Operations on multivalued functions

Let $A : \mathcal{X} \to 2^{\mathcal{X}}$ be a set-valued mapping, where $\mathcal{X} = \mathbb{R}^n$. We define the graph of A by $\mathrm{gr}A = \{(x, y) \,:\, y \in A(x)\}$. We define $A^{-1} : \mathcal{X} \to 2^{\mathcal{X}}$ as $A^{-1}(x) = \{y \,:\, x \in A(y)\}$ that is

$$\mathrm{gr}(A^{-1}) = \{(x, y) \,:\, (y, x) \in \mathrm{gr}A\} .$$

Its domain $\mathrm{dom}(A)$ is the set of $x \in \mathcal{X}$ such that $A(x) \neq \emptyset$.

Exercise 1. Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a proper closed convex function.

1. Write Fenchel-Young inequality and the conditions for equality.
2. Show the equivalence
$$\phi \in \partial f(x) \;\Leftrightarrow\; x \in \partial f^*(\phi) .$$
3. Deduce from it the equality $(\partial f)^{-1} = \partial f^*$.

Correction 1.

1. $\forall x, \phi, f(x) + f^*(\phi) \geq \langle x, \phi\rangle$ *and* $\phi \in \partial f(x) \Leftrightarrow f(x) + f^*(\phi) = \langle x, \phi\rangle$
2. *As f is proper closed and convex, we know by Fenchel-Moreau theorem that $f = f^{**}$. Now using twice the equality case in Fenchel-Young inequality*
$\phi \in \partial f(x) \Leftrightarrow f(x) + f^*(\phi) = \langle x, \phi\rangle \Leftrightarrow f^{**}(x) + f^*(\phi) = \langle x, \phi\rangle \Leftrightarrow x \in \partial f^*(\phi)$
3. *By definition of the inverse of a multivalued function,*
$x \in (\partial f)^{-1}(\phi) \Leftrightarrow \phi \in \partial f(x) \Leftrightarrow x \in \partial f^*(\phi)$
Hence, $(\partial f)^{-1} = \partial f^*$.

Exercise 2. Let $\gamma > 0$ and $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$.

1. Draw the graph of ∂f and the graph of $\gamma\,\partial f$.
2. Draw the graph of $I + \gamma\partial f$ where $I : x \mapsto x$ is the identity function.
3. Draw the graph of $(I + \gamma\partial f)^{-1}$. Whose function is it the graph ?

Exercise 3. Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a proper l.s.c. convex function. Let $x \in \mathcal{X}$ and $\gamma > 0$.

1. Show that $p \in (I + \gamma\partial f)^{-1}(x)$ if and only if
$$0 \in \partial f(p) + \frac{p - x}{\gamma} .$$

2. Show that $(I+\gamma\partial f)^{-1}(x)$ coincides with the set of minimizers of the function
$$y \mapsto f(y) + \frac{\|y-x\|^2}{2\gamma}.$$

3. Deduce that $(I+\gamma\partial f)^{-1}$ is single-valued (hence it is a fonction) and is given by
$$(I+\gamma\partial f)^{-1}(x) = \mathrm{prox}_{\gamma f}(x)\,.$$

Correction 3.

1. $p \in (I+\gamma\partial f)^{-1}(x) \Leftrightarrow x \in (I+\gamma\partial f)(p) \Leftrightarrow x \in p + \gamma\partial f(p) \Leftrightarrow 0 \in \partial f(p) + \frac{p-x}{\gamma}$
where the last equivalence comes from the fact that b belongs to the set A if and only if 0 belongs to the set $A - b = \{x \,:\, \exists a \in A, x = a - b\}$.

2. Let us define the function $g : y \mapsto f(y) + \frac{\|y-x\|^2}{2\gamma}$. As the domain of the squared norm is equal to $\mathcal{X}$, the subdifferential of g is the sum of the subdifferentials:
$\partial g(y) = \partial f(y) + \frac{y-x}{\gamma}$
Using Fermat's rule and the first question,
$p \in \arg\min g \Leftrightarrow 0 \in \partial g(p) \Leftrightarrow \partial f(p) + \frac{p-x}{\gamma} \Leftrightarrow p \in (I+\partial f)^{-1}(x)$

3. As g is strongly convex, it has at most one minimizer. It is equal to the proximity operator of γf at x because for $\gamma > 0$,
$\arg\min_y f(y) + \frac{\|y-x\|^2}{2\gamma} = \arg\min_y \gamma f(y) + \frac{1}{2}\|y-x\|^2$.

