Chapter 1

Some elements of convex analysis

1.1 Fenchel-Legendre Transform

Let us fix a Banach space X together with its dual X', and denote by $\langle \xi, x \rangle$ the duality between an element $\xi \in X'$ and $x \in X$. More generally, we could fix a pair of normed vector spaces on which we fix a bilinear form which plays the role of the duality between them.

Definition 1.1. We say that a function valued in $\mathbb{R} \cup \{+\infty\}$ is *proper* if it is not identically equal to $+\infty$. The set $\{f < +\infty\}$ is called the *domain* of f.

Definition 1.2. Given a vector space X and its dual X', and a proper function $f : X \to \mathbb{R} \cup \{+\infty\}$ we define its Fenchel-Legendre transform $f^* : X' \to \mathbb{R} \cup \{+\infty\}$ via

$$f^*(\xi) := \sup_x \langle \xi, x \rangle - f(x).$$

Remark 1.3. We observe that we trivially have $f^*(0) = -\inf_X f$.

We note that f^*, as a sup of affine continuous (in the sequel we will just say affine and mean affine and continuous, i.e. of the form $\ell(x) = \langle \xi, x \rangle + c$ for $\xi \in X'$ and $c \in \mathbb{R}$) functions, is both convex and l.s.c., as these two notions are stable by sup.

By abuse of notations, when considering functions defined on X' we will see their Fenchel-Legendre transform as a function defined on X (and not on X'': this is possible since $X \subset X''$ and we can restrict it to X, and by the way in most cases we will use only reflexive spaces).

We prove the following results.

Proposition 1.4.

1. *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and l.s.c. then there exists a continuous affine function ℓ such that $f \geq \ell$.*
2. *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and l.s.c. then it is a sup of continuous affine functions.*

3. *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and l.s.c. then there exists $g : X' \to \mathbb{R} \cup \{+\infty\}$ such that $f = g^*$.*

4. *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and l.s.c. then $f^{**} = f$.*

Proof. We consider the epigraph $\mathrm{Epi}(f) := \{(x,t) \in X \times \mathbb{R} : t \geq f(x)\}$ of f which is a convex and closed set in $X \times \mathbb{R}$. We take a point x_0 such that $f(x_0) < +\infty$ and consider the singleton $\{(x_0, f(x_0) - 1)\}$ which is a convex and compact set in $X \times \mathbb{R}$. The Hahn-Banach separation theorem provides the existence of a pair $(\xi, a) \in X' \times \mathbb{R}$ and a cosntant c such that $\langle \xi, x_0 \rangle + a(f(x_0) - 1) < c$ and $\langle \xi, x \rangle + at > c$ for every $(x,t) \in \mathrm{Epi}(f)$. Note that this last condition implies $a \geq 0$ since we can take $t \to \infty$. Moreover, we should also have $a > 0$ otherwise taking any point $(x,t) \in \mathrm{Epi}(f)$ with $x = x_0$ we have a contradiction. If we then take $t = f(x)$ for all x such that $f(x) < +\infty$ we obtain $af(x) \geq -\langle \xi, x \rangle + \langle \xi, x_0 \rangle + a(f(x_0) - 1)$ and, dividing by $a > 0$, we obtain the first claim.

We now take an arbitrary $x_0 \in X$ and $t_0 < f(x_0)$ and separate again the singleton $\{(x_0, t_0)\}$ from $\mathrm{Epi}(f)$, thus getting a pair $(\xi, a) \in X' \times \mathbb{R}$ and a constant c such that $\langle \xi, x_0 \rangle + at_0 < c$ and $\langle \xi, x \rangle + at > c$ for every $(x,t) \in \mathrm{Epi}(f)$. Again, we have $a \geq 0$. If $f(x_0) < +\infty$ we obtain as before $a > 0$ and the inequality $f(x) > -\frac{\xi}{a} \cdot (x - x_0) + t_0$. We then have an affine function ℓ with $f \geq \ell$ and $\ell(x_0) = x_0$. This shows that the sup of all affine functions smaller than f is, at the point x_0, at least t_0. Hence this sup equals f on $\{f < +\infty\}$. The same argument works for $f(x_0) = +\infty$ if for t_0 arbitrary large the corresponding coefficient a is strictly positive. If not, we have $\langle \xi, x_0 \rangle < 0$ and $\langle \xi, x \rangle \geq 0$ for every x such that $(x,t) \in \mathrm{Epi}(f)$ for at least one $t \in \mathbb{R}$, i.e. for $x \in \{f < +\infty\}$. Consider now $\ell_n = \ell - n\xi$ where ℓ is the affine function smaller than f previously found. We have $f \geq \ell \geq \ell - n\xi$ since ξ is non-negative on $\{f < +\infty\}$ and moreover $\lim_n \ell_n(x_0) = +\infty$. This shows that in such a point x_0 the sup of the affine functions smaller than f equals $+\infty$, and hence $f(x_0)$.

Once that we know that f is a sup of affine functions we can write

$$f(x) = \sup_\alpha \langle \xi_\alpha, x \rangle + c_\alpha$$

for a family of indexes α. We then set $c(\xi) := \sup\{c_\alpha : \xi_\alpha = \xi\}$. The set in the sup can be empty, which means $c(\xi) = -\infty$. Anyway, the sup is always finite: fix a point x_0 with $f(x_0) < +\infty$ and use since $c_\alpha \leq f(x_0) - \langle \xi, x_0 \rangle$. We then define $g = -c$ and we see $f = g^*$.

finally, before proving $f = f^{**}$ we prove that for any function f we have $f \geq f^{**}$ even if f is not convex or lsc. Indeed, we have $f^*(\xi) + f(x) \geq \langle \xi, x \rangle$ which allows to write $f(x) \geq \langle \xi, x \rangle - f^*(\xi)$, an inequality true for every ξ. Taking the sup over ξ we obtain $f \geq f^{**}$. We want now to prove that this inequality is an equality if f is convex and lsc. We write $f = g^*$ and transform this into $f^* = g^{**}$. We then have $f^* \leq g$ and, transforming this inequality (which changes its sign), $f^{**} \geq g^* = f$, which proves $f^{**} = f$. □

Corollary 1.5. *Given an arbitrary proper function $f : X \to \mathbb{R} \cup \{+\infty\}$ we have $f^{**} = \sup\{g : g \leq f,\, g$ is convex and lsc$\}$.*

Proof. Let us call h the function obtained as a sup on the right hand side. Since f^{**} is convex and lsc and smaller than f, we have $f^{**} \leq h$. Note that h, as a sup of convex and lsc functions, is also convex and lsc, and it is of course smaller than f. We write $f \geq h$ and double transform this inequality, which preserves the sign. We then have $f^{**} \geq h^{**} = h$, and the claim is proven. □

We finally discuss the relations between the behavior at ∞ of a fonction f and its Legendre transform. We give two definitions.

