MISCELLANEOUS CONCEPTS OF MATRIX ALGEBRA

1. Vector and matrix differentiation

1.1. **The gradient (or derivative) of f.** Let y = f(x) = f(x_1 x_2, ..., x_n) denote a real valued function of the vector x. We define the gradient of f at x as the vector ∇f(x) as

$$\nabla f(x) \;=\; \left(\frac{\partial f(x)}{\partial x_1},\; \frac{\partial f(x)}{\partial x_2}, ..., \frac{\partial f(x)}{\partial x_i}, ..., \frac{\partial f(x)}{\partial x_n} \right) \tag{1}$$

This is also sometimes denoted ∂f(x)/∂x. Note that in this case the gradient is written as a row vector. There is always some question about the proper notation for matrix derivatives since matrices are just ways to represent numbers. We will normally use the convention that the derivative of a function with respect to a column vector is a row vector and vice versa.

1.2. **Convention for writing the gradient of f.** Let y = ψ(x) be a general function of x. The dimensions of y are mx1 and the dimensions of x are nx1. The symbol

$$\frac{\partial y}{\partial x} \;=\; \left[\frac{\partial y_i}{\partial x_j} \right],\; i \,=\, 1,\, 2\,,\, ...\,,\, m,\; j = 1,\, 2\,, ...\,,\, n \tag{2}$$

will denote the matrix of first order partial derivatives of the transformation from x to y such that the ith row contains the derivatives (gradient) of the ith element of y with respect to the elements of x, viz.,

$$\left[\frac{\partial y_i}{\partial x_1},\; \frac{\partial y_i}{\partial x_2}, \frac{\partial y_i}{\partial x_3},\; ...\; \frac{\partial y_i}{\partial x_n} \right] \tag{3}$$

1.3. **Second derivative (Hessian) of f(x).** The Hessian of f with respect to the vector x is defined by

$$\nabla f^2(x) \;=\; \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \tag{4}$$

Note: The derivative of a real valued function with respect to a matrix, i.e., the derivative of a real valued function of a matrix is <u>defined</u> to be the matrix of derivatives of corresponding elements.

1.4. **Some results on matrix derivatives.** Let A = (a_{ij}) be an m x n matrix.

Let a = (a_1, a_2, a_3, . . . a_n)′ and $a \;=\; \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_n \end{bmatrix}$.

Date: February 18, 2008.

Let x = (x_1, x_2, x_3, . . . x_n)´ and z = (z_1, z_2, z_3, . . . z_n)´ so that

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \end{bmatrix} \quad z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ . \\ . \\ . \\ z_n \end{bmatrix} .$$

1.4.1. *Derivative of a scalar with respect to a vector.*

$$\frac{d(a'x)}{dx} = a' \tag{5}$$

1.4.2. *Derivative of $y = Ax$.*

$$\frac{\partial y}{\partial x} = A \tag{6}$$

1.4.3. *Derivative of $y = z'Ax$.*

$$\frac{\partial y}{\partial z} = x'A', \; \frac{\partial y}{\partial x} = z'A \tag{7}$$

1.4.4. *Derivative of $y = x'Ax$ (A is nxn).*

$$\frac{\partial y}{\partial x} = x'(A' + A) \tag{8}$$

1.4.5. *Derivative of $y = x'Ax$ with A symmetric.*

$$\begin{aligned} \frac{d(x'Ax)}{dx} &= 2x'A' = 2x'A \\ \frac{d(x'Ax)}{dx'} &= 2Ax \end{aligned} \tag{9}$$

1.4.6. *Second derivative of $y = x'Ax$.*

$$\frac{\partial^2(x'Ax)}{\partial x\, \partial x'} = A' + A \tag{10}$$

1.4.7. *Second derivative of $y = x'Ax$ with symmetric A.*

$$\frac{\partial^2(x'Ax)}{\partial x\, \partial x'} = 2A = 2A' \tag{11}$$

1.4.8. *Trace with A nxn.*

$$\frac{\partial \; tr \; (A)}{\partial A} = I \tag{12}$$

1.4.9. *Determinant of A.*

$$\begin{aligned} \frac{\partial |A|}{\partial A} &= |A| \; A^{-1} = A^{+}, \; if \; |A| \neq 0, \\ &= 0 \; if \; |A| = 0 \end{aligned} \tag{13}$$

