The Legendre transform

Jordan Bell

April 25, 2014

1 Convexity

Write $\overline{\mathbb{R}} = [-\infty, \infty]$. We define $\infty + \infty = \infty$, $-\infty - \infty = -\infty$, and $\infty - \infty$ is nonsense. If $a \in \mathbb{R}$, we define $a + \infty = \infty$ and $a - \infty = -\infty$.

If X is a vector space and $f : X \to \overline{\mathbb{R}}$ is a function, we define the *epigraph of f* to be the set

$$\text{epi } f = \{(x, \alpha) \in X \times \mathbb{R} : \alpha \geq f(x)\},$$

and if epi f is a convex subset of the vector space $X \times \mathbb{R}$, we say that f is *convex*. We define the *effective domain* of a convex function f to be

$$\text{dom } f = \{x \in X : f(x) < \infty\}.$$

We say that a convex function $f : X \to \overline{\mathbb{R}}$ is *proper* if dom $f \neq \emptyset$ and f does not take the value $-\infty$. If C is a convex subset of X and $f : C \to \mathbb{R}$ is a function, we extend f to X by defining $f(x) = \infty$ for $x \in X \setminus C$.

Lemma 1. *If X is a vector space, C is a convex subset of X, and $f : C \to \mathbb{R}$ satisfies*

$$f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2), \qquad x_1, x_2 \in C, \quad 0 \leq t \leq 1,$$

then $f : X \to \overline{\mathbb{R}}$ is convex.

Proof. Let $(x_1, \alpha_1), (x_2, \alpha_2) \in \text{epi } f$ and $0 \leq t \leq 1$. The fact that the pairs (x_i, α_i) belong to epi f means in particular that $f(x_i) < \infty$, and hence that $x_i \in C$, as otherwise we would have $f(x_i) = \infty$. But

$$(1-t)(x_1, \alpha_1) + t(x_2, \alpha_2) = ((1-t)x_1 + tx_2, (1-t)\alpha_1 + t\alpha_2),$$

and, as $x_1, x_2 \in C$,

$$f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2) \leq (1-t)\alpha_1 + t\alpha_2,$$

showing that $(1-t)(x_1, \alpha_1) + t(x_2, \alpha_2) \in \text{epi } f$, and hence that $f : X \to \overline{\mathbb{R}}$ is convex. $\square$

2   Definition of the Legendre transform

Let V be a locally convex space and let V^* be the dual space of V, i.e. the set of all continuous linear maps $V \to \mathbb{R}$. With the weak-* topology, V^* is itself a locally convex space and $V = (V^*)^*$, with the isomorphism of locally convex spaces $x \mapsto (\lambda \mapsto \lambda x)$. If $f : V \to \overline{\mathbb{R}}$ is a function, the *Legendre transform* or *convex conjugate* of f is the function $f^* : V^* \to \overline{\mathbb{R}}$ defined by

$$f^*(\lambda) = \sup\{\lambda v - f(v) : v \in V\} = \sup\{\lambda v - f(v) : v \in \operatorname{dom} f\}.$$

Like how the Fourier transform of a function from a locally compact abelian group to $\mathbb{C}$ is itself a function from the Pontryagin dual of the group to $\mathbb{C}$, the Legendre transform of a function from a locally convex space to $\overline{\mathbb{R}}$ is itself a function from the dual space to $\overline{\mathbb{R}}$.

Theorem 2. *If V is a locally convex space and $f : V \to \overline{\mathbb{R}}$ is convex, then its Legendre transform $f^* : V^* \to \overline{\mathbb{R}}$ is convex.*