Definition 1.6. A function $f : X \to \mathbb{R} \cup \{+\infty\}$ defined on a normed vector space X is said to be *coercive* if $\lim_{||x||\to\infty} f(x) = +\infty$; it is said to be *superlinear* if $\lim_{||x||\to\infty} \frac{f(x)}{||x||} = +\infty$.

We note that the definition of coercive does not include any speed of convergence to ∞, but that for onvex functions this should be at least linear:

Proposition 1.7. *A proper, convex, and l.s.c. function $f : X \to \mathbb{R} \cup \{+\infty\}$ is coercive if and only there exist two constants c_0, c_1 such that $f(x) \geq c_0||x|| - c_1$.*

Proof. We just need to prove that c_0, c_1 exist if f is coercive, the converse being trivial. Take a point x_0 such that $f(x_0) < +\infty$. Using $\lim_{||x||\to\infty} f(x) = +\infty$ we know that there exists a radius R such that $f(x) \geq f(x_0) + 1$ as soon as $||x - x_0|| \geq R$. By convexity, we have, for each x with $||x - x_0|| > R$, the inequality $f(x) \geq f(x_0) + ||x - x_0||/R$ (it is enough to use the definition of convexity on the three points x_0, x and $x_t = (1-t)x_0 + tx \in \partial B(x_0, R)$). Since f is bounded from below by an affine function, it is bounded from below by a constant on $B(x_0, R)$, so that we can write $f(x) \geq c_2 + ||x - x_0||/R$ for some $c_2 \in \mathbb{R}$ and all $x \in X$. We then use the triangle inequality and obtain the claim with $c_0 = 1/R$ and $c_1 = c_2 - ||x_0||/R$. □

Proposition 1.8. *A proper and convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ is coercive if and only if f^* is bounded in a neighboorhood of 0; it is superlinear if and only if f^* is bounded on each bounded ball of X'.*

Proof. We know that f is coercive if and only if there exist two constants c_0, c_1 such that $f \geq g_{c_0,c_1}$ where $g_{c_0,c_1}(x) := c_0||x|| - c_1$. Since both f and g_{c_0,c_1} are convex l.s.c., this inequality is equivalent to the opposite inequality for their transforms, i.e. $f^* \leq g^*_{c_0,c_1}$. We can compute the transform and obtain

$$g^*_{c_0,c_1}(\xi) = \begin{cases} c_1 & \text{if } ||\xi|| \leq c_0^{-1}, \\ +\infty & \text{if not.} \end{cases}$$

This shows that f is coercive if and only if there exist two constants R, C (with $R = c_0^{-1}$, $C = c_1$) such that $f^* \leq C$ on the ball of radius R of X', which is the claim.

We follow a similar procedure for the case of superlinear functions. We first note that a convex l.s.c. function f is superlinear if and only if for every c_0 there exists c_1 such that $f \geq g_{c_0,c_1}$. Indeed, it is clear that, should this condition be satisfied, we would have $\liminf_{||x||\to\infty} f(x)/||x|| \geq c_0$, and hence $\liminf_{||x||\to\infty} f(x)/||x|| = +\infty$ because c_0 is arbitrary, so that f would be superlinear. On the other, if f is superlinear, for every

c_0 we have $f(x) \geq c_0||x||$ for large $||x||$, say outside of $B(0,R)$. If we then choose $-c_1 := \min\{\inf_{B(0,R)} f - c_0 R, 0\}$ (a value which is finite since f is bounded from below by an affine function), the inequality $f(x) \geq c_0||x|| - c_1$ is true everywhere.

We then deduce that f is superlinear if and only if for every $R = c_0^{-1}$ there is a constant c_1 such that $f^* \leq c_1$ on the ball of radius R of X', which is, again, the claim. □

1.2 Subdifferentials

The above-the-tangent property of convex functions ispired the definition of an extension of the notion of differential, called sub-differential, as a set-valued map:

Definition 1.9. Given a function $f : X \to \mathbb{R} \cup \{+\infty\}$ we define its subdifferential at x as the set

$$\partial f(x) = \{\xi \in X' \, : \, f(y) \geq f(x) + \langle \xi, y - x \rangle \; \forall y \in X\}.$$

We observe that $\partial f(x)$ is always a closed and convex set, whatever is f. Moreover, if f is l.s.c. we easily see that the graph of the subdifferential multi-valued map is closed:

Proposition 1.10. *Suppose that f is l.s.c. and take a sequence $x_n \to x$. Suppose $\xi_n \rightharpoonup \xi$ and $\xi_n \in \partial f(x_n)$. Then $\xi \in \partial f(x)$.*

Proof. for every y we have $f(y) \geq f(x_n) + \langle \xi, y - x_n \rangle$. We can then use the strong convergence of x_n and the weak convergence of ξ_n, together with the lower semicontinuity of f, to pass to the limit and deduce $f(y) \geq f(x) + \langle \xi, y - x \rangle$, i.e. $\xi \in \partial f(x)$. □

Note that in the above proposition we could have exchanged strong convergence for x_n and weak for ξ_n for weak convergence for x_n (but f needed in this case to be weakly l.s.c.) and strong for ξ_n.

When dealing with arbitrary functions f, the subdifferential is in most cases empty, as there is no reason that the inequality defining $\xi \in \partial f(x)$ is satisfied for y very far from x. The situation is completely different when dealing with convex functions, which is the standard case here subdifferentials are defined. In this case we can prove that $\partial f(x)$ is never empty if x lies in the interior of the set $\{f < +\infty\}$ (note that outside $\{f < +\infty\}$ the subdifferential of a proper function is clearly empty).

We first provide an insight about the finite-dimensional case. In this case we simply write $\xi \cdot x$ for the duality product, which coincides with the Euclidean scalar product on $\mathbb{R}^N$.

