Chapter 12

Gradients and critical points

In this chapter we present an approach to optimization, called the "calculus of variations," that is well suited for optimizing the various functionals presented in the previous chapter.

The abstract idea of the calculus of variations is as follows. Suppose we have a collection of objects X and a functional $F\colon X \to \mathbb{R}$. We want to know when we should consider some object x in X to be a critical point of the functional F.

To do this we consider a parametrized path, called a ***variation***, in the space X that passes through x at parameter 0. (Because in some applications we already have time involved in our functionals, we do not use the word "time" for the parameter.) It is traditional to use the letter ϵ for the parameter and to denote the variation by x_ϵ. (In this course, all of the paths we consider take the form $x_\epsilon = x + \epsilon\eta$ where η is some "appropriate object" that represents the "direction" that the path moves. Precisely what makes η "appropriate" depends on what the space X is.)

We then make the following definition: An object x is a ***critical point*** of functional F if

$$\frac{d}{d\epsilon}\left[F(x_\epsilon)\right]_{\epsilon=0} = 0$$

for all variations x_ϵ.

To see why this definition makes sense, let's consider the case that $X = \mathbb{R}$. In this situation, our functional is simply a function $F\colon \mathbb{R} \to \mathbb{R}$ and our object x is some real number. A variation takes the form $x_\epsilon = x + \epsilon\eta$, where η is a real number. Thus the definition of critical point above reduced to

$$\frac{d}{d\epsilon}\left[F(x+\epsilon\eta)\right]_{\epsilon=0} = 0$$

for all numbers η. Using the chain rule, we can simply this to

$$F'(x)\eta = 0$$

for all numbers η, which clearly reduces to the definition of critical points we learned in first semester calculus.

The advantage of defining critical points via variations is that we don't need to have a well-defined sense of what F' means in order to define what it means for x to be a critical point of F. Reflecting upon some of the functionals in the previous chapter, we see that this is clearly an advantage – what, for instance, might "$E'[u]$" mean in the case where E is the Dirichlet energy?

In the next section we see that the definition of critical point above agrees with the definition we encountered in our multivariable calculus course. In fact, we see that this definition even allows us to compute the gradient of a function. Subsequently, we apply this definition to the functionals defined in the previous chapter.

12.1 Optimizing functions $f\colon \mathbb{R}^2 \to \mathbb{R}$

Consider some function $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let $\mathbf{x} = (x, y)$ be some point in $\mathbb{R}^2$ and let $\eta = \langle \eta_x, \eta_y \rangle$. We define the variation $\mathbf{x}_\epsilon = (x + \epsilon\eta_x, y + \epsilon\eta_y)$. Since

$$\frac{d\mathbf{x}_\epsilon}{d\epsilon} = \langle \eta_x, \eta_y \rangle = \eta,$$

we see that we can view η as the tangent vector to the path in $\mathbb{R}^2$ given by the variation $\mathbf{x}_\epsilon$.

Computing the derivative of f along the variation yields

$$\begin{aligned}\frac{d}{d\epsilon}\left[f(\mathbf{x}_\epsilon)\right] &= \frac{d}{d\epsilon}\left[f(x+\epsilon\eta_x, y+\epsilon\eta_y)\right] \\ &= \frac{\partial f}{\partial x}(x+\epsilon\eta_x, y_) + \epsilon\eta_y)\,\eta_x + \frac{\partial f}{\partial y}(x+\epsilon\eta_x, y+\epsilon\eta_y)\,\eta_y.\end{aligned}$$

Thus

$$\frac{d}{d\epsilon}\left[f(\mathbf{x}_\epsilon)\right]_{\epsilon=0} = \frac{\partial f}{\partial x}(x,y)\,\eta_x + \frac{\partial f}{\partial y}(x,y)\,\eta_y = \operatorname{grad} f(\mathbf{x})\cdot\eta. \tag{12.1}$$

There are two important consequences of (12.1). First, we see that a point $\mathbf{x}$ is a critical point of f if $\operatorname{grad} f(\mathbf{x})\cdot\eta = 0$ for all tangent vectors η. Note that if $\mathbf{y}$ is a vector such that $\mathbf{y}\cdot\eta = 0$ for all η, then it must be the case that $\mathbf{y} = \mathbf{0}$. Thus we conclude that

$$\mathbf{x} \text{ is a critical point of } f \quad \leftrightarrow \quad \operatorname{grad} f(\mathbf{x}) = \mathbf{0}.$$

This, of course, agrees with what we learned in our multivariable calculus course.