Proof. Let $(\lambda_1, \alpha_1), (\lambda_2, \alpha_2) \in \operatorname{epi} f^*$, and let $0 \le t \le 1$. We have

$$
\begin{aligned}
f^*((1-t)\lambda_1 + t\lambda_2) &= \sup\{((1-t)\lambda_1 + t\lambda_2)v - f(v) : v \in \operatorname{dom} f\} \\
&= \sup\{(1-t)(\lambda_1 v - f(v)) + t(\lambda_2 v - f(v)) : v \in \operatorname{dom} f\} \\
&\le \sup\{(1-t)(\lambda_1 v - f(v)) : v \in \operatorname{dom} f\} \\
&\quad + \sup\{t(\lambda_2 v - f(v)) : v \in \operatorname{dom} f\} \\
&= (1-t)f^*(\lambda_1) + tf^*(\lambda_2) \\
&\le (1-t)\alpha_1 + t\alpha_2,
\end{aligned}
$$

which means that $(1-t)(\lambda_1, \alpha_1) + t(\lambda_2, \alpha_2) \in \operatorname{epi} f^*$, and hence that f^* is convex. $\square$

3   Lower semicontinuity

If X is a topological vector space and $f : X \to \overline{\mathbb{R}}$ is a function, we say that f is *lower semicontinuous* if $\operatorname{epi} f$ is a closed subset of $X \times \mathbb{R}$.

Theorem 2 shows that the Legendre transform of a convex function is itself convex. The following lemma states that if a proper convex function is lower semicontinuous, then its Legendre transform is proper; one proves the lemma using the Hahn-Banach separation theorem.[1] We use this lemma in the proof of the theorem that comes after.

Lemma 3. *If V is a locally convex space and $f : V \to \overline{\mathbb{R}}$ is a lower semicontinuous proper convex function, then its Legendre transform $f^* : V^* \to \overline{\mathbb{R}}$ is proper.*

---

[1] Viorel Barbu and Teodor Precupanu, *Convexity and Optimization in Banach Spaces*, fourth ed., p. 78, Corollary 2.21.

Theorem 4. *If V is a locally convex space and $f : V \to \overline{\mathbb{R}}$ is a lower semicontinuous proper convex convex, then $f^{**} = f$.*

Proof. For any $\lambda \in V^*$ we have $f^*(\lambda) = \sup\{\lambda v - f(v) : v \in V\}$, and hence for any $v \in V$ we have $f^*(\lambda) \geq \lambda v - f(v)$. Thus, for any $v \in V$ and $\lambda \in V^*$ we have

$$\lambda v - f^*(\lambda) \leq f(v).$$

Using this, for any $v \in V$ we have

$$f^{**}(v) = \sup\{v\lambda - f^*(\lambda) : \lambda \in \mathrm{dom}\, f^*\} = \sup\{\lambda v - f^*(\lambda) : \lambda \in \mathrm{dom}\, f^*\} \leq f(v).$$

Suppose by contradiction that there were some $v_0 \in V$ for which $f^{**}(v_0) < f(v_0)$. First, by Lemma 3 we have that f^* is proper, so in particular $\mathrm{dom}\, f^* \neq \emptyset$, and this tells us that f^{**} does not take the value $-\infty$. Hence $-\infty < f^{**}(v_0) < f(v_0)$, which tells us that $(v_0, f^{**}(v_0)) \in V \times \mathbb{R} \setminus \mathrm{epi}\, f$. Therefore, epi f and the singleton $\{(v_0, f^{**}(v_0))\}$ are disjoint closed convex sets (epi f is closed because f is lower semicontinuous), and so we can apply the Hahn-Banach separation theorem: there is some $\Lambda \in (V \times \mathbb{R})^*$ and some $\gamma \in \mathbb{R}$ for which

$$\Lambda(v, \alpha) < \gamma < \Lambda(v_0, f^{**}(v_0)), \qquad (v, \alpha) \in \mathrm{epi}\, f.$$

As $\Lambda \in (V \times \mathbb{R})^*$, there is some $\lambda \in V^*$ and some $\beta \in \mathbb{R}^* = \mathbb{R}$ for which $\Lambda(v, \alpha) = \lambda v + \beta\alpha$, and so

$$\lambda v + \beta\alpha < \gamma < \lambda v_0 + \beta f^{**}(v_0), \qquad (v, \alpha) \in \mathrm{epi}\, f. \tag{1}$$

