

Summary of Matrix Calculus

15.1 Definitions, Notation and Preliminaries

- Derivative of scalar-valued function f with input $\mathbf{x} = (x_1, \dots, x_m)$
 - Interior point: $\{\mathbf{x} \in \mathcal{R}^{m \times 1} : \|\mathbf{x} - \mathbf{c}\| < r\}$ for some pos. const r
Applies to matrices too:
 $\{\mathbf{X} \in \mathcal{R}^{m \times n} : \|\mathbf{X} - \mathbf{C}\| < r\}$
 - **Def:** j^{th} (first) partial derivative (scalar-value function with vector input $\mathbf{x} = (x_1, \dots, x_m)^T$)
 $D_j f(\mathbf{c})$ denotes the j th part. deriv. of f at \mathbf{c} :
 $D_j f(\mathbf{c}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{c} + t\mathbf{u}_j) - f(\mathbf{c})}{t}$ where \mathbf{u}_j is the j th row of identity mx.
Alternative notation: $\frac{\partial f(\mathbf{x})}{\partial x_j}$
Alternative notation: At times, it may be more convenient to reshape vector x as matrix \mathbf{X} and denote its partial derivative wrt element x_{ij} as $\frac{\partial f(\mathbf{X})}{\partial x_{ij}}$. The way we treat this derivative depends on whether the elements of \mathbf{X} are dependent (*e.g.* symmetric matrix) or independent (Sec. 15.1.f)
 - **Def:** Vector of partial derivatives $\mathbf{D}f(\mathbf{c})$ denotes vector of all part. derivs of f at \mathbf{c} :
 $\mathbf{D}f(\mathbf{c}) = (D_1 f(\mathbf{c}), D_2 f(\mathbf{c}), \dots, D_m f(\mathbf{c}))'$
Alternative notation: $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}'}$
 - **Def:** $(\mathbf{D}f)'$ is called *gradient vector*
Alternative notation: $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$
 - **Def:** *continuously differentiable:*
Function f with domain $S \in \mathcal{R}^{m \times 1}$ is continuously differentiable at the interior pt $\mathbf{c} \in S$ if $D_1 f(\mathbf{x}), \dots, D_m f(\mathbf{x})$ exist and are continuous at every pt in some neighbourhood of \mathbf{c} .
In this case the following holds: $\lim_{\mathbf{x} \rightarrow \mathbf{c}} \frac{f(\mathbf{x}) - [f(\mathbf{c}) + \mathbf{D}f(\mathbf{c})(\mathbf{x} - \mathbf{c})]}{\|\mathbf{x} - \mathbf{c}\|} = 0$
 - **Def:** ij^{th} (second) partial derivative $D_{ij}^2 f(\mathbf{x})$
Alternative notation: $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$
 - **Def:** *Hessian Matrix* $\mathbf{H}f$
An $m \times m$ matrix whose ij th element is $D_{ij}^2 f(\mathbf{x})$
- Derivative of vector-valued fn $\mathbf{f} = (f_1, \dots, f_p)'$ where each f_i takes input $\mathbf{x} = (x_1, \dots, x_m)'$.
 - $D_j f_s(\mathbf{c})$: j th partial derivative of f_s
 - $D_j \mathbf{f}(\mathbf{c})$ is $p \times 1$ vector $D_j \mathbf{f}(\mathbf{c}) = [D_j f_1(\mathbf{c}), \dots, D_j f_p(\mathbf{c})]'$
Alternative notation $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}'}$ (row vector, $1 \times p$ – see Sec 15.1.c #287)
Alternative notation $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$ (column vector, $p \times 1$)
 - $\mathbf{D}f$ is $p \times m$ matrix: $\mathbf{D}f(\mathbf{c}) = [D_1 \mathbf{f}(\mathbf{c}), \dots, D_p \mathbf{f}(\mathbf{c})]$
 $\mathbf{D}f$ is called *Jacobian* of \mathbf{f} and it's the matrix whose sj th element is $D_j f_s$.
 $(\mathbf{D}f)'$ is called the *gradient (matrix)* of \mathbf{f} .
Alternative notation to Jacobian: $\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}'}$ and it's the matrix whose sj th element is $\frac{\partial f_s(\mathbf{x})}{\partial x_j}$
- Derivative of matrix of functions $F = f_{st}$ where \mathbf{F} is $p \times q$
 - It's preferable to keep the j th partial derivatives of \mathbf{F} as a separate $p \times q$ matrix, denoted as:
 $\frac{\partial \mathbf{F}(\mathbf{x})}{\partial x_j}$ or $D_j \mathbf{F}(\mathbf{x})$