The second important aspect of (12.1) is this – we can actually use this formula to define the gradient of a function. In particular, we can define the gradient of f at $\mathbf{x}$ to be whatever quantity satisfies

$$\frac{d}{d\epsilon}\left[f(\mathbf{x}_\epsilon)\right]_{\epsilon=0} = (\text{gradient goes here})\cdot\eta.$$

12.2 The L^2 gradient of a functional

We now consider the case of a functional F that takes inputs from $L^2(\Omega)$ and returns outputs in $\mathbb{R}$. Suppose that $u\colon \Omega \to \mathbb{R}$ is some function in $L^2(\Omega)$. We construct a variation $u_\epsilon = u + \epsilon\eta$, where $\eta \in C_0^\infty(\Omega)$ is a test function. Such variations are called ***compactly supported variations***.

The reason for considering compactly supported variations is that such variations respect Dirichlet (or other) boundary conditions that might need to be satisfied. For example, consider the boundary value problem (11.4). In this case we want each of the functions u_ϵ to satisfy the boundary condition $u_\epsilon|_\Gamma = b$. The easiest way to accomplish this is to require that u satisfy this condition and that η vanish near the boundary.

We now define the L^2 ***gradient of functional*** F ***at*** u to be the function $\operatorname{grad} F[u]$ satisfying

$$\frac{d}{d\epsilon}\left[F[u_\epsilon]\right]_{\epsilon=0} = \langle \operatorname{grad} F[u], \eta\rangle$$

for all variations u_ϵ of u.

As the following examples demonstrate, the gradient of a functional is frequently a differential operator.

Example 12.1. *Consider the functional* $L[u]$*, given by* (11.1)*, that describes the length of the graph of function* $u\colon [-1,1] \to \mathbb{R}$*. With* $u_\epsilon = u + \epsilon\eta$ *we compute*

$$\begin{aligned}
\frac{d}{d\epsilon}\left[L[u]\right] &= \frac{d}{d\epsilon}\int_{-1}^{1}\left(1 + (u'(x) + \epsilon\,\eta'(x))^2\right)^{1/2}\,dx \\
&= \int_{-1}^{1}\frac{u'_\epsilon}{\sqrt{1+(u'_\epsilon)^2}}\,\frac{d\eta}{dx}\,dx \\
&= \int_{-1}^{1} -\frac{d}{dx}\left[\frac{u'_\epsilon}{\sqrt{1+(u'_\epsilon)^2}}\right]\eta\,dx,
\end{aligned}$$

where we integrated by parts in the last step.

From this we conclude that the gradient of L is

$$\operatorname{grad} L[u] = -\frac{d}{dx}\left[\frac{u'}{\sqrt{1+(u')^2}}\right].$$

Recall from Chapter 7 that a function f is zero if and only if $\langle f, \eta\rangle = 0$ for all test functions $\eta \in C_0^\infty(\Omega)$. Combining this fact with the definition of the gradient of a functional we obtain the following.

Theorem 12.2. *Let* $F\colon L^2(\Omega) \to \mathbb{R}$ *be a functional. A function u is a critical point of* F *exactly when* $\operatorname{grad} F[u] = 0$.

We conclude this chapter by noting that, just as in the first-year calculus course, not all critical points are minimizers. There exists a number of techniques for determining whether or not critical points actually minimize a particular functional, including a generalization of the second derivative test. If you are interested in the general theory, you are encouraged to check out the book *The Calculus of Variations* by Bruce Van Brunt.