If $\beta > 0$ then we get a contradiction because for a fixed $v \in \mathrm{dom}\, f$ there are arbitrarily large positive α for which $(v, \alpha) \in \mathrm{epi}\, f$. Hence, $\beta \leq 0$. Assume by contradiction that $\beta = 0$, and hence that

$$\lambda v < \gamma < \lambda v_0, \qquad v \in \mathrm{dom}\, f. \tag{2}$$

If $v_0 \in \mathrm{dom}\, f$ then we get $\lambda v_0 < \gamma < \lambda v_0$, a contradiction. If $v_0 \notin \mathrm{dom}\, f$, we shall still obtain a contradiction. Let $\mu \in \mathrm{dom}\, f^*$. For any $h > 0$,

$$
\begin{aligned}
f^*(\mu + h\lambda) &= \sup\{(\mu + h\lambda)v - f(v) : v \in \mathrm{dom}\, f\} \\
&= \sup\{\mu v - f(v) + h\lambda v : v \in \mathrm{dom}\, f\} \\
&\leq \sup\{\mu v - f(v) : v \in \mathrm{dom}\, f\} + \sup\{h\lambda v : v \in \mathrm{dom}\, f\} \\
&= f^*(\mu) + h\sup\{\lambda v : v \in \mathrm{dom}\, f\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f^{**}(v_0) &\geq (\mu + h\lambda)v_0 - f^*(\mu + h\lambda) \\
&\geq (\mu + h\lambda)v_0 - f^*(\mu) - h\sup\{\lambda v : v \in \mathrm{dom}\, f\} \\
&= \mu v_0 - f^*(\mu) + h\big(\lambda v_0 - \sup\{\lambda v : v \in \mathrm{dom}\, f\}\big).
\end{aligned}
$$

But by (2) we have $\lambda v_0 - \sup\{\lambda v : v \in \mathrm{dom}\, f\} > 0$, and therefore the right-hand side of

$$f^{**}(v_0) \geq \mu v_0 - f^*(\mu) + h\big(\lambda v_0 - \sup\{\lambda v : v \in \mathrm{dom}\, f\}\big)$$

3

can be an arbitrarily large positive number (as $f^*(\mu) < \infty$), contradicting that $f^{**}(v_0) < \infty$. Therefore, $\beta < 0$, and dividing (1) by β then yields

$$\frac{1}{\beta}\lambda v + \alpha > \frac{\gamma}{\beta} > \frac{1}{\beta}\lambda v_0 + f^{**}(v_0), \qquad (v, \alpha) \in \text{epi } f.$$

Hence,

$$
\begin{aligned}
\left(-\frac{1}{\beta}\lambda\right) v_0 - f^{**}(v_0) \quad & > \quad \sup\left\{-\frac{1}{\beta}\lambda v - \alpha : (v, \alpha) \in \text{epi } f\right\} \\
& = \quad \sup\left\{-\frac{1}{\beta}\lambda v - f(v) : v \in \text{dom } f\right\} \\
& = \quad f^*\left(-\frac{1}{\beta}\lambda\right),
\end{aligned}
$$

which contradicts

$$\left(-\frac{1}{\beta}\lambda\right) v_0 - f^{**}(v_0) \le f^*\left(-\frac{1}{\beta}\lambda\right).$$

Therefore, there is no $v_0 \in V$ for which $f^{**}(v_0) < f(v_0)$, i.e., for all $v \in V$ we have

$$f(v_0) \le f^{**}(v_0).$$

$\square$

4   Example in $\mathbf{R}^n$

Let $V : \mathbb{R}^n \to \mathbb{R}$ be a function, let A be an $n \times n$ symmetric positive definite matrix, and define $L : T\mathbb{R}^n \to \mathbb{R}$ by

$$L(q, v) = \frac{1}{2}\langle v, Av \rangle - V(q).$$

Fix any $q \in \mathbb{R}^n$, let $X = T_q\mathbb{R}^n = \mathbb{R}^n$, and write $L_q(v) = L(q, v)$, for which $L_q : X \to \mathbb{R}$. The Legendre transform of L_q is $L_q^* : X^* \to \overline{\mathbb{R}}$, defined by