Exercise 4. The goal of this exercise is to show Moreau's identity: for every $x \in \mathbb{R}^n$,
$$\mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) = x\,.$$

1. Let $x \in \mathbb{R}^n$ and $p = \mathrm{prox}_f(x)$. Show that $x - p \in \partial f(p)$.

2. Using the result of Exercise 1, show that $p \in \partial f^*(x-p)$.

3. Prove Moreau's identity.

4. Show that Moreau's formula generalizes the famous identity $\Pi_E + \Pi_{E^\perp} = I$, where Π_E and $\Pi_{E^\perp}$ are the orthogonal projectors onto some linear subspace $E \subset \mathbb{R}^n$ and its supplementary space $E^\perp$ respectively.
Hint: choose f as the indicator function of a properly chosen set.

5. *Homework.* For $\gamma > 0$, generalize the identity to:
$$\mathrm{prox}_{\gamma f}(x) + \gamma\,\mathrm{prox}_{\gamma^{-1} f^*}(\frac{x}{\gamma}) = x\,.$$

Correction 4.

1. $p = \text{prox}_f(x) \Leftrightarrow p = \arg\min_y f(y) + \frac{1}{2}\|y-x\|^2 \Leftrightarrow 0 \in \partial f(p) + (p-x) \Leftrightarrow x-p \in \partial f(p)$

2. *$\partial f^* = (\partial f)^{-1}$ and $x-p \in \partial f(p)$ so $p \in \partial f^*(x-p)$.*

3. *Let us denote $q = x-p$. We have $x-q \in \partial f^*(q)$ so by Question 1, we deduce that $q = \text{prox}_{f^*}(x)$. We obtain $x = p+q = \text{prox}_f(x) + \text{prox}_{f^*}(x)$.*

4. *Let us choose $f = \iota_E$.*

   $\text{prox}_f(x) = \arg\min_{y\in\mathbb{R}^n} \iota_E(y) + \frac{1}{2}\|y-x\|^2 = \arg\min_{y\in E} \frac{1}{2}\|y-x\|^2 = \Pi_E(x)$.

   We now need to show that $f^* = \iota_{E^\perp}$.

   $f^*(x) = \sup_{y\in\mathbb{R}^n} \langle y, x\rangle - \iota_E(y) = \sup_{y\in E}\langle y, x\rangle$

   If $x \in E^\perp$, then for all $y \in E$, $\langle y, x\rangle = 0$ and so $f^*(x) = 0$.

   If $x \notin E^\perp$, the linear form $\varphi : E \to \mathbb{R}$, $\varphi(y) = \langle y, x\rangle$ is nonzero and so its supremum is $+\infty$. These two cases show that $f^* = \iota_{E^\perp}$.

   Using similar arguments as for f, we get $\text{prox}_{f^*} = \iota_{E^\perp}$.

   Applying Moreau's identity to f, we recover the equality

   $$\Pi_E + \Pi_{E^\perp} = I$$

5. $p = \text{prox}_{\gamma f}(x) \Leftrightarrow p = \arg\min_y f(y) + \frac{1}{2\gamma}\|y-x\|^2 \Leftrightarrow 0 \in \partial f(p) + \frac{p-x}{\gamma} \Leftrightarrow \frac{x-p}{\gamma} \in \partial f(p)$

   Hence, $p \in \partial f^*(\frac{x-p}{\gamma})$

   Let us denote $q = \frac{x-p}{\gamma}$. Since $p = x - \gamma q$, we have $x - \gamma q \in \partial f^*(q)$. This can be rewritten as $\frac{\gamma^{-1}x - q}{\gamma^{-1}} \in \partial f^*(q)$ and so $q = \text{prox}_{\gamma^{-1}f^*}(\gamma^{-1}x)$.

   We obtain $x = p + \gamma q = \text{prox}_{\gamma f}(x) + \gamma\,\text{prox}_{\gamma^{-1}f^*}(\gamma^{-1}x)$.