We start from the following property.

Proposition 1.11. *Given $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ suppose that f is differentiable at a point x_0. Then $\partial f(x_0) \subset \{\nabla f(x_0)\}$. If moreover f is convex, then $\partial f(x_0) = \{\nabla f(x_0)\}$.*

Proof. from the definition of sub-differential we see that $\xi \in \partial f(x_0)$ means that $y \mapsto f(y) - \xi \cdot y$ is minimal at $y = x_0$. Since we are supposing that f is differentiable at such a point, we obtain $\nabla f(x_0) - \xi = 0$, i.e. $\xi = \{\nabla f(x_0)$. This shows $\partial f(x_0) \subset \{\nabla f(x_0)\}$. The inclusion becomes an equality if f is convex since the above-the-tanent property of convex functions exactly provides $f(y) \geq f(x_0) + \nabla f(x_0) \cdot (y - x_0)$ for every y. □

Proposition 1.12. *Suppose that $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is convex and take a point x_0 in the interior of $\{f < +\infty\}$. Then $\partial f(x_0) \neq \emptyset$.*

Proof. It is well-known that convex functions in finite dimension are locally Lipschitz in the interior of their domain, and Lipschitz functions are differentiable Lebesgue-a.e. because of the Rademacher's theorem. We can then take a sequence of points $x_n \to x_0$ such that f is differentiable at x_n. We then have $\nabla f(x_n) \in \partial f(x_n)$ and the Lipschitz behavior of f around x_0 implies $|\nabla f(x_n)| \leq C$. It is then possible to extract a subsequence such that $\nabla f(x_n) \to v$. Proposition 1.10 implies $v \in \partial f(x_0)$. □

We can easily see, even from the 1D case, that different situations can occurr at the boundary of $\{f < +\infty\}$. If we take for instance the proper function f defined via

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases}$$

we see that we have $\partial f(0) = [-\infty, 0]$ so that the subdifferential can be "fat" on these boundary points. If we take, instead, the proper function f defined via

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } x \geq 0, \\ +\infty & \text{if } x < 0, \end{cases}$$

we see that we have $\partial f(0) = \emptyset$, a fact related to the infinite slope of f at 0, an inifnite slope that can of course only appaer on boundary points.

The proof of the fact that the sub-differential is non-empty[1] in the interior of the domain is more involved in the general (infinite-dimensional) case, and also based on the use of the Hahn-Banach theorem. It will require the function to be convex and l.s.c., this second assumption being useless in the finitdimensional case, since convex functions are locally Lipschitz in the interior of their domain. It also requires to discuss whether the function is indeed locally bounded around some points. We state the following clarifying proposition.

Proposition 1.13. *If $f : X \to \mathbb{R} \cup \{+\infty\}$ is a proper convex and l.s.c. function, the following facts are equivalent*

1. *f is locally Lipschitz continuous on the interior of its domain (i.e. for every point in the interior of the domain there exists a ball centered at such a point where f is Lipschitz continuous);*
2. *there exists a non-empty ball where f is finite-valued and Lipschitz continuous;*
3. *there exists a point where f is continuous and finite-valued;*
4. *there exists a point which has a neighborhood where f is bounded from above by a finite constant.*

[1] Note that, instead, we do not discuss the relations between subdifferential and gradients as we decided not to discuss the differentiability in the inifnite-dimensional case.

Moreover, all the above facts hold if X is a Banach space.

Proof. It is clear that 1. implies 2., which implies 3., which implies 4. Let us note that, if f is bounded from above on a ball $B(x_0,R)$, then necessarily f is Lipschitz continuous on $B(x_0,R/2)$. Indeed, we know (point 1 in Proposition 1.4) that f is also bounded from beow by an affine function, and hence by a constant on $B(0,R)$. Then, if there are two points $x_1,x_2 \in B(x_0,R/2)$ with an incremental ratio equal to L and $f(x_1) > f(x_2)$, then, following the half-line going from x_2 to x_1 we find two points $x_3 \in \partial B(x_0,R/2)$ and $x_4 \in \partial B(x_0,3R/4)$ with $|x_3 - x_4| \geq R/4$ and an incremental ration which is also at least L. This implies $f(x_4) > f(x_3) + LR/4$ but the upper and lower bounds on f on the ball $B(0,R)$ imply that L cannot be too large, so f is Lipschitz continuous on $B(x_0,R/2)$.

Hence, in order to prove that 4. implies 1. we just need to prove that every point in the interior of the domain admits a ball centered at such a point where f is bounded from above. We start from the existence of a ball $B(x_0,R)$ where f is bounded from above and we take another point x_1 in the interior of the domain of f. for small $\varepsilon > 0$, the point $x_2 := x_1 - \varepsilon(x_0 - x_1)$ also belongs to the domain of f and every point of the ball $B(x_1,r)$ with $r = \frac{\varepsilon R}{1+\varepsilon}$ can be written as a convex combination of x_2 and a point in $B(x_0,R)$: indeed we have

$$x_1 + v = \frac{1}{1+\varepsilon}x_2 + \frac{\varepsilon}{1+\varepsilon}(x_0 + \frac{1+\varepsilon}{\varepsilon}v)$$

so that $|v| < r$ implies $x_0 + \frac{1+\varepsilon}{\varepsilon}v \in B(x_0,R)$. Then, f is bounded from above on $B(x_1,r)$ by $\max\{f(x_2), \sup_{B(x_0,R)} f\}$ which shows the local bound around x_1.

finally, we want to prove that f is necessarily locally bounded around every point of the interior of its domain if X is complete. Consdier a closed ball $\overline{B}$ contained in $\{f < +\infty\}$ and write $\overline{B} = \bigcup_n \{f \leq n\} \cap \overline{B}$. Since f is l.s.c., each set $\{f \leq n\} \cap \overline{B}$ is closed. Since their countable union has non-empty interior Baire's theorem (which is valid in complete metric spaces) implies that at least one of these sets also has non-empty interior, so there exists a ball contained in a set $\{f \leq n\}$, hence a point where f is locally bounded from above. Then we satisfy condition 4., and consequently also condition 1., 2. and 3. □

We can now prove the following theorem.