$$L_q^*(\lambda) = \sup\{\lambda v - L_q(v) : v \in X\}.$$

Because A is symmetric, for any $v \in X$ we obtain

$$D(\lambda - L_q)(v) = \lambda - Av.$$

Hence, $D(\lambda - L_q)(v) = 0$ is equivalent to

$$v = A^{-1}\lambda,$$

and with this,

$$L_q(A^{-1}\lambda) = \frac{1}{2}\left\langle A^{-1}\lambda, AA^{-1}\lambda \right\rangle - V(q) = \frac{1}{2}\left\langle A^{-1}\lambda, \lambda \right\rangle - V(q),$$

and therefore

$$L_q^*(\lambda) = \lambda(A^{-1}\lambda) - \frac{1}{2}\left\langle A^{-1}\lambda, \lambda \right\rangle + V(q) = \frac{1}{2}\left\langle A^{-1}\lambda, \lambda \right\rangle + V(q).$$

4

5 Derivatives

Let Ω be a domain in $\mathbb{R}^n$ and let $f \in C^s(\Omega)$ for some $s \geq 2$. We define $\phi : \Omega \to \mathbb{R}^n$ by $\phi(x) = (Df)(x)$; we have $\phi \in C^{s-1}(\Omega, \mathbb{R}^n)$. Following Giaquinta and Hildebrandt,[2] we call ϕ a *gradient mapping*. The following theorem gives conditions under which $\phi : \Omega \to \phi(\Omega)$ is invertible.[3] (To be *locally invertible* means that for each point there is an open neighborhood such that the restriction of ϕ to that neighborhood is invertible.)

Theorem 5. *If*
$$\det(D^2 f)(x) \neq 0, \qquad x \in \Omega,$$
then ϕ is locally invertible on Ω. If Ω is convex and for all $x \in \Omega$ the matrix $(D^2 f)(x)$ is positive definite, then $\phi : \Omega \to \phi(\Omega)$ is a C^{s-1} diffeomorphism.

Proof. Because $\phi \in C^{s-1}(\Omega, \mathbb{R}^n)$ and $\det(D\phi)(x) = \det(D^2 f)(x) \neq 0$ for all $x \in \Omega$, by the inverse function theorem[4] we have that $\phi : \Omega \to \phi(\Omega)$ is a local C^{s-1} diffeomorphism.

Suppose that $\phi(x_1) = \phi(x_2)$ for some distinct $x_1, x_2 \in \Omega$. Put $x = x_2 - x_1$. Because $x_1, x_1 + x = x_2 \in \Omega$ and Ω is convex, for any $0 \leq t \leq 1$ we have $x_1 + tx \in \Omega$. Now define
$$A(t) = (D^2 f)(x_1 + tx), \qquad 0 \leq t \leq 1;$$
because $f \in C^s(\Omega)$ with $s \geq 2$, we have that A is continuous. Because
$$\frac{d}{dt}\phi(x_1 + tx) = (D\phi)(x_1 + tx)x = (D^2 f)(x_1 + tx)x = A(t)x,$$
we have
$$
\begin{aligned}
\langle x, \phi(x_2) - \phi(x_1)\rangle &= \left\langle x, \int_0^1 \frac{d}{dt}\phi(x_1 + tx)dt \right\rangle \\
&= \left\langle x, \int_0^1 A(t)dt \right\rangle \\
&= \int_0^1 \langle x, A(t)x\rangle \, dt.
\end{aligned}
$$

For each $0 \leq t \leq 1$ we have that $A(t)$ is a positive definite matrix, and because $x \neq 0$, this gives us that $\langle x, A(t)x\rangle > 0$. Moreover, $t \mapsto \langle x, A(t)x\rangle$ is continuous, so it follows that
$$\int_0^1 \langle x, A(t)x\rangle \, dt > 0.$$

But $\langle x, \phi(x_2) - \phi(x_1)\rangle = 0$, a contradiction. Therefore, ϕ is one-to-one. It is a fact that a local diffeomorphism that is one-to-one is a diffeomorphism, thus $\phi : \Omega \to \phi(\Omega)$ is a C^{s-1} diffeomorphism. $\square$

---

[2] Mariano Giaquinta and Stefan Hildebrandt, *Calculus of Variations II*, p. 6.

