i.e., there $\exists\mu$:

$$f^*(s-\mu) + g^*(\mu) + f(x) + g(x) - \langle s, x\rangle = 0 \qquad (1)$$

- by Fenchel-Young's inequality, we have

$$\begin{aligned} f^*(s-\mu) + f(x) - \langle s-\mu, x\rangle &\leq 0 \\ g^*(\mu) + g(x) - \langle \mu, x\rangle &\leq 0 \end{aligned}$$

- by (1), these must be equalities, i.e.:

$$s-\mu \in \partial f(x) \qquad\qquad \mu \in \partial g(x)$$

and

$$s = (s-\mu) + \mu \in \partial f(x) + \partial g(x)$$

Example – First-order optimality condition

- assume that ri dom $f \cap$ ri $C \neq \emptyset$
- an x optimizes $\inf_{x \in C} f(x)$ iff there exists $s \in \partial f(x)$ such that

$$\langle s, y - x\rangle \geq 0 \text{ for all } y \in C$$

  and $x \in C$
- proof: $\partial(f + \iota_c)(x) = \partial f(x) + N_C(x)$
- optimality condition: $0 \in \partial f(x) + N_C(x)$ or for any $\bar{s} \in \partial f(x)$:

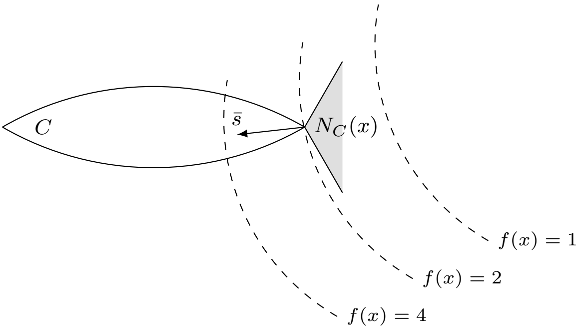
$$-\bar{s} \in N_C(x) = \{s \mid \langle s, y - x\rangle \leq 0 \text{ for all } y \in C\}$$

  for $x \in C$ (otherwise $N_C(x)$ is empty)

Graphical interpretation

- first-order optimality condition: there exists $\bar{s} \in \partial f(x)$ such that

$$-\bar{s} \in N_C(x) = \{s \mid \langle s, y - x \rangle \leq 0 \text{ for all } y \in C\}$$



Precomposition

- next, we cover the pure composition case
- assume that f is proper closed and convex and that

$$\text{Im } L \cap \text{ri dom } g \neq \emptyset \qquad \Longleftrightarrow \qquad \text{ri dom } (g \circ L) \neq \emptyset$$

then

$$\partial(g \circ L)(x) = L^* \circ \partial g(Lx)$$

for every $Lx \in$ dom g

Precomposition proof $\Leftarrow$

- assume that $p \in \partial g(Lx)$ with $Lx \in \text{dom } g$
- subdifferential definition with $z = Ly$:

$$\begin{aligned} & g(z) \geq g(Lx) + \langle p, z - Lx \rangle \\ \Longrightarrow \quad & g(Ly) \geq g(Lx) + \langle p, Ly - Lx \rangle \\ \Longrightarrow \quad & (g \circ L)(y) \geq (g \circ L)(x) + \langle L^* p, y - x \rangle \end{aligned}$$

that is, $L^* p \in \partial(g \circ L)(x)$ or $L^* \partial g(Lx) \subseteq \partial(g \circ L)(x)$

Precomposition proof $\Rightarrow$

- assume that $s \in \partial(g \circ L)(x)$, i.e., that

$$(g \circ L)^*(s) + (g \circ L)(x) - \langle s, x \rangle = 0 \tag{2}$$

- assumptions imply "Key result 1" can be used:

$$(g \circ L)^*(s) = (L^* g^*)(s) = \min_p \{g^*(p) \mid L^* p = s\} = g^*(\bar{p})$$

  where $L^* \bar{p} = s$

- therefore (2) becomes

$$0 = g^*(\bar{p}) + g(Lx) - \langle L^* \bar{p}, x \rangle = g^*(\bar{p}) + g(Lx) - \langle \bar{p}, Lx \rangle$$

- which implies $\bar{p} \in \partial g(Lx)$, $s \in L^* \partial g(Lx)$, and $\partial(g \circ L)(x) \subseteq L^* \partial g(Lx)$

Sum and composition

- adding the two previous results on f and $h = g \circ L$, we get:

$$\partial(f+h)(x) = \partial f(x) + \partial(g \circ L)(x) = \partial f(x) + L^* \partial g(Lx) \quad (3)$$

  provided that assumptions hold

- we assume:

  ri dom$g \cap$ ri L(dom$f$) $\neq \emptyset \quad \Longleftrightarrow \quad$ ri dom$(g \circ L) \cap$ ri dom $f \neq \emptyset$

- since ri dom $(g \circ L) =$ ri dom $h \neq \emptyset$, composition and sum OK!
- (will use (3) to derive optimality conditions)

Image function

- suppose that f is proper closed and convex
- suppose that $x \in \text{dom}(Lg) = L(\text{dom} g)$ and that there exists $\bar{y}$ with $L\bar{y} = x$ and $g(\bar{y}) = (Lg)(x)$, which holds e.g., if

$$\text{Im} L^* \cap \text{ri dom } g^* \neq \emptyset$$

- then

$$\partial(Lg)(x) = \{s \mid L^* s \in \partial g(\bar{y})\}$$

Proof

- recall $\exists \bar{y}$ with $L\bar{y} = x$ and $g(\bar{y}) = (Lg)(x)$
- that $s \in \partial(Lg)(x)$ is, by Fenchel-Young, equivalent to that

$$(Lg)^*(s) + (Lg)(x) - \langle s, x \rangle = 0$$

or

$$(Lg)^*(s) + g(\bar{y}) - \langle s, L\bar{y} \rangle = 0$$

since $(Lg)^* = g^* \circ L^*$, i.e., $(Lg)^*(s) = g^*(L^*s)$ we have

$$g^*(L^*s) + g(\bar{y}) - \langle L^*s, \bar{y} \rangle = 0$$

or equivalently $L^*s \in \partial g(\bar{y})$, i.e. $\partial(Lg) = \{s \mid L^*s \in \partial g(\bar{y})\}$