The Legendre Transform

Arthur Jaffe

September 26, 2024

I Background

The Legendre transform $\mathbb{L}$ takes a convex function f defined on a convex set $\mathcal{D}$ into another convex function $\mathbb{L}f$ defined on a convex set $\mathbb{L}\mathcal{D}$. In particular

$$(\mathbb{L}f)(p) = \sup_{v\in\mathcal{D}} \left(\langle p, v\rangle - f(v)\right) \ . \tag{1}$$

For a convex function f on a convex set, the Legendre transform has period two, namely

$$\mathbb{L}^2 f = f \ . \tag{2}$$

The Legendre transform is non-linear (unlike the Fourier or Laplace transforms). It has interesting, if subtle, properties. In particular, the non-linearity makes the Legendre transformation more difficult to deal with, but nevertheless the transformation is very useful. In these notes we consider an elementary form of the Legendre transform, in which we transform a function f of a single, vector-valued variable v. In certain fields of physics one encounters "higher order" Legendre transforms, associated with functions of more than one vector-valued variable.

I.1 Convex Sets and Convex Functions

A set $\mathcal{D}$ is said to be convex, if given two points $x, y \in \mathcal{D}$, the interpolating line $tx+(1-t)y \in \mathcal{D}$ for $0 \leq t \leq 1$. A real-valued function $f(v)$ defined for v in a convex set $\mathcal{D}$ is convex, if

$$f(tv + (1-t)w) \leqslant tf(v) + (1-t)f(w) \;, \quad \text{for all} \quad v, w \in \mathcal{D} \;, \text{ and } 0 \leqslant t \leqslant 1 \;. \tag{3}$$

In other words, the graph of the function lies below a straight line interpolating between the value of the function at two given points. One says that f is *strictly convex* if the inequality (3) is strict for all $v \neq w$ and $t \in (0,1)$.

The simplest case occurs when $\mathcal{D} \subset \mathbb{R}$ is a closed interval on the line and the function f is twice differentiable. Then taking $t = 1/2$, multiplying by $(v-w)^{-2}$ and letting $w \to v$ in (3) shows that convexity ensures $f''(v) \geqslant 0$. In fact positivity of the second derivative is also sufficient for f to be convex. The second derivative test gives a straightforward way to determine whether a real-valued function $f(v)$ of one variable is convex.

In the case of a real-valued function of many variables, one needs to generalize the second derivative test to a condition on the matrix of possible second derivatives. If $v \in \mathbb{R}^N$, consider the $N \times N$ real, symmetric matrix $\mathfrak{h}$ with entries

$$\mathfrak{h}_{ij}(v) = \frac{\partial^2 f(v)}{\partial v_i \partial v_j} \;.$$

The matrix $\mathfrak{h}$ is called the Hessian matrix of $f(v)$. This matrix is real and symmetric, so $\mathfrak{h}$ has real eigenvalues. The condition that generalizes $0 \leqslant f''$, is the condition that all the eigenvalues of $\mathfrak{h}$ are non-negative. One says that $\mathfrak{h}$ is "positive" and writes

$$0 \leqslant \mathfrak{h} \;. \tag{4}$$

Equivalently, for any vector $a \in \mathbb{R}^N$, the expectation of $\mathfrak{h}$ in a satisfies, $0 \leqslant \langle a, \mathfrak{h}a \rangle$.[1]

Generally we want this inequality to be strictly positive unless $a = 0$. In that case every eigenvalue of $\mathfrak{h}$ is strictly positive, and the matrix $\mathfrak{h}$ is said to be "positive definite" which one writes as

$$0 < \mathfrak{h} \;, \tag{5}$$

or $0 < \langle a, \mathfrak{h}a \rangle$ unless $a = 0$. Any positive definite matrix has an inverse.