Theorem 1.14. *Suppose that $f : X \to \mathbb{R} \cup \{+\infty\}$ is a proper convex and l.s.c. function and take a point x_0 in the interior of $\{f < +\infty\}$. Also suppose that there exists at least a point x_1 (possibly different from x_0) where f is continuous. Then $\partial f(x_0) \neq \emptyset$.*

Proof. Let us consider the set A given by the interior of $\mathrm{Epi}(f)$ in $X \times \mathbb{R}$ and $B = \{(x_0, f(x_0))\}$. They are two disjoint convex sets, and a is open. Hence, there exists a pair $(\xi, a) \in X' \times \mathbb{R}$ and a constant such that $\langle \xi, x_0 \rangle + af(x_0) \leq c$ and $\langle \xi, x \rangle + at > c$ for every $(x,t) \in A$.

The assumption on the exisntence of a point x_1 where f is continuous implies that f is bounded (say by a constant M) on a ball around x_1 (say $B(x_1,R)$ and hence the set A is non-empty since it includes $B(x_1,R) \times (M,\infty)$. Then, the convex set $\mathrm{Epi}(f)$ has non-empty interior and it is thus the closure of its interior (to see this it is enough to connect every point of a closed convex set to the center of a ball contained in the set, and see that there is a cone composed of small balls contained in the convex set approximating such a point). In particular, since $(x_0, f(x_0))$ belongs to $\mathrm{Epi}(f)$, it is in the closure of A, so necessarily we have $\langle \xi, x_0 \rangle + af(x_0) \geq c$ and hence $\langle \xi, x_0 \rangle + af(x_0) = c$.

As we did iin Proposition 1.4, we must have $a > 0$. Indeed, we use Proposition 1.13 to see that f is also locally bounded around x_0 so that points of the form $(x,t) = (x_0,t)$ for t large enough should belong to A and should satisfy $a(t - f(x_0)) > 0$, which implies $a > 0$.

Then, we can write $\langle \xi, x \rangle + at > c = \langle \xi, x_0 \rangle + af(x_0)$, dividing by a and using $\tilde{\xi} := \xi / a$ as

$$\langle \tilde{\xi}, x \rangle + t > \langle \tilde{\xi}, x_0 \rangle + f(x_0) \text{ for every } (x,t) \in A.$$

The inequality becomes large on the clouse of A but implies, when applied to $(x,t) = (x, f(x)) \in \mathrm{Epi}(f) = \overline{A}$,

$$\langle \tilde{\xi}, x \rangle + f(x) \geq \langle \tilde{\xi}, x_0 \rangle + f(x_0),$$

which exactly means $-\tilde{\xi} \in \partial f(x_0)$. □

Of course, thanks to the last claim in Proposition 1.13, when x is a Banach space we obtain $\partial f(x_0) \neq \emptyset$ for every x_0 in the interior of $\{f < +\infty\}$.

We list some other properties of subdifferentials.

Proposition 1.15. *1. A point x_0 solves $\min\{f(x) : x \in X\}$ if and only if $0 \in \partial f(x_0)$.*

2. The subdifferential satisfies the monotonicity property

$$\xi_i \in \partial f(x_i) \textit{ for } i = 1,2 \ \Rightarrow\ \langle \xi_1 - \xi_2, x_1 - x_2 \rangle \geq 0.$$

3. If f is convex and l.s.c., the subdifferentials of f and f^* are related through

$$\xi \in \partial f(x) \Leftrightarrow x \in \partial f^*(\xi) \Leftrightarrow f(x) + f^*(\xi) = \langle \xi, x \rangle.$$

Proof. 1. is a straightforward consequence of the definition of subdifferential, since $0 \in \partial f(x_0)$ means that for every y we have $f(y) \geq f(x_0)$.

2. is also straightforward if we sum up the inequalities

$$f(x_2) \geq f(x_1) + \langle \xi, x_2 - x_1 \rangle; \quad f(x_1) \geq f(x_2) + \langle \xi, x_1 - x_2 \rangle.$$

for part 3., once we know that for convex and l.s.c. functions we have $f^{**} = f$, it is enough to prove $\xi \in \partial f(x) \Leftrightarrow f(x) + f^*(\xi) = \langle \xi, x \rangle$ since then, by symmetry, we can also obtain $x \in \partial f^*(\xi) \Leftrightarrow f(x) + f^*(\xi) = \langle \xi, x \rangle$. We now look at the definition of subdifferential, and we have

$$\begin{aligned}
\xi \in \partial f(x) \quad & \Leftrightarrow \quad \text{for every } y \in X \text{ we have } f(y) \geq f(x) + \langle \xi, y - x \rangle \\
& \Leftrightarrow \quad \text{for every } y \in X \text{ we have } \langle \xi, x \rangle - f(x) \geq \langle \xi, y \rangle - f(y) \\
& \Leftrightarrow \quad \langle \xi, x \rangle - f(x) \geq \sup_y \langle \xi, y \rangle - f(y) \\
& \Leftrightarrow \quad \langle \xi, x \rangle - f(x) \geq f^*(\xi).
\end{aligned}$$

This shows that $\xi \in \partial f(x)$ is equivalent to $\langle \xi, x \rangle \geq f(x) + f^*(\xi)$, which is in turn equivalent to to $\langle \xi, x \rangle = f(x) + f^*(\xi)$, since the opposite inequality is always true by definition of f^*. □

We also state another property, but we prefer to stick to the finite-dimensional case for simplicity.

Proposition 1.16. *A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is strictly convex if and only if $\partial f(x_0) \cap \partial f(x_1) = \emptyset$ for all $x_0 \neq x_1$.*

Proof. If we have $\partial f(x_0) \cap \partial f(x_1) \neq \emptyset$ for two points $x_0 \neq x_1$, take $\xi \in \partial f(x_0) \cap \partial f(x_1)$ and define $\tilde{f}(x) := f(x) - \langle \xi, x\rangle$. Then both x_0 and x_1 minimize $\tilde{f}$ and this implies that f is not strictly convex. If, instead, we suppose that f is not strictly convex, we need to find two points with a same vector in the subdifferential. Consider two points x_0 and x_1 on which the strict convexity fails, i.e. f is affine on $[x_0, x_1]$. We then have a segment $S = \{(x, f(x)) : x \in [x_0, x_1]\}$ in the graph of f and we can separate it from the interior A of $\mathrm{Epi}(f)$ exactly as we did in Theorem 1.14. Up to restricting the space where f is defined to the minimal affine space containing its domain, we can always suppose that the domain has non-empty interior and hence the epigraph as well. This is not restrictive, since if we want then to define subdifferentials in the original space it is enough to add arbitrary components orthogonal to the affine space containing the domain.

