1 Warning

I'm going to use only a physicist's level of rigour in the following. I'm also going to assume a little familiarity with the Legendre transform and the dual of an optimisation problem.

2 Two views of convex regions

There are two complementary ways we can look at convex regions: in terms of the points at the boundary, or as an intersection of half planes. The Legendre transform allows us to move back and forth between these dual representations in the case when the convex region is the *epigraph* of a function, the region above its graph. (Note that we can think of a half plane as defined by a linear function on a space. Points in the half plane are those for which the liner function evaluates to zero or less. These linear functions are *duals* in the sense of linear algebra. So the dual of convex analysis is related to that of linear algebra, despite frequent claims to the contrary.)

3 The Legendre and Fourier transforms

The Fourier transform also allows us to move back and forth between two representations of functions. One is as a sum of Dirac delta functions, ie.

$$f(x) = \int_{-\infty}^{\infty} dx f(a)\delta(x-a)$$

and one is as a sum of plane waves

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \hat{f}(\omega) \exp(i\omega x)$$

The Legendre transform allows us to move between two representations in a similar way. We can define a "convex" analogue of the Dirac delta function

$$\sigma(x) = \begin{cases} 0 & x = 0 \\ \infty & x \neq 0 \end{cases}$$

so we have

$$f(x) = inf_a f(a)\sigma(x-a)$$

This allows us to think of functions as infima of infinite collections of downward spikes, similar to the above picture of a function as a sum of infinitely many upward Dirac spikes. The Legendre transform f^* gives a function f as the supremum of infinitely many straight lines:

$$f(x) = \sup_{\lambda} (\lambda x - f^*(\lambda))$$

In the case of the Fourier representation, we have a sum of plane waves, each scaled by a possibly different amount. In the Legendre representation we have the sup of linear functions, each one shifted up or down by a possibly different amount.

There is a dictionary that allows us to translate between the language of Legendre transforms and the language of Fourier transforms. Here is part of the dictionary:
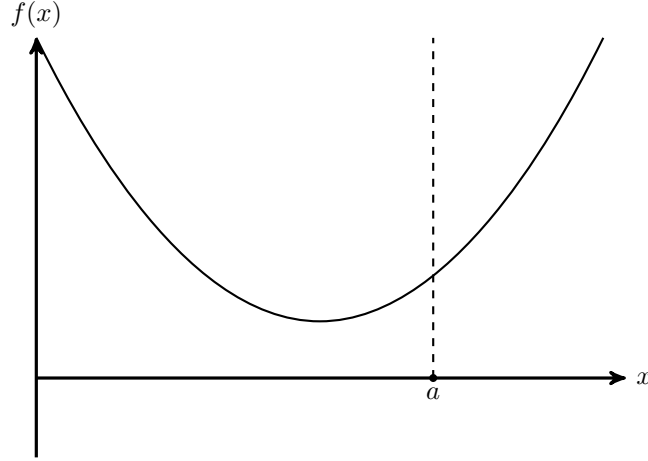
Figure 1: A convex function

Binary operations	$+, \cdot, 0, 1$	$\min, +, \infty, 0$
Integration	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}$	inf
"Dirac" form	$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(p)\delta(x-p)dx$	$f(x) = \inf_p f(p)\sigma(x-p)$
Diagonalisation	$\exp(i\omega(x+a)) = \exp(i\omega a)\exp(i\omega x)$	$p(x+a) = pa + px$
	$g(x) = f(x+a)$	$g(x) = f(x+a)$
	$G(\omega) = \exp(i\omega a)F(\omega)$	$g^*(p) = pa + f^*(p)$
Fourier-Legendre transform	$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)\exp(-i\omega x)dx$	$f^*(p) = \sup_x(xp - f(x))$
Inverse transform	$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega)\exp(i\omega x)d\omega$	$f(x) = \sup_p(px - f^*(p))$
Scaling	$g(x) = f(x+a)$	$g(x) = f(x+a)$
	$G(\omega) = \frac{1}{|a|}G(\omega/a)$	$g^*(p) = f^*(p/a)$
Unitarity/Fenchel duality	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \bar{F}(\omega)G(\omega)d\omega = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \bar{f}(x)g(x)dx$	$\inf_x(f(x)+g(x)) = \inf_p(f^*(p)+g^*(-p))$
Plancherel's theorem	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |F(\omega)|^2 \, d\omega = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |f(x)|^2 \, dx$	$2\inf_x f(x) = \inf_p(f^*(p)+f^*(-p))$
Gaussian/quadratic	$f(x) = \exp(-ax^2/2)$	$f(x) = ax^2/2$
	$F(\omega) = \frac{1}{\sqrt{a}}\exp(-\omega^2/2a)$	$f^*(p) = p^2/2a$
(Inf-)convolution	$h(x) = (f \otimes g)(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y)g(x-y)dy$	$h(x) = (f \oplus g)(x) = \inf_y f(y)g(x-y)$
	$H(\omega) = F(\omega)G(\omega)$	$h^*(p) = f^*(p) + g^*(p)$
	Quantum mechanics	Classical mechanics
	Feynman path integral	Hamilton's principle of least action

my goal here is to add a couple of new entries that arise when constrained optimisation is considered.