1.4.10. *Logarithm of det A.*

$$\frac{\partial \; \ln |A|}{\partial A} = A^{-1} \tag{14}$$

1.5. **Some examples.**

1.5.1. *Example 1.*

a: y = f(x) = $3x_1 + 4x_2$
b: a′ = [3, 4]
c: x′ = [x_1, x_2]

$$\frac{\partial(a'x)}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \ \frac{\partial f}{\partial x_2}\right] = [3, \ 4]$$
$$\frac{\partial(a\acute{}x)}{\partial x\acute{}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \tag{15}$$

1.5.2. *Example 2.*

a: y = x′Ax
b: x′ = [x_1, x_2]

$$A = \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix}$$

$$\begin{aligned} y &= [x_1, \ x_2] \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= [x_1, \ x_2] \begin{bmatrix} 3x_1 + 5x_2 \\ 5x_1 + 7x_2 \end{bmatrix} \\ &= 3x_1^2 + 5x_2x_1 + 5x_1x_2 + 7x_2^2 \\ &= 3x_1^2 + 10x_1x_2 + 7x_2^2 \end{aligned} \tag{16}$$

$$\Rightarrow \ \frac{\partial y}{\partial x} = \left[\frac{\partial y}{\partial x_1}, \ \frac{\partial y}{\partial x_2}\right] = [6x_1 + 10x_2, \ 10x_1 + 14x_2]$$

Now if we use the differentiation formulas directly we obtain

$$\begin{aligned} \frac{\partial y}{\partial x} &= 2x'A \\ &= 2\,[x_1, \ x_2] \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix} \\ &= 2\,[3x_1 + 5x_2, \ 5x_1 + 7x_2] \\ &= [6x_1 + 10x_2, \ 10x_1 + 14x_2] \end{aligned} \tag{17}$$

and

$$\begin{aligned}
\frac{\partial y}{\partial x'} &= 2Ax \\
&= 2 \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= 2 \begin{bmatrix} 3x_1 + 5x_2 \\ 5x_1 + 7x_2 \end{bmatrix} \\
&= \begin{bmatrix} 6x_1 + 10x_2 \\ 10x_1 + 14x_2 \end{bmatrix}
\end{aligned} \tag{18}$$

2. Kronecker Products of Matrices

2.1. **Definition of a Kronecker Product.** Let A be mxn and B be pxq. Then the Kronecker product of A and B is denoted by A ⊗ B and is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

A ⊗ B is of dimension (mp)x(nq)

2.2. **Properties of Kronecker Products.**

2.2.1. *1. If $D_1 = A_1 \otimes B$ and $D_2 = A_2 \otimes B$ then $D_1 + D_2 = (A_1 + A_2) \otimes B$. Similarly, if $E_1 = A \otimes B_1$ and $E_2 = A \otimes B_2$ then $E_1 + E_2 = A \otimes (B_1 + B_2)$.*

2.2.2. $\alpha(A \otimes B) = (\alpha A) \otimes B = A \otimes (\alpha B)$.

2.2.3. $(A \otimes C)(B \otimes D) = (AB) \otimes (CD)$.

2.2.4. $(P \otimes Q)^{-1} = P^{-1} \otimes Q^{-1}$.

2.2.5. $M \otimes N)' = M' \otimes N'$.

