# Matrix Recipes

November 27, 2011

# 1 Notation

An $m \times n$ matrix is a real valued matrix with $m$ rows and $n$ columns. $R^{m \times n}$ represents the space of such matrices. $a'$ is the transpose of matrix $a$. $I_n$ is the $n \times n$ identity matrix. An $m$ dimensional vector is a matrix with $m$ rows and 1 column.

# 2 Cross Products

For $a, b \in \Re^3$

$$a \times b \stackrel{\text{def}}{=} R[a]b \tag{1}$$

where the linear function $R : \Re^3 \to \Re^3 \otimes \Re^3$ is defined as follows

$$R[x] = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} \tag{2}$$

**Properties**

- $$a \times b = |a||b| \sin(\theta)\, n \tag{3}$$

  where $\theta$ is the angle between $a$, and $b$, and $n$ is a unit vector orthogonal to the plane defined by $a, b$ and oriented according to the *right hand rule*: Forefinger points to $a$, middle finger points to $b$ and $n$ points in the direction of the thumb.

- $a \times b = -b \times a$
- $a \times (b + c) = a \times b + a \times c$
- $a \times (b \times c) \neq (a \times b) \times c$ where the $\neq$ means that equality does not always hold.
- $a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0$

- If $a \times b = a \times c$ and $a = 0$ this does not mean that $b = c$ To see why note $a \times b - a \times c = a \times (b - c)$ and we do not need $b = c$ for this to be zero, we just need $b - c$ parallel to $a$.
- $a \times (b \times c) = b(a \cdot c) - c(a \cdot b)$ Also known as the BAC minus CAB rule
- $|a \times b|^2 = (a \cdot b)^2 + |a|^2 + |b|^2$
- $r(a \times b) = (ra) \times (rb)$ where $r$ is a rotation matrix.

  Pf. Since rotations do not change vector lenghts and vector angles, thus $a \times b = r(|a||b| \sin(\theta(a, b))n(a, b)) = |ra||rb| \sin(\theta(ra, rb)n(ra, rb)$ and since $n(a, b)$ is a unit vector orthogonal to $a, b$ then $(rn)'(rb) = n'r'rb = n'b = 0$, i.e., $rn$ is orthogonal to $rb$, and by similar argument to $ra$. Thus $rn = n(ra, rb)$.

# 3 Determinants

- $|ab| = |a||b|$
- $|a^{-1}| = \frac{1}{|a|}$
- $|I + xy'| = 1 + x'y,\ x, y \in \Re^n$
- $|a + xy'| = |a|(1 + (a^{-1}x)'y)$
- $|e^a| = e^{\mathrm{Tr}(a)}$

# 4 Trace

- $\mathrm{Tr}(a + b) = \mathrm{Tr}(a) + \mathrm{Tr}(b)$
- $\mathrm{Tr}(a) = \mathrm{Tr}(a')$
- $\mathrm{Tr}(abc) = \mathrm{Tr}(cab) = \mathrm{Tr}(bca)$
- $\mathrm{Tr}(xy'a) = y'ax,\ x \in \Re^m, y \in \Re^n$
- $\mathrm{Tr}(a'rb') = a'rb$ *need to confirm whether $r$ does not need rotation*
- If a is $m \times n$ and b is $n \times m$ then $\mathrm{Tr}(ab) = \mathrm{Tr}(ba) = \mathrm{Tr}(a'b')$

## 5 Matrix Inverses

- Woodbury Identity: $(a + ubv)^{-1} = a^{-1} - a^{-1}u(b^{-1} + va^{-1}u)^{-1}va^{-1}$ where $a, b, u, v$ are matrices.
- Special case of Woodbury Identity:
  $(a + \alpha x x')^{-1} = a^{-1} - \frac{\alpha}{1+\alpha x' y} y y'$
  where $a$ is a symmetric matrix, $\alpha$ is a scalar, $x$ is a vector, and $y = a^{-1}x$.
- $(a^{-1} + b^{-1})^{-1} = a(a + b)^{-1}b$
- $(ab)^{-1} = b^{-1}a^{-1}$
- $(a + b)^{-1} = a^{-1} \sum_{i=0}^{\infty} (-1)^i (ba^{-1})^i$
  for $ba^{-1}$ with spectral radius smaller than 1.
- $(a + \epsilon b)^{-1} \approx a^{-1} - \epsilon a^{-1}ba^{-1}$ with error

  $$-\epsilon a^{-1}b((a + \epsilon b)^{-1} - a^{-1}) \tag{4}$$

  where $\epsilon > 0$ is a small positive constant.

The $(a+b)^{-1}$ infinite series and the the Woodbury identity are proven in the Appendix. It is useful in recursive least square problems. In such cases we typically have an equation of the form

$$\sigma_{t+1} = (\sigma_t^{-1} + \alpha x_t x_t')^{-1} \tag{5}$$

where $\alpha$ is a positive constant, $x$ is a vector and $\sigma_t$ is known. Using Woodbury's identity we get

$$\sigma_{t+1} = \sigma_t - \frac{\alpha}{1 + \alpha x_t' \sigma_t x_t} \sigma_t x_t x_t' \sigma_t \tag{6}$$

Note $1 + \alpha x_t' \sigma_t x_t$ is a scalar so basically we avoided having to do a matrix inversion.

Regarding the matrix inverse approximation note

$$(a + \epsilon b)^{-1} = (a(I + \epsilon a^{-1}b))^{-1} = (I + \epsilon a^{-1}b)^{-1}a^{-1} \approx (I - \epsilon a^{-1}b)a^{-1} \tag{7}$$

with error of order $\epsilon^2$.

