2.2. Nondynamic Function Optimization

©Erik I. Verriest*

November 19, 2008

1 The Unconstrained Problem

Consider the problem of optimizing $J = \int_0^T F(u(t), t)\, dt$ over the class of smooth functions. A necessary condition for an extremum is that for all variations δu we have

$$\delta J = \int_0^T \frac{\partial F}{\partial u}\, \delta u\, dt = 0. \tag{1}$$

But then, since δu is completely arbitrary, it must hold that

$$\frac{\partial F}{\partial u} = 0. \tag{2}$$

Example: Consider the functional

$$J = \int_{\pi/3}^{\pi/2} [\cos t\, u(t) - \sin t\, u^2(t)]\, dt$$

We find

$$\frac{\partial}{\partial u}[\cos t\, u(t) - \sin t\, u^2(t)] = \cos t - 2 \sin t\, u(t)$$

and thus

$$u_o(t) = \frac{1}{2} \cot t.$$

Thus giving

$$J_o = \int_{\pi/3}^{\pi/2} \left[\frac{\cos^2 t}{2 \sin t} - \frac{\cos^2 t}{4 \sin t} \right] dt = \frac{1}{4} \int_{\pi/3}^{\pi/2} \frac{\cos^2 t}{\sin t}\, dt = \frac{1}{2}[\log 3 - 1].$$

This stationary solution is a maximum since

$$J = \int_0^T \left. \frac{\partial^2 F}{\partial u^2} \right|_o (\delta u)^2\, dt = \int_{\pi/3}^{\pi/2} [-2 \sin t]\, dt = -1 \leq 0.$$

*School of ECE, Georgia Tech, Atlanta, GA, USA. Presently with SCD, ESAT, KULeuven.

2 The Constrained Problem

In this section we add a number of constraints to the problem. First let us require that the optimizing function should satisfy the constraint

$$\int_0^T G(u(t),t)\,dt = 0. \tag{3}$$

At a stationary point we must still have (1), but now this does not imply (2) anymore since the variation δu cannot be taken arbitrary, but must be consistent with the constraint (3). This means that the variation has to satisfy

$$\int_0^T \left.\frac{\partial G}{\partial u}\right|_o \delta u\,dt = 0. \tag{4}$$

Note that this is an orthogonality condition in the function space. Thus we can imbed all such admissible variations in the class of *arbitrary* variations as follows Take any arbitrary variation, δv, but subtract the component in line with $g(t) = \left.\frac{\partial G}{\partial u}\right|_o$. This gives an admissible δu

$$\delta u = \delta v - \nu g, \tag{5}$$

where the constant ν is chosen to satisfy the orthogonality condition, i.e.

$$\int_0^T g[\delta v - \nu g]\,dt = 0, \tag{6}$$

or

$$\nu = \frac{\int_0^T g\delta v\,dt}{\int_0^T g^2\,dt}. \tag{7}$$

Let's now *choose* a variation $\delta v = \epsilon f$, where $f = \left.\frac{\partial F}{\partial u}\right|_o$. We find

$$\delta J = \int_0^T f\left(\epsilon f - \epsilon g\frac{\int_0^T fg\,dt}{\int_0^T g^2\,dt}\right)\,dt \tag{8}$$

Reorganizing terms, we find

$$\delta J = \epsilon\left[\int_0^T f^2\,dt - \frac{(\int_0^T fg\,dt)^2}{\int_0^T g^2\,dt}\right] \tag{9}$$

Recall Schwarz's inequality. If f and g are not proportional, the term between brackets is nonzero. Hence according to its sign, we can choose ϵ positive or negative to decrease (or increase as the case might be) the value of J. But this contradicts that u_0 was an extremizing solution. Consequently, the coefficient of δv must have been zero, i.e., at the extremizing solution:

$$\int_0^T \left[f\,\delta v - \frac{\int_0^T g\delta v\,dt}{\int_0^T g^2\,dt}g\right]\,dt = 0 \tag{10}$$

or, rearranged,

$$\int_0^T \left[f - \frac{\int_0^T fg\,dt}{\int_0^T g^2\,dt} g \right] \delta v\,dt = 0 \tag{11}$$

implies

$$f - \frac{\int_0^T fg\,dt}{\int_0^T g^2\,dt} g = 0 \tag{12}$$

Formally, we define a Hamiltonian function (for all t)

$$H(u(t), \nu, t) = F(u(t), t) + \nu\, G(u(t), t) \tag{13}$$

and state the necessary condition

$$\frac{\partial H}{\partial u} = 0 \tag{14}$$

where the Lagrange multiplier ν is chosen to satisfy the constraint

$$\int_0^T G(u(t), t)\,dt = 0. \tag{15}$$

We state now the general case as an exercise:

Exercise: Derive necessary conditions for extremal solutions to

$$J = \int_0^T F(u(t), t)\,dt = 0, \tag{16}$$

where $u(t)$ is an n-dimensional vector function, each component belonging to the class of smooth functions, and such that

$$\int_0^T G(u(t), t)\,dt = \Gamma. \tag{17}$$

for G a p-dimensional vector function.

3 Application: Entropy Maximization

This is a problem in statistical mechanics. Suppose we know that a probability density of a nonnegative random variable x has the expected value $\mathbf{E}\,x = m$. What is its most likely distribution? The notion of *most likely* means that it is the least prejudiced distribution, given the information of the mean: i.e., it is the *most random* distribution. Information theory tells us that the random ness of a distribution is specified by its entropy (differential entropy in the continuous case). The (differential) entropy of a density $f(x)$ is defined as the functional

$$\mathcal{H}(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx. \tag{18}$$

The most random density on the positive real line with given mean m is then given by the maximizer f_o of $\mathcal{H}(f)$ subject to the two constraints

$$\int_0^\infty f(x)\,dx = 1 \tag{19}$$

$$\int_0^\infty x f(x)\,dx = m. \tag{20}$$

With the Hamiltonian,

$$H(f, x, \lambda, \mu) = -f \log f + \lambda f + \mu x f, \tag{21}$$

The stationarity condition, $dH/df = 0$, yields

$$-\log f - 1 + \lambda + \mu x = 0. \tag{22}$$

This says that $f(x)$ must be exponential in x. Regrouping constants, the normalization condition says that μ must be negative, and thus

$$f(x) = |\mu| e^{-|\mu| x}, \tag{23}$$

while mu itself follows then from

$$m = \int_0^\infty x|\mu|e^{-|\mu|x}\,dx = \frac{1}{|\mu|}. \tag{24}$$

Thus finally, the most random density on $[0, \infty]$ with mean $m > 0$ is: $f_o(x) = \frac{1}{m}e^{-x/m}\chi(x)$, where $\chi(x)$ is the indicator function for the positive real line (also known as the Heaviside step function). We check that the result is indeed a valid density (we did not actually put the constraint $f(x) \geq 0$ in the problem formulation to keep it simple, but were lucky that the answer came out a probability density.

Exercise: Determine the most random probability density on the real line, with mean m and variance σ^2.