Chapter 2

Calculus of variations

2.1 Optimization in $\mathbb{R}^n$

FD-List

Exercise 2.1.1. Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$. We make the following definitions:

- Gradient
$$\operatorname{grad} f(\mathbf{v}) = \left\langle \frac{\partial f}{\partial x}(\mathbf{v}), \frac{\partial f}{\partial y} f(\mathbf{v}) \right\rangle$$
- Hessian
$$\operatorname{hess} f(\mathbf{v}) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(\mathbf{v}) & \dfrac{\partial^2 f}{\partial x \partial y}(\mathbf{v}) \\ \\ \dfrac{\partial^2 f}{\partial x \partial y}(\mathbf{v}) & \dfrac{\partial^2 f}{\partial y^2}(\mathbf{v}) \end{pmatrix}$$

Compute the gradient and Hessian of the following functions:

1. $u(x, y) = 3x^2 - 2xy + 3y^2$
2. $u(x, y) = 6xy$
3. $u(x, y) = \frac{1}{1+x^2+y^2}$
4. $u(x, y) = \frac{x}{1+x^2+y^2}$
5. $u(x, y) = \frac{x^2}{1+x^2+y^2}$
6. $u(x, y) = \frac{x^3}{1+x^2+y^2}$

Exercise 2.1.2. Here's an alternate way to define the gradient and Hessian of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$.

1. Suppose we have a path in $\mathbb{R}^2$ starting at point $\mathbf{v}$ and moving in direction $\mathbf{w}$. Show that this path can be described by $\mathbf{p}(\varepsilon) = \mathbf{v} + \varepsilon\mathbf{v}$. (Here ε is the "time" parameter.) Draw a picture!
2. Let $\phi(\varepsilon)$ be the function $\mathbb{R} \to \mathbb{R}$ given by $\phi(\varepsilon) = f(\mathbf{p}(\varepsilon))$. Show that:
    - $\phi(0) = f(\mathbf{v})$,
    - $\phi'(0) = \operatorname{grad} f(\mathbf{v}) \cdot \mathbf{w}$,
    - $\phi''(0) = \operatorname{hess} f(\mathbf{v})\mathbf{w} \cdot \mathbf{w}$

   Explain how we could have *defined* the gradient and hessian to be the objects such that
    - $\frac{d}{d\varepsilon}\left[f(\mathbf{v} + \varepsilon\mathbf{w})\right]_{\varepsilon=0} = \operatorname{grad} f(\mathbf{v}) \cdot \mathbf{w}$
    - $\frac{d^2}{d\varepsilon^2}\left[f(\mathbf{v} + \varepsilon\mathbf{w})\right]_{\varepsilon=0} = \operatorname{hess} f(\mathbf{v})\mathbf{w} \cdot \mathbf{w}$
3. Show that the Taylor series for ϕ is

   $$\phi(\varepsilon) = \phi(0) + \phi'(0)\varepsilon + \frac{1}{2}\varepsilon^2\phi''(0) + \dots$$

   and therefore that the Taylor expansion for f is

   $$f(\mathbf{v} + \varepsilon\mathbf{w}) = f(\mathbf{v}) + \varepsilon \operatorname{grad} f(\mathbf{v}) \cdot \mathbf{w} + \frac{\varepsilon^2}{2}(\operatorname{hess} f(\mathbf{v})\mathbf{w}) \cdot \mathbf{w} + \dots$$

4. This method of computing (using paths) is called the *method of variations*, because we are using the path to "vary" the point $\mathbf{v}$.

   Pick a few functions from the list in Exercise 2.1. Compute the gradient and Hessian using the method of variations. Confirm that you get the same thing as before.

Exercise 2.1.3. Here we explore the function $f(x, y) = xy - 2y + x - 2$.

1. What is the Calc III definition of a critical point of f?
2. Find the one critical point of the function; we'll call this point $\mathbf{v}_*$.

3. Let $\mathbf{w} = \langle 2, 1\rangle$; consider the path $\mathbf{v}_* + \varepsilon\mathbf{w}$. Make a plot of the one-dimensional function $f(\mathbf{v}_* + \varepsilon\mathbf{w})$. Show that $\varepsilon = 0$ is a critical point of this function. Is $\varepsilon = 0$ a min/max/neither?
4. Repeat for the following vectors: $\mathbf{y} = \langle 0, 1\rangle$, $\mathbf{y} = \langle 1, 0\rangle$, $\mathbf{y} = \langle 1, 1\rangle$, $\mathbf{y} = \langle 1, -1\rangle$.
5. Based on all these paths, what type of critical point do you think $\mathbf{v}_*$ is? Explain.
6. Make a plot of the function g and confirm your conjecture.

Exercise 2.1.4. Now we focus on the optimization problem. Suppose we have a function f as in the previous exercise and, for some vectors $\mathbf{v}$ and $\mathbf{w}$, we build the function ϕ. Our plan is to view $\mathbf{v}$ as "fixed" and let $\mathbf{w}$ "vary."

1. Explain why

   $\mathbf{v}$ is an optimizer for f

   is equivalent to

   0 is an optimizer for ϕ for every $\mathbf{w}$.
2. We say that $\mathbf{v}$ is a critical point (in the Calc IV sense) of f if

$$\frac{d}{d\varepsilon}\left[f(\mathbf{v} + \varepsilon\mathbf{w})\right]_{\varepsilon=0} = 0$$

   for every $\mathbf{w}$. Explain (1) why this is a reasonable definition and (2) why it agrees with our previous definition.
3. What is the Calc 1 version of the second derivative test? How might one use this to obtain a second derivative test for f in terms of the Hessian?
4. Find all critical points of the functions listed in Exercise 2.1.