Fenchel-Legendre transforms and subdifferentials

Exercise 5. Calculate f^*, where $f : \mathbb{R}^n \to \mathbb{R}$ is provided below.

1. $f(x) = \|x\|_1$
2. $f(x) = \frac{1}{2}x^\top Qx$ where Q is a symmetric positive definite matrix.
3. $f(x) = \frac{1}{2}x^\top Qx$ where Q is a symmetric positive semi-definite matrix.
Hint: Distinguish between the cases $\varphi \in Im(Q)$ and $\varphi \notin Im(Q)$.

Correction 5.

1. $f^*(q) = \sup_{x\in\mathbb{R}^n} \langle x, q\rangle - \|x\|_1 = \sup_{x\in\mathbb{R}^n} \sum_{i=1}^n x_i q_i - |x_i| = \sum_{i=1}^n \sup_{x_i\in\mathbb{R}^n} x_i q_i - |x_i|$
Now, if $q_i \in [-1, 1]$, then $q_i \in \partial\|\cdot\|_1$ and so $0 \in \arg\max_{x_i\in\mathbb{R}^n} x_i q_i - |x_i|$.
If $q_i < -1$ or $q_i > 1$, we can let $x_i q_i - |x_i|$ go to $+\infty$ by letting x_i go to $-\infty$ (respectively $+\infty$).
Hence, $f^*(q) = \sum_{i=1}^n \iota_{[-1,1]}(q_i) = \iota_{B_\infty}(q)$ where $B_\infty = \{x : \|x\|_\infty \le 1\}$.
2. $f^*(s) = \sup_{x\in\mathbb{R}^n} s^\top x - \frac{1}{2}x^\top Qx = \sup_{x\in\mathbb{R}^n} \phi_s(x)$
ϕ_s is differentiable and $\nabla\phi_s(x) = s - Qx$. As Q is invertible, the maximum of ϕ_s is attained at $Q^{-1}s$ and we have
$f^*(s) = s^\top Q^{-1}s - \frac{1}{2}s^\top Q^{-\top}QQ^{-1}s = \frac{1}{2}s^\top Q^{-1}s.$
3. *We still have $f^*(s) = \sup_{x\in\mathbb{R}^n} s^\top x - \frac{1}{2}x^\top Qx = \sup_{x\in\mathbb{R}^n} \phi_s(x)$ where ϕ_s is differentiable and $\nabla\phi_s(x) = s - Qx$. However, now, Q may not be invertible.*
If $s \in Im(Q)$, then there exists $x_s \in \mathbb{R}^n$ such that $s = Qx_s$. One may choose for instance x_s given by $x_s = Q^+s$ where Q^+ is the Moore-Penrose generalized inverse of Q. Then, $f^*(s) = s^\top Q^+s - \frac{1}{2}s^\top Q^+QQ^+s = \frac{1}{2}s^\top Q^+s$ because $Q^+QQ^+ = Q^+$.
If $s \notin Im(Q)$, then $\nabla\phi_s$ never vanishes and so ϕ_s does not have any maximizer. A quadratic function is either unbounded or attains its maximum so ϕ_s is unbounded and $f^*(s) = +\infty$.
To summarize, $f^*(s) = \frac{1}{2}s^\top Q^+s + \iota_{Im(Q)}(s)$.

Exercise 6. Let C be a convex subset of $\mathbb{R}^n$. The *normal cone* of C at x is the set

$$N_C(x) = \{\varphi \in \mathbb{R}^n : \forall y \in C,\ \langle \varphi, y - x\rangle \le 0\},$$

if $x \in C$, and $N_C(x) = \emptyset$ otherwise.

1. Show that $N_C = \partial\iota_C$.
2. Let $x \in \mathrm{int}(C)$. Prove that $N_C(x) = \{0\}$.
3. Let $C = \{(u, v) \in (-\infty, 0]^2 : u + v \ge 0\}$. Draw C and $N_C((0, 0))$.

4. Let $C = \{x \in \mathbb{R}^n \ : \ \|x\|_2 \leq 1\}$. Calculate $N_C(x)$ for all x.

Correction 6.