# 6 Matrix Exponentials

- $|e^a| = e^{\mathrm{Tr}}(a)$
- if $a = p\Lambda p^{-1}$ then $e^a = pe^{\Lambda}p^{-1}$
- $\frac{de^{ta}}{dt} = ae^{ta}$
- $\frac{d}{da}x'e^{ta}y = te^{ta}xy'$

## 6.1 Proofs:

$$\begin{aligned} \frac{\partial x'e^{ta}}{\partial a_{ij}} &= \frac{d}{d\delta}x'e^{t(a+\delta 1_i 1'_j)}y\Big|_{\delta=0} = x'e^{ta}\left(\frac{d}{d\delta}e^{t\delta 1_i 1'_j}\right)y\Big|_{\delta=0} \\ &= x'e^{ta}t1_i1'_je^{\delta 1_i 1'_j}y\Big|_{\delta=0} = t(e^{ta}x)_iy_j \end{aligned} \tag{8}$$

Thus

$$\frac{\partial x'e^{ta}}{\partial a} = te^{ta}xy' \tag{9}$$

# 7 Matrix Logarithms

Let $a$ a real or complex square matrix of ordeer $n$ with positive eigenvalues. Then there is a unique matrix $b$ such that (1) $a = e^b$ and (2) the imagignary part of the eigenvalues is in $[-\pi, \pi]$. We call $b$ the principle logarithm of $a$.

- if $a = p\Lambda p^{-1}$ then $\log(a) = p\log(\Lambda)p^{-1}$

# 8 Kronecker Product, Kronecker Sum, Vec, Vec-T, Vec-H

Provide a way to deal with derivatives of matrix functions without having to use cubix or quartix (i.e., matrices with 3 or 4 dimensions). Instead of working with matrix functions we work with vectorized versions of matrix functions. This gives rise to the Kronecker product, or tensor product.

**Definition 1. Kronecker product**

$$a \otimes b = \begin{pmatrix} a_{11}b & \cdots & a_{1n}b \\ \vdots & \ddots & \vdots \\ a_{m1}b & \cdots & a_{mn}b \end{pmatrix} \tag{10}$$

**Definition 2. Kronecker sum** Let $a$ be an $m \times m$ matrix and $b$ an $n \times n$ matrix. Then

$$a \oplus b = (I_m \times a) + (b \otimes I_n) \tag{11}$$

Note the Kronecker sum is not commutative

**Definition 3. Vec operator** For a matrix $a$ the Vec operator creates a column vector Vec$[a]$ by stacking the columns of $x$ bellow one another

$$\text{Vec}[a] = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ a_{12} \\ \vdots \\ a_{m2} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \tag{12}$$

**Definition 4. Vec Transpose** The vec-transpose matrix $\mathcal{T}_{m,n}$ of an $m \times m$ matrix $a$ is defined as follows

$$\mathcal{T}_{m,n}\text{Vec}[a] = \text{Vec}[a'] \tag{13}$$

Note $\mathcal{T}_{m,n}$ is an $(mn) \times (mn)$ permutation matrix, i.e., it is made out of zeros and 1s, an each row and each column has a single 1. For example for a $2 \times 3$ matrix

$$\text{Vec}[a] = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \quad \text{Vec}[a'] = \begin{pmatrix} a'_{11} \\ a'_{21} \\ a'_{12} \\ a'_{22} \\ a'_{13} \\ a'_{23} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ a_{31} \\ a_{32} \end{pmatrix} \tag{14}$$

Thus

$$\mathcal{T}_{3,2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{15}$$

It is easy to confirm that for an $m \times n$ matrix $a$, the element $k$ in Vec$[a]$ maps into element

$$\text{mod}[\frac{k-1}{m}]n + \text{int}[\frac{k-1}{m}] + 1 \tag{16}$$

where mod$[\cdot]$ gives the remainder in $(k-1)/m$ and int$[\cdot]$ gives the integer part. Below is matlab code to compute the $\mathcal{T}_{m,n}$ vec-transpose matrix.

```
d = m*n;
Tmn = zeros(d,d);
i = 1:d;
rI = 1+m.*(i-1)-(m*n-1).*floor((i-1)./n);
I1s = sub2ind([d d],rI,1:d);
Tmn(I1s) = 1;
Tmn = Tmn';
```

**Definition 5. Vech** The Vech operator vectorizes a matrix with the upper portion excluded. It is typically used when the matrix is symetric

$$\text{Vech}(a) = \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \\ a_{2,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{n-1,n} \\ a_{n,n} \end{pmatrix} \tag{17}$$

The Vech operator can be seen as a matrix $S_n$ times the Vec operator

$$\text{Vech}(a) = S_n \text{Vec}(a) \tag{18}$$

## 8.1 Properties

If $a$, and $b$ have one of the properties below, then $a \otimes b$ has that property too: (1) nonsingular, square upper triangular, square lower triangular, banded, symmetric, positive definite, stochastic, Toeplitz, orthogonal.

If $a$, $b$ are square

$$(a \otimes b)^n = a^n \otimes b^n \tag{19}$$

$$a \otimes b \otimes c = (a \otimes b) \otimes c = a \otimes (b \otimes c) \tag{20}$$

provided the dimensions of the matrices allows for all the expressions to exist.

$$(a + b) \otimes (c + d) = a \otimes c + a \otimes d + b \otimes c + b \otimes d \tag{21}$$

$$(a \otimes b)(c \otimes d) = (ac) \otimes (bd) \tag{22}$$

From the previous property letting $c = I_n$, $b = I_p$ it follows that

$$a \otimes d = (a \otimes I_p)(I_n \otimes d) \tag{23}$$

$$(a \otimes b)(b \otimes d) = (ac) \otimes (bd) \tag{24}$$

$$(a \otimes b)' = a' \otimes b' \tag{25}$$

$$(a \otimes b)^{-1} = a^{-1} \otimes b^{-1} \tag{26}$$

$$\text{Vec}\,[abc] = (c' \otimes a)\,\text{Vec}\,[b] \tag{27}$$

A useful corollary. If $a$ is an $n \times m$ matrix and $b$ an $m \times p$

$$\text{Vec}(ab) = \text{Vec}[abI_p] = (I_p \otimes a)\text{Vec}[b] \tag{28}$$

$$\text{Vec}(ab) = \text{Vec}[I_m ab] = (b' \otimes I_m)\text{Vec}[a] \tag{29}$$

$$\text{Vec}(ab) = \text{Vec}[aI_n b] = (b' \otimes a)\text{Vec}[I_n] \tag{30}$$

If $x, y$ are vectors then

$$xy' = x \otimes a' \tag{31}$$

$$\nabla_b \text{Vec}[abc] = \nabla_b (c' \otimes a)\text{Vec}[b] = (c' \otimes a)' = c \otimes a' \tag{32}$$

