Optimization-Based Data Analysis

Recitation 9

1. Let $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable with $f''(x) \leq M$ for some $M \in \mathbb{R}$. If $f(0)$ and $f'(0)$ are known, give an upper bound on $f(h)$ for fixed $h \in \mathbb{R}$.

   Solution. By Taylor's theorem we have

   $$f(h) = f(0) + f'(0)h + \frac{1}{2}f''(\eta)h^2,$$

   for some η between 0 and h. Applying our bound shows

   $$f(h) \leq f(0) + f'(0)h + \frac{1}{2}Mh^2.$$

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

   (a) Show that
   $$f(\vec{x} + \vec{h}) = f(\vec{x}) + \nabla f(\vec{x})^T \vec{h} + \frac{1}{2}\vec{h}^T \nabla^2 f(\vec{\eta})\vec{h},$$
   where $\vec{\eta}$ is on the line segment between $\vec{x}$ and $\vec{x} + \vec{h}$.

   (b) Suppose all eigenvalues of $\nabla^2 f(\vec{x})$ lie in the interval $[m, M]$ where $0 < m \leq M$ for all $\vec{x} \in \mathbb{R}^n$. Use this to give bounds on $f(\vec{x} + \vec{h})$.

   Solution.

   (a) Define $g : [0, 1] \to \mathbb{R}$ by $g(s) = f(\vec{x} + s\vec{h})$. Then we have

   $$f(\vec{x} + \vec{h}) = g(1) = g(0) + g'(0) + \frac{1}{2}g''(\xi),$$

   where $\xi \in (0, 1)$. This gives

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

   so let $\vec{\eta} = \vec{x} + \xi\vec{h}$.

   (b) By the above we have

   $$f(\vec{x}) + \nabla f(\vec{x})^T \vec{h} + m\|\vec{h}\|_2^2/2 \leq f(\vec{x} + \vec{h}) \leq f(\vec{x}) + \nabla f(\vec{x})^T \vec{h} + M\|\vec{h}\|_2^2/2.$$

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with all eigenvalues of $\nabla^2 f(\vec{x})$ lying in the interval $[m, M]$ where $0 < m \leq M$ for all $\vec{x} \in \mathbb{R}^n$. We want to use gradient descent to minimize f.

(a) Explain why having the upper bound M is useful when proving gradient descent converges quickly.

(b) Explain why having the lower bound m is useful when proving gradient descent converges quickly.

Solution.

(a) If our step is $-s\nabla f(\vec{x})$ then we have

$$f(\vec{x} - s\nabla f(\vec{x})) \leq f(\vec{x}) - s\|\nabla f(\vec{x})\|_2^2 + Ms^2\|\nabla f(\vec{x})\|_2^2/2.$$

This quadratic can be minimized over $s = 1/M$. Intuitively, if there is no upper bound on the Hessian then the gradient can change quickly, and thus we need to take very short steps.

(b) Note that

$$f(\vec{x} + \vec{h}) \geq f(\vec{x}) + \nabla f(\vec{x})^T\vec{h} + \frac{m}{2}\|\vec{h}\|_2^2.$$

The righthand side is a quadratic in $\vec{h}$ that is minimized (by computing the gradient) at $\vec{h} = -\frac{1}{m}\nabla f(\vec{x})$. This gives

$$f(\vec{x} + \vec{h}) \geq f(\vec{x}) - \frac{1}{2m}\|\nabla f(\vec{x})\|_2^2.$$

If we let $\vec{x} + \vec{h} = \vec{x}^*$, the unique global minimizer of f (since f is strictly convex) we see

$$2m(f(\vec{x}) - f(\vec{x}^*)) \leq \|\nabla f(\vec{x})\|_2^2.$$

Combining with the previous part shows, for $s = 1/M$, that

$$f(\vec{x} - s\nabla f(\vec{x})) - f(\vec{x}^*) \leq f(\vec{x}) - f(\vec{x}^*) - \frac{1}{2M}\|\nabla f(\vec{x})\|_2^2 \leq (f(\vec{x}) - f(\vec{x}^*))(1 - m/M).$$

Intuitively, this says if there is no lower bound on the Hessian, then $\nabla f(\vec{x})$ can be small, and thus we may not make much progress each step.

4. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(\vec{x}) = \frac{1}{2}\vec{x}^TA\vec{x}$ where A is symmetric.

(a) Under what conditions is f convex/concave?

(b) Under what conditions does f have a local minimum/maximum at 0?

(c) What are the possible shapes for the contour lines of f?

Solution.

(a) Convex if A is positive semidefinite, concave if negative semidefinite.

(b) If convex/concave.

(c) Elliptical (if strictly convex or concave), hyperbolic (if indefinite), or lines (if non-strictly convex or concave).

5. Show that if $A \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite then $A = BB^T$ for some $B \in \mathbb{R}^{n \times n}$.

   Solution. Did this on homework. Write $A = QDQ^T$ and let $B = Q\sqrt{D}$. Alternatively, let $B = Q\sqrt{D}Q^T$ where B is symmetric.

6. Suppose $A \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite.

   (a) Must the trace of A be non-negative?

   (b) Must every diagonal element of A be non-negative?

   (c) Suppose every diagonal element of A is 1 and $n > 2$.

      i. Must A be positive definite?

      ii. Are there any restrictions on the value of $A_{12} = A_{21}$?

      iii. Suppose $A_{12} = A_{21} = 1$. Must A not be positive definite?

   Solution.

   (a) Note that $\operatorname{tr} A = \operatorname{tr} QDQ^T = \operatorname{tr} DQQ^T = \operatorname{tr} D$, where $A = QDQ^T$ is the spectral decomposition.

   (b) Note that $D_{ii} = \operatorname{tr} e_i^T A e_i \geq 0$.

      i. No. For example, if the entire matrix was 1 it wouldn't be invertible.

      ii. Yes, $A_{12} \in [-1, 1]$ by Cauchy-Schwarz. To see this, write $A = B^T B$ showing that every column of B has norm 1.

      iii. Yes, the first two rows of A must be equal by Cauchy-Schwarz.