[1] Here we have dropped using the vector sign on a and write $\langle a, b \rangle = \sum_j a_j b_j$ instead of $\vec{a} \cdot \vec{b} = \sum_j a_j b_j$. Both notations are common, and for real vectors the two notations are interchangeable. In the case of complex vectors, one generally defines $\langle a, b \rangle = \sum_j \bar{a}_j b_j$, where $\bar{a}_j$ denotes the complex conjugate of a_j.

II The Legendre Transform and its Dual Domain

We start from a convex domain $\mathfrak{D}$ and a real-valued, convex function $f(v)$ defined for $v \in \mathcal{D}$. Then one defines the Legendre transform $\mathbb{L}f$ of f as (1), namely

$$(\mathbb{L}f)(p) = \sup_{v\in\mathcal{D}} \left(\langle p, v\rangle - f(v)\right) \ . \tag{6}$$

Here $\langle p, v\rangle = \sum_{j=1}^{N} p_j\, v_j$. We can think of the domain $\mathbb{L}\mathcal{D}$ of $\mathbb{L}f$ as a dual to the domain $\mathcal{D}$. This is just the set of p such that (1) is well defined (finite).

One says that $f(v)$ is an even function of v if two things hold:

E1. The domain $\mathcal{D}$ of f is even. This means that $v \in \mathcal{D}$, if and only if $-v \in \mathcal{D}$.

E2. The function f has the property $f(-v) = f(v)$.

In case f is even, then $\mathbb{L}f$ is also even. This follow from

$$\begin{aligned}(\mathbb{L}f)(-p) &= \sup_{v\in\mathcal{D}} \left(\langle -p, v\rangle - f(v)\right) = \sup_{v\in\mathcal{D}} \left(\langle p, -v\rangle - f(-v)\right) \\ &= \sup_{v\in\mathcal{D}} \left(\langle p, v\rangle - f(v)\right) = (\mathbb{L}f)(p) \ .\end{aligned} \tag{7}$$

We claim that $\mathbb{L}\mathcal{D}$ is a convex set and that $\mathbb{L}f$ is a convex function on $\mathbb{L}\mathcal{D}$. Let $p_1, p_2 \in \mathbb{L}\mathcal{D}$ and take $0 \leqslant t \leqslant 1$. Define $p(t) = tp_1 + (1-t)p_2$. Then

$$\langle p(t), v\rangle - f(v) = t\left(\langle p_1, v\rangle - f(v)\right) + (1-t)\left(\langle p_2, v\rangle - f(v)\right) \ .$$

Taking the sup over v then yields

$$\sup_{v\in\mathcal{D}} \left(\langle p(t), v\rangle - f(v)\right) \leqslant t\mathbb{L}f(p_1) + (1-t)\mathbb{L}f(p_2) \ . \tag{8}$$

As the right side is bounded, this shows $(\mathbb{L}f)(p(t))$ is well-defined by the left side of (8). Hence $p(t) \in \mathbb{L}\mathcal{D}$, and $\mathbb{L}\mathcal{D}$ is a convex set. Furthermore (8) shows that

$$(\mathbb{L}f)(p(t)) \leqslant t(\mathbb{L}f)(p_1) + (1-t)(\mathbb{L}f)(p_2) \ , \tag{9}$$

so the Legendre transform $\mathbb{L}f$ is a convex function on $\mathbb{L}\mathcal{D}$.

II.1 An Interesting Inequality

Note that at the point where (1) holds, the function $(\mathbb{L}f)(p)$ is the maximum of the function $\langle p, v\rangle - f(v)$. But at any other point v, one still has the relation

$$\langle p, v\rangle \leqslant f(v) + (\mathbb{L}f)(p) \; . \tag{10}$$

This inequality enters some applications of the Legendre transform.