15.2 Differentiation of Scalar-valued Functions

- Lem 15.2.1 – If $f(\mathbf{x})$ does not vary wrt x_j at \mathbf{c} then $D_j f(\mathbf{c}) = 0$
- Lem 15.2.2 – Let l, h, r be functions defined as: $l = af + bg$, $h = fg$ and $r = f/g$. The rules of derivative for single-variable calculus function apply for the j th partial derivative of l, h, g .

15.3 Differentiation of Linear and Quadratic Forms

- Let $\mathbf{a} = (a_1, \dots, a_m)$ be constant (or fn of \mathbf{x} that is invariant wrt x_j), and \mathbf{A} be an $m \times m$ constant matrix (or matrix of functions invariant wrt x_j). Then:

$$- \frac{\partial \mathbf{a}' \mathbf{x}}{\partial x_j} = a_j \text{ (see \#294)}$$

The core idea is to see that $\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$

$$- \frac{\partial \mathbf{a}' \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \text{ or } \frac{\partial \mathbf{a}' \mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}'$$

$$- \frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial x_j} = \sum_{i=1}^m a_{ij} x_i + \sum_{k=1}^m a_{jk} x_k \text{ (see \#295)}$$

The core idea is again a similar piecewise (4-case) function as above

$$- \frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}') \mathbf{x} \text{ — if } \mathbf{A} \text{ symmetric then } \frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x}$$

$$- \frac{\partial^2 \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial x_s \partial x_j} = a_{sj} + a_{js} \text{ (see \#295)}$$

$$- \frac{\partial^2 \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}^2} = (\mathbf{A} + \mathbf{A}') \text{ — if } \mathbf{A} \text{ symmetric then } \frac{\partial^2 \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}^2} = 2\mathbf{A}$$

15.4 Differentiation of Matrix Sums, Products and Transposes

Now we consider to function-valued matrices $\mathbf{F}, \mathbf{G}, \mathbf{H}$

- $\frac{\partial a\mathbf{F} + b\mathbf{G}}{\partial x_i} = a \frac{\partial \mathbf{F}}{\partial x_i} + b \frac{\partial \mathbf{G}}{\partial x_i}$
- $\frac{\partial \mathbf{F}\mathbf{G}}{\partial x_j} = \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_j} + \frac{\partial \mathbf{F}}{\partial x_j} \mathbf{G}$
- $\frac{\partial \mathbf{F}\mathbf{G}\mathbf{H}}{\partial x_j} = \mathbf{F}\mathbf{G} \frac{\partial \mathbf{H}}{\partial x_j} + \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_j} \mathbf{H} + \mathbf{F}\mathbf{G} \frac{\partial \mathbf{H}}{\partial x_j}$
- if g is fn of \mathbf{x} : $\frac{\partial g\mathbf{F}}{\partial x_i} = \frac{\partial g}{\partial x_i} \mathbf{F} + g \frac{\partial \mathbf{F}}{\partial x_i}$

Section 15.5 – 15.7

- 15.5 Differentiation of Vector/Matrix \mathbf{x}, \mathbf{X} wrt its Elements x_j, x_{ij}
 - $\frac{\partial \mathbf{x}}{\partial x_i} = \mathbf{u}_i$
 - if \mathbf{X} matrix of independent variables x_{ij} :
 $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j'$
 - if \mathbf{X} symmetric matrix:
 $\frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{u}_i \mathbf{u}_j' + \mathbf{u}_j \mathbf{u}_i'$
 - The above are derived by constructing piecewise functions for $\frac{\partial x_{st}}{\partial x_{ij}}$ by considering all cases for s, i, j, t (i.e. when they are equal, unequal etc.)
 - $\frac{\partial a\mathbf{F} + b\mathbf{G}}{\partial x_i} = a \frac{\partial \mathbf{F}}{\partial x_i} + b \frac{\partial \mathbf{G}}{\partial x_i}$
 - $\frac{\partial \mathbf{F}\mathbf{G}}{\partial x_j} = \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_j} + \frac{\partial \mathbf{F}}{\partial x_j} \mathbf{G}$
 - $\frac{\partial \mathbf{F}\mathbf{G}\mathbf{H}}{\partial x_j} = \mathbf{F}\mathbf{G} \frac{\partial \mathbf{H}}{\partial x_j} + \mathbf{F} \frac{\partial \mathbf{G}}{\partial x_j} \mathbf{H} + \mathbf{F}\mathbf{G} \frac{\partial \mathbf{H}}{\partial x_j}$