1. $\forall x \in \mathbb{R}^n$, $\partial \iota_C(x) = \{\varphi \in \mathbb{R}^n : \forall y \in \mathbb{R}^n, \iota_C(y) \geq \iota_C(x) + \langle \varphi, y - x\rangle\}$

If $x \notin C$, then $\partial \iota_C(x) = \emptyset = N_C(x)$.

If $x \in C$, then $\iota_C(x) = 0$ and for $y \notin C$, the inequality $\iota_C(y) = +\infty \geq \iota_C(x) + \langle \varphi, y - x\rangle = \langle \varphi, y - x\rangle$ is trivial and does not bear any information. Hence,

$$\begin{aligned}\partial \iota_C(x) &= \{\varphi \in \mathbb{R}^n : \forall y \in C, \iota_C(y) \geq \langle \varphi, y - x\rangle\} = \{\varphi \in \mathbb{R}^n : \forall y \in C, 0 \geq \langle \varphi, y - x\rangle\} \\ &= N_C(x).\end{aligned}$$

Combining both cases, we get $\partial \iota_C = N_C$.

2. *Suppose that $x \in \operatorname{int} C$. Then for all $h \in \mathbb{R}^n$, there exists $t_0 > 0$ such that for all $t \in [-t_0, t_0]$, $x + th \in C$.*

First, $0 \in N_C(x)$ because $\langle 0, y - x\rangle = 0 \leq 0$ for all $y \in C$. Then, let us consider $\varphi \in N_C(x)$. For $h \in \mathbb{R}^n$ and $t \in [-t_0, t_0]$, $x + th \in C$. Thus,

$\langle \varphi, x + th - x\rangle = t\langle \varphi, h\rangle \leq 0.$

Taking $t = t_0$ and $t = -t_0$ in this inequality, we get $\langle \varphi, h\rangle = 0$ for all $h \in \mathbb{R}^n$. This implies that $\varphi = 0$.

This shows that $N_C(x) = \{0\}$ if $x \in \operatorname{int} C$.

3. *The only element in C is $(0, 0)$, so $C = \{(0, 0)\}$.*

$N_C((0, 0)) = \{\varphi \in \mathbb{R}^2 : \langle \varphi, 0 - 0\rangle \leq 0\} = \mathbb{R}^2.$

4. $C = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$.

If $\|x\| < 1$, then $x \in \operatorname{int} C$ and so $N_C(x) = \{0\}$.

If $\|x\| > 1$, then $x \notin C$ and so $N_C(x) = \emptyset$.

Let us suppose now that $\|x\| = 1$. And let $\varphi \in N_C(x)$. We know that $\forall y \in C$, $\langle \varphi, y - x\rangle \leq 0$.

Taking $y = 0$ in this inequality, we get $\langle \varphi, x\rangle \geq 0$.

We now consider $h \in x^\perp$ (that is a h such that $\langle x, h\rangle = 0$) and a sequence (t_n) of positive real numbers that converges to 0. We consider $y_n = \frac{x+t_nh}{\|x+t_nh\|_2}$. $y_n \in C$ so $\langle \varphi, y_n - x\rangle \leq 0$.

Dividing by $t_n > 0$, we get

$$0 \geq \Big\langle \varphi, \frac{1}{t_n}\big(\frac{1}{\|x + t_nh\|_2} - 1\big)x + \frac{1}{\|x + t_nh\|_2}h\Big\rangle$$

For the second term we use the continuity of norm to deduce $\|x + t_nh\|_2 \to \|x\| = 1$.

For the first term, we recognise a directional derivative. Indeed, if we denote $g(x) = \frac{1}{\|x\|_2} = (\|x\|_2^2)^{-1/2}$, *g is differentiable for all* $x \neq 0$ *and we have*

$$\lim_{t_n \to 0} \frac{1}{t_n}\Big(\frac{1}{\|x + t_n h\|_2} - \frac{1}{\|x\|_2}\Big) = \langle \nabla g(x), h \rangle$$

$\nabla g(x) = -\frac{1}{2}\frac{1}{\|x\|_2^{3/2}} 2x$ *so by our choice of* $h \in x^\perp$, $\langle \nabla g(x), h \rangle = 0$.