In the case that f is even, then $\mathbb{L}f$ is also even by (7). Hence

$$\langle -p, v\rangle \leqslant f(v) + (\mathbb{L}f)(-p) = \langle p, -v\rangle \leqslant f(v) + (\mathbb{L}f)(p) \; .$$

Therefore we conclude that if f is even, then one has,

$$|\langle p, v\rangle| \leqslant f(v) + (\mathbb{L}f)(p) \; . \tag{11}$$

II.2 Another Formula for $\mathbb{L}f$

Since the function $(\mathbb{L}f)(p)$ maximizes the expression $\langle p, v\rangle - f(v)$ as a function of v, any point $v = v_0$ for which the maximum occurs is a critical point of $\langle p, v\rangle - f(v)$. In other words,

$$\frac{\partial}{\partial v_j}\left(\langle p, v\rangle - f(v)\right)\Big|_{v=v_0} = 0 \; , \quad j = 1, \ldots, N \; .$$

In case f is strictly convex, then there is a unique point v_0 that maximizes this expression. One simply writes

$$(\mathbb{L}f)(p) = \langle p, v\rangle - f(v) \; , \quad \text{where} \quad p_j = \frac{\partial f(v)}{\partial v_j} \; , \text{ for } \quad j = 1, \ldots, N \; . \tag{12}$$

This is the form of the transformation from the Lagrangian to the Hamiltonian, where we suppress the dependence of both sides on the variables q and possibly t. In this case $f(v) = \mathcal{L}(q, v, t)$, where $\mathcal{L}$ is the Lagrangian and $\mathbb{L}\mathcal{L} = H$, the Hamiltonian.

II.3 Remark about $\mathfrak{h} = 0$

Note that in case the function $f(v)$ is linear, we then have $\frac{\partial^2 f}{\partial v_i \partial v_j} = \mathfrak{h}_{ij} \equiv 0$ for all v. In this case (and likewise in the case that f is piecewise linear) the equations $p_j = \frac{\partial f}{\partial v_j}$ cannot be solved for v_j as a function of p. Thus we cannot use the definition (12) of the Legendre transform. But we can still use the definition (1), and we take that one.

II.4 The Relation $\mathfrak{h}\widetilde{\mathfrak{h}}$ =Id

There is a close relationship between the Hessian matrix $\mathfrak{h}(v)$ of $f(v)$ and the Hessian matrix $\widetilde{\mathfrak{h}}(p)$ of the Legendre transform $(\mathbb{L}f)(p)$ of f. In fact when one evaluates $\widetilde{\mathfrak{h}}(p)$ at the corresponding point $p = \nabla_v f(v)$, these matrices are inverses of one another, namely $\mathfrak{h}(v) = \widetilde{\mathfrak{h}}(p)^{-1}$. In order to see this, notice that

$$\mathfrak{h}_{jk}(v) = \frac{\partial^2 f(v)}{\partial v_j \partial v_k} , \quad \text{and} \quad \widetilde{\mathfrak{h}}_{jk}(p) = \frac{(\partial^2 \mathbb{L} f)(p)}{\partial p_j \partial p_k} . \tag{13}$$

At the point

$$p_j = \frac{\partial f(v)}{\partial v_j} , \tag{14}$$

one thus has

$$\mathfrak{h}_{jk}(v) = \frac{\partial^2 f(v)}{\partial v_j \partial v_k} = \frac{\partial p_j}{\partial v_k} . \tag{15}$$

The relation (15) states that the Hessian $\mathfrak{h}(v)$ of $f(v)$ is the Jacobian matrix of the transformation $v \mapsto p$ at the point (14). It also shows that the Jacobian matrix equals its transpose.

Also at the point (14), one has $(\mathbb{L}f)(p) = \langle p, v\rangle - f(v)$, so

$$\frac{(\partial \mathbb{L} f)(p)}{\partial p_i} = v_i + \sum_j p_j \frac{\partial v_j}{\partial p_i} - \sum_j \frac{\partial f(v)}{\partial v_j} \frac{\partial v_j}{\partial p_i} = v_i . \tag{16}$$

Notice that in the Lagrangian-Hamiltonian case, relation (16) says $\dot{q}_i = \dfrac{\partial H(q,p,t)}{\partial p_i}$. From (16) we infer that

$$\widetilde{\mathfrak{h}}_{ij}(p) = \frac{\partial^2 (\mathcal{L} f)(p)}{\partial p_i \partial p_j} = \frac{\partial v_i}{\partial p_j} . \tag{17}$$

So the relation (17) says that the Hessian $\widetilde{\mathfrak{h}}(p)$ is the Jacobian matrix of the inverse transformation $p \mapsto v$.

