Optimization and Computational Linear Algebra – Brett Bernstein

Recitation 11

It may be helpful if after recitation you try to re-solve these problems by yourself, and use them as additional study problems for the class.

1. Let $f(x) = -e^{-x^2/2}$. Is f convex?

   Solution. No. $f'(x) = xe^{-x^2/2}$ and $f''(x) = (1-x^2)e^{-x^2/2}$ which is negative for $x > 1$.

2. Prove that $e^x \geq 1 + x$ and $e^x \geq ex$ for all $x \in \mathbb{R}$.

   Solution. Note that e^x is convex since if $f(x) = e^x$ then $f''(x) = e^x > 0$. Also note that $f'(0) = 1$ and $f'(1) = e$. Thus the tangent lines given by $1 + x$ and e^x must globally underestimate the function.

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by $f(x) = \log(1 + x^T Ax)$ where $A \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite. Compute $\nabla f(x)$.

   Solution. We can apply the chain rule. Let $g(t) = \log(t)$ and $h(x) = 1 + x^T Ax$ so that $f(x) = g(h(x))$. Then we have

   $$\nabla f(x) = g'(h(x))\nabla h(x) = \left(\frac{1}{1 + x^T Ax}\right)(2Ax).$$

   To see that $\nabla h(x) = 2Ax$ we can either apply the homework result or use the following calculation:

   $$h(x+v) = 1 + (x+v)^T A(x+v) = 1 + x^T Ax + 2x^T Av + v^T Av = h(x) + 2x^T Av + v^T Av.$$

   Thus $\nabla h(x)^T = 2x^T A$ since

   $$\left|\frac{h(x+v) - h(x) - 2x^T Av}{\|v\|}\right| = \frac{|v^T Av|}{\|v\|} \leq \frac{\|A\|\|v\|^2}{\|v\|} \to 0.$$

4. Let $f : \mathbb{R} \to \mathbb{R}$ be convex and differentiable. Prove that for all $x, h \in \mathbb{R}$ that

   $$f(x+h) \geq f(x) + f'(x)h.$$

   Solution.

   Proof. Suppose, for contradiction, there is an x, h with

   $$f(x+h) < f(x) + f'(x)h.$$

We assume $h > 0$ (a similar proof holds for $h < 0$, or we could just apply the result proven below to $g(x) = f(-x)$, which is also convex). Define $\alpha > 0$ such that

$$f(x+h) = f(x) + (f'(x) - \alpha)h.$$

In other words

$$\alpha := -\frac{f(x+h) - f(x) - f'(x)h}{h} > 0.$$

By convexity we have, for all $t \in (0,1)$,

$$\begin{aligned} f(x+th) &= f((1-t)x + t(x+h)) \\ &\leq (1-t)f(x) + tf(x+h) \\ &= (1-t)f(x) + tf(x) + t(f'(x) - \alpha)h \\ &= f(x) + t(f'(x) - \alpha)h. \end{aligned}$$

Rearranging we have

$$\frac{f(x+th) - f(x)}{th} < f'(x) - \alpha$$

for all t. Letting $t \to 0$ shows

$$f'(x) < f'(x) - \alpha,$$

a contradiction. □

5. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Prove that for all $x, h \in \mathbb{R}^n$ that

$$f(x+h) \geq f(x) + \nabla f(x)^T h.$$

Solution.

Proof. Suppose, for contradiction, there is an $x, h \in \mathbb{R}^n$ with

$$f(x+h) < f(x) + \nabla f(x)^T h.$$

Define $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = f(x+th)$. Then g is convex and the above statement says that

$$g(1) < g(0) + g'(0).$$

But this is impossible by the previous problem. □

6. Suppose $A \in \mathbb{R}^{2\times 2}$ is symmetric with positive eigenvalues. Describe geometrically the contour lines of $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = x^T Ax$. Recall that the contour line for value γ is given by

$$\{x \in \mathbb{R}^2 : f(x) = \gamma\}.$$

Solution. First suppose A is diagonal:

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Then $x^T Ax = \lambda_1 x_1^2 + \lambda_2 x_2^2$. Thus solving $f(x) = \gamma$ is the equation for an elippse. By the spectral theorem, any symmetric matrix is diagonal up to a rotation. Thus generally we obtain a rotated ellipse.

7. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by $f(x) = x_1^2 + 100x_2^2$. Explain what issues this may pose for gradient descent.

Solution. The contour lines of f are very eccentric ellipses. Gradient descent started at $(-1, 0.1)$ will take many steps to converge.

8. (a) Give a quadratic approximation $q(h)$ to $f(x + h)$ at x.
   (b) What is the minimizer of the quadratic approximation assuming the Hessian is positive definite (has positive eigenvalues)?

Solution.

(a)

$$\begin{aligned} f(x+h) &\approx f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h \\ &= f(x) + \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j \\ &=: q(h). \end{aligned}$$

To get the result from class, we let $h = y - x$ to obtain

$$f_x^2(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x).$$

(b) As $q(h)$ is quadratic, we can apply the homework to see that

$$\nabla q(h) = \nabla f(x) + \nabla^2 f(x) h,$$

which is zero when

$$h = -(\nabla^2 f(x))^{-1} \nabla f(x).$$

9. What is the third order approximation to f at x?

Solution.

$$f(x+h) \approx f(x) + \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j + \frac{1}{6} \sum_{i,j,k=1}^{n} \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k.$$

As above we obtain $f_x^3(y)$ letting $h = y - x$.

10. ($\star$) Let $f : \mathbb{R} \to \mathbb{R}$ be convex and let X be a random variable.

    (a) Assuming X only takes finitely many values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$ that sum to 1. Prove that $E[f(X)] \geq f(E[X])$.

    (b) Prove $E[f(X)] \geq f(E[X])$ for an arbitrary random variable X.

    Solution.

    (a) Consider the random vector $(X, f(X))$ taking values in $\mathbb{R}^2$ on the graph of f. Then we have

    $$E[(X, f(X))] = \sum_{i=1}^{n} p_i(x_i, f(x_i)) = (E[X], E[f(X)])$$

    which is a convex combination of points on the graph of f, and thus must lie in the epigraph of f. This immediately gives

    $$E[f(X)] \geq f(E[X]).$$

    (b) Although we only prove this for differentiable functions in class, for any convex function $f : \mathbb{R} \to \mathbb{R}$ and any $v \in \mathbb{R}$ there is a tangent line $g(x) = f(v) + a(x - v)$ (for some $a \in \mathbb{R}$) that lies below f (i.e., $g(x) \leq f(x)$ for all $x \in \mathbb{R}$). Letting $v = E[X]$ we obtain

    $$f(E[X]) + a(x - E[X]) \leq f(x)$$

    for all $x \in \mathbb{R}$, so we plug in X for x and take expectations to obtain

    $$E[f(X)] \geq E[f(E[X]) + a(X - E[X])] = f(E[X]).$$

11. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be smooth (at least twice continuously differentiable) and assume that $x \in \mathbb{R}$ has $\nabla f(x) = 0$. What do each of the following conditions on the Hessian $\nabla^2 f(x)$ imply?

    (a) $\nabla^2 f(x)$ has two positive eigenvalues.

    (b) $\nabla^2 f(x)$ has two negative eigenvalues.

    (c) $\nabla^2 f(x)$ has one positive and one negative eigenvalue.

    Solution.

(a) Local minimum.

(b) Local maximum.

(c) Saddle point.

We will prove the first of these statements to show how these types of results are justified.

Lemma 1 (Multivariate Taylor's Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2$. Then for all $x, h \in \mathbb{R}^n$ we have*

$$f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x+sh)h,$$

for some $s \in (0,1)$.