If $abc$ is a vector then

$$\nabla_{\text{Vec}[b]} abc = c \otimes a' \tag{33}$$

If $\{\lambda_i, u_i\}$ are eigenvalues/eigenvectors of $a$ and $\{\delta_i, v_i\}$ are eigenvalues/eigenvectors of $b$ then $\{\lambda_i \delta_j, u_i \otimes v_j\}$ are eigenvalues/eigenvectors ofr $a \otimes b$

$$\det[a \otimes b] = \det[a]^m \det[b]^n \tag{34}$$

where $a, b$ are of order $n \times n$ and $m \times m$ respectively.

$$\text{Tr}[a \otimes b] = \text{Tr}[a]\,\text{Tr}[b] \tag{35}$$

$$\text{Tr}[ab] = \text{Vec}[a']'\text{Vec}[b] \tag{36}$$

$$\text{Tr}[abc] = \text{Vec}[a']'(I \otimes b)\text{Vec}[c] \tag{37}$$

$$\text{Vec}[a \otimes b] = (I_n \otimes \mathcal{T}_{q,m} I_p)(\text{Vec}[a] \otimes \text{Vec}[b]) \tag{38}$$

Let $a$ be $n \times n$ matrix and $b$ be a $m \times m$ matrix. Let $\oplus$ be teh Kronecker sum

$$a \oplus b = a \otimes I_m + I_n \otimes b \tag{39}$$

$$\text{rank}[a \otimes b] = \text{rank}[a]\text{rank}[b] \tag{40}$$

$$e^{a \oplus b} = e^a \otimes e^b \tag{41}$$

$$\begin{aligned} \nabla^2_{\text{Vec}[x]}(axbx') &= \nabla^2_{\text{Vec}[x]} \text{Vec}(x')(b' \otimes a)\text{Vec}x \\ &= b \otimes a' + b' \otimes a\text{Tr}(a \otimes b) \\ &= \text{Tr}[a]\,\text{Tr}[b] \end{aligned} \tag{42}$$

where $a, b, x$ are matrices. Regarding the properties of the vec transpose matrix. First note

$$\mathcal{T}_{1,m} = \mathcal{T}_{m,1} = I_m \tag{43}$$

Given an $m \times n$ matrix a

$$\mathcal{T}_{n,m}\mathcal{T}_{m,n}\text{Vec}[a] = \text{Vec}[a] \tag{44}$$

Thus

$$\mathcal{T}_{m,n}^{-1} = \mathcal{T}_{n,m} \tag{45}$$

In addition it can be shown

$$\mathcal{T}_{n,m} = (\mathcal{T}_{m,n})' \tag{46}$$

Thus $\mathcal{T}_{m,n}$ is orthonormal.

Given an $m \times n$ matrix $a$ and an $p \times q$ matrix $b$

$$a \otimes b = \mathcal{T}_{m,p}(b \otimes a)\mathcal{T}_{q,n} \tag{47}$$

To see why let $c$ be an arbitrary $q \times n$ matrix and note

$$\begin{aligned} \mathcal{T}_{m,p}(b \otimes a)\mathcal{T}_{q,n}\text{Vec}[c] &= \mathcal{T}_{m,p}(b \otimes a)\text{Vec}[c'] = \mathcal{T}_{m,p}\text{Vec}[ac'b'] \\ &= \text{Vec}[bca'] = (a \otimes b)\text{Vec}[c] \end{aligned} \tag{48}$$

From the previous property it easily follows that

$$\mathcal{T}_{p,m} a \otimes b = (b \otimes a)\mathcal{T}_{q,n} \tag{49}$$

$$a \otimes b\mathcal{T}_{n,q} = \mathcal{T}_{m,p}(b \otimes a) \tag{50}$$

Moreover if either $a$ or $b$ are vectors, i.e., $n = 1$ or $q = 1$ then

$$a \otimes b = \mathcal{T}_{m,p}(b \otimes a) \tag{51}$$

# 9 Hadamard product

The Hadamard product $a \circ b$ of two $m \times n$ matrices is defined as follows

$$(a \circ b)_{i,j} = a_{i,j}b_{i,j} \text{for } i = 1 \cdots m, \; j = 1 \cdots n \tag{52}$$

**Properties:**

$$\text{Vec}[a \circ b] = \text{Diag}[\text{Vec}[a]]\text{Vec}[b] = \text{Diag}[\text{Vec}[b]]\text{Vec}[a] \tag{53}$$

where Diag$[v]$ of a vector $v$ is a diagonal matrix whose diagonal is the vector $v$.

# 10 Matrix Calculus

**Partial Derivatives:** The partial derivative of a $p \times q$ matrix $f(x)$ with respect to a $m \times n$ matrix $x$ is a $(mp) \times (nq)$ matrix takes the following form

$$\frac{\partial f(x)}{\partial x} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial f(x)}{\partial x_{11}} & \cdots & \frac{\partial f(x)}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_{m1}} & \cdots & \frac{\partial f(x)}{\partial x_{mn}} \end{pmatrix} \tag{54}$$

where

$$\frac{\partial f(x)}{\partial x_{ij}} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial f(x)_{11}}{\partial x_{ij}} & \cdots & \frac{\partial f(x)_{1q}}{\partial x_{ij}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)_{p1}}{\partial x_{ij}} & \cdots & \frac{\partial f(x)_{pq}}{\partial x_{ij}} \end{pmatrix} \tag{55}$$