Thus at the point (14) the two Hessian matrices $\mathfrak{h}(v)$ and $\widetilde{\mathfrak{h}}(p)$ are the Jacobian matrices of inverse transformations. Therefore the two Hessian matrices are inverses of each other. In other words at the point (14),

$$\left(\widetilde{\mathfrak{h}}\mathfrak{h}\right)_{ik} = \sum_j \widetilde{\mathfrak{h}}_{ij}(p)\, \mathfrak{h}_{jk}(v) = \sum_j \frac{\partial v_i}{\partial p_j}\frac{\partial p_j}{\partial v_k} = \frac{\partial v_i}{\partial v_k} = \delta_{ik} \ . \tag{18}$$

II.5 The Relation $\mathbb{L}^2 f = f$

The key to showing that the Legendre transform squares to the identity is the relation (16). For if this relation holds, then one can evaluate the supremum of the convex function of p, namely

$$\langle v, p\rangle - (\mathbb{L}f)(p) \ , \tag{19}$$

at the point where $v_i = \dfrac{\partial(\mathbb{L}f)(p)}{\partial p_i}$, to obtain the function $(\mathbb{L}^2 f)(v)$. Assuming that we can invert the relationship $v \mapsto p$, the point (16) is the same as the point (14). But at the point (14), one has $(\mathbb{L}f)(p) = \langle p, v\rangle - f(v)$. Thus as $\langle p, v\rangle = \langle v, p\rangle$, we conclude that

$$(\mathbb{L}^2 f)(v) = \sup_p \left(\langle v, p\rangle - ((\mathbb{L}f)(p)\right) = f(v) \ . \tag{20}$$

In the context of classical mechanics, performing the Legendre transform on the variable $\dot{q} \leftrightarrow p$ we have,

$$(\mathbb{L}\mathcal{L})(q, \dot{q}, t) = H(q, p, t) \ , \quad \text{and} \quad (\mathbb{L}H)(q, p, t) = \mathcal{L}(q, \dot{q}, t) \ . \tag{21}$$

III Some Examples

Here are several instances of the Legendre transformation in practice. We give no explanation here, but only suggest that these examples indicate that the Legendre transformation is ubiquitous.

- **Classical Mechanics:** The correspondence between the Lagrangian $\mathcal{L}(q,\dot{q})$ and the Hamiltonian $H(q,p)$.

  Lagrangian Homogeneous 2^{nd}-Order in Velocity. The first case we consider is that the Lagrangian has the form $\mathcal{L} = T - V$ where $T = T(q,\dot{q})$ is homogeneous, quadratic function of the velocity, meaning $T(q,\lambda\dot{q}) = \lambda^2 T(q,\dot{q})$, and $V = V(q)$. Such a T has the form

$$T = \frac{1}{2}\left\langle \dot{q}, M(q)\dot{q}\right\rangle = \frac{1}{2}\sum_{i,j=1}^{n} \dot{q}_i M(q)_{ij}\dot{q}_j \ . \tag{22}$$

  Here $M = M(q)$ is an $n \times n$ symmetric matrix with entries $M_{ij} = M(q)_{ij}$, called the *mass matrix*. In the case $n = 1$ this reduces to $T = \frac{1}{2}m\dot{q}^2$, with $m = M(q)_{11}$. It is no loss of generality to assume that M is a symmetric matrix, as the kinetic energy expression is a quadratic function of $\dot{q}$, and hence any skew-symmetric M would contribute 0 to T. As any matrix M can be written as a sum of symmetric and skew-symmetric parts, $M = \frac{1}{2}\left(M + M^{\text{tr}}\right) + \frac{1}{2}\left(M - M^{\text{tr}}\right)$, where M^{tr} is the transpose of M. Thus we can always replace M by its symmetric part, without changing T. The Lagrangian we are considering is then

$$\mathcal{L}(q,\dot{q}) = \frac{1}{2}\left\langle \dot{q}, M(q)\dot{q}\right\rangle - V(q) \ . \tag{23}$$

