EECS 127/227AT Optimization Models in Engineering
Spring 2020

Discussion 2

1. Gradients and Hessians

(a) The *Gradient* of a scalar-valued function $g : \mathbb{R}^n \to \mathbb{R}$, is the column vector of length n, denoted ∇g, containing the derivatives of components of g with respect to the variables:

$$(\nabla g(\vec{x}))_i = \frac{\partial g}{\partial x_i}(\vec{x}),\ i = 1, \ldots n.$$

Compute the gradient, $\nabla g(\vec{x})$, of:

i. $g(\vec{x}) = \vec{c}^\top \vec{x}$

ii. $g(\vec{x}) = \vec{x}^\top \vec{x}$

iii. $g(\vec{x}) = \ln\left(\sum\limits_{i=1}^{n} e^{x_i}\right)$

Solution:

i. Note $g(\vec{x}) = \vec{c}^\top \vec{x} = \sum_{i=1}^n c_i x_i$. Then $\frac{\partial g}{\partial x_i}(\vec{x}) = c_i$. It follows that $\nabla g(\vec{x}) = \vec{c}$.

ii. Note $g(\vec{x}) = \vec{x}^\top \vec{x} = \sum\limits_{i=1}^{n} x_i^2$. Then $\frac{\partial g}{\partial x_i}(\vec{x}) = 2x_i$. It follows that $\nabla g(\vec{x}) = 2\vec{x}$.

iii. $\nabla g(\vec{x}) = \frac{\vec{z}}{Z}$, where $\vec{z} = \begin{bmatrix} e^{x_1} & \cdots & e^{x_n} \end{bmatrix}^\top$, and $Z = \sum\limits_{i=1}^{n} z_i$

(b) The *Hessian* of a scalar-valued function $g : \mathbb{R}^n \to \mathbb{R}$, is the $n \times n$ matrix, denoted as $\nabla^2 g$, containing the second derivatives of components of g with respect to the variables:

$$(\nabla^2 g(\vec{x}))_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}(\vec{x}),\ \ i = 1, \ldots, n,\ \ j = 1, \ldots, n.$$

Compute the Hessian, $\nabla^2 g(\vec{x})$, of:

i. g($\vec{x}$)= $\vec{c}^\top \vec{x}$

ii. g($\vec{x}$) $= \vec{x}^\top \vec{x}$.

iii. g($\vec{x}$) $= \vec{x}^\top \mathrm{A} \vec{x}$.

Solution:

i. From a), $\nabla g(\vec{x}) = \vec{c}$. Since $\nabla g(\vec{x})$ is not a function of $\vec{x}$, the second order derivatives with respect to x_i are all zero and it follows that $\nabla^2 g = 0_{n\times n}$ where $0_{n\times n}$ denotes a $n \times n$ matrix of all zeros.

ii. In part a) we saw that, $g(\vec{x}) = \vec{x}^\top \vec{x} = \sum_{i=1}^n x_i^2$.
Note for $i \neq j$, we have,

$$\frac{\partial^2 g}{\partial x_i \partial x_j} = 0,$$

and for $i = j$ we have,

$$\frac{\partial^2 g}{\partial x_i^2} = 2.$$

Hence it follows that $\nabla^2 g(\vec{x}) = 2I_n$ where I_n is the $n \times n$ identity matrix.

iii. Let $A = [\vec{a}_1, \ \vec{a}_2, \ \ldots \ , \vec{a}_n]$ where $\vec{a}_i$ is the i-th column of A. Similarly, let $\vec{a}_i^\top$ be the i-th row of $A^\top$. For notational convenience, let $\vec{\alpha}_i^T$ denote the i-th row of A. Finally, let a_{ij} denote the (i, j)th entry of A. Then

$$\begin{aligned} g(x) &= \vec{x}^\top A\vec{x} \\ &= \vec{x}^\top [\vec{a}_1, \ \vec{a}_2, \ \ldots \ , \vec{a}_n]\vec{x} \\ &= \vec{x}^\top (\vec{a}_1 x_1 + \vec{a}_2 x_2 + \ \ldots \ + \vec{a}_n x_n) \\ &= \sum_{i=1}^{n} (\vec{x}^\top \vec{a}_i) x_i. \end{aligned}$$

Then,

$$\begin{aligned} \frac{\partial g}{\partial x_j}(\vec{x}) &= \frac{\partial}{\partial x_j}\left[(\vec{x}^\top \vec{a}_j)x_j + \sum_{i \neq j} (\vec{x}^\top \vec{a}_i) x_i\right] \\ &= \vec{x}^\top \vec{a}_j + a_{jj} x_j + \sum_{i \neq j} a_{ji} x_i \\ &= \vec{a}_j^\top \vec{x} + \vec{\alpha}_j^\top \vec{x}. \end{aligned}$$

It follows that $\nabla g(\vec{x}) = (A + A^\top)\vec{x}$. Note if A is symmetric this reduces to $2A\vec{x}$. Based on the definition of the Hessian, it follows that the ith column of the Hessian is the ith column of $A + A^\top$. Thus $\nabla^2 g(\vec{x}) = A + A^\top$.