4   An elementary constrained optimisation
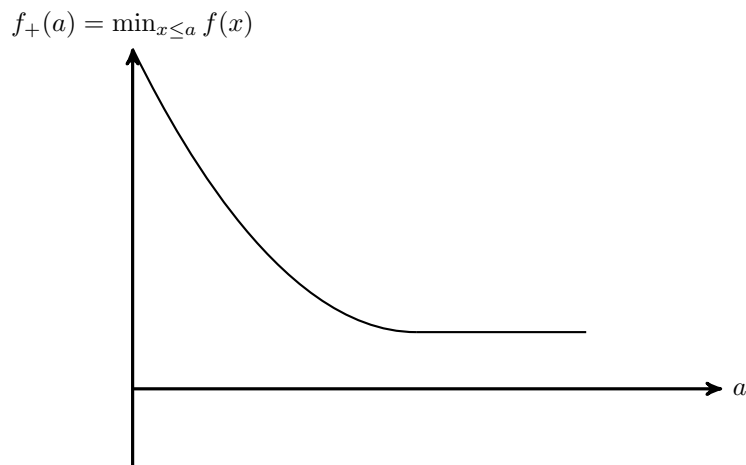
Consider the constrained optimisation problem

$$f_+(a) = \min_{x \le a} f(x)$$

Figure 2: The constrained minimum of a convex function

$$\begin{aligned} \text{minimise} \quad & f(x) \\ \text{subject to} \quad & x \le a \end{aligned}$$

We can form the Lagrangian in the usual way (for example see [5]):

$$L(a, x, \lambda) = f(x) + \lambda(x - a) \text{ defined for } \lambda \ge 0.$$

Minimising this with respect to x gives us a new function:

$$g(a, \lambda) = \inf_x \left(f(x) + \lambda(x - a) \right)$$

which gives us the dual optimisation problem:

$$\begin{aligned} \text{maximise} \quad & g(a, \lambda) \\ \text{subject to} \quad & \lambda \ge 0 \end{aligned}$$

In this case, strong duality holds (see [4]) so both of these problems achieve the same optimal value.

Now consider the optimal value as a function of a, $f_+(a) = \min_{\lambda \ge 0} g(a, \lambda)$. We plot typical examples of $f(x)$ and $f_+(a)$ in Figures 1 and 2.

We can look at these graphs another way. The function f_+ is a version of f where all the parts with positive gradient have been "removed", keeping only the non-increasing part. Think about this in terms of the *epigraph* of f, the set of points above the curve. The dual view of the epigraph, illustrated in Figure 3, is that it is the intersection of an infinite collection of half-planes, each one just touching f. To construct the epigraph of f_+ we take only those half-planes whose boundaries have non-positive gradient. This is illustrated in Figure 4. This is precisely what the construction of the dual problem achieves. It forms the Legendre transform of the original function. But instead of reconstructing the original function using the full Legendre transform, instead it reconstructs it using only lines for which $\lambda \ge 0$.

We could have defined f_+ directly using the primal problem but I chose to go via the dual path to investigate what happens when we translate this to the Fourier domain.
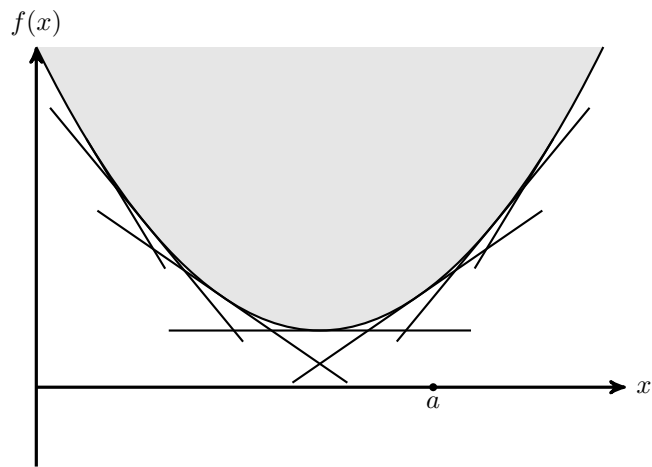
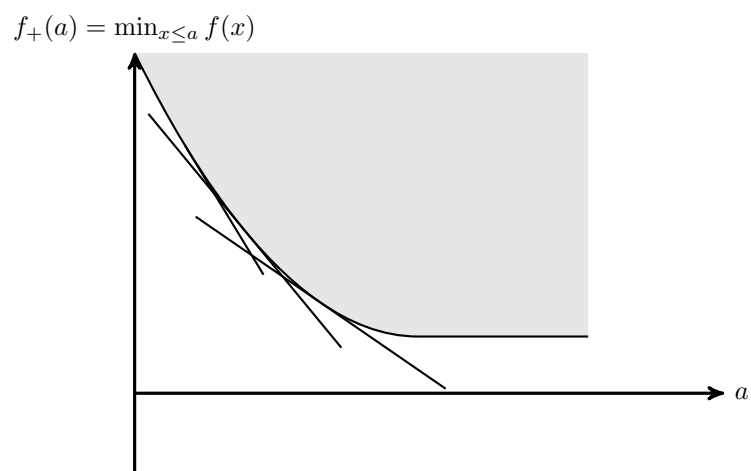Figure 3: The epigraph of a convex function as the intersection of half-planes