Thus

$$\frac{\partial f(x)}{\partial x} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial f(x)_{11}}{\partial x_{11}} & \cdots & \frac{\partial f(x)_{11}}{\partial x_{1n}} & \cdots & \frac{\partial f(x)_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f(x)_{1q}}{\partial x_{1n}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \frac{\partial f(x)_{11}}{\partial x_{m1}} & \cdots & \frac{\partial f(x)_{11}}{\partial x_{mn}} & \cdots & \frac{\partial f(x)_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f(x)_{1q}}{\partial x_{mn}} \\ & \vdots & & \cdots & & \vdots & \\ \frac{\partial f(x)_{p1}}{\partial x_{11}} & \cdots & \frac{\partial f(x)_{p1}}{\partial x_{1n}} & \cdots & \frac{\partial f(x)_{pq}}{\partial x_{11}} & \cdots & \frac{\partial f(x)_{pq}}{\partial x_{1n}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \frac{\partial f(x)_{p1}}{\partial x_{m1}} & \cdots & \frac{\partial f(x)_{p1}}{\partial x_{mn}} & \cdots & \frac{\partial f(x)_{pq}}{\partial x_{m1}} & \cdots & \frac{\partial f(x)_{pq}}{\partial x_{mn}} \end{pmatrix} \tag{56}$$

**Matrix Derivatives:** $f(x)$ can be a scalar, vector or matrix, and the argument $x$ can be a scalar, vector or matrix. It is useful to define the matrix derivative differently depending on the types for $f$ and $x$:
If $x$ is a matrix and $f(x)$ a scalar or if $x$ is a scalar and $f(x)$ us a matrix, we define the matrix derivative $df/dx$ as follows

$$\frac{df(x)}{dx} = D_x f(x) \stackrel{\text{def}}{=} \frac{\partial f(x)}{\partial x'} = \begin{pmatrix} \frac{\partial f(x)_1}{\partial x_1} & \cdots & \frac{\partial f(x)_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)_p}{\partial x_1} & \cdots & \frac{\partial f(x)_p}{\partial x_m} \end{pmatrix} \tag{57}$$

otherwise we define it as follows

$$D_x f(x) = \frac{df(x)}{dx} \stackrel{\text{def}}{=} \frac{\partial \text{Vec}[f(x)]}{\partial \text{Vec}[x]'} \tag{58}$$

**Jacobians, Gradients:** The **Jacobian** matrix is another word as the matrix derivative (as defined here). Sometimes we use the symbol $J_x$ to represent it

$$J_x f(x) = \frac{df(x)}{dx} = D_x f(x) \tag{59}$$

The **gradient** is the transpose of the Jacobian, i.e. the transpose of the matrix derivative. We represent it as $\nabla_x$

$$\nabla_x f(x) = (J_x(x))' \tag{60}$$

**Hessians:** We define the Hessian of a scalar function $f(x)$ of a vector $x$ as the derivative of the transpose of the derivative of $x$. We represent it with a variety of symbols, incluyding : $H_x f(x), \nabla_x^2 f(x)$ as seen below

$$\begin{aligned} H_x f(x) = \nabla_x^2 f(x) = D_x(D_x f(x)') = D_{xx'} f(x) = \frac{d}{dx}(\frac{d}{dx} f(x))' \\ = \frac{d^2 f(x)}{dx dx'} = \frac{\partial}{\partial x'} \left( \frac{\partial f(x)}{\partial x'} \right)' = \frac{\partial^2 f(x)}{\partial x' \partial x} \\ \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix} \end{aligned} \tag{61}$$

## 10.1 Chain Rule

Let $z = g(y)$, $y = f(x)$, be matrix functions of matrice

$$D_x z = D_y z \; D_x y \tag{62}$$

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \tag{63}$$

or in partial derivative notation

$$\frac{\partial \mathrm{Vec}[z]}{\partial x'} = \frac{\partial \mathrm{Vec}[z]}{\partial y'} \frac{\partial \mathrm{Vec}[y]}{\partial x'} \tag{64}$$

Taking the transpose we get th chain rule in gradient notation

$$\nabla_x z = \nabla_x y \nabla_y z \tag{65}$$

**Example 1.** Let $x$, $y$ vectors

$$\frac{d\ e^{y'x}}{dx} = \frac{d\ e^{y'x}}{dy'x}\frac{d\ y'x}{dx} = e^{y'x}\ y' \tag{66}$$

## 10.2 Product Rules

Let $f(x)$ be an $m \times n$ matrix and $g(x)$ a $n \times p$ matrix then

$$\frac{df(x)g(x)}{dx} = (g(x)' \otimes I_m)\frac{df(x)}{dx} + (I_p \otimes f(x))\frac{dg(x)}{dx} \tag{67}$$

The rule is derived in Example 2

**Corollary 1.** *Inner Products: If $f(x), g(x)$ are vectors then applying the product rule*

$$\frac{df(x)'g(x)}{dx} = g(x)'\frac{df(x)}{dx} + f(x)'\frac{dg(x)}{dx} \tag{68}$$

*since $f(x)'$ is an $1 \times n$ matrix (i.e., not a column vector) then the derivative is defined as follows*

$$\frac{df(x)'}{dx} = \frac{d\,Vec[f(x)]}{Vec[x]} = \frac{f(x)}{dx} \tag{69}$$

*Thus*

$$\frac{df(x)'g(x)}{dx} = g(x)'\frac{df(x)}{dx} + f(x)'\frac{dg(x)}{dx} \tag{70}$$

**Corollary 2. Scalar products** *If $f(x)$ is a vector function and $g(x)$ a scalar function of a vector $x$. Then applying the product rule*

$$\frac{df(x)g(x)}{dx} = (g(x)' \otimes I_m)\frac{df(x)}{dx} + f(x)\frac{dg(x)}{dx} \tag{71}$$

*or using the standard definition for the product of a scalar times a matrix*

$$\frac{df(x)g(x)}{dx} = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx} \tag{72}$$

**Corollary 3. Derivatives with respect to scalars:** *If $f(x)g(x)$ is a matrix and $x$ a scalar. Then it can be shown that the product rule becomes as follows*

$$\frac{df(x)g(x)}{dx} = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx} \tag{73}$$

## 10.3 Matrix Differential Calculus

Working with matrix differentials provides an elegant way to prove useful rules for matrix derivatives. We define the differential $df(x, dx)$ of a matrix $f(x)$ at $x$ with increment $dx$ as the part of $f(x + dx) - f(x)$ that is linear on $dx$. When the context permits, we use the symbol $df(x)$ as short-hand notation for $df(x, dx)$.

