Peeter Joot
peeterjoot@protonmail.com

Jacobian and Hessian matrices

1.1 Motivation

In class this Friday the Jacobian and Hessian matrices were introduced, but I did not find the treatment terribly clear. Here is an alternate treatment, beginning with the gradient construction from [2], which uses a nice trick to frame the multivariable derivative operation as a single variable Taylor expansion.

1.2 Multivariable Taylor approximation

The Taylor series expansion for a scalar function $g : \mathbb{R} \rightarrow \mathbb{R}$ about the origin is just

$$g(t) = g(0) + t g'(0) + \frac{t^2}{2} g''(0) + \cdots \tag{1.1}$$

In particular

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(0) + \cdots \tag{1.2}$$

Now consider $g(t) = f(\mathbf{x} + \mathbf{a}t)$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(0) = f(\mathbf{x})$, and $g(1) = f(\mathbf{x} + \mathbf{a})$. The multivariable Taylor expansion now follows directly

$$f(\mathbf{x} + \mathbf{a}) = f(\mathbf{x}) + \left. \frac{df(\mathbf{x} + \mathbf{a}t)}{dt} \right|_{t=0} + \frac{1}{2} \left. \frac{d^2 f(\mathbf{x} + \mathbf{a}t)}{dt^2} \right|_{t=0} + \cdots \tag{1.3}$$

The first order term is

$$\begin{aligned} \left. \frac{df(\mathbf{x} + \mathbf{a}t)}{dt} \right|_{t=0} &= \sum_{i=1}^n \frac{d(x_i + a_i t)}{dt} \left. \frac{\partial f(\mathbf{x} + \mathbf{a}t)}{\partial (x_i + a_i t)} \right|_{t=0} \\ &= \sum_{i=1}^n a_i \frac{\partial f(\mathbf{x})}{\partial x_i} \\ &= \mathbf{a} \cdot \boldsymbol{\nabla} f. \end{aligned} \tag{1.4}$$

Similarly, for the second order term

$$
\begin{aligned}
\left.\frac{d^2 f(\mathbf{x}+\mathbf{a}t)}{dt^2}\right|_{t=0} &= \left.\left(\frac{d}{dt}\left(\sum_{i=1}^{n} a_i \frac{\partial f(\mathbf{x}+\mathbf{a}t)}{\partial(x_i+a_i t)}\right)\right)\right|_{t=0} \\
&= \left.\left(\sum_{j=1}^{n}\frac{d(x_j+a_j t)}{dt}\sum_{i=1}^{n} a_i \frac{\partial^2 f(\mathbf{x}+\mathbf{a}t)}{\partial(x_j+a_j t)\partial(x_i+a_i t)}\right)\right|_{t=0} \\
&= \sum_{i,j=1}^{n} a_i a_j \frac{\partial^2 f}{\partial x_i \partial x_j} \\
&= (\mathbf{a}\cdot\boldsymbol{\nabla})^2 f.
\end{aligned}
\tag{1.5}
$$

The complete Taylor expansion of a scalar function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is therefore

$$
f(\mathbf{x}+\mathbf{a}) = f(\mathbf{x}) + \mathbf{a}\cdot\boldsymbol{\nabla} f + \frac{1}{2}(\mathbf{a}\cdot\boldsymbol{\nabla})^2 f + \cdots, \tag{1.6}
$$

so the Taylor expansion has an exponential structure

$$
f(\mathbf{x}+\mathbf{a}) = \sum_{k=0}^{\infty}\frac{1}{k!}(\mathbf{a}\cdot\boldsymbol{\nabla})^k f = e^{\mathbf{a}\cdot\boldsymbol{\nabla}} f. \tag{1.7}
$$

Should an approximation of a vector valued function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be desired it is only required to form a matrix of the components

$$
\mathbf{f}(\mathbf{x}+\mathbf{a}) = \mathbf{f}(\mathbf{x}) + [\mathbf{a}\cdot\boldsymbol{\nabla} f_i]_i + \frac{1}{2}[(\mathbf{a}\cdot\boldsymbol{\nabla})^2 f_i]_i + \cdots, \tag{1.8}
$$

where $[.]_i$ denotes a column vector over the rows $i \in [1, m]$, and f_i are the coordinates of $\mathbf{f}$.