  The Hessian of $\mathcal{L}$ with respect to the velocity is

$$\mathfrak{h} = M(q) \ , \tag{24}$$

  so the requirement that $\mathfrak{h}$ is positive is the assumption that M is positive. We also assume that the matrix M is invertible (no zero eigenvalues) and has inverse M^{-1}.

  In this case the canonical momentum vector is $p = M\dot{q}$ or $\dot{q} = M^{-1}p$. Then we can express T as a function of p, namely

$$T = \frac{1}{2}\left\langle p, M^{-1}p\right\rangle \ , \tag{25}$$

  and

$$\begin{aligned} H(q,p) &= (\mathbb{L}\mathcal{L})(q,p) = \langle p,\dot{q}\rangle - T + V = \left\langle p, M^{-1}p\right\rangle - T + V = 2T - T + V \\ &= \frac{1}{2}\left\langle p, M(q)^{-1}p\right\rangle + V(q) \ . \end{aligned} \tag{26}$$

Lagrangian Quadratic in Velocity. The difference in this case from the previous one is that we allow a linear term in the kinetic energy, rather than it being strictly homogeneous, of degree 2. In particular let us now consider replacing (23) by

$$\mathcal{L}(q,\dot{q}) = \frac{1}{2}\langle \dot{q}, M(q)\dot{q}\rangle + \langle K(q),\dot{q}\rangle - V(q)\;, \tag{27}$$

where $K = K(q)$ is a vector function of q but not $\dot{q}$.

Since the new term in linear in the velocities, the Hessian is unchanged, and the convexity property corresponds to positive M. For the canonical momentum

$$p = M\dot{q} + K\;, \quad \text{so} \quad \dot{q} = M^{-1}\,(p-K)\;. \tag{28}$$

Then $\langle p,\dot{q}\rangle = \langle p, M^{-1}\,(p-K)\rangle$, and $\langle K,\dot{q}\rangle = \langle K, M^{-1}\,(p-K)\rangle$, so the Hamiltonian becomes

$$\begin{aligned} H &= \mathbb{L}\mathcal{L} = \langle p,\dot{q}\rangle - \mathcal{L} \\ &= \big\langle p, M^{-1}\,(p-K)\big\rangle - \frac{1}{2}\big\langle M^{-1}\,(p-K), M(q)M^{-1}\,(p-K)\big\rangle \\ &\qquad - \big\langle K, M^{-1}\,(p-K)\big\rangle + V(q) \\ &= \frac{1}{2}\big\langle (p-K), M^{-1}(p-K)\big\rangle + V(q)\;. \end{aligned} \tag{29}$$

Particle in an Electromagnetic Field. The Lagrangian (27) applies to a single particle with coordinate $q = x$ (a 3-vector), velocity $v = \dot{x}$ (another 3-vector), and charge q moving in an electromagnetic field. We suppose that the electric field E and the magnetic field B both depend on space x and time t, and that they are given according to Maxwell's equations by a vector potential $A(x,t)$ and a scalar potential $\Phi(x,t)$. (Here we use units consistent with Goldstein, Poole, and Safko. Other units introduce factors of c.) Then the identification $K = qA$ and $V = q\Phi$ lead to the Lagrangian of the form (27),

$$\mathcal{L} = \frac{1}{2}m\,\langle v,v\rangle + q\,\langle A,v\rangle - q\Phi\;. \tag{30}$$

The corresponding Hamiltonian of the form (29) is

$$H = \frac{1}{2m}\,\langle (p-qA),(p-qA)\rangle + q\Phi\;. \tag{31}$$

This form of the interaction between a charge and an electromagnetic potential is called "minimal coupling."