Following the same arguments as in Theorem 1.14, we obtain the existence of a pair $(\xi, a)i \in X' \times \mathbb{R}$ such that

$$\langle \xi, x\rangle + af(x) < \langle \xi, x'\rangle + at \text{ for every } (x', t) \in A \text{ and every } x \in [x_0, x_1].$$

We then prove $a > 0$, divide by a, pass to a large inequality on the closure of A, and deduce $-\xi/a \in \partial f(x)$ for every $x \in [x_0, x_1]$. □

Limiting once more to the finite-dimensional case (also because we do not want to discuss differentiability in other settings) we can deduce from the previous proposition and from part 3. of Proposition 1.15 the following fact.

Proposition 1.17. *Take two proper, convex and l.s.c. conjugate functions f and f^* (with $f = f^{**}$); then f is a real-valued C^1 function on $\mathbb{R}^N$ if and only if f^* is strictly convex and superlinear.*

Proof. We first prove that a convex function is C^1 if and only if $\partial f(x)$ is a singleton for every x. If it is C^1, and then differentiable, we already saw that this implies $\partial f(x) = \{\nabla f(x)\}$. The converse implication can be proven as follows: first we observe that, if we have a map $v : \mathbb{R}^N \to \mathbb{R}^N$ with $\partial f(x) = \{v(x)\}$, then v is necessarily locally bounded and continuous. Indeed, the function f is necessaril finite everywhere (since the subdifferential is non-empty at every point) and hence locally Lipschitz; local boundedess of v comes from $v(x) \cdot e \leq f(x+e) - f(x) \leq C|e|$ (where we use the Local Lipschitz behavior of f): using e oriented as $v(x)$ we obtain a bound on $|v(x)|$. Continuity comes from local boundedness and from Proposition 1.10: when $x_n \to x$ then $v(x_n)$ admits a converging subsequence, but the limit can only belong to $\partial f(x)$, i.e. it must equal $v(x)$. Then we prove $v(x) = \nabla f(x)$, and for this we use the definition of $\partial f(x)$ and $\partial f(y)$ so as to get

$$v(y) \cdot (y - x) \geq f(y) - f(x) \geq v(x) \cdot (y - x).$$

We then write $v(y)\cdot(y-x) = v(x)\cdot(y-x) + o(|y-x|)$ and we obtain the first-order development $f(y)-f(x) = v(x)\cdot(y-x)+o(|y-x|)$ which characterizes $v(x) = \nabla f(x)$.

This shows that f is C^1 as soon as subdifferentials are singletons. On the other hand, point 3. in 1.15 shows that this is equivalent to having, for each $x \in \mathbb{R}^N$, exacty one point ξ with $x \in \partial f^*(\xi)$. The fact that no more than one point ξ has the same vector in the subdifferential is equivalent (Proposition 1.16) to being strictly convex. The fact that each point is taken at least once as a subdifferential is, instead, equivalent to being superlinear (see Lemma 1.18 below) □

Lemma 1.18. *A convex and l.s.c. function $f : \mathbb{R}^N \to \mathbb{R}\cup\{+\infty\}$ is superlinear if and only if ∂f is surjective.*

Proof. Let us suppose that f is convex, l.s.c, and superlinear. Then ofr every ξ the function $\tilde f$ given by $\tilde f(x) := f(x) - \xi\cdot x$ is also l.s.c, and superlinear, and it admitds a minizer x_0. Such a point satisfies $\xi \in \partial f(x_0)$, which proves that ξ is in the image of ∂f, which is then surjective.

Let us suppose now that ∂f is surjective. Let us fix a number $L > 0$ and take the $2N$ vectors $\xi_i^\pm := \pm L e_i$, where the vectors e_i are the canonical basis of $\mathbb{R}^N$. Since each of these vectors belong to the image of the subdifferential, there exist points $x_i^\pm$ such that $\xi_i^\pm \in \partial f(x_i^\pm)$. This implies that f satisfies $2N$ inequalities of the form $f(x) \ge \xi_i^\pm\cdot x - C_i^\pm$ for some constants $C_i^\pm$. We then obtain $f(x) \ge L||x||_\infty - C$, where $||x||_\infty = \max\{|x_i|\} = \max_i x\cdot(\pm e_i)$ is the norm on $\mathbb{R}^N$ given by the maximal modulus of the components, and $C = \max C_i^\pm$. from the equivalence of the norms in finite dimension we get $f(x) \ge C(N)L||x|| - C$, which shows $\liminf_{||x||\to\infty} \frac{f(x)}{||x||} \ge C(N)L$. The arbitrariness of L concludes the proof. □

1.3 Formal duality for constrained and penalized optimization problems

In this section we introduce the notion of dual problem of a convex optimization problem through an inf-sup exchange procedure. This often requires to write possible constraints as a sup penalization, and we will then see how to adapt to more general problems. No proof of duality results will be given, and we will analyze in details in the next section the most relevant examples in the calculus of variations. The proofs presented in Section 4.3 will then inspire Section 4.5 for a more general theory.

We start from the following problem:

$$\min\{f(x)\ :\ x \in X, Ax = b\},$$

where $A : X \to Y$ is a linear map between two normed vector spaces, $b \in Y$ is a fixed vector and $f : X \to \mathbb{R}\cup\{+\infty\}$ is a given convex and l.s.c. function. We will denote by A^t the transpose operator of A, a linear mapping defined on Y', the dual of Y, and valued into X', dual of X, and characterized by

$$\langle A^t\xi, x\rangle := \langle \xi, Ax\rangle \ \text{ for all } \xi \in Y' \text{ and } x \in X.$$

We can see that the above problem is equivalent to

$$\min\left\{f(x)+\sup_{\xi\in Y'}\langle\xi,Ax-b\rangle \;:\; x\in X\right\},$$

since we can compute the value of the expression $\sup_{\xi\in Y'}\langle\xi,Ax-b\rangle$ by distinguishing two cases: either $Ax=b$, in which case $\langle\xi,Ax-b\rangle=0$ for every ξ and the sup equals 0, or $Ax\neq b$, in which case there exists an element $\xi\in Y'$ such that $\langle\xi,Ax-b\rangle\neq 0$ and, by multiplying ξ times arbitrarily large constants, positive or negative depending on the sign of $\langle\xi,Ax-b\rangle$, we can see that the sup is $+\infty$. Hence, adding this sup means adding 0 if the constraint is satisfied or adding $+\infty$ if not; since in a minimization problem the value $+\infty$ is the same as a constraint, we can see the equivalence between the problem with the constraint $Ax=b$ and the problem with the sup over ξ.