2.2.6. $|A \otimes B| = |A|^n \, |B|^m$ *where A is m x m and B is n x n.*

2.2.7. *tr(A ⊗ B) = tr(A)tr(B) where A and B are square.*

3. Vectorization of Matrices

3.1. **Definition of Vectorization.** Let A be an n x m matrix; the notation vec(A) will mean the nm-element column vector whose first n elements are the first column of A, $a_{\cdot 1}$; the second n elements, the second column of A, $a_{\cdot 2}$, and so on. Thus

$$\begin{aligned}
(\vec{A}) &= (a'_{\cdot 1},\ a'_{\cdot 2}, ..., a'_{\cdot m})' \\
&= \begin{bmatrix} a_{\cdot 1} \\ a_{\cdot 2} \\ \vdots \\ a_{\cdot m} \end{bmatrix}
\end{aligned} \tag{19}$$

3.2. Properties of the vec operator.

3.2.1. *vec(AB).* Let A, B be n x m, m x q respectively. Then vec(AB) = (B′ ⊗ I) vec(A) = (I ⊗ A) vec(B).

3.2.2. *Let A_1, A_2, A_3 be suitably dimensioned matrices.* Then

$$\begin{aligned}(A_1 \vec{A_2} A_3) &= (I \otimes A_1 A_2)\,(\vec{A_3}) \\ &= (A_3' \otimes A_1)\,(\vec{A_2}) \\ &= (A_3' A_2' \otimes I)\,(\vec{A_1})\end{aligned}$$

3.2.3. *Let A_1, A_2, A_3, A_4 be suitably dimensioned matrices.* Then

$$\begin{aligned}(A_1 \overrightarrow{A_2 A_3} A_4) &= (I \otimes A_1 A_2 A_3)\,(\vec{A_4}) \\ &= (A_4' \otimes A_1 A_2)\,(\vec{A_3}) \\ &= (A_4' A_3' \otimes A_1)\,(\vec{A_2}) \\ &= (A_4' A_3' A_2' \otimes I)\,(\vec{A_1})\end{aligned}$$

3.2.4. *Let A, B be m x n.* Then vec(A + B) = vec(A) + vec(B).

3.2.5. *Let A, B, C, D be suitably dimensioned matrices.* Then

$$\begin{aligned}vec[(A+B)(C+D)] =& [(I \otimes A) + (I \otimes B)][vec(C) + vec(D)] \\ =& [(C' \otimes I) + (D' \otimes I)][vec(A) + vec(B)]\end{aligned}$$

3.2.6. *Let A, B be suitably dimensioned matrices.* Then

$$\begin{aligned}tr(AB) =& (\vec{A'})'(\vec{B}) \\ =& vec(B')'(\vec{A}).\end{aligned}$$

4. Partitioned Matrices

4.1. Definition of a Partitioned Matrix. Using the concept of submatrices, we can group the elements of a matrix into submatrices. For example consider the matrix A below.

$$A = \begin{pmatrix} 3 & 4 & 7 \\ 2 & 5 & 2 \\ 1 & 0 & 4 \end{pmatrix} \tag{20}$$

We can partition it into 4 submatrices. The 2x2 matrix in the upper left hand corner can be denoted A_{11}

$$A_{11} = A(\{1,2\},\{1,2\}) = \begin{pmatrix} 3 & 4 \\ 2 & 5 \end{pmatrix} \tag{21}$$

Similarly the 2x1 matrix to the right of this matrix is given by

$$A_{12} = A(\{1,2\},\{3\}) = \begin{pmatrix} 7 \\ 2 \end{pmatrix} \tag{22}$$

The 2x1 matrix in the lower left is given by

$$A_{21} \;=\; A(\{3\},\{1,2\}) = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{23}$$

And finally the matrix in the lower right is given by

$$A_{22} \;=\; A(\{3\},\{3\}) = \begin{pmatrix} 4 \end{pmatrix} \tag{24}$$

We can then write A as

$$A = \begin{pmatrix} 3 & 4 & 7 \\ 2 & 5 & 2 \\ 1 & 0 & 4 \end{pmatrix} \;=\; \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{25}$$

A partitioned matrix is called block diagonal if the submatrices not on the main diagonal are zero matrices. For example the following matrix is block diagonal.

$$B = \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \tag{26}$$

4.2. **Addition and multiplication.** Consider two partitioned matrices, A and B, that are conformably partitioned. Then we have the following relation for addition.