[3] Mariano Giaquinta and Stefan Hildebrandt, *Calculus of Variations II*, p. 6, Lemma 1.

[4] Jerrold E. Marsden and Michael J. Hoffman, *Elementary Classical Analysis*, second ed., p. 393, Theorem 7.1.1.

Suppose that the gradient mapping $\phi : \Omega \to \phi(\Omega)$ is a C^{s-1} diffeomorphism. We write $\psi = \phi^{-1}$, so $\psi : \phi(\Omega) \to \Omega$ is a C^{s-1} diffeomorphism. The following theorem gives an explicit expression for the Legendre transform of certain functions.[5]

Theorem 6. *If Ω is a convex domain in $\mathbb{R}^n$, $f \in C^2(\Omega)$, and for all $x \in \Omega$ the matrix $(D^2 f)(x)$ is positive definite, then*

$$f^*(\xi) = \xi \psi(\xi) - f(\psi(\xi)), \qquad \xi \in \phi(\Omega).$$

Proof. Fix $\xi \in \phi(\Omega)$ and define $g : \Omega \to \mathbb{R}$ by

$$g(x) = \xi x - f(x).$$

We have $g \in C^2(\Omega)$, and we have $(Dg)(x) = \xi - (Df)(x)$ and $(D^2 g)(x) = -(D^2 f)(x)$. Thus, for each $x \in \Omega$, the matrix $(D^2 g)(x)$ is negative definite. It follows that if there is a point $x_0 \in \Omega$ at which $(Dg)(x_0) = 0$, then for all other $x \in \Omega$ we have $g(x) < g(x_0)$. To have $(Dg)(x_0) = 0$ is equivalent $(Df)(x_0) = \xi$, and because $\phi : \Omega \to \phi(\Omega)$ is a bijection, there is indeed a unique $x_0 \in \Omega$ for which $(Df)(x_0) = \phi(x_0) = \xi$. Therefore,

$$
\begin{aligned}
f^*(\xi) &= \sup\{\xi x - f(x) : x \in \Omega\} \\
&= \sup\{g(x) : x \in \Omega\} \\
&= g(x_0) \\
&= \xi x_0 - f(x_0) \\
&= \xi \psi(\xi) - f(\psi(\xi)).
\end{aligned}
$$

$\square$

Using the above expression for the Legendre transform of a C^2 function with positive definite Hessian, we show in the following theorem that the Legendre transform of a C^s function with positive definite Hessian is itself a C^2 function.[6]

Theorem 7. *If Ω is a convex domain in $\mathbb{R}^n$, $f \in C^s(\Omega)$ with $s \geq 2$, and for all $x \in \Omega$ the matrix $(D^2 f)(x)$ is positive definite, then $Df^* = \psi$ and $f^* \in C^s(\phi(\Omega))$.*

Proof. For all $\xi \in \phi(\Omega)$ we have by Theorem 6,

$$f^*(\xi) = \xi \psi(\xi) - f(\psi(\xi)).$$

Thus,

$$
\begin{aligned}
(Df^*)(\xi) &= \psi(\xi) + \xi(D\psi)(\xi) - (Df)(\psi(\xi))(D\psi)(\xi) \\
&= \psi(\xi) + \xi(D\psi)(\xi) - \phi(\psi(\xi))(D\psi)(\xi) \\
&= \psi(\xi) + \xi(D\psi)(\xi) - \xi(D\psi)(\xi) \\
&= \psi(\xi).
\end{aligned}
$$

---

[5] Mariano Giaquinta and Stefan Hildebrandt, *Calculus of Variations II*, p. 9.
[6] Mariano Giaquinta and Stefan Hildebrandt, *Calculus of Variations II*, p. 7, Lemma 2.