We get now to a problem of the form

$$\inf_x\sup_\xi L(x,\xi),\qquad \text{where } L(x,\xi)=f(x)+\langle\xi,Ax-b\rangle.$$

This is an inf-sup problem, and we can associate with it a second optimization problem, obtained by switching the order of the inf and the sup. We can consider

$$\sup_\xi\inf_x L(x,\xi),$$

which means maximizing over ξ the function obtained as the value of the inf over x. Remember that we have forgotten the constraint $Ax=b$, since it took part in the definition of L, and we now minimize over all x. We can then give a better expression to this new problem, that we will call dual problem. Indeed we have

$$\sup_\xi\inf_x L(x,\xi)=\sup_\xi -\langle\xi,b\rangle+\inf_x f(x)+\langle\xi,Ax\rangle.$$

We then rewrite $\langle\xi,Ax\rangle$ as $\langle A^t\xi,x\rangle$ and change the ign in the inf so as to write it as a sup. We do obtain

$$\sup_\xi\inf_x L(x,\xi)=\sup_\xi -\langle\xi,b\rangle-\sup_x -f(x)+\langle -A^t\xi,x\rangle.$$

We now recognize in the sup over x the form of a Legendre transform and we finally obtain

$$\sup_\xi\inf_x L(x,\xi)=\sup_\xi -\langle\xi,b\rangle-f^*(-A^t\xi).$$

This is a convex optimization problem in the variable ξ (the maximization of the sum of a linear functional and the opposite of a convex function, f^*, applied to a linear function of ξ, involving the Legendre transform of the original objective function f.

We would like the two above optimization problems ("inf sup" and "sup inf") to be related to each other, and for instance their values to be the same. Given an arbitrary function L the values of inf sup and of sup inf are in genrel different, as w can see from this very simple example: take $L:A\times B\to\mathbb{R}$ with $A=B=\{\pm1\}$ and $L(a,b)=\mathrm{sign}(ab)$. In this case we have $\inf\sup=1>\sup\inf=-1$. Indeed, we always have an inequality, that we prove here.

Proposition 1.19. *Given an arbitrary function $L = A \times B \to \mathbb{R}$ we have*

$$\inf_a \sup_b L(a,b) \geq \sup_b \inf_a L(a,b).$$

Proof. Take $(a_0,b_0) \in A \times B$ and write $L(a_0,b_0) \geq \inf_a L(a,b_0)$. We then take the sup over b_0 on both sides, thus obtaining $\sup_{b_0} L(a_0,b_0) \geq \sup_{b_0} \inf_a L(a,b_0)$. We have now a number on the right-hand side, and a function of a_0 on the left-hand side. We then take the inf over a_0 and get

$$\inf_{a_0} \sup_{b_0} L(a_0,b_0) \geq \sup_{b_0} \inf_a L(a,b_0),$$

which is exactly the same as the claim up to renaming the variables. □

If in general it is not possible to connect the two problems obtained as inf-sup and sup-inf of a same function, it can be the case when some conditions are met. The main tool to do it is a theorem by Rockafellar (see [**?**], Section 37) requiring concavity in the variable on which we maximize, convexity in the other one, and some compactness assumption. In our precise case concavity and convexity are met, since L is convex in x and linear in ξ, and hence concave. Yet, Rockafellar's statement concerns finite-dimensional spaces, and moreover we should still deal with the compactness properties we would need. Hence, we will not provide any proof here that $\min\{f(x) : Ax = b\}$ and $\max\{-\langle \xi, b\rangle - f^*(-A^t\xi) : \xi \in Y'\}$ are equal, and we will wait till the next section for a proof in a very particular case.

We only discuss here some consequences and some variants of this duality approach.

A first consequence concerns sufficient optimality conditions. Suppose to consider two optimization problems issued by an inf-sup/sup-inf procedure, i.e. $\min\{f(a) : a \in A\}$ and $\max\{g(b) : b \in B\}$ with $f(a) := \sup_b L(a,b)$ and $g(b) := \inf_a L(a,b)$. Then, if $a_0 \in A$ and $b_0 \in B$ are such that $f(a_0) = g(b_0)$, automatically a_0 minimizes f and b_0 maximizes g, just because Proposition 1.19 guarantees $f(a_0) \geq \inf f \geq \sup g \geq g(b_0)$ and all inequalities must be equalities here. In our precise case the functional f should be replaced with f plus the constraint $Ax = b$, and g is given by $g(\xi) := -\langle \xi, b\rangle - f^*(-A^t\xi)$

A second consequence concers instead necessary optimality conditions. We need now to believe in $\inf f = \sup g$, and we suppose that we have a pair (a_0,b_0) where a_0 minimizes f and b_0 maximizes g. Then we deduce $f(a_0) = g(b_0)$, but this equality is a very strong piece of information in many cases. for instance, in our case this means that, if x_0 and ξ_0 are optimal, then we have

$$f(x_0) = -\langle \xi_0, b\rangle - f^*(-A^t\xi_0) \quad \text{and} \quad Ax_0 = b.$$

This can be re-written as

$$f(x_0) + f^*(-A^t\xi_0) = -\langle \xi_0, b\rangle = -\langle \xi_0, Ax_0\rangle = -\langle A^t\xi_0, x_0\rangle,$$

i.e. we have equality in the inequality $f(x) + f^*(y) \geq \langle x, y\rangle$. This is equivalent to

$$x_0 \in \partial f^*(-A^t\xi_0) \quad \text{and} \quad -A^t\xi_0 \in \partial f(x_0).$$

We can note the similarity with Lagrange multipliers, where optimizing a function f under a linear constraint of the form $Ax = b$ can be translated into the fact that ∇f should belong to a subspace, orthogonal to the affine space of the constraints, which is indeed the image of A^t.

