3rd October Lecture Notes

MAT 180

4.3.1 Matrix Differentiation

Consider $f : \mathbb{R}^n \to \mathbb{R}^m$. This can be rewritten as

$$f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $1 \leq i \leq m$. Define the Jacbobian of f to be

$$J(f) = \frac{\partial}{\partial \vec{x}} f(\vec{x}) = \left(\frac{\partial f_i(\vec{x})}{\partial x_j} \right)_{i,j=1,1}^{m,n}$$

Example. Consider $f(x, y, z) = \begin{bmatrix} x^2 + yz \\ xyz \end{bmatrix} \in \mathbb{R}^2$. So $f_1(\vec{x}) = x^2 + xyz$ and $f_2(\vec{x}) = xyz$. Then

$$J(f) = \frac{\partial}{\partial \vec{x}} f(\vec{x}) = \begin{bmatrix} \frac{\partial}{\partial x} f_1 & \frac{\partial}{\partial y} f_1 & \frac{\partial}{\partial z} f_1 \\ \frac{\partial}{\partial x} f_2 & \frac{\partial}{\partial y} f_2 & \frac{\partial}{\partial z} f_2 \end{bmatrix}$$

Remark. If $f : \mathbb{R}^m \to \mathbb{R}$ then $\dfrac{\partial f(\vec{x})}{\vec{x}} = \nabla_x f(\vec{x})^T$

Example. $f(\vec{x}) = x^2 + y^2 + z^2$. Since $f : \mathbb{R}^3 \to \mathbb{R}$, then

$$\frac{\partial f}{\partial \vec{x}}(\vec{x}) = (2x, 2y, 2z) = \nabla_{\vec{x}} f(\vec{x})^T$$

Example. $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and define $f(\vec{x}) = A\vec{x} + \vec{b}$, where $f : \mathbb{R}^n \to \mathbb{R}$. We want to compute the Jacobian. First, note that

$$f_i(\vec{x}) = A_{i,*}\vec{x} + b_i = \left(\sum_{i,j} A_{i,j}x_j\right) + b_i$$

With this in mind,

$$\frac{\partial f}{\partial \vec{x}}(\vec{x}) = (A_{i,j})_{i,j=1,1}^{m,n} = A$$

Corollary. For $\vec{x} \in \mathbb{R}^n$, we have that $\dfrac{\partial}{\partial \vec{x}}\vec{x} = \vec{I}_{n\times n}$

Rules

1. $\dfrac{\partial}{\partial \vec{x}}(A\vec{x} + \vec{b}) = A$ ($A, \vec{b}$ constant)

2. $\dfrac{\partial}{\partial \vec{x}}(\vec{x}^T A) = A^T$ (A constant)

3. $\dfrac{\partial}{\partial \vec{x}}(\vec{x}^T A\vec{x}) = \vec{x}^T(A + A^T)$ (A constant)

4. For f, g functions with compatible dimension,
   $\dfrac{\partial}{\partial \vec{x}}f(\vec{x})^T g(\vec{x}) = g(\vec{x})^T \dfrac{\partial}{\partial \vec{x}}f(\vec{x}) + f(\vec{x})^T \dfrac{\partial}{\partial g(\vec{x})}$

5. For f, g functions with compatible dimension,
   $\dfrac{\partial}{\partial \vec{x}}f(\vec{x})^T A g(\vec{x}) = g(\vec{x})^T A^T \dfrac{\partial}{\partial \vec{x}}f(\vec{x}) + f(\vec{x})^T A \dfrac{\partial}{\partial \vec{x}}g(\vec{x})$

More rules can be found in “Matrix Calculus” on Canvas.

Returning to an edifying calculation from the PCA lecture:

We had that $\nabla_{\vec{y}}(-2\vec{x}^T D\vec{y} + \vec{y}^T \vec{y}) = 0$. The Jacobian of this expression is

$$
\begin{aligned}
&\frac{\partial}{\partial \vec{y}}(-2\vec{x}^T D\vec{y} + \vec{y}^T \vec{y}) = 0 \\
&\implies -2\frac{\partial}{\partial \vec{y}}(\vec{x}^T D\vec{y}) + \frac{\partial}{\partial \vec{y}}(\vec{y}^T \vec{y}) = 0 \\
&\implies {\scriptstyle\text{(Rule 5)}} - 2\left(\vec{y}^T D^T \frac{\partial}{\partial \vec{y}}\vec{x} + \vec{x}^T D \frac{\partial}{\partial \vec{y}}\vec{y}\right) + \frac{\partial}{\partial \vec{y}}(\vec{y}^T y) = 0 \\
&\implies -2\vec{x}^T DI + \vec{y}^T(I + I) = 0 \\
&\implies -2\vec{X}^T D = -2y^T \\
&\implies y = D^T \vec{x}
\end{aligned}
$$

The **Hessian matrix** of $f : \mathbb{R}^n \to \mathbb{R}$ is

$$J(\nabla_{\vec{x}} f(\vec{x})) = \frac{\partial}{\partial \vec{x}}\frac{\partial}{\partial \vec{x}} f(\vec{x}) = \left(\frac{\partial^2}{\partial x_i \partial x_j} f(\vec{x})\right)_{i,j=1,1}^{n,n}$$

Remark. At points $\vec{x}$ such that all 2nd partials of f are continuous,

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(\vec{x}) = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(\vec{x})$$

so $H(f)(\vec{x})$ is symmetric, so $H(f)(\vec{x})$ has an orthonormal eigenvector decomposition and has real eigenvalues.

Another remark. For $\vec{u}$ a unit vector,

$$\vec{u}^T H(f)(\vec{x})\vec{u}$$

is the 2nd derivative of $f(\vec{x})$ in the direction of $\vec{u}$.

Returning to the Concept of Gradient Descent.

Our objective is to find $\text{argmin}_{\vec{x}} f(\vec{x})$. Let $\vec{g} = \nabla_{\vec{x}} f(\vec{x})$ and $H = H(f)(\vec{x}_0)$. Let the unit vector $\vec{u}$ be defined as $\vec{u} = \dfrac{\vec{x} - \vec{x}_0}{||\vec{x} - \vec{x}_0||}$.
Then applying the Taylor expansion for $f(\vec{x})$,

$$f(\vec{x}) \approx f(\vec{x}_0) + (\vec{x} - \vec{x}_0)^T \vec{g} + \frac{1}{2}(\vec{x} - \vec{x}_0)^T H (\vec{x} - \vec{x}_0)$$