Lecture 1 Exercises

Exercise 1.1

Make sure you understand how to derive formula (1.7) using function $\varphi(t) = f(x + td)$.

Solution: Formula (1.7) is the Mean Value Theorem. We have $\varphi(t) = f(x + td)$. Then $\varphi(1) = f(x + d)$ and $\varphi(0) = f(x)$. By the Mean Value Theorem, there exists $z \in [0, 1]$ such that:

$$\varphi(1) = \varphi(0) + \varphi'(z)(1 - 0)$$

$$f(x + d) = f(x) + \varphi'(z)$$

Now, $\varphi'(t) = \frac{d}{dt} f(x + td) = \langle \nabla f(x + td), d \rangle$. So, $\varphi'(z) = \langle \nabla f(x + zd), d \rangle$. Let $\tilde{z} = x + zd$, then $f(x + d) = f(x) + \langle \nabla f(\tilde{z}), d \rangle$, where $\tilde{z} \in [x, x + d]$.

Exercise 1.2

Let $f(x) = \frac{1}{2}\|Ax - b\|^2$ for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Compute the gradient in two ways: using the definition (1.5) and using the chain rule.

Solution:

Method 1: Using Definition (1.5)

$$\begin{aligned}
f(x + td) &= \frac{1}{2}\|A(x + td) - b\|^2 \\
&= \frac{1}{2}\|Ax + tAd - b\|^2 \\
&= \frac{1}{2}\langle Ax + tAd - b, Ax + tAd - b \rangle \\
&= \frac{1}{2}\langle Ax - b, Ax - b \rangle + t\langle Ax - b, Ad \rangle + \frac{t^2}{2}\langle Ad, Ad \rangle \\
&= f(x) + t\langle A^T(Ax - b), d \rangle + \frac{t^2}{2}\|Ad\|^2
\end{aligned}$$

Thus, $\nabla f(x) = A^T(Ax - b)$.

Method 2: Using Chain Rule

Let $g(x) = Ax - b$ and $h(y) = \frac{1}{2}\|y\|^2$. Then $f(x) = h(g(x))$. We have $\nabla h(y) = y$ and $g'(x) = A$. By the chain rule, $\nabla f(x) = g'(x)^T \nabla h(g(x)) = A^T(Ax - b)$.

Exercise 1.3

Find the gradient and the Hessian of $f(x) = \frac{1}{2}\langle Qx, x \rangle - \langle b, x \rangle + c$, where $x, b \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$, and $c \in \mathbb{R}$.

Solution:

Gradient

$$\begin{aligned}
f(x+d) &= \frac{1}{2}\langle Q(x+d), x+d\rangle - \langle b, x+d\rangle + c \\
&= \frac{1}{2}\langle Qx, x\rangle + \langle Qx, d\rangle + \frac{1}{2}\langle Qd, d\rangle - \langle b, x\rangle - \langle b, d\rangle + c \\
&= f(x) + \langle Qx - b, d\rangle + \frac{1}{2}\langle Qd, d\rangle
\end{aligned}$$

Thus, $\nabla f(x) = Qx - b$. (Assuming Q is symmetric)

Hessian

Since $\nabla f(x) = Qx - b$, we have $\nabla^2 f(x) = Q$.

Exercise 1.4

Make sure you understand what equation (1.10) means and why that derivation is correct.

Solution: Equation (1.10) states that the gradient of a function at a point is the direction of the local fastest increase of the function. The derivation is correct because it follows from the definition of the directional derivative and the Cauchy-Schwarz inequality. The directional derivative is maximized when the direction vector is aligned with the gradient.

Exercise 1.5

When is the function defined in Exercise 1.3 (i) convex; (ii) strongly convex?

Solution: (i) The function $f(x) = \frac{1}{2}\langle Qx, x\rangle - \langle b, x\rangle + c$ is convex if and only if $Q \succeq 0$ (Q is positive semi-definite). (ii) The function is strongly convex if and only if $Q \succ 0$ (Q is positive definite).

Exercise 1.6

Prove equivalence in (ii) in Lemma 1.2.

Solution: We need to show that f is μ-strongly convex if and only if $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}||y - x||^2$.