Hence $Df^* = \psi$. Because $\psi \in C^{s-1}(\phi(\Omega))$, it follows that $f \in C^s(\phi(\Omega))$. $\square$

For all $x \in \Omega$, because $\psi(\phi(x)) = x$ we have $(D\psi)(\phi(x))(D\phi)(x) = I$, i.e. $(D^2 f^*)(\phi(x))(D^2 f)(x) = I$, so

$$(D^2 f)(x) = ((D^2 f^*)(\phi(x)))^{-1}.$$

For all $\xi \in \phi(\Omega)$, because $\phi(\psi(\xi)) = \xi$, we have $(D\phi)(\psi(\xi))(D\psi)(\xi) = I$, i.e. $(D^2 f)(\psi(\xi))(D^2 f^*)(\xi) = I$, so

$$(D^2 f^*)(\xi) = ((D^2 f)(\psi(\xi)))^{-1}.$$

6 Example in $\mathbf{R}^2$

Suppose that Ω is a convex domain in $\mathbb{R}^2$, that $f \in C^2(\Omega)$, and that for all $x \in \Omega$, the matrix $(D^2 f)(x)$ is positive definite. Write $\rho(x) = \det(D^2 f)(x)$; $\rho(x) > 0$ for all $x \in \Omega$. Because $(D^2 f)(x) = ((D^2 f^*)(\phi(x)))^{-1}$ for all $x \in \Omega$, we have

$$
\begin{aligned}
\begin{pmatrix} f_{11}(x) & f_{12}(x) \\ f_{21}(x) & f_{22}(x) \end{pmatrix} &= \begin{pmatrix} f_{11}^*(\phi(x)) & f_{12}^*(\phi(x)) \\ f_{21}^*(\phi(x)) & f_{22}^*(\phi(x)) \end{pmatrix}^{-1} \\
&= \frac{1}{\det(D^2 f^*)(\phi(x))} \begin{pmatrix} f_{22}^*(\phi(x)) & -f_{12}^*(\phi(x)) \\ -f_{21}^*(\phi(x)) & f_{11}^*(\phi(x)) \end{pmatrix} \\
&= \rho(x) \begin{pmatrix} f_{22}^*(\phi(x)) & -f_{12}^*(\phi(x)) \\ -f_{21}^*(\phi(x)) & f_{11}^*(\phi(x)). \end{pmatrix}
\end{aligned}
$$

Giaquinta and Hildebrandt[7] give the following consequence of what we have just written out. If f satisfies the above conditions and satisfies the equation

$$(1 + f_2^2) f_{11} - 2 f_1 f_2 f_{12} + (1 + f_1^2) f_{22} = 2H(1 + f_1^2 + f_2^2)^{3/2},$$

on Ω, where H is some constant, then

$$
\begin{aligned}
&(1 + f_2(x)^2) \rho(x) f_{22}^*(\phi(x)) + 2 f_1(x) f_2(x) \rho(x) f_{12}^*(\phi(x)) \\
&+ (1 + f_1(x)^2) \rho(x) f_{11}^*(\phi(x)) \\
=&2H(1 + f_1(x)^2 + f_2(x)^2)^{3/2}
\end{aligned}
$$

for all $x \in \Omega$. Therefore,

$$
\begin{aligned}
&(1 + \xi_2^2) \rho(\psi(\xi)) f_{22}^*(\xi) + 2\xi_1 \xi_2 \rho(\psi(\xi)) f_{12}^*(\xi) + (1 + \xi_1^2) \rho(\psi(\xi)) f_{11}^*(\xi) \\
=&2H(1 + \xi_1^2 + \xi_2^2)^{3/2}
\end{aligned}
$$

for all $\xi \in \phi(\Omega)$, $\xi = (\xi_1, \xi_2)$. In the case where $H = 0$, then, dividing by $\rho(\psi(\xi))$, which is > 0, we obtain

$$(1 + \xi_2^2) f_{22}^*(\xi) + 2\xi_1 \xi_2 f_{12}^*(\xi) + (1 + \xi_1^2) f_{11}^*(\xi) = 0.$$

---

[7]Mariano Giaquinta and Stefan Hildebrandt, *Calculus of Variations II*, p. 14.

