1. **[0 points] Gradients and Hessians**

Recall that a matrix $A \in \mathbb{R}^{n \times n}$ is *symmetric* if $A^T = A$, that is, $A_{ij} = A_{ji}$ for all i, j. Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

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

$x \in \mathbb{R}^n$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}\left\{\frac{1}{2} x^T A x + b^T x\right\}$$

dot product

$$= \frac{\partial}{\partial x} \frac{1}{2} x^T A x + \frac{\partial}{\partial x} b^T x$$

$$= \frac{1}{2}(2Ax) + b$$

$$= Ax + b$$

$$\nabla f(x) = Ax + b$$

1. **[0 points] Gradients and Hessians**

Recall that a matrix $A \in \mathbb{R}^{n \times n}$ is *symmetric* if $A^T = A$, that is, $A_{ij} = A_{ji}$ for all i, j. Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

(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 $\nabla f(x)$?

$h = A^T x + c'$ $\quad A \in \mathbb{R}^n$

$g(z) = c''(z) + c'''$ $\quad c \in \mathbb{R}$

$\frac{\partial h}{\partial} = A$

$f(x) = g(h(x))$

$f'(x) = g'(h(x))\, h'(x)$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial g(h(x))}{\partial x} = \underbrace{\frac{\partial g(h(x))}{\partial h(x)}}_{\in \mathbb{R}} \times \underbrace{\frac{\partial h(x)}{\partial (x)}}_{\in \mathbb{R}^n}$$

1. **[0 points] Gradients and Hessians**

Recall that a matrix $A \in \mathbb{R}^{n\times n}$ is *symmetric* if $A^T = A$, that is, $A_{ij} = A_{ji}$ for all i, j. Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

(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 $\nabla^2 f(x)$?

(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 $\nabla^2 f(x)$? (*Hint:* your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

(c) $\frac{\partial f(x)}{\partial x} = Ax + b$

$\frac{\partial f(x)'}{\partial x} = A$

1. **[0 points] Gradients and Hessians**

Recall that a matrix $A \in \mathbb{R}^{n\times n}$ is *symmetric* if $\boxed{A^T = A, \text{ that is, } A_{ij} = A_{ji} \text{ for all } i, j.}$ Also recall the gradient $\nabla f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is the n-vector of partial derivatives

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \quad \text{where} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

The hessian $\nabla^2 f(x)$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ symmetric matrix of twice partial derivatives,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2^2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n^2} f(x) \end{bmatrix}.$$

$\partial^2 g(x) \neq \partial g(x) \partial g(x)$

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

(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 $\nabla^2 f(x)$? (*Hint:* your expression for $\nabla^2 f(x)$ may have as few as 11 symbols, including $'$ and parentheses.)

$$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial g(a^T x)}{\partial x} = \frac{\partial g(t)}{\partial t} \times \frac{\partial t}{\partial x} = g'(t) \times \frac{\partial (a^T x)}{\partial x}$$

$$= g'(t) \times a$$

$$= g'(a^T x) \times a = \frac{\partial g(a^T x)}{\partial a^T x} \times a$$

$h(x) = a^T x$

$f(x) = g(h(x))$

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

$$= \frac{\partial \partial g(h(x))}{\partial h(x) \partial h(x)} \times a_i a_j$$

$$= \frac{\partial g'(h(x))}{\partial h(x)} \times a_i a_j$$

$$= g''(h(x)) \times a_i a_j$$

$$= g''(a^T x) a_i a_j$$