Before moving on to variants of the previous pair of dual problems we want to insist that writing an equality constraint as a sup over test elements of a dual space is exactly what is always done in the weak formulation of PDEs. In the next section we will see as an example what happens when the constraint is of the form $\nabla \cdot \mathbf{v} = f$, which can be written as $\int \nabla\phi \cdot \mathbf{v} + \phi f = 0$ for every test function ϕ, and it is very natural to replace the constraint with a sup over ϕ. In this case, the dual problem turns out to be a maximization over scalar functions ϕ. Moreover, the transpose of the divergence operator $\nabla\cdot$ is the opposite of the gradient, since $\int \phi \nabla \cdot \mathbf{v} = -\int \nabla\phi \cdot \mathbf{v}$ by integration by parts, as soon as boundary conditions are taken care of. In this way, the functional $f^*(-A^t\phi)$ will be a very classical functional in calculus of variations.

Among variants, we first want to discuss the case of inequality constraints instead of equalities. A constraint of the form $Ax \le b$ only has a meaning if we give a notion of inequality among vectors, which is general is not canonically defined. In finite dimension a general convention, mainly used by computer scientists in optimization problem, is that we can consider the inequality component-wise. In calculus of variations, we can expect both Ax and b to be functions in a certain functional space, and we can require the inequality to be satisfied pointwise (or a.e.). What is important is that we should characterize the inequality in terms of test functions. for instance, the inequality $f \le g$ a.e. is equivalent to $\int \phi(f-g) \le 0$ for every $\phi \ge 0$ and a similar equivalence can be stated in the finite-dimensional componentwise case. We then write

$$\min\{f(x) \,:\, x \in X, Ax \le b\} = \min\left\{ f(x) + \sup_{\xi \in Y', \xi \ge 0} \langle \xi, Ax - b\rangle \,:\, x \in X \right\},$$

since

$$\sup_{\xi \in Y', \xi \ge 0} \langle \xi, y\rangle = \begin{cases} 0 & \text{if } y \le 0, \\ +\infty & \text{if not.} \end{cases}$$

We can then go on with the very same procedures and obtain the dual problem

$$\max\{-\langle \xi, b\rangle - f^*(-A^t\xi) \,:\, \xi \in Y, \xi \ge 0\}.$$

Finally, once that we know how to build dual problems out of constrained optimization problems, we could consider a more general case, such as

$$\min\{f(x) + g(Ax)\},$$

where $g = \mathbb{1}_{\{b\}}$ corresponds to the previous example. In this case we do not have constraints to write as a sup, but we can decide to write one of the two functions f or g as a sup thanks to the double Legendre transform. We then set

$$L(x,\xi) := f(x) + \langle \xi, Ax\rangle - g^*(\xi)$$

and we easily see that we have

$$\min\{f(x)+g(Ax)\} = \inf_x \sup_\xi L(x,\xi).$$

We then interchange inf and sup thus obtaining the dual problem

$$\begin{aligned}\sup_\xi \inf_x L(x,\xi) &= \sup_\xi -g^*(\xi) + \inf_x f(x) + \langle \xi, Ax\rangle \\ &= \sup_\xi -g^*(\xi) - \sup_x -f(x) + \langle -A^t\xi, x\rangle = \sup_\xi -g^*(\xi) - f^*(-A^t\xi).\end{aligned}$$

As we said, the equality constraint $Ax=b$ corresponds to $g=\mathbb{1}_{\{b\}}$, so that we have $g^*(\xi)=\langle \xi, b\rangle$.

The duality between

$$\min\{f(x)+g(Ax)\} \quad \text{and} \quad \sup_\xi -g^*(\xi) - f^*(-A^t\xi),$$

is a classical object in convex analysis and a theorem guaranteeing, under some conditions, that the the values are actually equal is known as Fenchel-Rockafellar's theorem. We will see in Section 4.5 a proof, in a simplified setting, of this theorem, inspired by the precise proof of a concrete duality result presented in Section 4.3.

1.4 A proof of Fenchel-Rockafellar's duality

In this section we want to take advantage of the technique developed in Section 4.3 for the precisa case of minimal flow problems in order to prove a general abstract version of the Fenchel-Rockafellar duality theorem. For simplicity, we will stick to the case where all spaces are reflexive, so that the role of the function in the primal and in the dual problems are completely symmetric. We will start from the following statement.

Theorem 1.20. *Suppose that X and Y are separable reflexive normed vector spaes, that $f : X \to \mathbb{R}\cup\{+\infty\}$ and $g : Y \to \mathbb{R}\cup\{+\infty\}$ are convex and lower-semicontinuous functions, and that $A : X \to Y$ is a continuous linear mapping. Suppose that g is bounded from below and f coercive. Then we have*

$$\min\{f(x)+g(Ax) : x\in X\} = \sup\left\{-g^*(\xi) - f^*(-A^t\xi) : \xi \in Y'\right\},$$

where the existence of the minimum on the right-hand side is part of the claim.

Proof. We will define a function $\mathscr{F} : Y \to \mathbb{R}\cup\{+\infty\}$ via

$$\mathscr{F}(p) := \min\{f(x)+g(Ax+p) : x\in X\}.$$

The existence of the minimum is a consequence of the following fact: for any sequence (x_n, p_n) with $f(x_n)+g(Ax_n+p_n)\le C$ the sequence x_n is bounded. This boundedness comes from the lower bound on g and from the coercive behavior of f. Once we know this, we can use $p_n = p$, take a minimizing sequence x_n for fixed p, and extract a weakly converging subsequence $x_n \rightharpoonup x$ using the Banach-Alaoglu Theorem. We also have $Ax_n + p \rightharpoonup Ax + p$ and the semicontinuity of f and g provide the minimality of x (since, being convex, f and g are both l.s.c. for the strong and the weak convergence, in X and Y, respectively).

We now compute $\mathscr{F}^* : Y' \to \mathbb{R} \cup \{+\infty\}$:

$$\begin{aligned}
\mathscr{F}^*(\xi) &= \sup_p \langle \xi, p\rangle - \mathscr{F}(p) \\
&= \sup_{p,x} \langle \xi, p\rangle - f(x) - g(Ax+p) \\
&= \sup_{y,x} \langle \xi, y - Ax\rangle - f(x) - g(y) \\
&= \sup_{y} \langle \xi, y\rangle - g(y) + \sup_x \langle -A^t\xi, x\rangle - f(x) \\
&= g^*(\xi) + f^*(-A^t\xi).
\end{aligned}$$

Now we use, as we did in Section 4.3, $\mathscr{F}^{**}(0) = \sup -\mathscr{F}^*$, which proves the claim, as soon as we prove that $\mathscr{F}$ is convex and l.s.c.