- if g is fn of \mathbf{x} : $\frac{\partial g \mathbf{F}}{\partial x_i} = \frac{\partial g}{\partial x_i} \mathbf{F} + g \frac{\partial \mathbf{F}}{\partial x_i}$
- 15.6 Differentiation of a Trace of a Matrix – again trace is critical for more complicated differentiations
 - Some trace properties:
 - * $\text{tr}(AB) = \text{tr}(BA)$
 - * $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
 - $\frac{\partial \text{tr} \mathbf{F}}{\partial x_j} = \text{tr} \left(\frac{\partial \mathbf{F}}{\partial x_j} \right)$
 - $\frac{\partial (\mathbf{A} \mathbf{X})}{\partial x_{ij}} = a_{ji}$
 - $\frac{\partial (\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}'$
 - if \mathbf{X} is symmetric:

$$\frac{\partial (\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}' - \text{diag}(a_{11}, a_{22}, \dots, a_{mm})$$
 - Regardless whether \mathbf{X} is symmetric or not, $\frac{\partial \text{tr} \mathbf{X}}{\partial \mathbf{X}} = \mathbf{I}$
- 15.7 The Chain Rule
 - Thm 15.7.1: Let $\mathbf{h} = \{h_i\}$ be an $n \times 1$ vector of functions of variables $\mathbf{x} = (x_1, \dots, x_m)$. Let g be a scalar-valued function of a vector of $\mathbf{y} = (y_1, \dots, y_n)$. Define $f(\mathbf{x}) = g[\mathbf{h}(\mathbf{x})]$. Then, the j th partial derivative of f :

$$D_j f(\mathbf{c}) = \sum_{i=1}^n D_i g[\mathbf{h}(\mathbf{c})] D_j h_i(\mathbf{c}) = \mathbf{D}g[\mathbf{h}(\mathbf{c})] D_j \mathbf{h}(\mathbf{c})$$
 - Alternative notation: $\frac{\partial f}{\partial x_j} = \sum_{i=1}^n \frac{\partial g}{\partial y_i} \frac{\partial h_i}{\partial x_j} = \frac{\partial g}{\partial \mathbf{y}'} \frac{\partial \mathbf{h}}{\partial x_j}$
 - The vector of all first partial derivatives:

$$\mathbf{D}f(\mathbf{c}) = \sum_{i=1}^n D_i g[\mathbf{h}(\mathbf{c})] \mathbf{D}h_i(\mathbf{c}) = \mathbf{D}g[\mathbf{h}(\mathbf{c})] \mathbf{D}(\mathbf{h}(\mathbf{c}))$$
 Alternative notation: $\frac{\partial f}{\partial \mathbf{x}'} = \sum_{i=1}^n \frac{\partial g}{\partial y_i} \frac{\partial h_i}{\partial \mathbf{x}'} = \frac{\partial g}{\partial \mathbf{y}'} \frac{\partial \mathbf{h}}{\partial \mathbf{x}'}$
 - For vector-valued function \mathbf{f} :

$$D_j \mathbf{f}(\mathbf{c}) = \sum_{i=1}^n D_i \mathbf{g}[\mathbf{h}(\mathbf{c})] D_j h_i(\mathbf{c}) = \mathbf{D}g[\mathbf{h}(\mathbf{c})] D_j \mathbf{h}(\mathbf{c})$$
 - For all partial derivatives of \mathbf{f} :

$$\mathbf{D}f(\mathbf{c}) = \sum_{i=1}^n D_i \mathbf{g}[\mathbf{h}(\mathbf{c})] \mathbf{D}h_i(\mathbf{c}) = \mathbf{D}g[\mathbf{h}(\mathbf{c})] \mathbf{D}h(\mathbf{c})$$
 Alternative notation: $\frac{\partial \mathbf{f}}{\partial \mathbf{x}'} = \sum_{i=1}^n \frac{\partial \mathbf{g}}{\partial y_i} \frac{\partial h_i}{\partial \mathbf{x}'} = \frac{\partial \mathbf{g}}{\partial \mathbf{y}'} \frac{\partial \mathbf{h}}{\partial \mathbf{x}'}$