**Properties of Matrix differentials**

$$da = 0, \text{ for } a \text{ constant with respect to } x \tag{74}$$
$$d(f(x)g(x)) = (df(x))g(x) + f(x)dg(x) \tag{75}$$
$$d(f(x) \otimes g(x)) = (df(x)) \otimes g(x) + f(x) \otimes dg(x) \tag{76}$$
$$d(f(x) \circ g(x)) = (df(x)) \circ g(x) + f(x) \circ dg(x) \tag{77}$$
$$dx^{-1} = -x^{-1}dxx^{-1} \tag{78}$$
$$d|x| = |x|\mathrm{Tr}[x^{-1}dx] \tag{79}$$
$$d\mathrm{Tr}[x] = \mathrm{Tr}[dx] \tag{80}$$
$$d[o(x)] = o(d[x]) \tag{81}$$

where $o$ is a linear function, i.e. if $x, y$ are matrices and $\alpha, \beta$ are scalars then

$$o(\alpha x) + o(\alpha y) \tag{82}$$

Example of such linear functions are the trace,, the transpose, the Vec or any operators that rearranges or reshapes $x$.

**Identification Theorem 1:** The identification theorems allow us to work with matrix differentials, which is convenient, and then extract the corresponding matrix derivatives:

$$\mathrm{Vec}[df(x, dx)] = a(x)d\mathrm{Vec}[x] \iff \frac{df(x)}{dx} = a(x) \tag{83}$$

or in simplified notation

$$d\mathrm{Vec}[f(x)] = a(x)d\mathrm{Vec}[x] \iff \frac{df(x)}{dx} = a(x) \tag{84}$$

**Identification Corollary 1:** If $f(x)$ is a scalar

$$df(x) = \mathrm{Tr}[a(x)dx] \rightarrow \frac{df(x)}{dx} = \mathrm{Vec}[a(x)']' \tag{85}$$

**Chain Rule:** The identification theorem can be seen as a chain rule.

$$\mathrm{Vec}[df(x,dx)] = \frac{df(x)}{dx}\mathrm{Vec}[dx] \tag{86}$$

We can then combine it with the chain rule of derivatives. Thus, if $z = h(x) = g(y)$ where $y = f(x)$ then

$$\mathrm{Vec}[dh(x,dx)] = \frac{dz}{dx}\mathrm{Vec}[dx] = \frac{dz)}{dy}\frac{dy}{dx}\mathrm{Vec}[dx] \tag{87}$$

From this follows the **Cauchy rule of invariance:**

$$dh(x,dx) = dg(x, df(x,dx)) \tag{88}$$

*Proof.*

$$dh(x,dx) = \frac{dg(y)}{dy}\frac{dy}{dx}\mathrm{Vec}[dx] = \frac{dg(y)}{dy}\mathrm{Vec}[dy] = dg(y,dy) \tag{89}$$

□

### 10.3.1 Proofs of differential properties

Here we show examples of how to prove the properties of matrix differentials

**Product rule:**

$$d(f(x)g(x)) = (df(x))g(x) + f(x)dg(x) \tag{90}$$

Using linear properties of differentials

$$\begin{aligned}(d(f(x)g(x)))_{ij} = d(f(x)g(x))_{ij} = d(\sum_k f_k(x)g_k(x)) \\ = \sum_k d(f_k(x)g_k(x))\end{aligned} \tag{91}$$

And since to first order

$$f_k(x+dx)g_k(x+dx) = f_k(x)\frac{dg_k(x)}{dx}dx + g_k(x)\frac{df_k(x)}{dx}dx \tag{92}$$

then

$$d(f_k(x)g_k(x)) = f_k(x)dg_k(x) + (df_k(x))g_k(x) \tag{93}$$

**Inverse rule:**

$$d(x^{-1}x) = x^{-1}dx + dx^{-1}x \tag{94}$$

and since $xx' = I$ and the differntial of a constant is zero, then

$$dx^{-1} = -x^{-1}dxx^{-1} \tag{95}$$

**Identification Corollary 1:** From the properties of the Vec operator

$$\text{Tr}[a(x)dx] = \text{Vec}[a(x)']'\text{Vec}[dx] \tag{96}$$

Thus considering $f(x)$ is a scalar, if

$$df(x) = \text{Tr}[a(x)dx] \tag{97}$$

then

$$d\text{Vec}(fx) = \text{Vec}[a(x)']'d\text{Vec}[x] \tag{98}$$

## 10.4 Matrix derivative rules

Here we see examples of how to use matrix differential rules to obtain some useful matrix derivatives rules

**Example 2. Product Rule:** Let $f(x) \in \Re^{m\times n}$, $g(x) \in \Re^{n\times p}$

$$d(f(x)g(x)) = (df(x))g(x) + f(x)dg(x) = I_m(df(x))g(x) = f(x)d(g(x))I_p \tag{99}$$

Thus

$$\begin{aligned}\text{Vec}[d(f(x)g(x)] &= g(x)' \otimes I_m \text{Vec}[df(x)] + I_p \otimes f(x)\text{Vec}[dg(x)] \\ &= g(x)' \otimes I_m \frac{df(x)}{dx}\text{Vec}[d(x)] + I_p \otimes f(x)\frac{dg(x)}{dx}\text{Vec}[d(x)]\end{aligned} \tag{100}$$

Thus, using the identification theorem

$$\frac{d(f(x)g(x))}{dx} = (g(x)' \otimes I_m)\frac{df(x)}{dx} + (I_p \otimes f(x))\frac{dg(x)}{dx} \tag{101}$$

