Optimization on Manifolds

Conor Rowan

March 2024

1 Brief Notes

This is meant to explore what energy minimization for physics problems look like when the solution is discretized in a nonlinear fashion. A usual discretization of a solution is $u(x) = \sum_i w_i N_i(x)$ where w_i are unknown parameters or "coordinates" for the problem. In general, the solution could depend nonlinearly on the parameters. The non-linear degree of freedom means that the solution lives on a manifold, as opposed to in a linear vector space. To illustrate how to deal with this, we will consider taking derivatives of vector-valued functions defined over parameterized spaces. For example, we may want to find the extremum of the function

$$f(x, y, z) = xyz$$

constrained to the unit sphere centered at the origin. Thus, the solution space is parameterized by two coordinates (defining a surface). The unit sphere can be parameterized with two angular coordinates by

$$\underline{x} = \begin{bmatrix} \cos(\theta_1)\sin(\theta_2) \\ \sin(\theta_1)\sin(\theta_2) \\ \cos(\theta_2) \end{bmatrix}$$

The extremum of the function constrained to the surface is defined as a point at which a small "nudge" along the constraint surface produces no change in the function value. This says that the gradient of the function if normal to the constraint surface. An admissible nudge is one that stays within the space defined by the parameterized surface (respects constraints imposed by parameterization of the solution). The nudge is therefore tangent to the parameterized surface. When this surface has curvature, admissible nudges live in the local tangent plane and thus depend on where we are on the surface. For the current example, this condition can be stated as

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[f\left(\underline{x}(\theta_1, \theta_2) + \epsilon \left(a \frac{\partial \underline{x}}{\partial \theta_1} + b \frac{\partial \underline{x}}{\partial \theta_2} \right) \right) - f(\underline{x}(\theta_1, \theta_2)) \right] = 0 \quad \forall a, b$$

A pair (θ_1, θ_2) that satisfies this condition extremizes the function f and lives on the parameterized surface. The arbitrary coefficients a and b are used to parameterize all possible nudges in the tangent plane. Because $\frac{\partial \underline{x}}{\partial \theta_1}$ and $\frac{\partial \underline{x}}{\partial \theta_2}$ are a basis for the tangent plane, it suffices to show that

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[f\left(\underline{x}(\theta_1, \theta_2) + \epsilon \frac{\partial \underline{x}}{\partial \theta_1} \right) - f(\underline{x}(\theta_1, \theta_2)) \right] = 0$$

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[f\left(\underline{x}(\theta_1, \theta_2) + \epsilon \frac{\partial \underline{x}}{\partial \theta_2} \right) - f(\underline{x}(\theta_1, \theta_2)) \right] = 0$$

This is equivalent to showing that the directional derivatives of f are zero in the direction of bases for plane tangent to the parameterized solution space:

$$\nabla f \cdot \frac{\partial \underline{x}}{\partial \theta_1} = \nabla f \cdot \frac{\partial \underline{x}}{\partial \theta_2} = 0$$

This derivation serves to motivate that of functionals. Now, assume we have a functional

$$\Pi = \int f(u, u_x) dx$$

where the solution lives in a parameterized space

$$u(x) = u(x; \theta_1, \ldots, \theta_n)$$

We can think of these parameters constructing an approximation of the solution in terms of unknown degrees of freedom. However, the solution need not have the typical linear approximation

$$u(x) = \sum_i \theta_i N_i(x)$$

An example of a parameterized solution that with non-linear degrees of freedom is

$$u(x) = \sum_i \theta_i N_i(x; \theta_i)$$

The structure of the parameterized solution space is more complicated in the example of non-linear degrees of freedom. Extrema of the functional constrained to a parameterized solution space are defined analogously to the vector-valued function:

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\Pi\left(u(x) + \epsilon \sum_i a_i \frac{\partial u}{\partial \theta_i} \right) - \Pi(u(x)) \right] = 0$$

where the a_i are arbitrary. As before, this can be split into a single condition for each basis of the tangent space

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\Pi\left(u(x) + \epsilon \frac{\partial u}{\partial \theta_i} \right) - \Pi(u(x)) \right] = 0 \quad \forall i$$

A typical calculus of variations derivation then leads to

$$\int \frac{\partial f}{\partial u} \left(\frac{\partial u}{\partial \theta_i} \right) dx + \int \frac{\partial f}{\partial u_x} \frac{\partial}{\partial x} \left(\frac{\partial u}{\partial \theta_i} \right) dx = 0 \quad \forall i$$

When the approximation of the solution is linear, the test functions $\partial u/\partial \theta_i$ have a particularly simple form. Nonlinear parameterizations of the solution will lead to nonlinear systems of equations, even when the energy functional is linear. These can be solved with Newton's Method, or the energy can be minimized using gradient descent. Note that one issue with this method is that the system of equations arising from the condition for a minimum can be satisfied for many choices of parameters. This is equivalent to saying that a minimum of the energy can be found for any linear basis expansion of the solution. Thus it is not clear we extract an optimal representation of the solution out of the residual equations. When we directly minimize the energy, there may be local minima but at least there is a notion of continuing to decrease the energy, which ideally leads to optimal representations of the solution.