One can check that his leads to the equation of motion being given by the Lorentz force $F = q(E + v \times B)$, namely

$$m\ddot{x} = q(E + v \times B) \ . \tag{32}$$

Here E and B are related to the potential by

$$B = \nabla \times A \ , \quad E = -\nabla\Phi - \frac{\partial A}{\partial t} \ . \tag{33}$$

Free Relativistic Particle. As a special case, let us work out the properties of the Legendre transform for the Lagrangian of a single relativistic particle of rest mass m, moving freely (without the action of forces) according to the physical laws of special relativity. Let us take the motion to be along one axis of the Cartesian coordinate system, so we use the notation $q = x$ and $\dot{q} = v$. Then the Lagrangian is

$$\mathcal{L}(v) = -mc^2\sqrt{1 - \left(\frac{v}{c}\right)^2} \ , \tag{34}$$

with the domain $\mathcal{D} = (-c, c)$ for v. This Lagrangian is even.

The canonical momentum corresponding to $\mathcal{L}$ is

$$p = \frac{\partial \mathcal{L}}{\partial v} = \frac{mv}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \ . \tag{35}$$

Sometimes one defines a dynamical mass M in terms of the rest mass m as

$$M = \frac{m}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \ . \tag{36}$$

In this case, one could write

$$p = Mv \ . \tag{37}$$

The Hessian matrix $\mathfrak{h}$ of $\mathcal{L}$ is one-dimensional; it equals the second derivative

$$\mathfrak{h} = \frac{\partial^2 \mathcal{L}}{\partial v^2} = \frac{m}{\left(1 - \left(\frac{v}{c}\right)^2\right)^{3/2}} > 0 \ . \tag{38}$$

So the Lagrangian $\mathcal{L}(v)$ is convex.

Now let us consider the Legendre transform of this Lagrangian. The relation (35) and the definition of $\mathcal{L}(v)$ show that

$$pv - \mathcal{L}(v) = \frac{mv^2}{\sqrt{1-\left(\frac{v}{c}\right)^2}} + mc^2\sqrt{1-\left(\frac{v}{c}\right)^2} = \frac{mv^2 + mc^2\left(1-\left(\frac{v}{c}\right)^2\right)}{\sqrt{1-\left(\frac{v}{c}\right)^2}} = \frac{mc^2}{\sqrt{1-\left(\frac{v}{c}\right)^2}} \ . \tag{39}$$

Using (35), we also see that

$$\frac{1}{1-\left(\frac{v}{c}\right)^2} = 1 + \frac{\left(\frac{v}{c}\right)^2}{1-\left(\frac{v}{c}\right)^2} = 1 + \left(\frac{p}{mc}\right)^2 \ . \tag{40}$$

Thus (39) can be rewritten as a function purely of p, namely

$$pv - \mathcal{L}(v) = mc^2\sqrt{1+\left(\frac{p}{mc}\right)^2} = \sqrt{(pc)^2 + (mc^2)^2} \ . \tag{41}$$

Thus we see that the Legendre transform of the Lagrangian $\mathcal{L}(v)$ given by (34) is the Hamiltonian (energy)[2]

$$H(p) = (\mathbb{L}\mathcal{L})(p) = \sqrt{(pc)^2 + (mc^2)^2} \ . \tag{43}$$

One can see that these expressions are also convex functions of p. The Hessian $\mathfrak{h}$ of $H(p)$ is

$$\frac{\partial^2 H(p)}{\partial p^2} = \frac{m^2c^6}{\left(p^2c^2 + m^2c^4\right)^{3/2}} > 0 \ . \tag{44}$$

Our discussion demonstrates that the Lagrangian $\mathcal{L}(v)$ in (34) leads to the famous Einstein relation that the energy H and momentum p of a free particle lie on the mass hyperboloid, $H^2 - (pc)^2 = m^2c^4$.