**Example 3.** Let $a \in \Re^{m\times n}$ be a constant with respect to $x \in \Re^{n\times p}$. Then using the product rule

$$dax = adx \tag{102}$$

Moreover

$$\text{Vec}[d(ax)] = \text{Vec}[a(dx)I_p] = I_p \otimes a\text{Vec}[dx] \tag{103}$$

Thus

$$\frac{dax}{dx} = I_p \otimes a \tag{104}$$

Let $x \in \Re^{m\times n}$, $a \in \Re^{n\times p}$

$$d(xa) = (dx)a = I_m(dx)a \tag{105}$$

Thus

$$\text{Vec}[d(xa)] = a' \otimes I_m\text{Vec}[dx] \tag{106}$$

and

$$\frac{dxa}{dx} = a' \otimes I_m \tag{107}$$

**Example 4.** Let $x \in \Re^{m\times n}$,

$$dxx' = (dx)x' + x(dx') = I_m(dx)x' + x(dx')I_m \tag{108}$$

Thus

$$\begin{aligned} \text{Vec}[dxx'] &= (x \otimes I_m)dx + (I_m \otimes x)dx' \\ &= (x \otimes I_m)dx + (I_m \otimes x)\mathcal{T}_{n,m}dx \end{aligned} \tag{109}$$

Thus, using the identification theorem

$$\frac{dxx'}{dx} = (x \otimes I_m) + (I_m \otimes x)\mathcal{T}_{n,m} \tag{110}$$

Moreover, since

$$(I_m \otimes x)\mathcal{T}_{n,m} = \mathcal{T}_{m,m}(x \otimes I_m) \tag{111}$$

then

$$\frac{dxx'}{dx} = (x \otimes I_m) + \mathcal{T}_{m,m}(x \otimes I_m) \tag{112}$$

Note if $m = 1$, then $vv_{m,m} = I_m = 1$ and

$$\frac{dxx'}{dx} = 2x \tag{113}$$

**Example 5.** Let $x, a$ be symmetric matrices

$$d(xax) = I(dx)(ax) + xa(dx)I \tag{114}$$

$$\text{Vec}[d(xax)] = ((ax)' \otimes I)\text{Vec}[dx] + (I \otimes (xa))\text{Vec}[dx] = (I \otimes (xa))'\text{Vec}[dx] + (I \otimes (xa))\text{Vec}[dx \tag{115}$$

Thus

$$\frac{d(xax)}{dx} = (I \otimes (xa))' + (I \otimes (xa)) \tag{116}$$

Note for $x, a$ symmetric, the product $xa$ is symmetric only if $xa = ax$.

**Example 6.** Using the product rule

$$\frac{dx^{-1}x}{dx} = (x' \otimes I_n)\frac{dx^{-1}}{dx} + I_n \otimes x^{-1} = 0 \tag{117}$$

Thus

$$\frac{dx^{-1}}{dx} = -(x' \otimes I_n)^{-1}(I_n \otimes x^{-1}) = -(x' \otimes x^{-1}) \tag{118}$$

**Example 7.**

$$d\text{Tr}[axb] = \text{Tr}[adxb] = \text{Tr}[badx] \tag{119}$$

Thus

$$\frac{d\text{Tr}[axb]}{dx} = ba \tag{120}$$

**Example 8.**

$$
\begin{aligned}
d\mathrm{Tr}[(axb-c)'e(axb-c)] &= \mathrm{Tr}[(a(dx)b)'e(axb-c)] \\
&+\mathrm{Tr}[(axb-c)ea(dx)b]
\end{aligned} \tag{121}
$$

Note

$$
\mathrm{Tr}[(a(dx)b)'e(axb-c)] = \mathrm{Tr}[(axb-c)'e'a(dx)b] = \mathrm{Tr}[b(axb-c)'e'adx] \tag{122}
$$

and

$$
\mathrm{Tr}[(axb-c)'ea(dx)b] = \mathrm{Tr}[b(axb-c)'eadx] \tag{123}
$$

Thus

$$
d\mathrm{Tr}[(axb-c)'e(axb-c)] = \mathrm{Tr}[b(axb-c)'(e+e')adx] \tag{124}
$$

and using the identification theorem

$$
\frac{d\mathrm{Tr}[(axb-c)'e(axb-c)]}{dx} = b(axb-c)'(e+e')a \tag{125}
$$

**Example 9.** Let $a, b$ be constant. Then using the product rule

$$
d(axb) = (dax)b = a(dx)b \tag{126}
$$

Using the rules for Kronecker products

$$
d\mathrm{Vec}[axb] = \mathrm{Vec}[a(dx)b] = (b' \otimes a)\mathrm{Vec}[dx] \tag{127}
$$

Thus, using the identification theorem

$$
\frac{daxb}{dx} = b' \otimes a \tag{128}
$$

Note if $a, b$ are vectors then

$$
\frac{da'xb}{dx} = ab' = b'a \tag{129}
$$

**Example 10.** Let $a, b$ be constant. Then using the inverse rule

$$d(ax^{-1}b) = -ax^{-1}(dx)x^{-1}b \tag{130}$$

Using the rules for Kronecker products

$$d\mathrm{Vec}[ax^{-1}b] = -(b'x^{-1} \otimes (ax^{-1})) \otimes \mathrm{Vec}[dx] \tag{131}$$

Thus, using the identification theorem

$$\frac{dax^{-1}b}{dx} = -(x^{-1}b)' \otimes (ax^{-1}) \tag{132}$$

Note, if $a, b$ are vectors and $x$ is symmetric

$$\frac{da'x^{-1}b}{dx} = -x^{-1}ba'x^{-1} \tag{133}$$

**Example 11.**

$$\begin{aligned} d(f(x)'g(x)) &= (df(x)')g(x) + f(x)(dg(x)) = g(x)'df(x) + f(x)dg(x) \\ &= \Big(g(x)'\frac{df(x)}{dx} + f(x)\frac{dg(x)}{dx}\Big)dx \end{aligned} \tag{134}$$