($\implies$) Suppose f is μ-strongly convex. Then for $\lambda \in [0, 1]$:

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) - \frac{\mu\lambda(1-\lambda)}{2}||y - x||^2$$

Let $z = \lambda y + (1-\lambda)x$, then $y = x + \frac{z-x}{\lambda}$. Substitute to the formula and take $\lambda \to 0$:

$$f(x) + \langle \nabla f(x), z - x\rangle \leq f(z)$$

Also, from strong convexity we have:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}||y - x||^2$$

($\Longleftarrow$) Assume $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}||y - x||^2$. Let $z = \lambda x + (1-\lambda)y$. We have $f(x) \geq f(z) + \langle \nabla f(z), x - z \rangle + \frac{\mu}{2}||x - z||^2$ and $f(y) \geq f(z) + \langle \nabla f(z), y - z \rangle + \frac{\mu}{2}||y - z||^2$

Multiply the first inequality by λ, and the second by $(1-\lambda)$ and add them: $\lambda f(x) + (1-\lambda)f(y) \geq f(z) + \langle \nabla f(z), \lambda x + (1-\lambda)y - z \rangle + \frac{\mu\lambda}{2}||x - z||^2 + \frac{\mu(1-\lambda)}{2}||y - z||^2$ Since $z = \lambda x + (1-\lambda)y$, then $\langle \nabla f(z), \lambda x + (1-\lambda)y - z \rangle = 0$. Also, $x - z = (1-\lambda)(x-y)$ and $y - z = \lambda(y-x)$. Then: $\lambda f(x) + (1-\lambda)f(y) \geq f(z) + \frac{\mu\lambda(1-\lambda)^2}{2}||x - y||^2 + \frac{\mu(1-\lambda)\lambda^2}{2}||x - y||^2$

$$\lambda f(x) + (1-\lambda)f(y) \geq f(z) + \frac{\mu\lambda(1-\lambda)}{2}||x - y||^2$$

Rearranging terms, we get the definition of strong convexity:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu\lambda(1-\lambda)}{2}||x - y||^2$$

Exercise 1.7

One may expect that in Lemma 1.3, strict convexity is equivalent to $\nabla^2 f(x) \succ 0$. Show that this is not true.

Solution: Consider the function $f(x) = x^4$ on $\mathbb{R}$. We have $f''(x) = 12x^2$. Thus, $f''(x) \geq 0$ for all x, and $f''(x) = 0$ only at $x = 0$. Thus, $f(x)$ is convex, but not strictly convex (since $f''(0) = 0$).

However, it is strictly convex (show it using the definition). But $f''(0) = 0$, which means that $f''(0) \not\succ 0$. This counterexample shows that strict convexity is not equivalent to $\nabla^2 f(x) \succ 0$.

Exercise 1.8

Prove Lemma 1.4.

Solution: (i) If f is convex, then every local minimum is a global one, and the set of minima is convex.

Suppose x is a local minimum, but not a global one. Then there exists y such that $f(y) < f(x)$. For any $\lambda \in (0,1)$, by convexity:

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

Since x is a local minimum, there exists a neighborhood around x where f(x) ¡= f(z) for all z in that neighborhood. However, we just showed that for any convex combination of x and y, the function value is less than f(x), contradicting that x is a local minimum. Therefore, a local minimum must be a global minimum.

To prove that the set of minima is convex, let x_1 and x_2 be two global minima, so $f(x_1) = f(x_2) = f*$. Then for any $\lambda \in [0,1]$, let $x = \lambda x_1 + (1-\lambda)x_2$. By convexity:

$f(x) \leq \lambda f(x_1) + (1-\lambda)f(x_2) = \lambda f * + (1-\lambda)f* = f*$

Since $f*$ is the global minimum value, it must be that $f(x) = f*$, so any convex combination of two global minima is also a global minimum.

(ii) If f is strictly convex, then it has at most one minimum.

Assume for contradiction that there exist two minima $x_1 \neq x_2$. Then $f(x_1) = f(x_2) = f*$. Since f is strictly convex, for any $\lambda \in (0, 1)$:

$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) = f*$,

which is a contradiction, because $f*$ is the global minimum value. Thus, there can exist at most one minimum.

(iii) If f is strongly convex, then a minimum always exists.

Strong convexity means that the function is "bowl-shaped" and grows at least quadratically at infinity. Specifically, there exists $\mu > 0$ such that:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}||y - x||^2$$

Choose some point x_0. Consider the set $S = \{x : ||x - x_0|| \leq R\}$. Then for any x outside of the set S (i.e., $||x - x_0|| > R$):

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{\mu}{2}||x - x_0||^2 \geq f(x_0) - ||\nabla f(x_0)|| ||x - x_0|| + \frac{\mu}{2}||x - x_0||^2$$

$$f(x) \geq f(x_0) - ||\nabla f(x_0)||R + \frac{\mu}{2}R^2$$

We can choose R large enough so that the right-hand side is greater than $f(x_0)$. This means that the function values outside the set S are always greater than the value at the point x_0.

Since f is continuous on the compact set S, by the Weierstrass theorem, it attains its minimum on S. Moreover, since the values of f outside S are always greater than $f(x_0)$, the minimum on S is also a global minimum.