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix} \tag{27}$$

For multiplication we obtain

$$\begin{aligned} AB \;&=\; \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ &=\; \begin{pmatrix} A_{11}\,B_{11} \;+\; A_{12}\,B_{21} & A_{11}\,B_{12} \;+\; A_{12}\,B_{22} \\ A_{21}\,B_{11} \;+\; A_{22}\,B_{21} & A_{21}\,B_{12} \;+\; A_{22}\,B_{22} \end{pmatrix} \end{aligned} \tag{28}$$

4.3. **Determinants of partitioned matrices.** The determinant of a block diagonal matrix is easy to compute as follows

$$|\,B\,| \;=\; \left| \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \right| \;=\; |\,B_{11}\,|\,|\,B_{22}\,| \tag{29}$$

In general the determinant of a partitioned matrix is given by

$$\begin{aligned} |\,A\,| \;=\; \left| \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \right| \;&=\; |\,A_{22}\,|\,|\,A_{11} \;-\; A_{12}\,A_{22}^{-1}\,A_{21}\,| \\ &=\; |\,A_{11}\,|\,|\,A_{22} \;-\; A_{21}\,A_{11}^{-1}\,A_{12}\,| \end{aligned} \tag{30}$$

4.4. **Inverses of partitioned matrices.** The inverse of a block diagonal matrix is computeed as follows

$$B^{-1} = \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix}^{-1} = \begin{pmatrix} B_{11}^{-1} & 0 \\ 0 & B_{22}^{-1} \end{pmatrix} \tag{31}$$

In general the inverse of a partitioned matrix is given by

$$\begin{aligned} A^{-1} &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1}\left(I + A_{12} F_2 A_{21} A_{11}^{-1}\right) & -A_{11}^{-1} A_{12} F_2 \\ -F_2 A_{21} A_{11}^{-1} & F_2 \end{pmatrix} \\ F_2 &= \left(A_{22} - A_{21} A_{11}^{-1} A_{12}\right)^{-1} \end{aligned} \tag{32}$$

To see how this formula is derived first premultiply A by a matrix which will zero out lower left hand matrix as follows.

$$\begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & -A_{21} A_{11}^{-1} A_{12} + A_{22} \end{pmatrix} \tag{33}$$

Then postmultiply by a matrix to eliminate the upper right triangle creating a block diagonal matrix.

$$\begin{aligned} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} &= \begin{pmatrix} A_{11} & A_{12} \\ 0 & -A_{21}A_{11}^{-1} A_{12} + A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A_{11} & 0 \\ 0 & -A_{21} A_{11}^{-1} A_{12} + A_{22} \end{pmatrix} \end{aligned} \tag{34}$$

This expression in useful because the inverse of a block diagonal matrix is just the inverse of each block diagonal element. Now take the inverse of both sides of equation 34.

$$\begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix}^{-1} = \begin{pmatrix} A_{11} & 0 \\ 0 & -A_{21} A_{11}^{-1} A_{12} + A_{22} \end{pmatrix}^{-1} \tag{35}$$

Now premultiply both sides by the inverse of the first matrix on the left of equation 35to obtain

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & -A_{21} A_{11}^{-1} A_{12} + A_{22} \end{pmatrix}^{-1} \tag{36}$$

Now postmultiply both sides by the inverse of the second matrix on the left of equation 36 to obtain

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & -A_{21} A_{11}^{-1} A_{12} + A_{22} \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix}$$
$$= \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix} \tag{37}$$

Now simply multiply out equation 37 as follows

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I & -A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1} A_{12} \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \\ 0 & \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21} A_{11}^{-1} & I \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \\ -\left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} A_{21} A_{11}^{-1} & \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1} \end{pmatrix} \tag{38}$$

Now let

$$F_2 = \left(A_{22} - A_{21} A_{11}^{-1} A_{12} \right)^{-1}$$

and rewrite equation 38 as

$$A^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} F_2 A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} F_2 \\ -F_2 A_{21} A_{11}^{-1} & F_2 \end{pmatrix} \tag{39}$$