Optimization

Adam Kelly (`ak2316@cam.ac.uk`)

December 8, 2021

This is a short description of the course. It should give a little flavour of what the course is about, and what will be roughly covered in the notes.

This article constitutes my notes for the 'Optimization' course, held in Easter 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered.

1 Convex Optimization

1.1 Optimization Problems

In this course, we care about problems that look like 'minimize $f(x)$ subject to $x \in X$', or problems with constraints that look like 'minimize[1] $f(x)$ subject to $x \in X$ where $h(x) = b$'.

In such problems, we call f the **objective function**, and if $x \in \mathbb{R}^n$, we call the components **decision variables**. Constraints such as the one above can be given as a **functional constraint** $h(x) = b$ and a **regional constraints** $x \in X$. If both of the constraints hold for a given x, we say it is in the **feasible set** $X(b)$. A problems is feasible if the feasible set is non-empty and bounded if the minimum of this set is bounded.

A point $x^* \in X(b)$ is **optimal** if it minimizes f over (b). The value $f(x^*)$ is the **optimal cost**.

1.2 Unconstrained Optimization of Convex Functions

We will first study optimization problems involving convex functions without constraints. In particular, we are going to study the problem:

Minimize $f(x)$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

The reason we will care specifically about convex functions is, as you will see, that local information about the function gives us global information about the function.

It is likely that you are already familiar with the notions of convex sets and convex functions.

Definition 1.1 (Convex Set)**.** A set $S \in \mathbb{R}^n$ is **convex** if for all $x, y \in S$ and all $\lambda \in [0, 1]$ we have $\lambda x + (1 - \lambda)y \in S$.

---

[1]We will almost always talk about minimizing, since maximizing is basically the same thing.

Informally, a set is convex if the line segment between any two points is totally contained in the set. We can also define convexity for functions.

Definition 1.2 (Convex Function). A function $f : S \to \mathbb{R}$ is **convex** if S is convex and for all $x, y \in S$ and $\lambda \in [0, 1]$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

If this inequality is strict then the function is **strictly convex**. A function is called **concave** if $-f$ is convex.

Again informally, this says that if we consider the graph of the function then the chord between two points will always be above the function.

In general, checking this definition by hand is hard, and we typically use some other conditions to check for convexity.

Proposition 1.3 (First-Order Conditions for Convexity). *A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for all $x, y \in \mathbb{R}^n$ we have*

$$f(y) \geq f(x) + (y - x) \cdot \nabla f(x).$$

Proof. We will first check that convexity implies that this condition holds by showing it holds for $n = 1$ and using that for the general case.

If $f : \mathbb{R} \to \mathbb{R}$ is convex, and $x, y \in \mathbb{R}$ then for any $t \in [0, 1]$ we have

$$
\begin{aligned}
f((1 - t)x + ty) &= f(x + t(y - x)) \\
&\leq (1 - t)f(x) + tf(y) \\
\implies f(y) &\geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t}.
\end{aligned}
$$

Taking the limit as $t \to 0$, we get that $f(x) \geq f(x) + f'(x)(y - x)$.

Now for $f : \mathbb{R}^n \to \mathbb{R}$, taking any $x, y \in \mathbb{R}^n$ we can define the restricted function $g(t) = f((1 - t)x + ty)$. Since f is convex, g is also convex. We can then compute

$$g'(t) = (y - x) \cdot \nabla f((1 - t)x + ty)$$

So using the $n = 1$ case, we have $g(1) \geq g(0) + g'(0)(1 - 0)$, that is, $f(y) \geq f(x) + (y - x) \cdot \nabla f(x)$, so the condition holds as required.

Now we will show that the first-order conditions imply convexity. For some $t \in [0, 1]$, we define $x_t = (1 - t)x + ty$. Then we know that

$$
\begin{aligned}
f(x) &\geq f(x_t) + (x - x_t) \cdot \nabla f(x_t), \\
f(y) &\geq f(x_t) + (y - x_t) \cdot \nabla f(x_t).
\end{aligned}
$$

With a small amount of algebraic manipulation, summing gives

$$(1 - t)f(x) + tf(y) \geq f(x_t),$$

as required. $\square$

A possibly more familiar condition for convexity is the second-order conditions. This involves the *Hessian*, the generalization of the second derivative to $\mathbb{R}^n$.

Definition 1.4 (Hessian). If $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable function, then the **Hessian matrix** $\nabla^2 f$ is the symmetric matrix of square derivatives given by

$$(\nabla^2 f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

In particular, we care about a specific property of the Hessian matrix.

Definition 1.5 (Positive Semidefinite). A matrix A is **positive semidefinite** if for all $x \in \mathbb{R}^n$ we have $x \cdot Ax \geq 0$. Equivalently, if all of the eigenvalues of A are non-negative.

Proposition 1.6 (Second-Order Conditions for Convexity). *A twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\nabla^2 f(x)$ is positive semidefinite at all $x \in \mathbb{R}^n$.*

Proof. Using the Taylor expansion of f, we have

$$f(y) = f(x) + (y - x) \cdot \nabla f(x) + \frac{1}{2}(y - x) \cdot (\nabla^2 f(z)(y - x)),$$

where $z = (1 - \lambda)x + \lambda y$ for some $\lambda \in [0, 1]$. Then by the first order conditions, f is convex. $\square$