Figure 4: The epigraph of the constrained minimum

4

We can also define $f_-(x) = \min_{x \geq a} f(x)$ (or similarly through a dual problem). Our original function f is now given by

$$f(x) = \max(f_+(x), f_-(x))$$

We've "projected" all of the descending bits into one function and all of the ascending bits into another.

5 The same construction in the Fourier domain

Using the dictionary, the analogue of f_+ is given by

$$g(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, f(x) \exp(-\lambda(x-a))$$

The dual optimisation problem is now translated into

$$I(a) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} d\lambda \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-\lambda(x-a))$$

Let's swap the order of those integrals so

$$I(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{2\pi}} \int_0^{\infty} d\lambda \exp(-\lambda(x-a))$$

The λ integral is essentially the Fourier transform of the Heaviside theta function [1]. If

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

then

$$\hat{\Theta}(\lambda) = \sqrt{\frac{\pi}{2}} \delta(\lambda) - \frac{i}{\sqrt{2\pi}(x-a)}$$

That expression is to be interpreted as a distribution so it needs to be evaluated in an integral. The $\frac{1}{\lambda}$ requires the principal part of the integral. We get

$$I(a) = \frac{1}{2} f(x) + \frac{i}{2\pi} P \int_{-\infty}^{\infty} \frac{f(x) dx}{a-x}$$

with $P\int$ being the principal part of the integral. We can write this as

$$I(a) = \frac{1}{2} f(x) + i\frac{1}{2} H f(x)$$

where H is the Hilbert transform [2]. The Hilbert transform is used to solve a certain type of Riemann-Hilbert [3] problem. This is the problem of taking a function f on the real line and writing it as

$$f(x) = F_+(x) - F_-(x)$$

with F_+ holomorphic (ie. analytic with no poles) on the upper half plane (UHP) and F_- holomorphic on the lower half plane (LHP). In fact

$$F_+(x) = \frac{1}{2} f(x) + \frac{i}{2} H f(x)$$

Compare with the end of Section 4. Both versions allow us to "project" functions onto a pair of subspaces. In the first case it was non-increasing and non-decreasing functions. In the second case it was functions without poles in the upper half plane and functions without poles in the lower half plane. That suggests adding these two rows to the dictionary:

Non-increasing convex functions	functions whose analytic continuation to the UHP has no poles
Non-decreasing convex functions	functions whose analytic continuation to the LHP has no poles

6 Notes

I plan to finish this document one day. When I do, here are some things I'll mention:

The analogy quantum-classical mechanics isn't quite perfect. If the Lagrangian is quadratic then Hamilton's principle is a convex optimisation with a unique local minimum. However, more general physics problems aren't convex and we need to consider all stationary solutions, minima and saddle points included. The analogy works slightly differently in these cases.

Laplace transforms are closely related to Fourier transforms and this means we also have a Laplace-Legendre dictionary too. This is why the Legendre transform arises in thermodynamics.

There are some "operations" that take us between the two domains. For example Maslov dequentization. And the stationary phase approximation.

I learnt about this stuff from John Baez but haven't decided on the best reference to give.

7 Summary

Forming the dual to an optimisation constrained by an inequality, and computing the Hilbert transform, are almost the related operations. In one case we form the Legendre transform, throw away "half" of the result and then invert the Legendre transform. In the other we do something similar with the Fourier transform.

References

[1] Heaviside step function. `http://en.wikipedia.org/wiki/Heaviside_step_function#Fourier_transform`. Accessed: 2014-01-14.

[2] Hilbert transform. `http://en.wikipedia.org/wiki/Hilbert_transform`. Accessed: 2014-01-14.

[3] Riemann-hilbert problem. `http://en.wikipedia.org/wiki/Riemann-Hilbert_problem`. Accessed: 2014-01-14.

[4] Strong duality. `http://en.wikipedia.org/wiki/Strong_duality`. Accessed: 2014-01-14.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.