The convexity of $\mathscr{F}$ is easy. We just need to take $p_0, p_1 \in Y$, and define $p_t := (1-t)p_0 + tp_1$. Let x_0, x_1 be optimal in the definition of $\mathscr{F}(p_0)$ and $\mathscr{F}(p_1)$, i.e. $\int f(x_i) + g(Ax_i + p_i) = \mathscr{F}(p_i)$, and set $x_t := (1-t)x_0 + tx_1$. We have

$$\mathscr{F}(p_t) \leq f(x_t) + g(Ax_t + p_t) \leq (1-t)\mathscr{F}(p_0) + t\mathscr{F}(p_1),$$

and the convexity is proven.

For the semicontinuity, we take a sequence $p_n \to p$ in Y. We can suppose $\mathscr{F}(p_n) \leq C$ otherwise there is nothing to prove. Take the corresponding optimal points x_n and, applying the very first observation of this proof, we obtain $||x_n|| \leq C$. We can extract a subsequence such that $\lim_k \mathscr{F}(p_{n_k}) = \liminf_n \mathscr{F}(p_n)$ and $x_{n_k} \rightharpoonup x$. The semicontinuity of f and g provides

$$\mathscr{F}(p) \leq f(x) + g(Ax+p) \leq \liminf_k f(x_{n_k}) + g(Ax_{n_k} + p_{n_k}) = \lim_k \mathscr{F}(p_{n_k}) = \liminf_n \mathscr{F}(p_n),$$

which gives the desired result. □

We now note that, if g is not bounded from below, it is always possible to remove a suitable linear function from it so as to make it bounded from below, since all convex and l.s.c. functions are bounded from below by an affine function. We can then define $\tilde{g}$ via $\tilde{g}(y) = g - \langle \xi_0, y\rangle$ for a suitable ξ_0, and guarantee $\inf \tilde{g} > -\infty$. In order not to change the value of the primal problem we also need to modify f into $\tilde{f}$ defined via $\tilde{f}(x) := f + \langle A^t\xi_0, x\rangle$, so that

$$\tilde{f}(x) + \tilde{g}(Ax) = f(x) + \langle A^t\xi_0, x\rangle + g(Ax) - \langle \xi_0, Ax\rangle = f(x) + g(Ax).$$

Moreover, we can compute what changes in the dual problem. Is it true that we have $\sup_\xi -g^*(\xi) - f^*(-A^t\xi) = \sup_\xi -\tilde{g}^*(\xi) - \tilde{f}^*(-A^t\xi)$?

In order to do this, we need to compute the Legendre transform of $\tilde{f}$ and $\tilde{g}$. A general, and easy, fact, that is proposed as an exercise (see Exercise **??**) states that subtracting a linear function translates into a translation on the Legendre transform. We then have

$$\tilde{g}^*(\xi) = g^*(\xi + \xi_0); \quad \tilde{f}^*(\zeta) = f^*(\zeta - A^t \xi_0)$$

and then

$$\tilde{g}^*(\xi) + \tilde{f}^*(-A^t \xi) = g^*(\xi + \xi_0) + f^*(-A^t(\xi + \xi_0))$$

and a simple change of variable $\xi \mapsto \xi + \xi_0$ shows that the sup has not changed. This shows that the duality result is not affected by this reformulation in terms of $\tilde{f}$ and $\tilde{g}$. It is then enough, for the duality to hold, that the assumptions of Theorem 1.20 are satisfied by $(\tilde{f}, \tilde{g})$ instead of (f, g). Since we chose ξ_0 on purpose in order to have $\tilde{g}$ lower bounded, we only need now to require that $\tilde{f}$ is coercive. Not that this would be the case if f was superlinear, as it would stay superlinear after adding any linear function, but it is not automatic when speaking of a generic coercive function.

The condition on ξ_0 such that, at the same time $\tilde{g}$ is bounded from below and $\tilde{f}$ superlinear can be more easily translated in terms of f^* and g^*. We can indeed state the following propositon.

Proposition 1.21. *Suppose that X and Y are separable reflexive normed vector spaes, that $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$ are convex and lower-semicontinuous functions, and that $A : X \to Y$ is a continuous linear mapping. Suppose that there exists $\xi_0 \in Y'$ such that $g^*(\xi_0) < +\infty$ and that f^* is continuous and finite at $A^t \xi_0$. Then we have*

$$\min\{f(x) + g(Ax) : x \in X\} = \sup\left\{-g^*(\xi) - f^*(-A^t \xi) : \xi \in Y'\right\},$$

where the existence of the minimum on the right-hand side is part of the claim.

Proof. The condition $g^*(\xi_0) < +\infty$ means $\tilde{g}^*(0) < +\infty$, which means that $\tilde{g}$ is bounded from below.

The condition on f^* at $A^t \xi_0$ translates into the same condition for $\tilde{f}^*$ at 0, and we know that a function is coercive if and only if its Legendre transform is bounded on a neighborhood of 0. This means that $\tilde{f}$ is coercive.

We then conclude that we can apply Theorem 1.20 to $(\tilde{f}, \tilde{g})$ instead of (f, g), which provides the desired result. □

We can also deduce the following statement, which is probably the most standard formulation of the Fenchel-Rockafellar duality theorem, even if we only state it for reflexive spaces.

Theorem 1.22. *Suppose that X and Y are separable reflexive normed vector spaes, that $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$ are convex and lower-semicontinuous functions, and that $A : X \to Y$ is a continuous linear mapping. Suppose that there exists $x_0 \in X$ such that $f(x_0) < +\infty$ and that g is continuous and finite at Ax_0. Then we have*

$$\inf\{f(x) + g(Ax) : x \in X\} = \max\left\{-g^*(\xi) - f^*(-A^t \xi) : \xi \in Y'\right\},$$

where the existence of the minimum on the right-hand side is part of the claim.

Proof. The proof is straightforward once we realize that we can interchange f with g^*, g with f^*, and A with A^t in the statement of Proposition 1.21. □