1.3 The Jacobian matrix

In [1] the Jacobian $D\mathbf{f}$ of a function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined in terms of the limit of the l_2 norm ratio

$$
\frac{\|\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{x}) - (D\mathbf{f})(\mathbf{z}-\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2}, \tag{1.9}
$$

with the statement that the function $\mathbf{f}$ has a derivative if this limit exists. Here the Jacobian $D\mathbf{f} \in \mathbb{R}^{m\times n}$ must be matrix valued.

Let $\mathbf{z} = \mathbf{x} + \mathbf{a}$, so the first order expansion of eq. (1.8) is

$$
\mathbf{f}(\mathbf{z}) = \mathbf{f}(\mathbf{x}) + [(\mathbf{z}-\mathbf{x})\cdot\boldsymbol{\nabla} f_i]_i. \tag{1.10}
$$

With the (unproven) assumption that this Taylor expansion satisfies the norm limit criteria of eq. (1.9), it is possible to extract the structure of the Jacobian by comparison

$$
\begin{aligned}
(D\mathbf{f})(\mathbf{z}-\mathbf{x}) &= \left[(\mathbf{z}-\mathbf{x})\cdot\boldsymbol{\nabla} f_i\right]_i \\
&= \left[\sum_{j=1}^{n}(z_j - x_j)\frac{\partial f_i}{\partial x_j}\right]_i \\
&= \left[\frac{\partial f_i}{\partial x_j}\right]_{ij}(\mathbf{z}-\mathbf{x}),
\end{aligned}
\tag{1.11}
$$

so

$$\boxed{(D\mathbf{f})_{ij} = \frac{\partial f_i}{\partial x_j}} \tag{1.12}$$

Written out explictly as a matrix the Jacobian is

$$D\mathbf{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\nabla} f_1)^{\mathrm{T}} \\ (\boldsymbol{\nabla} f_2)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{\nabla} f_m)^{\mathrm{T}} \end{bmatrix}. \tag{1.13}$$

In particular, when the function is scalar valued

$$Df = (\boldsymbol{\nabla} f)^{\mathrm{T}}. \tag{1.14}$$

With this notation, the first Taylor expansion, in terms of the Jacobian matrix is

$$\boxed{\mathbf{f}(\mathbf{z}) \approx \mathbf{f}(\mathbf{x}) + (D\mathbf{f})\,(\mathbf{z} - \mathbf{x})\,.} \tag{1.15}$$

1.4 The Hessian matrix

For scalar valued functions, the text expresses the second order expansion of a function in terms of the Jacobian and Hessian matrices

$$f(\mathbf{z}) \approx f(\mathbf{x}) + (Df)\,(\mathbf{z} - \mathbf{x}) + \frac{1}{2}\,(\mathbf{z} - \mathbf{x})^{\mathrm{T}}\,(\boldsymbol{\nabla}^2 f)\,(\mathbf{z} - \mathbf{x})\,. \tag{1.16}$$

Because $\boldsymbol{\nabla}^2$ is the usual notation for a Laplacian operator, this $\boldsymbol{\nabla}^2 f \in \mathbb{R}^{n\times n}$ notation for the Hessian matrix is not ideal in my opinion. Ignoring that notational objection for this class, the structure of the Hessian matrix can be extracted by comparison with the coordinate expansion

$$\mathbf{a}^{\mathrm{T}}(\boldsymbol{\nabla}^2 f)\mathbf{a} = \sum_{r,s=1}^{n} a_r a_s \frac{\partial^2 f}{\partial x_r \partial x_s} \tag{1.17}$$

so

$$\boxed{(\boldsymbol{\nabla}^2 f)_{ij} = \frac{\partial^2 f_i}{\partial x_i \partial x_j}.} \tag{1.18}$$

In explicit matrix form the Hessian is

$$\boldsymbol{\nabla}^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}. \tag{1.19}$$

Is there a similar nice matrix structure for the Hessian of a function $f : \mathbb{R}^n \to \mathbb{R}^m$?

Bibliography

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 1.3

[2] D. Hestenes. *New Foundations for Classical Mechanics*. Kluwer Academic Publishers, 1999. 1.1