2. Gradients with respect to matrices (OPTIONAL)

Assume that $A \in \mathbb{R}^{p\times m}, C, X \in \mathbb{R}^{m\times n}$, $\Sigma \in \mathbb{R}^{m\times m}$ and $\vec{a} \in \mathbb{R}^m, \vec{b} \in \mathbb{R}^n$. Find the following gradients and specify the dimensions of the gradients.

(a) $\nabla_X \mathrm{tr}(X^\top C)$

Solution: We have

$$\frac{\partial \mathrm{tr}(X^\top C)}{\partial X_{ij}} = \frac{\partial \sum_i \sum_j X_{ij} C_{ij}}{\partial X_{ij}} = C_{ij},$$

and therefore,

$$\nabla_X \mathrm{tr}(X^\top C) = \nabla_X \langle X, C\rangle = C, \tag{1}$$

and the dimension of the gradient is $m \times n$, which is equal to the dimension of X.

(b) $\nabla_X(\vec{a}^\top X\vec{b})$

Solution: For each X_{ij},

$$\frac{\partial \vec{a}^\top X\vec{b}}{\partial X_{ij}} = \frac{\partial \sum_i \sum_j a_i b_j X_{ij}}{\partial X_{ij}} = a_i b_j.$$

Therefore, we have

$$\nabla_X(\vec{a}^\top X\vec{b}) = \vec{a}\vec{b}^\top,$$

and the dimension of the gradient is $m \times n$. Note that the derivative is the transpose of the gradient.

(c) $\nabla_{\Sigma^{-1}} \text{tr}(X^\top \Sigma^{-1} X)$

Solution: Note that by the properties of trace, we have

$$\text{tr}(X^\top \Sigma^{-1} X) = \text{tr}(\Sigma^{-1} X X^\top) = \langle \Sigma^{-1}, X X^\top \rangle.$$

Then, based on the analogy to (1), we can write:

$$\nabla_{\Sigma^{-1}} \langle \Sigma^{-1}, X X^\top \rangle = X X^\top,$$

and the dimension of the gradient is $m \times m$. The fact that Σ^{-1} is the inverse of a matrix is of no consequence; it is only a dummy variable.

(d) $\nabla_X \|AX\|_F^2$

Solution: Note that $\|AX\|_F^2 = \langle AX, AX \rangle = \text{tr}(X^\top A^\top AX)$.

Remember the product rule for univariate functions: given two functions $f, g : \mathbb{R} \mapsto \mathbb{R}$, we have

$$\begin{aligned}
\frac{d}{dx}(f(x)g(x)) &= \frac{d}{dx_1}(f(x_1)g(x))\bigg|_{x_1=x} + \frac{d}{dx_2}(f(x)g(x_2))\bigg|_{x_2=x} \\
&= g(x)\frac{df(x)}{dx} + f(x)\frac{dg(x)}{dx}.
\end{aligned}$$

Similarly, we can use the product rule for gradients:

$$\begin{aligned}
\nabla_X \text{tr}(X^\top A^\top AX) &= \nabla_{X_1} \text{tr}(X_1^\top A^\top AX)\Big|_{X_1=X} + \nabla_{X_2} \text{tr}(X^\top A^\top AX_2)\Big|_{X_2=X} \\
&= A^\top AX + A^\top AX = 2A^\top AX.
\end{aligned}$$

The dimension of the gradient is $m \times n$.

3. Jacobians (OPTIONAL)

The *Jacobian* of a vector-valued function $g : \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix, denoted as Dg, containing the derivatives of the components of g with respect to the variables:

$$(Dg)_{ij} = \frac{\partial g_i}{\partial x_j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

(a) Compute the Jacobian of $g(\vec{x}) = A\vec{x}$

Solution: Note $g_i(\vec{x}) = \vec{\alpha_i}^\top \vec{x}$ where $\vec{\alpha_i}^\top$ is the i-th row of A. Then $\frac{\partial g_i}{\partial x_j} = \alpha_{ij}$ which is simply the (i, j) entry of A. It follows that $Dg(\vec{x}) = A$.