Appendix B: Notation and Taylor Series

The following is a generic review of Taylor series and the type of notation we use in certain parts of the course.

REAL-VALUED FUNCTIONS: Let $h(\boldsymbol{x}, \alpha)$ be a real-valued function of a vector $\boldsymbol{x}$ (which is irrelevant to the developments here) and a $(r \times 1)$ vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)^T$. We write

$$h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}) = \partial/\partial\boldsymbol{\alpha}\, h(\boldsymbol{x}, \boldsymbol{\alpha}) = \begin{pmatrix} \partial/\partial\alpha_1 h(\boldsymbol{x}, \boldsymbol{\alpha}) \\ \vdots \\ \partial/\partial\alpha_r h(\boldsymbol{x}, \boldsymbol{\alpha}) \end{pmatrix} \quad (r \times 1). \tag{B.1}$$

The vector $h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha})$ of partial derivatives of $h(\boldsymbol{x}, \boldsymbol{\alpha})$ with respect to the elements of $\boldsymbol{\alpha}$ is referred to as the ***gradient*** (vector). Of course, $h_\alpha^T(\boldsymbol{x}, \boldsymbol{\alpha})$ denotes its transpose, a $(1 \times r)$ vector.

We can extend this notation to second partial derivatives with respect to the elements of $\boldsymbol{\alpha}$. This is done by writing the $(r \times r)$ symmetric matrix of second partial derivatives of $h(\boldsymbol{x}, \boldsymbol{\alpha})$ as follows.

$$h_{\alpha\alpha}(\boldsymbol{x}, \boldsymbol{\alpha}) = \partial/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T h(\boldsymbol{x}, \boldsymbol{\alpha}) = \begin{pmatrix} \partial^2/\partial\alpha_1^2 h(\boldsymbol{x}, \boldsymbol{\alpha}) & \partial^2/\partial\alpha_1\partial\alpha_2 h(\boldsymbol{x}, \boldsymbol{\alpha}) & \cdots & \partial^2/\partial\alpha_1\partial\alpha_r h(\boldsymbol{x}, \boldsymbol{\alpha}) \\ & \partial^2/\partial\alpha_2^2 h(\boldsymbol{x}, \boldsymbol{\alpha}) & \cdots & \partial^2/\partial\alpha_2\partial\alpha_r h(\boldsymbol{x}, \boldsymbol{\alpha}) \\ & & \ddots & \vdots \\ & & & \partial^2/\partial\alpha_r^2 h(\boldsymbol{x}, \boldsymbol{\alpha}) \end{pmatrix} \tag{B.2}$$

This is often called the ***Hessian***.

More generally, suppose that $h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ is a real-valued function of two vectors $\boldsymbol{\alpha}$ $(r \times 1)$ and $\boldsymbol{\delta}$ $(s \times 1)$. Then define

$$h_{\alpha\delta}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = \partial/\partial\boldsymbol{\alpha}\partial\boldsymbol{\delta}^T h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = \begin{pmatrix} \partial^2/\partial\alpha_1\partial\delta_1 h & \partial^2/\partial\alpha_1\partial\delta_2 h & \cdots & \partial^2/\partial\alpha_1\partial\delta_s h \\ \partial^2/\partial\alpha_2\partial\delta_1 h & \partial^2/\partial\alpha_2\partial\delta_2 h & \cdots & \partial^2/\partial\alpha_2\partial\delta_s h \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2/\partial\alpha_r\partial\delta_1 h & \cdots & \cdots & \partial^2/\partial\alpha_r\partial\delta_s h \end{pmatrix}, \tag{B.3}$$

where we have written $h = h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ for brevity. This is a $(r \times s)$ matrix; it follows that, by reducing the roles of $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$, $h_{\delta\alpha}$ is a $(s \times r)$ matrix, and is the transpose of $h_{\alpha\delta}$ in (B.3).

VECTOR-VALUED FUNCTIONS: Now suppose that $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\delta})$ is a vector-valued function of dimension n of a vector $\boldsymbol{x}$ and parameter $\boldsymbol{\delta}$ $(s \times 1)$; i.e.

$$\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\delta}) = \begin{pmatrix} h_1(\boldsymbol{x}, \boldsymbol{\delta}) \\ \vdots \\ h_n(\boldsymbol{x}, \boldsymbol{\delta}) \end{pmatrix}.$$

Thus, the vector-valued function $\boldsymbol{h}$ has real-valued component functions $h_1, \ldots, h_n$.

