Summary of Matrix Calculus

15.1 Definitions, Notation and Preliminaries

- Derivative of scalar-valued function f with input $\mathbf{x} = (x_1, \ldots, x_m)$
  - 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* $\mathbf{x} = (x_1, \ldots, x_m)^T$
    $D_j f(\mathbf{c})$ denotes the jth part. deriv. of f at $\mathbf{c}$:
    $Df_j(\mathbf{c}) = \lim_{t\to 0} = \frac{f(\mathbf{c}+t\mathbf{u}_j)-f(\mathbf{c})}{t}$ where $\mathbf{u}_j$ is the jth 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 $\frac{\partial f(\mathbf{X})}{x_{ij}}$. 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 $\mathbf{c}$:
    $\mathbf{D}f(\mathbf{c}) = (D_1 f(\mathbf{c}), D_2 f(\mathbf{c}), \ldots, D_m f(\mathbf{c}))'$
    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 $D_1 f(\mathbf{x}), \ldots, D_m f(\mathbf{x})$ exist and are continuous at every pt in some neighbourhood of $\mathbf{c}$.
    In this case the following holds: $\lim_{\mathbf{x}\to\mathbf{c}} \frac{f(\mathbf{x})-[f(\mathbf{c})+\mathbf{D}f(\mathbf{c})(\mathbf{x}-\mathbf{c})]}{||\mathbf{x}-\mathbf{c}||}$
  - **Def:** *ij^{th} (second) partial derivative* $D^2_{ij} f(\mathbf{x})$
    Alternative notation: $\frac{\partial^2 f(\mathbf{x})}{\partial x_i x_j}$
  - **Def:** *Hessian Matrix* $\mathbf{H}f$
    An $m \times m$ matrix whose ijth element is $D^2_{ij} f(\mathbf{x})$
- Derivative of vector-valued fn $\mathbf{f} = (f_1, \ldots, f_p)'$ where each f_i takes input $\mathbf{x} = (x_1, \ldots, x_m)'$.
  - $D_j f_s(\mathbf{c})$: jth partial derivative of f_s
  - $D_j \mathbf{f}(\mathbf{c})$ is $p \times 1$ vector $D_j \mathbf{f}(\mathbf{c}) = [D_j f_1(\mathbf{c}), \ldots, D_j f_p(\mathbf{c})]'$
    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, $p \times 1$)
  - $\mathbf{Df}$ is $p \times m$ matrix: $\mathbf{Df}(\mathbf{c}) = [D_1 \mathbf{f}(\mathbf{c}), \ldots, D_p \mathbf{f}(\mathbf{c})]$
    $\mathbf{Df}$ is called *Jacobian* of $\mathbf{f}$ and it's the matrix whose sjth element is $D_j f_s$.
    $(\mathbf{Df})'$ 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 sjth 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 jth 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 jth partial derivative of l, h, g.

15.3 Differentiation of Linear and Quadratic Forms

- Let $\mathbf{a} = (a_1, \ldots, a_m)$ 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:
  - $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial x_j} = a_j$ (see #294)
  - $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$ or $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}'$
  - $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_j} = \sum\limits_{i=1}^{m} a_{ij} x_i + \sum\limits_{k=1}^{m} a_{jk} x_k$ (see #295)
  - $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$ — if $\mathbf{A}$ symmetric then $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$
  - $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_s x_j} = a_{sj} + a_{js}$ (see #295)
  - $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}^2} = (\mathbf{A} + \mathbf{A}')$ — if $\mathbf{A}$ symmetric then $\frac{\partial^2 \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}^2} = 2\mathbf{A}$