We could also write (43) in terms of the dynamical mass M of (36) instead of the rest mass m, by using the middle expression in (41) with (40). This gives another famous Einstein formula for the energy E, namely

$$H = E = Mc^2 \ . \tag{45}$$

[2]In situations where one uses the units corresponding to $c = 1$, one writes

$$H(p) = \sqrt{p^2 + m^2} \ . \tag{42}$$

- **Thermodynamics:** The correspondence between the Gibbs free energy and the Helmholtz free energy. In thermodynamics, one defines the Helmholz free energy $F(T, V)$ as a function of temperature T and volume V. On the other hand, the Gibbs free energy $G(T, P)$ depends on temperature T and pressure P. The relationship between the two is a Legendre transformation,

$$G(T, P) = PV + F(T, V) . \tag{46}$$

One generally eliminates V on the right side, by solving for $V = V(P, T)$.

- **Mathematics, Interval and Angle:** An elementary example is the transform of a straight-line interval function $f(v)$. Let $0 \leq t \leq 1$, and $v = (1-t)v_1 + tv_2$ be a point in the interval $[v_1, v_2]$. Define $f(v) = (1-t)f_1 + tf_2$, where $f_1 = f(v_1)$ and $f_2 = f(v_2)$. Let us assume that $f_1 < f_2$. Then $f(v)$ is a straight line defined on the interval $[v_1, v_2]$ and going from f_1 to f_2.

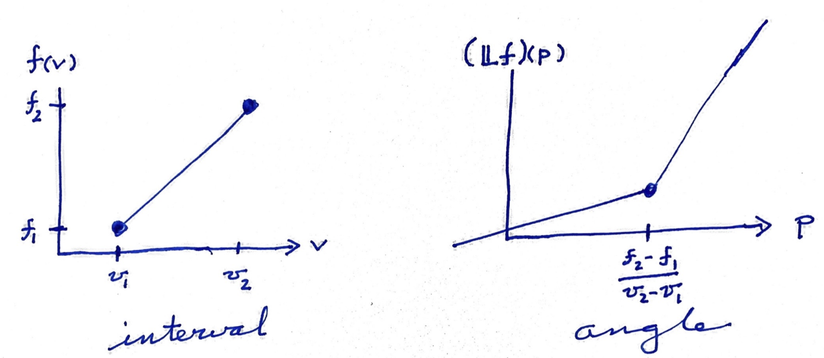

Figure 1: The Legrendre Transform of an Interval f is an Angle $\mathbb{L}f$.

Let us see why the Legendre transform of this interval is an angle. With the definitions above, the Legendre transform of f is

$$(\mathbb{L}f)(p) = \sup_{v \in [v_1, v_2]} (pv - f(v)) = \sup_{0 \leq t \leq 1} ((1-t)(pv_1 - f_1) + t(pv_2 - f_2)) . \tag{47}$$

Suppose p is such that $pv_1 - f_1 < pv_2 - f_2$, namely $p(v_1 - v_2) < f_1 - f_2$ or $p > \frac{f_2 - f_1}{v_2 - v_1}$, the slope of the function f. Then the sup over t is reached by taking the larger term, namely setting $t = 1$ and $(\mathbb{L}f)(p) = pv_2 - f_2$, a straight line with slope v_2. On the other hand, if $p < \frac{f_2 - f_1}{v_2 - v_1}$, the slope of the function f, then the sup is obtained by taking $t = 0$ and $(\mathbb{L}f)(p) = pv_1 - f_1$, a straight line with slope v_1. The transition takes place at the point $p = \frac{f_2 - f_1}{v_2 - v_1}$.

Thus the Legendre transform of an interval is an angle, with the slope of the interval becoming the location of the vertex of the angle. The endpoints of the interval correspond to the slopes of the lines forming the angle. In particular, the two slopes of the lines forming the angle are $v_1 = \tan\theta_1$ and $v_2 = \tan\theta_2$ respectively. Thus the change in angle between the two lines of the Legendre transform $\mathbb{L}f$ is $\theta_2 - \theta_1 = \arctan v_2 - \arctan v_2$, and the angle itself is $\theta = \pi - (\theta_2 - \theta_1)$; see Figure 2.