We will write

$$\boldsymbol{h}_\delta(\boldsymbol{x}, \boldsymbol{\delta}) = \partial/\partial\boldsymbol{\delta}^T \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\delta}) = \begin{pmatrix} \partial/\partial\delta_1 h_1(\boldsymbol{x}, \boldsymbol{\delta}) & \cdots & \partial/\partial\delta_s h_1(\boldsymbol{x}, \boldsymbol{\delta}) \\ \vdots & \vdots & \vdots \\ \partial/\partial\delta_1 h_n(\boldsymbol{x}, \boldsymbol{\delta}) & \cdots & \partial/\partial\delta_s h_n(\boldsymbol{x}, \boldsymbol{\delta}) \end{pmatrix},$$

a $(n \times s)$ matrix. In particular, note that if we have a real-valued function $h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ for α $(r \times 1)$ and $\boldsymbol{\delta}$ $(s \times 1)$, then

$$h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = \begin{pmatrix} \partial/\partial\alpha_1 h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) \\ \vdots \\ \partial/\partial\alpha_r h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) \end{pmatrix} = \begin{pmatrix} h_{\alpha_1}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) \\ \vdots \\ h_{\alpha_r}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) \end{pmatrix} \quad (r \times 1).$$

Thus, $h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ is a vector-valued function. Applying the above definition of the partial derivative of a vector-valued function with respect to a vector parameter to $h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$, we may conclude that the result of differentiating $h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is the $(r \times s)$ matrix $h_{\alpha\delta}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ defined in (B.3); that is,

$$\partial/\partial\boldsymbol{\delta}^T h_\alpha(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = \partial/\partial\boldsymbol{\alpha}\partial\boldsymbol{\delta}^T h(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\delta}).$$

We now consider various forms of Taylor's theorem.

UNIVARIATE TAYLOR'S THEOREM: Assume $h(\alpha)$ has $(k-1)$ continuous derivatives in $[a, b]$ and finite kth derivative in (a, b), where α is univariate and h is a real-valued function. Let $\alpha_0 \in [a, b]$. For each $\alpha \in [a, b]$, $\alpha \neq \alpha_0$, there exists α_* interior to the interval joining α_0 and α such that

$$h(\alpha) = h(\alpha_0) + \sum_{\ell=1}^{k-1} \frac{1}{\ell!}\{\partial^\ell/\partial\alpha^\ell h(\alpha)\}_{\alpha=\alpha_0}(\alpha-\alpha_0)^\ell + \frac{1}{k!}\{\partial^k/\partial\alpha^k h(\alpha)\}_{\alpha=\alpha_*}(\alpha-\alpha_0)^k.$$

Taylor's theorem is used heavily in making large sample arguments in statistics. In such arguments, we usually deal with vector-valued functions of vector-values parameters, for which a multivariate version of Taylor's theorem is required.

MULTIVARIATE TAYLOR'S THEOREM: First consider a real-valued function $h(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is $(r \times 1)$. We I state the theorem for general k, but write the form of the representation of $h(\boldsymbol{\alpha})$ for $k = 2$ only, as things get messy pretty quickly. One should be able to deduce the form for larger k by analogy to the univariate version.

Assume that $h(\boldsymbol{\alpha})$ on $\mathcal{R}^r$ has continuous partial derivatives of order k at each point in an open set $S \subset \mathcal{R}^r$. Let $\boldsymbol{\alpha}_0 \in S$. For each $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ such that the line segment joining $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_0$ lies in S, there exists $\boldsymbol{\alpha}_*$ in the interior of this line segment such that, in the case $k = 2$,

$$h(\boldsymbol{\alpha}) = h(\boldsymbol{\alpha}_0)+\sum_{\ell=1}^{r}\{\partial/\partial\boldsymbol{\alpha}_\ell h(\boldsymbol{\alpha})\}_{\alpha=\alpha_0}(\boldsymbol{\alpha}_\ell-\boldsymbol{\alpha}_{0,\ell})+(1/2)\sum_{\ell=1}^{r}\sum_{t=1}^{r}\{\partial^2/\partial\boldsymbol{\alpha}_\ell\partial\boldsymbol{\alpha}_t h(\boldsymbol{\alpha})\}_{\alpha=\alpha_*}(\boldsymbol{\alpha}_\ell-\boldsymbol{\alpha}_{0,\ell})(\boldsymbol{\alpha}_t-\boldsymbol{\alpha}_{0,t}).$$

Using the shorthand notation above, we can write this more succinctly as

$$h(\boldsymbol{\alpha}) = h(\boldsymbol{\alpha}_0) + h_\alpha^T(\boldsymbol{\alpha}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + (1/2)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T h_{\alpha\alpha}(\boldsymbol{\alpha}_*)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0).$$

Here, we have used a further shorthand

$$h_\alpha(\boldsymbol{\alpha}_0) = \{\partial/\partial\boldsymbol{\alpha} h(\boldsymbol{\alpha})\}_{\alpha=\alpha_0}.$$

We use similar notation for $h_{\alpha\alpha}(\boldsymbol{\alpha})$ and other expressions.

