Curve optimization example (February 7, 2008)

In many applications, one needs to construct a smooth curve passing through a sequence of points. In one of the simplest cases, we want to find a function $y(t)$ for $t \in [0 \ldots 1]$, such that it starts and ends at zero with horizontal tangents, and passes through a point y_f at $t = t_f$. These conditions can be written as

$$y(0) = y(1) = y'(0) = y'(1) = 0;\ y(t_f) = y_f \tag{1}$$

These conditions are not sufficient to determine the whole curve. Typically, one uses a measure of curve quality to define it uniquely. For example, if we want the curve to be as smooth as possible, the curvature magnitude $|y''|/(1+(y')^2)^{3/2}$ can be used as a measure of quality. However, it is nonlinear in the unknown function y, and generally difficult to deal with, so the second derivative is often used instead. Note that for small y', the curvature is approximated well by the second derivative.

The derivative $y''(t)$ tells us how smooth the curve for a given value of t. To obtain a single number which we can minimize to get "the best" curve, we integrate the derivative squared over the whole interval, to obtain a *functional* $F(y)$, depending on the unknown function y:

$$F(y) = \frac{1}{2}\int_0^1 y''(t)dt \tag{2}$$

Our problem is now can be stated formally: *find the function $y(t)$ minimizing the functional (2) subject to the boundary conditions (1)*

To solve this optimization problem, we use one of two approaches: (1) derive a differential equation equivalent to solving the minimization problem, then discretize and solve the differential equation (2) discretize the functional $F(y)$, and derive optimality conditions for the discrete form of the functional.

Approach 1. This approach is based on *variational calculus* and the equation we get is called the *Euler-Lagrange* equation for the problem. The idea is similar to deriving conditions for an extremum of a function. We present a somewhat informal derivation (although a bit more formal than what we did in class); a complete rigorous theory is beyond the scope of this class.

Suppose y_* is the optimal function, that is, $F(y_*) \le F(y)$ for any y sufficiently close to y_*, satisfying the same boundary conditions. We assume that it is as many times differentiable as necessary. Consider a small perturbation $y(t) = y_*(t) + \epsilon(t)$ of y_*.

More formally, consider $\epsilon(t) = \delta g(t)$, where $g(t)$ is an arbitrary function, at least twice continuously differentiable, and $\delta > 0$. We assume that the perturbation $\delta g(t)$ vanishes at $t = 0, 1, t_f$, and $g'(t)$ vanishes at $t = 0, 1$; then $y(t)$ also satisfies the boundary conditions. In this way, $y(t)$ also satisfies the same boundary conditions.

As $F(y_*) \le F(y)$, the difference $F(y_* + \epsilon) - F(y_*)$ has to be positive for any choice of the perturbation $\epsilon(t)$.

Consider $(1/\delta)(F(y_* + \delta g) - F(y_*))$. Expanding $(y_*'' + \delta g'')^2$, and taking limits, we obtain the L^2*-gradient of $F(y)$ with respect to y, evaluated at g:*

$$\lim_{\delta \to 0} \frac{F(y_* + \delta g) - F(y_*)}{\delta} = \int (y_*'')g''dt$$

We integrate this expression by parts:

$$\int_0^1 (y_*'')g''dt = (y_*'')g'|_0^1 - \int_0^1 (y_*''')g'dt = -\int_0^1 (y_*''')g'dt$$

(the first term vanishes because $g'(0) = g'(1) = 0$). Repeating integration by parts, we obtain

$$\int_0^1 (y_*'')g''dt = \int_0^1 (y_*'''')gdt$$

This integral should be nonnegative. However, if we are free to choose arbitrary g, we can always choose it to be $-y_*''''$. In this case, the integral is non-positive. This is only possible if the integral is zero for any g, *that is if $y_*'''' = 0$. This is the differential equation for y_* which defines it uniquely when combined with the boundary conditions above.*

The fourth derivative can be converted to a discrete equation using finite differences. We will discuss finite differences in more detail later; for now, we observe that if for a function f we denotef_i its values at points ih, then at a point $x = jh$, the approximation to the first derivative is $(f_{j+1} - f_j)/h$. Taking approximations to first derivatives at points j and $j+1$, we obtain the approximation to the second *derivative, taking the difference of these approximations, and dividing by h. In terms of f_i, this second derivative approximation is*

$$\frac{(f_{j+2} - 2f_j + f_j)}{h^2}$$

similarly, the third and fourth derivative approximation can be obtained. We will see later that these approximations actually converge to the correct derivatives as $h \to 0$, if the function is continuously differentiable sufficiently many times.

The coefficients in the divided differences can be seen to be the Pascal triangle numbers taken with alternating signs. It is also possible to shift the indices of f_i used in the formulas without changing convergence (in fact, as we have seen, the rate of convergence may improve for symmetric (central) finite differences). This leads us to the approximation to the fourth derivative:

$$\frac{f_{j-2} - 4f_{j-1} + 6f_j - 4f_{j+1} + f_{j+2}}{h^4} \tag{3}$$

If we split $[0,1]$ into N subintervals, of length $h = 1/N$ each, and denote $t_i = ih$, $i = 0 \dots N+1$, we can write the equation $y'''' = 0$ in discrete form above for $N-3$ points $i = 2 \dots N-2$, with $N+1$ unknowns $y_i = y(t_i)$. The boundary conditions yield additional equations $y_0 = 0$, $y_N = 0$, $(y_1 - y_0)/h = 0$ and $(y_N - y_{N-1})/h = 0$. This adds four more equations, and we get a square system of linear equations for y_i. The last condition, $y(t_f) = y_f$, requires removing one of the equations (e.g., for $j = \lfloor t_f \rfloor$), and replacing it with $y_j = y_f$.

Approach 2. *Instead, we can simply replace the integral by a finite sum and the second derivative by its approximation:*

$$\frac{1}{2}\int_0^1 (f'')^2 dt \approx \frac{1}{2}\sum_{i=1}^{N-1} h_i \left(\frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}\right)^2 \tag{4}$$

where $h_i = h = 1/N$ for $i = 2 \dots N-2$, and $3/2h$ otherwise, to account for the absence of estimates for the second derivative at t_0 and t_1.

We have replaced the functional $F(y)$ with a smooth function $F^{approx}(y_0, y_1, \dots y_N)$ of a finite (although potentially very large) number of variables. The necessary condition for an extremum of this function is that its gradient is zero; in other words, $\partial F^{approx}/\partial y_i$ is zero for all i.

For $2 < i < N-2$, y_i is present in three summation terms, which yields exactly the same (up to a scalar factor) equation as we have obtained from $y'''' = 0$!

Warnings.

- *In general it is* not *sufficient to use a convergent derivative discretization for the solution of the system of linear equations for discrete variables to converge to the solution of the original differential equation as $h \to 0$ (it works in our case).*
- *Similarly, the discretization of the derivative and the integral approximation by a sum we have used in the second approach need to satisfy certain conditions.*
- *In general, approach 1 and approach 2 may not yield the same system of equations.*

The exact conditions for convergence of this type of numerical methods are mostly beyond the scope of this class and are discussed in more advanced numerical methods courses.