Proof. Fix $x, h \in \mathbb{R}^n$ and define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(t) = f(x+th).$$

Then by the one-dimensional version of Taylor's theorem we have

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(s),$$

for some $s \in (0,1)$. Using the definition of g and applying the chain rule we have

$$g'(t) = \nabla f(x+th)^T h \quad \text{and} \quad g''(t) = h^T \nabla^2 f(x+th)h.$$

Plugging in we obtain

$$f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x+sh)h.$$

□

Lemma 2. *Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric. Let λ_n^A and λ_n^B denote the smallest eigenvalues of A and B, respectively. Then $|\lambda_n^A - \lambda_n^B| \le \|A - B\|$.*

Proof. For any $x \in \mathbb{R}^n$ with $\|x\| = 1$ we have

$$|x^T Bx - x^T Ax| = |x^T (B-A)x| \le \|B - A\|.$$

Since

$$\lambda_n^A = \min_{\|x\|=1} x^T Ax$$

we have that

$$\lambda_n^B = \min_{\|x\|=1} x^T Bx \le \min_{\|x\|=1} x^T Ax + x^T (B-A)x \le \lambda_n^A + \|B - A\|.$$

Reversing the roles of A, B completes the proof. □

Lemma 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2$. If $\nabla^2 f(x) \succ 0$ for some $x \in \mathbb{R}^n$ then there is a $\delta > 0$ such that $\nabla^2 f(y) \succ 0$ for all y with $\|x - y\| < \delta$.*

Proof. Fix $x \in \mathbb{R}^n$. Since $f \in C^2$, each of the second partial derivatives

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

are continuous. Thus for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\|\nabla^2 f(y) - \nabla^2 f(x)\|_F^2 < \epsilon$$

whenever $\|x - y\| < \delta$. [To see this note that for each $\frac{\partial^2 f}{\partial x_i \partial x_j}$ there is a δ_{ij} such that $|\frac{\partial^2 f(y)}{\partial x_i \partial x_j} - \frac{\partial^2 f(x)}{\partial x_i \partial x_j}|^2 < \epsilon/n^2$. Thus the bound holds for $\|x - y\|_\infty < \min_{i,j} \delta_{ij}$. But $\|x - y\|_\infty \leq \|x - y\|_2$, so we can let $\delta = \min_{i,j} \delta_{ij}$.] Let $\epsilon = \lambda_n^{\nabla^2 f(x)}/2$, where $\lambda_n^{\nabla^2 f(x)}$ is the smallest eigenvalue of $\nabla^2 f(x)$. Since the Frobenius norm upper bounds the spectral norm, the previous lemma applies proving that

$$\lambda_n^{\nabla^2 f(y)} \geq \lambda_n^{\nabla^2 f(x)} - \epsilon = \lambda_n^{\nabla^2 f(x)}/2 > 0.$$

□

Theorem 4 (Second Derivative Test for a Local Minimum)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2$ and suppose there is an $x \in \mathbb{R}^n$ with $\nabla f(x) = 0$ and $\nabla^2 f(x) \succ 0$. Then x is a local minimizer of f.*

Proof. By the previous lemma, there is a $\delta > 0$ such that $\nabla^2 f(y) \succ 0$ for all y with $\|y - x\| < \delta$. For any such y we can apply Taylor's theorem (letting $h = y - x$) to obtain

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x) \\ &= f(x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x) \\ &\geq f(x), \end{aligned}$$

for some $t \in (0, 1)$. The final inequality follows since $\|x - (x + t(y - x))\| = \|t(y - x)\| < \delta$ so the Hessian is positive definite and

$$\frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x) \geq 0.$$

□

12. If f, g are convex, is $h(x) = f(g(x))$ convex?

Solution. No. Let $f, g : \mathbb{R} \to \mathbb{R}$ with $f(x) = -x$ and $g(x) = x^2$. The statement is true if f is also increasing.

13. What is the relationship between $b \in \mathbb{R}^n$ and $S = \{x \in \mathbb{R}^n : b^T x \geq a\}$?

    Solution. S is a half-space on one side of a hyperplane. b is orthogonal to the hyperplane pointing into the half-space.

14. Let $S = \{x \in \mathbb{R}^2 : (Ax)_i \geq 0 \text{ for } i = 1, \ldots, m\}$ where $A \in \mathbb{R}^{m \times n}$. Give a geometric description of S.

    Solution. It is an intersection of half planes, also called a polyhedral set. If the resulting set is bounded we call it a convex polygon in $\mathbb{R}^2$, or a convex polytope in higher dimensions.