In the case $H = 0$, the equation satisfied by f is called the *minimal surface equation*, and we have thus found a partial differential equation satisfied by the Legendre transform of a solution of the minimal surface equation that satisfies the conditions we imposed at the start of the example. Writing the equation satisfied by f^* in the form

$$Af_{11}^* + 2Bf_{12}^* + Cf_{22}^* = 0,$$

we have $A = (1 + \xi_1^2)$, $B = \xi_1\xi_2$, $C = (1 + \xi_2^2)$, with which

$$B^2 - AC = \xi_1^2\xi_2^2 - (1 + \xi_1^2)(1 + \xi_2^2) = -1 - \xi_1^2 - \xi_2^2 \le -1,$$

which means that partial differential equation satisfied by f^* is elliptic.

7 Lagrangians and Hamiltonians

Theorem 6 states that if Ω is a convex domain in $\mathbb{R}^n$, $f \in C^2(\Omega)$, and for all $x \in \Omega$ the matrix $(D^2 f)(x)$ is positive definite, then

$$f^*(\xi) = \xi\psi(\xi) - f(\psi(\xi)), \qquad \xi \in \phi(\Omega).$$

Suppose that $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ is a function such that for each $q \in \mathbb{R}^n$ and $t \in \mathbb{R}$, the function $v \mapsto L(q, v, t)$ satisfies the above conditions. Fix $q \in \mathbb{R}^n$ and $t \in \mathbb{R}$. With $DL = (\frac{\partial L}{\partial q}, \frac{\partial L}{\partial v}, \frac{\partial L}{\partial t})$ and $\phi(v) = \frac{\partial L}{\partial v}(q, v, t)$, $\psi = \phi^{-1}$, we have

$$L^*(q, p, t) = p\psi(p) - L(q, \psi(p), t), \qquad p \in \phi(\mathbb{R}^n),$$

or with $H = L^*$,

$$H(q, p, t) = p\psi(p) - L(q, \psi(p), t), \qquad p \in \phi(\mathbb{R}^n).$$

We have

$$\frac{\partial H}{\partial q}(q, p, t) = -\frac{\partial L}{\partial q}(q, \psi(p), t),$$

and

$$
\begin{aligned}
\frac{\partial H}{\partial p}(q, p, t) &= \psi(p) + p(D\psi)(p) - \frac{\partial L}{\partial v}(q, \psi(p), t)(D\psi)(p) \\
&= \psi(p) + p(D\psi)(p) - \phi(\psi(p))(D\psi)(p) \\
&= \psi(p) + p(D\psi)(p) - p(D\psi)(p) \\
&= \psi(p),
\end{aligned}
$$

and

$$\frac{\partial H}{\partial t}(q, p, t) = -\frac{\partial L}{\partial t}(q, \psi(p), t).$$

For a path $(q(t), v(t), t)$ to satisfy the *Euler-Lagrange equation* means that

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v}(q(t), v(t), t)\right) = \frac{\partial L}{\partial q}(q(t), v(t), t).$$

With $p(t) = \phi(v(t))$, this yields

$$\frac{dp}{dt}(t) = \frac{\partial L}{\partial q}(q(t), \psi(p(t)), t),$$

and hence

$$\frac{dp}{dt}(t) = -\frac{\partial H}{\partial q}(q(t), p(t), t).$$

8   Physical units

First, if a Lagrangian L, $L(q, v, t)$, has units J, then the Hamiltonian $H = L^*$, $H(q, p, t)$, has the same units J, and it follows that $p\psi(p)$ has units J. Second, $[\psi(p)] = [v]$, and so $[H] = [p][\psi(p)] = [p][v]$. Therefore, $[p][v] = $ J. If we take $[v] = $ m/s, then this implies that $[p] = $ kg m/s.

9   More books

V. I. Arnold, *Mathematical Methods of Classical Mechanics*, second ed., p. 61, §14; Ralph Abraham and Jerrold E. Marsden, *Foundations of Mechanics*, second ed., p. 218, §3.6; Jerrold E. Marsden and Tudor S. Ratiu, *Introduction to Mechanics and Symmetry*, second ed., p. 183, §7.2; Jürgen Jost and Xianqing Li-Jost, *Calculus of Variations*, chapter 4; David Yang Gao, *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*.