Section 15.8 – Derivs of Determinants, Inverses, Adjugates and Generalized inverses

- $\frac{\partial \det(\mathbf{X})}{\partial x_{ij}} = \xi_{ij}$ where ξ_{ij} is the ij th cofactor of \mathbf{X} and \mathbf{X} is a matrix of variables (and not functions)
- $\frac{\partial \mathbf{X}}{\partial \mathbf{X}} = [\text{adj}(\mathbf{X})]'$ where $\text{adj}()$ is adjugate
- $\frac{\partial \det(\mathbf{F})}{\partial x_j} = \text{tr} \left[\text{adj}(\mathbf{F}) \frac{\partial \mathbf{F}}{\partial x_j} \right] \stackrel{(a)}{=} |\mathbf{F}| \text{tr} \left(\mathbf{F}^{-1} \frac{\partial \mathbf{F}}{\partial x_j} \right)$ where \mathbf{F} is a matrix of functions (a) follows only if \mathbf{F} is nonsingular and differentiable.
- $\frac{\partial \log \det(\mathbf{X})}{\partial x_{ij}} = \text{tr}(\mathbf{X}^{-1} \mathbf{u}_i \mathbf{u}_j') = \mathbf{u}_j' \mathbf{X}^{-1} \mathbf{u}_i = y_{ji}$ where y_{ji} is j th element of \mathbf{X}^{-1}
- $\frac{\partial \log \det(\mathbf{F})}{\partial x_j} = \frac{1}{|\mathbf{F}|} \frac{\partial \det(\mathbf{F})}{\partial x_j} = \frac{1}{|\mathbf{F}|} \text{tr} \left[\text{adj}(\mathbf{F}) \frac{\partial \mathbf{F}}{\partial x_j} \right] = \text{tr} \left(\mathbf{F}^{-1} \frac{\partial \mathbf{F}}{\partial x_j} \right)$
- if \mathbf{X} symmetric matrix of variables:

- $\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = 2\text{adj}(\mathbf{X}) - \text{diag}(\xi_{11}, \xi_{22}, \dots, \xi_{mm})$
- $\frac{\partial \log \det(\mathbf{X})}{\partial \mathbf{X}} = 2\text{adj}(\mathbf{X}^{-1}) - \text{diag}(y_{11}, y_{22}, \dots, y_{mm})$ where y_{ij} are elements of \mathbf{X}^{-1}
- $\frac{\partial \text{adj}(\mathbf{F})}{\partial x_j} = \frac{\partial |\mathbf{F}|}{\partial x_j} \mathbf{F}^{-1} + |\mathbf{F}| \frac{\partial \mathbf{F}^{-1}}{\partial x_j} = \dots$ (can be completed with info above)

Section 15.9 - Second-order derivatives of Determinants and Inverses

Differentiation of Generalized Inverses

The section considers derivatives of generalized inverses of possibly singular or non-square matrix of functions \mathbf{F} . The idea is to re-order the rows and cols of \mathbf{F} and then partition the new matrix so the leading principal submatrix is nonsingular.

Summary of Theorem 15.10.1 Let \mathbf{P}, \mathbf{Q} be permutation matrices that yield matrix $\mathbf{B} = \mathbf{PFQ}$ such that the leading principal submatrix \mathbf{B}_{11} in partitioning $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$ is an $r \times r$ nonsingular matrix

where $r = \text{rank}(\mathbf{F})$. Then, there exists a generalized inverse of \mathbf{G} of \mathbf{F} such that: $\mathbf{G} = \mathbf{Q} \begin{bmatrix} \mathbf{B}_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{P}$ and its derivative is:

$$\frac{\partial \mathbf{G}}{\partial x_j} = -\mathbf{Q} \begin{bmatrix} \mathbf{B}_{11}^{-1} (\partial \mathbf{B}_{11} / \partial x_j) \mathbf{B}_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{P} = -\mathbf{G} \frac{\partial \mathbf{F}}{\partial x_j} \mathbf{G} \quad (1)$$