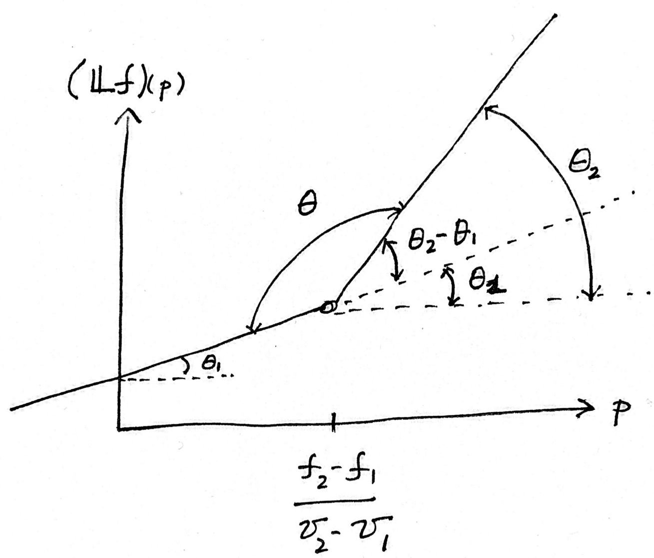

Figure 2: Details of the Angle $\mathbb{L}f$.

If one repeats the Legendre transform, one sees that the Legendre transform of an angle is an interval.

One can modify the original interval f, replacing it by a convex sequence of intervals; then one obtains as the Legendre transform a convex sequence of angles. In the limit of an infinite number of transitions, closer and closer together, one sees that the Legendre transform of a convex function is a convex function.

- **Mathematics, Hölder's inequality:** Suppose $1 < \alpha, \beta$ are related by $\alpha^{-1}+\beta^{-1} = 1$. For $\|f\|_\alpha = \left(\int |f|^\alpha dx\right)^{1/\alpha}$, Hölder's inequality states that

$$\|fg\|_1 \leqslant \|f\|_\alpha \|g\|_\beta \ . \tag{48}$$

The Schwarz inequality is the case $\alpha = \beta = 2$.

Suppose $0 \leqslant v, p$ are positive numbers, and define $f(v) = \frac{1}{\alpha}v^\alpha$. Then $(\mathbb{L}f)(p) = \frac{1}{\beta}p^\beta$, and the Legendre transform inequality (10) says that

$$vp \leqslant \frac{1}{\alpha}v^\alpha + \frac{1}{\beta}p^\beta \ . \tag{49}$$

For $\alpha = \beta = 2$, this says that the geometric mean is bounded by the arithmetic mean. Replace v and p by their positive square root, so

$$\sqrt{vp} \leqslant \frac{1}{2}(v+p) \ .$$

To obtain (48), replace $f(x)$ by $f(x)/\|f\|_\alpha$ and $g(x)$ by $g(x)/\|g\|_\beta$. Thus one sees that it is sufficient to establish (48) in case that $\|f\|_\alpha = \|g\|_\beta = 1$, namely when the right side of (48) is 1. Note that (49) ensures that

$$|f((x)g(x)| \leqslant \frac{1}{\alpha}|f(x)|^\alpha + \frac{1}{\beta}|g(x)|^\beta \ .$$

Integrating this inequality over x and using $\|f\|_\alpha = \|g\|_\beta = 1$, yields

$$\|fg\|_1 \leqslant \frac{1}{\alpha}\|f\|_\alpha^\alpha + \frac{1}{\beta}\|g\|_\beta^\beta = \frac{1}{\alpha} + \frac{1}{\beta} = 1 \ .$$

- **Statistical Mechanics and Quantum Fields:** The relation between connected Feynman diagrams and "one-particle irreducible" diagrams.