Thus, using the identification theorem

$$\frac{df(x)'g(x)}{dx} = g(x)'\frac{df(x)}{dx} + f(x)\frac{dg(x)}{dx} \tag{135}$$

**Example 12.** Using the inner product rule

$$\begin{aligned} dg(x)'ah(x) &= (dg(x))'ah(x) + g(x)'adh(x) \\ &= \Big(h(x)'a'\frac{dg(x)}{dx} + g(x)'a\frac{dh(x)}{dx}\Big)dx \end{aligned} \tag{136}$$

Thus

$$\frac{dg(x)'ah(x)}{dx} = h(x)'a'\frac{dg(x)}{dx} + g(x)'a\frac{dh(x)}{dx} \tag{137}$$

**Example 13.**

$$d|x| = \mathrm{Tr}[|x|x^{-1}dx] \tag{138}$$

Thus, since $|x|$ is a scalar, using the identification theorem

$$\frac{d|x|}{dx} = |x|x^{-1} \tag{139}$$

**Example 14.**

$$\begin{aligned} d|x'x| &= \mathrm{Tr}[|x'x|(x'x)^{-1}d(x'x)] \\ &= \mathrm{Tr}[|x'x|(x'x)^{-1}x'dx + (dx')x] \end{aligned} \tag{140}$$

Considering $(x'x)^{-1}$ is symmetric

$$\begin{aligned} \mathrm{Tr}[(x'x)^{-1}(dx')x] &= \mathrm{Tr}[((x'x)^{-1}(dx')x)'] \\ &= \mathrm{Tr}[(x'dx)((x'x)^{-1})'] = \mathrm{Tr}[(x'x)^{-1}x'dx] \end{aligned} \tag{141}$$

Thus

$$d|x'x| = 2\mathrm{Tr}[|x'x|(x'x)^{-1}x; dx] \tag{142}$$

and, since $|x'x|$ is a scalar

$$\frac{d|x'x|}{dx} = 2|x'x|(x'x)^{-1}x \tag{143}$$

**Example 15.**

$$\begin{aligned} d\mathrm{Tr}[a'(x'x)^{-1}a] = \mathrm{Tr}[a'(d(x'x)^{-1}a] &= -\mathrm{Tr}[a'(x'x)^{-1}d(x'x)(x'x)^{-1}a] \\ &= -\mathrm{Tr}[a'(x'x)^{-1}(x'dx + (dx')x)(x'x)^{-1}a] \end{aligned} \tag{144}$$

Note

$$\mathrm{Tr}[a'(x'x)^{-1}x'dx(x'x)^{-1}a] = \mathrm{Tr}[(x'x)^{-1}aa'(x'x)^{-1}x'dx] \tag{145}$$

and

$$\begin{aligned} \mathrm{Tr}[a'(x'x)^{-1}(dx')x(x'x)^{-1}a] &= \mathrm{Tr}[(dx')x(x'x)^{-1}aa'(x'x)^{-1}] \\ &= \mathrm{Tr}[(x(x'x)^{-1}aa'(x'x)^{-1})dx] \\ &= \mathrm{Tr}[(x'x)^{-1}aa'(x'x)^{-1}x'dx] \end{aligned} \tag{146}$$

Thus

$$\frac{d}{dx}\mathrm{Tr}[a'(x'x)^{-1}a] = -2(x'x)^{-1}aa'(x'x)^{-1}x' \tag{147}$$

**Example 16.** Let $a = xx'$ where $x$ is an $n \times n$ lower triangular matrix.

$$da = xdx' + (dx)c' = cdx'I_n + I_n(dx)c' \tag{148}$$

$$\text{Vec}[da] = \Big(I_n \otimes)\mathcal{T}_{n,n} + (c\otimes)\Big)\text{Vec}[dx] \tag{149}$$

$$\frac{da}{dx} = \frac{d\text{Vec}[a]}{d\text{Vec}[x]}\big(I_n \otimes)\mathcal{T}_{n,n} + (c\otimes)\big) \tag{150}$$

Note

$$\text{Vech}[a] = S_n \text{Vec}[a] \tag{151}$$

Using the chain rule

$$\frac{d\text{Vech}[a]}{d\text{Vech}[x]} = \frac{d\text{Vech}[a]}{da}\frac{da}{dx}\frac{dx}{d\text{Vech}[x]} = S_n \frac{da}{dx} S_n' \tag{152}$$

**Example 17. Matrix powers**
Let $k$ a non-negative integer

$$\frac{dx^k}{dx} = \frac{dx^{k-1}x}{dx} = (x' \otimes I)\frac{dx^{k-1}}{dx} + (I \otimes x^{k-1}) \tag{153}$$

Thus, unfolding through time

$$\frac{dx^k}{dx} = \sum_{i=1}^{k} (x')^{i-k} \otimes x^{i-1} \tag{154}$$

**Example 18. Matrix exponentials**
Let

$$e^x = \sum_{k=0}^{\infty} \frac{a^k}{k!} \tag{155}$$

Thus

$$\frac{de^x}{x} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=1}^{k} (x')^{k-i} \otimes x^{i-1} \tag{156}$$

**Example 19. Matrix logarithms**

$$\log(x) = \sum_{k=1}^{\infty} \frac{1}{k}(I - x)^k \tag{157}$$

Thus

$$\frac{de^x}{x} = \sum_{k=1}^{\infty} \frac{1}{k} \sum_{i=1}^{k} ((I - x)')^{k-i} \otimes (I - x)^{i-1} \tag{158}$$

**Example 20. Hadamard Products**

$$\begin{aligned} \mathrm{Vec}[d(f(x) \circ g(x))] &= \mathrm{Vec}[(df(x)) \circ g(x)] + \mathrm{Vec}[f(x) \circ dg(x)] \\ &= \mathrm{Diag}[\mathrm{Vec}[g(x)]]\mathrm{Vec}[df(x)] + \mathrm{Diag}[\mathrm{Vec}[f(x)]]\mathrm{Vec}[dg(x)] \\ &= \mathrm{Diag}[\mathrm{Vec}[g(x)]]\frac{df(x)}{dx}\mathrm{Vec}[dx] \\ &+ \mathrm{Diag}[\mathrm{Vec}[f(x)]]\frac{dg(x)}{dx}\mathrm{Vec}[dx] \end{aligned} \tag{159}$$