10   History

As best as I can tell, the thing we call the Legendre transform is named after Legendre because of the following paper: Adrien-Marie Legendre, *Mémoire sur l'intégration de quelques Équations aux différences partielles*, Histoire de l'Académie royale des sciences (1787), 309–351. The following is a partial bibliography of works that refer to this paper of Legendre's. No historical summary of the Legendre transform exists in the literature, and the following is presented as an aid to the preparation of one. To properly tell the story of the Legendre transform, one would be well served by carefully digging through sources and attentively reading Legendre's original paper, and also by making oneself comfortable with how it appears in convex analysis, minimal surfaces, contact geometry, thermodynamics, etc. Such a comprehensive history would require meticulously scanning through Legendre's monumental *Traite* on elliptic integrals lest relevant material is included there. The best biography of Legendre that exists is the one by Itard in the *Dictionary of Scientific Biography*, who mentions that something relevant to the Legendre transform appears in volume II of the 1826 *Traite*, concerning arc lengths. One should also scan through the work of Lagrange, including his 1788 *Méchanique analitique*, and the work of Euler on the calculus of variations.

Correspondance de Leonhard Euler avec A. C. Clairaut, J. d'Alembert et J. L. Lagrange, pp. 440–441, Note 6; S. S. Demidov, *The study of partial differential equations of the first order in the 18th and 19th centuries*, Arch. Hist. Exact Sci. **26** (1982), no. 4, 325–350; Erwin Kreyszig, *On the Theory of Minimal Surfaces*, The Problem of Plateau (Themistocles M. Rassias, ed.), 1992, 138–164, p. 145; Julian Lowell Coolidge, *A History of Geometrical Methods*, p. 377; Alfred Enneper, *Bemerkungen über einige Flächen von constantem Krümmungsmaaß*, Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen (1876), 597–619, p. 614; Alfred Enneper, *Ueber Flächen mit besonderen Meridiancurven*, Abhandlungen der Königlichen Gesellschaft der Wissenschaften in Göttingen **29** (1882), 3–87, p. 6; Gaston Darboux, *Leçons sur la théorie générale des surfaces*, vol. 1, p. 271, §177; Édouard Goursat, *Leçons sur l'intégration des équations aux dérivées partielles du second ordre, à deux variables indépendantes*, tome 2, p. 32, chapter V, §113; René Taton, *L'œuvre scientifique de Monge*, p. 262; Karin Reich, *Die Geschichte der Differentialgeometrie von Gauß bis Riemann (1828–1868)*, Arch. Hist. Exact Sci. **11** (1973), no. 4, 273–376, p. 315; Ivor Grattan-Guinness, *Convolutions in French Mathematics, 1800-1840*, vol. I, p. 152; Morris Kline, *Mathematical Thought From Ancient to Modern Times*, chapter 22; João Caramalho Domingues, *Lacroix and the Calculus*, p. 223; Ernst Hairer, Syvert Paul Nørsett and Gerhard Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, p. 32; Paul Mansion, *Théorie des équations aux dérivées partielles du premier ordre*, p. 76; A. R. Forsyth, *A Treatise on Differential Equations*, sixth ed., pp. 418, 476; Lagrange, *Méchanique analitique* (1788), tome 1, partie 2, §IV; Ernesto Pascal, *Die Variationsrechnung*, p. 125; Bernhard Riemann, *Ueber die Fläche vom kleinsten Inhalt bei gegebener Begrenzung*; Courant and Hilbert, vol. II; Cornelius Lanczos, *The Variational Principles of Mechanics*, fourth ed., §VI.1; Ed. Combescure, *Remarques sur un Mémoire de Legendre*, Comptes rendus hebdomadaires des séances de l'Académie des sciences **74** (1872), 798–802; Johannes C. C. Nitsche, *Vorlesungen über Minimalflächen*, p. 147; A. W. Conway and J. L. Synge (ed.), *The Mathematical Papers of Sir William Rowan Hamilton*, vol. I, (1931), p. 474.