If h is a function of two vectors $\boldsymbol{\alpha}$ $(r \times 1)$ and $\boldsymbol{\delta}$ $(s \times 1)$, we can define the single, "stacked" vector $(\boldsymbol{\alpha}^T, \boldsymbol{\delta}^T)^T$ $(r + s \times 1)$ and apply the theorem to obtain a representation of $h(\boldsymbol{\alpha}, \boldsymbol{\delta})$ about some value $(\boldsymbol{\alpha}_0^T, \boldsymbol{\delta}_0^T)^T$. This is handy when we want to maintain the distinction between two arguments of a function and treat them separately. Using the above, it is easy to show that, for $k = 2$, we can write

$$\begin{aligned}
h(\boldsymbol{\alpha}, \boldsymbol{\delta}) &= h(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0) + \{h_\alpha^T(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + h_\delta^T(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\} \\
&+(1/2)\{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T h_{\alpha\alpha}(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + 2(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T h_{\alpha\delta}(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*)(\boldsymbol{\delta} - \boldsymbol{\delta}_0) + \\
&(\boldsymbol{\delta} - \boldsymbol{\delta}_0)^T h_{\delta\delta}(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\}.
\end{aligned}$$

It is often necessary to apply Taylor's theorem to vector-valued functions. This can appear complicated, but it is mainly a matter of notation. Suppose $\boldsymbol{h}(\boldsymbol{\alpha}) = [h_1(\boldsymbol{\alpha}), \ldots, h_v(\boldsymbol{\alpha})]^T$ $(v \times 1)$, where again $\boldsymbol{\alpha}$ is $(r \times 1)$.

If we apply the multivariate Taylor theorem to each component of $\boldsymbol{h}$ (each of which is a real-valued function), it can be shown that the expansion of $\boldsymbol{h}(\boldsymbol{\alpha})$ about $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ for $k = 2$ can be written compactly as

$$\boldsymbol{h}(\boldsymbol{\alpha}) = \boldsymbol{h}(\boldsymbol{\alpha}_0) + \begin{pmatrix} h_{1\alpha}^T(\boldsymbol{\alpha}_0) \\ \vdots \\ h_{v\alpha}^T(\boldsymbol{\alpha}_0) \end{pmatrix} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + (1/2)\{\boldsymbol{I}_v \otimes (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T\}\boldsymbol{H}^*_{\alpha\alpha}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0),$$

where $\boldsymbol{H}^*_{\alpha\alpha}$ is the $(rv \times r)$ matrix consisting of the $(r \times r)$ matrices $h_{1\alpha\alpha}(\boldsymbol{\alpha}_*), \ldots, h_{v\alpha\alpha}(\boldsymbol{\alpha}_*)$ stacked vertically; $\boldsymbol{I}_v$ is a $(v \times v)$ identity matrix, and $\otimes$ denotes ***Kronecker product***.

Finally, if $\boldsymbol{h}(\boldsymbol{\alpha}, \boldsymbol{\delta})$ $(v \times 1)$ is a vector-valued function depending on $\boldsymbol{\alpha}$ $(r \times 1)$ and $\boldsymbol{\delta}$ $(s \times 1)$, and we wish to separate explicitly the terms involving $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$, the above extends in the obvious way; for $k = 1$, we have, using obvious notation,

$$\boldsymbol{h}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \boldsymbol{h}(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0) + \begin{pmatrix} h_{1\alpha}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \\ \vdots \\ h_{v\alpha}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \end{pmatrix} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \begin{pmatrix} h_{1\delta}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \\ \vdots \\ h_{v\delta}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \end{pmatrix} (\boldsymbol{\delta} - \boldsymbol{\delta}_0).$$

This can, of course, be written more compactly as

$$\boldsymbol{h}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \boldsymbol{h}(\boldsymbol{\alpha}_0, \boldsymbol{\delta}_0) + \begin{pmatrix} h_{1\alpha}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) & h_{1\delta}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \\ \vdots & \vdots \\ h_{v\alpha}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) & h_{v\delta}^T(\boldsymbol{\alpha}_*, \boldsymbol{\delta}_*) \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} - \boldsymbol{\alpha}_0 \\ \boldsymbol{\delta} - \boldsymbol{\delta}_0 \end{pmatrix}.$$