Calculus of Variations in a Thimble

Consider the problem of minimizing the functional

$$v[y(x)] = \int_{x_0}^{x_1} F(x, y(x), y'(x), \ldots, y^{(k)}(x))dx$$

we would like to find a function $y(x)$ to solve this problem. To do this we consider a one-parameter family of functions

$$y(x, \alpha) = y(x) + \alpha\delta y(x)$$

where δy denotes some perturbation of the function $y(k)$ and α is a scalar denoting the magnitude of the perturbation. Obviously,

$$y(x, 0) = y(x)$$

so if we limit our optimization only to curves of the form $y(x, \alpha)$, then, we would require,

$$\frac{d}{d\alpha} v[y(x, \alpha)] = 0$$

otherwise we could improve by moving y in the direction δy. Thus,

$$\delta v = \int_{x_0}^{x_1} (F_y \delta y + F_{y'} \delta y' + \ldots + F_{y^{(k)}} \delta y^{(k)})dx = 0$$

Integrate the second summand,

$$\int F_{y'} \delta y' dx = F_{y'} \delta y|_{x_0}^{x_1} - \int \frac{d}{dx} F_{y'} \delta y$$

Similarly integrating twice

$$\int F_{y''} \delta y'' dx = F_{y''} \delta y'|_{x_0}^{x_1} - \frac{d}{dx} F_{y''} \delta y + \int \frac{d^2}{dx^2} F_{y''} \delta y dx$$

and so forth, so the k^{th} term is,

$$\int F_{y^{(k)}} \delta y^{(k)} dx = F_{y^k} \delta y^{k-1}|_{x_0}^{x_1} - \frac{d}{dx} F_{y^{(k)}} \delta y^{(k-2)}|_{x_0}^{x_1}$$

$$\ldots + (-1)^n \int \frac{d^k}{dx^k} F_{y^{(k)}} \delta y dx.$$

If we insist that the boundary conditions be satisfied for $y(x)$, then it follows that $\delta y = \delta y' = \delta y'' = \ldots = \delta y^{k-1} = 0$ at both x_0 and x_1, so this drastically simplifies to,

$$\delta v = \int_{x_0}^{x_1} (F_y - \frac{d}{dx} F_{y'} + \frac{d^2}{dx^2} F_{y''} + \ldots + (-1)^k \frac{dk}{dx^k} F_{y^{(k)}})\delta y dx$$

But δy is arbitrary, so the only way that this can hold is if the factor in parentheses is identically zero. So we require

$$F_y - \frac{d}{dx}F_{y'} + \frac{d^2}{dx^2} + \frac{d^2}{dx^2}F_{y''} + \ldots + (-1)^n \frac{d^k}{dx^k}F_{y^{(k)}} = 0.$$

This is the Euler or Euler-Poisson equation. It plays the same role as the usual first-order conditions for finite dimension optimization. In effect it is a (not-so-obvious) generalization of the usual requirement that directional derivatives should be zero in all directions, provided that the objective function is suitably smooth. In non-smooth cases then we need more sophisticated optimality criteria, obviously.

Example 0 In the Epanechnikov case,

$$\min \int K^2(x)dx \quad \text{s.t.} \quad \int K(x) = 1, \int xK(x) = 0, \int x^2K(x) = 1$$

or

$$\min \int (K^2 + \mu_0 K + \mu_1 xK + \mu_2 x^2 K)dx$$

The Euler condition is simple since there are no derivatives,

$$\begin{aligned} & 2K + \mu_0 + \mu_1 x + \mu_2 x^2 = 0 \\ \Rightarrow \quad & K(x) = a + bx + cx^2 \end{aligned}$$

and now we need to find a, b, c to satisfy our constraints. Note that there is, as we said in class, some technical issues about the bounded support of the K solution.

Example 1

$$v[y] = \int_0^1 (1 + (y''(x))^2)dx \quad \text{s.t.} \quad y(0), y'(0) = 1, y(1) = 1, y'(1) = 1$$

The Euler condition is

$$\frac{d^2}{dx^2}2y''(x) = 0 \qquad \text{or} \qquad y^{(4)} = 0$$

so this implies that $y(x)$ is cubic and the boundary conditions require that $y(x) = x^3$.

Example 2

$$v[y] = \int_0^{\pi/2} ((y'')^2 - y^2 + x^2)dx \text{ s.t. } y(0) = 1, y'(0) = 0, y(\pi/2) = 0, y'(\pi/2) = -1,$$

which has Euler condition,

$$y^{(4)} - y = 0$$

which has the general solution,

$$y = C_1 e^x + C_2 e^{-x} + C_3 \cos x + C_4 \sin x$$

but when one enforces the boundary conditions it can be seen that $C_1 = C_2 = C_4 = 0$, so the solution is simply, $y = \cos x$.

Example 3 A more serious statistical example yields Huber's (1964) M-estimator. Huber posed the following problem: suppose we have iid observations from the contaminated distribution,

$$F(x - \theta) = (1 - \epsilon)\Phi(x - \theta) + \epsilon H(x - \theta)$$

where Φ is the standard normal df, and H is an arbitrary, symmetric distribution. We would like to estimate θ, the center of symmetry, what is the least favorable H? That is, what is the H that makes it most difficult? This is a decidedly non-trivial problem, and Huber's solution came as something of a surprise to the statistical community. It might be expected that the least favorable distribution would be very heavy tailed, but Huber's solution is only mildly heavy tailed.

It is difficult to find an elementary derivation of Huber's solution, the development in Huber (1981) makes very clever use of Cauchy-Schwartz, but is rather opaque, at least from my perspective. The following argument, suggested to me by Jiaying Gu, provides the essential insights.

Let f denote the density of the contaminated distribution F. We would like to minimize the Fisher information for location over possible f's, that is, to solve,

$$\min\{I(f) \mid \int f = 1\}$$

or in the formalism of the prior development, to minimize,

$$v[f] = \int ((f'/f)^2 f + \lambda f) dx.$$

The Euler condition is thus,

$$-\left(\frac{f'}{f}\right)^2 + \lambda - \frac{d}{dx}\frac{2f'}{f} = -\left(\frac{f'}{f}\right)^2 + \lambda - 2\frac{f''f - (f')^2}{f^2} = 0.$$

Now, note that this differential equation is solved for $f(x) = \frac{1}{2}e^{-|x|}$, or any scale dilation of this f, with $\lambda = 1$, or another scale constant, if f has been rescaled. Thus, the least favorable model is the double exponential (Laplace) model. Note that this f has heavier tails than the Gaussian, like e^{-x} rather than $e^{-x^2/2}$, but much lighter than the usual algebraic (Pareto) tails we might have expected.

But we are not done yet since the Laplace model is not of the require contamination form. How are we to find an H, or its corresponding density h to satisfy this requirement? Since $h(x)$ must be non-negative for all x, we have the additional constraint that, for all x,

$$f(x) \geq (1 - \epsilon)\phi(x)$$

so it is now "clear" that all the mass of h should concentrate in the tails of f, and it is just a matter of solving for Huber's mysterious k where the Gaussian and Laplacian pieces join together. This is a matter of carefully adjusting k so we get the "right" amount of contamination. The resulting least favorable density has exponential tails beyond $\pm k$ and satisfies,

$$f(x) = (1 - \epsilon)\phi(x)$$

on the interval $[-k, k]$. The required h puts no mass on this inner interval and just enough in the outer intervals to raise the tails from Gaussian to Laplacian.