Gathering the two limits, we get $0 \geq \langle \varphi, h \rangle$. *This is true also for* $-h$, *so* $\langle \varphi, h \rangle = 0$.

Hence, $\varphi \in (x^\perp)^\perp = \mathrm{span}(x)$. *As we also have* $\langle \varphi, x \rangle \geq 0$, *we can see that there exists* $\lambda \geq 0$ *such that* $\phi = \lambda x$.

For the other inclusion, let us show that if $\lambda \geq 0$, *then* $\lambda x \in N_C(x)$.

$\langle \lambda x, y - x \rangle = \lambda(\langle x, y \rangle - \|x\|_2^2) \leq \lambda(\|x\|_2 \|y\|_2 - 1) \leq 0$, *which shows that* $\lambda x \in N_C(x)$.

Exercise 7. Calculate the subdifferential of $f(x) = \sum_{i=1}^n g_i(x_i)$.

Correction 7.

$$\begin{aligned} \phi \in \partial f(x) &\Leftrightarrow \forall y \in \mathbb{R}^n, f(y) \geq f(x) + \langle \phi, y - x \rangle \\ &\Leftrightarrow \forall y \in \mathbb{R}^n, \sum_{i=1}^n g_i(y_i) \geq \sum_{i=1}^n g_i(x_i) + \sum_{i=1}^n \phi_i(y_i - x_i) \end{aligned}$$

For $i \in \{1, \dots n\}$, *take* y *such that* $y_j = x_j$ *if* $j \neq i$ *and* y_i *is free. We obtain that*
$\forall y_i \in \mathbb{R}, g_i(y_i) \geq g_i(x_i) + \phi_i(y_i - x_i)$
Hence, $\phi_i \in \partial g_i(x_i)$ *for all* i.
Conversely, if $\phi_i \in \partial g_i(x_i)$ *for all* i, *we obviously have*
$\forall y \in \mathbb{R}^n, \sum_{i=1}^n g_i(y_i) \geq \sum_{i=1}^n g_i(x_i) + \sum_{i=1}^n \phi_i(y_i - x_i)$.
To conclude,

$$\partial f = \partial g_1 \times \partial g_2 \times \ldots \times \partial g_n.$$

Exercise 8. Let $f : \mathbb{R}^n \to \mathbb{R}$.

1. Show that if f is differentiable at x, then $\partial f(x) \subseteq \{\nabla f(x)\}$.
2. Show that if f is convex and differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

Correction 8.

1. *Let* f *be a differentiable function, let us consider* $x \in \mathbb{R}^n$ *and, if it exists,* $\varphi \in \partial f(x)$.

   $\forall y \in \mathbb{R}^n, f(y) \geq f(x) + \langle \varphi, y - x \rangle$ *so this is also true if* (t_n) *is a positive sequence converging to 0,* $h \in \mathbb{R}^n$ *and we take* $y_n = x + t_n h$. *Using the differentiability of* f, *we get that there exists a function* $\epsilon(t_n)$ *such that* $\epsilon(t_n) \to 0$ *and*

   $f(x) + t_n \langle \nabla f(x), h \rangle + t_n \epsilon(t_n) \geq f(x) + t_n \langle \varphi, h \rangle$.

   Dividing by t_n *and removing* $f(x)$ *on both sides, we get* $\langle \nabla f(x), h \rangle + \epsilon(t_n) \geq \langle \varphi, h \rangle$.

Passing to the limit yields $\langle \nabla f(x), h \rangle \geq \langle \varphi, h \rangle$ *for all* h*. Doing the same for* $-h$*, we get* $\langle \nabla f(x), h \rangle = \langle \varphi, h \rangle$ *and so* $\nabla f(x) = \varphi$*.*

Hence, either $\partial f(x) = \emptyset$ *or* $\partial f(x) = \{\nabla f(x)\}$*.*

2. *If* f *is differentiable and convex, then*

$$\forall y \in \mathbb{R}^n, \forall x \in \mathbb{R}^n, f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

This shows that $\nabla f(x) \in \partial f(x)$*. Combined with the first question, we have the equality* $\{\nabla f(x)\} = \partial f(x)$*.*