Problem Set 0

Jonathan Qin

June 13, 2024

1 Gradients and Hessians

(a) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla f(x)$?

$\nabla f(x) = \nabla(\frac{1}{2}x^T Ax) + \nabla(b^T)$. Let $m(x) = x^T Ax$, $n(x) = b^T x$

$$m(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j$$
$$n(x) = \sum_{i=1}^{n} b_i x_i$$

$$\begin{aligned}
\frac{\partial}{\partial x_k}(x^T Ax) &= \frac{\partial}{\partial x_k}(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(\frac{\partial x_i}{\partial x_k} A_{ij} x_j + x_i A_{ij}\frac{\partial x_j}{x_k}) \\
&= \sum_{j=1}^{n} A_{kj} x_j + \sum_{i=1}^{n} x_i A_{ik}\ (\frac{\partial x_i}{\partial x_k} = \delta_{ik}, \text{ Kronecker delta, is 1 if } i = k, \text{ 0 otherwise}) \\
&= 2\sum_{i=1}^{n} A_{ki} x_i\ (A \text{ is symmetric, } A_{ij} = A_{ji}) \\
&= 2Ax \\
\frac{\partial}{\partial x_k}(b^T x) &= \frac{\partial}{\partial x_k}(\sum_{i=1}^{n} b_i x_i) \\
&= b
\end{aligned}$$

Hence, $\nabla f(x) = \frac{1}{2} \cdot 2Ax + b = Ax + b$

(b) Let $f(x) = g(h(x))$,where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

Chain rule:

$$\begin{aligned}\frac{\partial g(h(x))}{\partial x_i} &= \frac{\partial g(h(x))}{\partial h(x)}\frac{\partial h(x)}{\partial x_i} \\ &= g'(h(x))\frac{\partial h(x)}{\partial x_i} \nabla f(x) \quad = \nabla g(h(x)) = g'(h(x))\nabla h(x)\end{aligned}$$

(c) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is symmetric and $b \in \mathbb{R}^n$ is a vector. what is $\nabla^2 f(x)$?

$$\begin{aligned}\nabla^2 f(x) &= \nabla(Ax + b) \\ &= [\frac{\partial}{\partial x_1}f(x)\frac{\partial}{\partial x_2}f(x)\cdots\frac{\partial}{\partial x_n}f(x)] \\ &= \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn}\end{bmatrix} = A\end{aligned}$$

(d) Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$?

$$\begin{aligned}\nabla f(x) &= \nabla g(a^T x) \\ &= g'(a^T x) \cdot \nabla(a^T x) \; (a^T x = \sum_{i=1}^{n} a_i x_i) \\ &= g'(a^T x) \cdot a\end{aligned}$$

$$\frac{\partial^2 g(h(x))}{\partial x_i \partial x_j} = \frac{\partial^2 g(h(x))}{\partial (h(x))^2}\frac{\partial h(x)}{\partial x_i}\frac{\partial h(x)}{\partial x_j} = g''(h(x))\frac{\partial h(x)}{\partial x_i}\frac{\partial h(x)}{\partial x_j}$$

$$\frac{\partial^2 g(a^T x)}{\partial x_i \partial x_j} = g''(a^T x)\frac{\partial(a^T x)}{\partial x_i}\frac{\partial(a^T x)}{\partial x_j} = g''(a^T x) a_i a_j$$

$$\nabla^2 f(x) = \nabla^2 g(a^T x) = g''(a^T x)\begin{bmatrix} a_1a_1 & a_1a_2 & \dots & a_1a_n \\ a_2a_1 & a_2a_2 & \dots & a_2a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_na_1 & a_na_2 & \dots & a_na_n\end{bmatrix} = g''(a^T x)aa^T$$

2 Positive definite matrices

(a) Let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semidefinite.

$A^T = (zz^T)^T = (z^T)^T z^T = zz^T = A$

$x^T zz^T x = x^T z(x^T z)^T = (x^T z)^2 >= 0$

(b) Let $z \in \mathbb{R}^n$ be a non-zero n-vector. Let $A = zz^T$. What is the null-space of A? What is the rank of A?

$Ax = 0 \rightarrow zz^T x = 0$

Since z is a non-zero vector, $z^T x = 0, Null(A) = \{x \in \mathbb{R}^n : x^T z = 0\}$

$Rank(A) = 1$. It is the number of independent columns, which in this case, all columns are a multiple of z.

(c) Let $A \in \mathbb{R}^{n \times n}$ be positive semidefinite and $B \in \mathbb{R}^{m \times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so, prove it. If not, give a counterexample with explicit A, B.

$Proof:$

$(BAB^T)^T = (B^T)^T A^T B^T = BAB^T$(since A is symmetric)

$x^T BAB^T x = (x^T B)A(x^T B)^T \geq 0$

3 Eigenvectors, eigenvalues, and the spectral theorem

(a) Suppose that the matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $A = T\Lambda T^{-1}$ for an invertible matrix $T \in \mathbb{R}^{n \times n}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Use the notation $t^{(i)}$ for the columns of T, so that $T = \begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix}$, where $t^{(i)} \in \mathbb{R}^n$. Show that $At^{(i)} = \lambda_i t^{(i)}$, so that the eigenvalues/eigenvector pairs of A are $(t^{(i)}, \lambda_i)$

$$A = T\Lambda T^{-1}, AT = \Lambda T$$

$$A[t^{(1)}, t^{(2)}, \ldots, t^{(n)}] = [t^{(1)}, t^{(2)}, \ldots, t^{(n)}] * \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$
$$[At^{(1)}, At^{(2)}, \ldots, At^{(n)}] = [\lambda_1 t^{(1)}, \lambda_2 t^{(2)}, \ldots, \lambda_n t^{(n)}]$$
$$At^{(i)} = \lambda_i t^{(i)}$$

(b) Let A be symmetric. Show that if $U = \begin{bmatrix} u^{(1)} & \cdots & u^{(n)} \end{bmatrix}$ is orthogonal, where $u^{(i)} \in \mathbb{R}^n$ and $A = U\Lambda U^T$, then $u^{(i)}$ is an eigenvector of A and $Au^{(i)} = \lambda_i u^{(i)}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

$$AU = U\Lambda U^T U = U\Lambda$$
$$A \begin{bmatrix} u^{(1)} & u^{(2)} & \cdots & u^{(n)} \end{bmatrix} = \begin{bmatrix} u^{(1)} & u^{(2)} & \cdots & u^{(n)} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$
$$\begin{bmatrix} Au^{(1)} & Au^{(2)} & \cdots & Au^{(n)} \end{bmatrix} = \begin{bmatrix} \lambda_1 u^{(1)} & \lambda_2 u^{(2)} & \cdots & \lambda_n u^{(n)} \end{bmatrix}$$
$$Au^{(i)} = \lambda_i u^{(i)}$$

(c) Show that if A is PSD, then $\lambda_i(A) \geq 0$ for each i.

$$At^{(i)} = \lambda_i t^{(i)}$$
$$(t^(i))^T At^{(i)} = \lambda_i ||t^{(i)}||_2 = \lambda_i \geq 0$$