Thus

$$\begin{aligned} \frac{d(f(x) \circ g(x))}{dx} &= \mathrm{Diag}[\mathrm{Vec}[g(x)]]\ \frac{df(x)}{dx} \\ &+ \mathrm{Diag}[\mathrm{Vec}[f(x)]]\ \frac{dg(x)}{dx} \end{aligned} \tag{160}$$

**Example 21. Kronecker Products** Let $f(x)$ be an $m \times n$ matrix and $g(x)$ a $p \times q$ matrix

$$\begin{aligned} \mathrm{Vec}[d(f(x) \otimes g(x))] &= \mathrm{Vec}[(df(x)) \otimes g(x)] + \mathrm{Vec}[f(x) \otimes dg(x)] \\ &= (I_n \otimes \mathcal{T}_{q,m} \otimes I_p)\Big((\mathrm{Vec}[df(x)] \otimes \mathrm{Vec}[g(x)]) + (\mathrm{Vec}[f(x)] \otimes \mathrm{Vec}[dg(x)])\Big) \\ &= (I_n \otimes \mathcal{T}_{q,m} \otimes I_p)\Big((g(x)' \otimes I_{mn})\mathrm{Vec}[df(x)] + (I_{pq} \otimes f(x))\mathrm{Vec}[dg(x)]\Big) \\ &= (I_n \otimes a)\mathrm{Vec}[df(x)] + (b \otimes I_p)\mathrm{Vec}[dg(x)] \end{aligned} \tag{161}$$

where

$$a = (\mathcal{T}_{q,m} \otimes I_q)(I_m \mathrm{Vec}[g(x)]) \tag{162}$$

$$b = (I_n \otimes \mathcal{T}_{q,m})(\mathrm{Vec}[f(x)] \otimes I_q) \tag{163}$$

Thus

$$\frac{d(f(x) \otimes g(x))}{dx} = (I_n \otimes a)\frac{df(x)}{dx} + (b \otimes I_p)\frac{dg(x)}{dx} \tag{164}$$

## 10.5 Matrix Calculus Summary

**Chain Rule:** If $z = g(y)$, is a vector function of a vector $y$ and $y = f(x)$ is a vector function of a vector $x$ then

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \tag{165}$$

**Product Rule:** Let $f(x)$ be an $m \times n$ matrix and $g(x)$ a $n \times p$ matrix then

$$\frac{df(x)g(x)}{dx} = (g(x)' \otimes I_m)\frac{df(x)}{dx} + (I_p \otimes f(x))\frac{dg(x)}{dx} \tag{166}$$

**Useful Formulae**

$$\frac{dcx}{dx} = \frac{x'c}{dx} = c \tag{167}$$

$$\frac{daxb}{dx} = b' \otimes a \tag{168}$$

$$\frac{dax^{-1}b}{dx} = -(x^{-1}b)' \otimes (x^{-1}a) \tag{169}$$

$$\frac{d}{dx}\mathrm{Tr}[a'(x'x)^{-1}a] = -2(x'x)^{-1}aa'(x'x)^{-1}x' \tag{170}$$

$$\frac{d}{dx}\mathrm{Tr}[(axb-c)'e(axb-c)] = b(axb-c)'(e+e')a \tag{171}$$

$$\frac{d}{dx}\mathrm{Tr}[axb] = ba \tag{172}$$

$$\frac{d|x|}{dx} = |x|\ x^{-1} \tag{173}$$

$$\frac{d|x'x|}{dx} = 2|x'x|(x'x)^{-1}x' \tag{174}$$

$$\nabla^2_{\mathrm{Vec}[x]}(axbx') = \mathrm{Tr}[a]\mathrm{Tr}[b] \tag{175}$$

# 11 Matrix Differential Equations

$$\frac{d(a_t + b_t)}{dt} = \frac{da_t}{dt} + \frac{db_t}{dt} \tag{176}$$

$$\frac{d(a_t b_t)}{dt} = \frac{da_t}{dt} b_t + a_t \frac{db_t}{dt} \tag{177}$$

$$\frac{d(a_t^2)}{dt} = \frac{da_t}{dt} a_t + a_t \frac{da_t}{dt} \tag{178}$$

$$\frac{da_t a_t^{-1}}{dt} = 0 = \frac{da_t}{dt} a_t^{-1} + a_t^{-1} \frac{da_t}{dt} \tag{179}$$

$$\frac{a_t^{-1}}{dt} = -a_t^{-1} \frac{da_t}{dt} a^{-1} \tag{180}$$

$$\tag{181}$$

It is only possible to expect $de^{at}dt = a_t e^{at}$ under conditions of full commutativity. Types of linear matrix equations

$$\frac{da_t}{dt} = b a_t c \tag{182}$$

$$\tag{183}$$

with special cases when $b$ or $c$ are the identity matrix. The following tricks allow moving the coefficients to the left, or right

$$\frac{da_t}{dt} = a_t c \tag{184}$$

then

$$\frac{da_t'}{dt} = c' a_t' \tag{185}$$

$$\tag{186}$$

and if

$$\frac{da_t}{dt} = b a_t \tag{187}$$

then

$$\frac{da_t^{-1}}{dt} = -a^{-1} \frac{da_t}{dt} a^{-1} = -a^{-1} b a_t a_t^{-1} = -a_t^{-1} b \tag{188}$$

$$\tag{189}$$

It can be shown that if $a_0$ is non-singular, then the solution $a_t$ is non-singular for all $t$.

## 12 Appendix

**Lemma 1.** *Let $a, b$ matrices such that spectral radius of $ba^{-1}$ is less than one. Then*

$$(a+b)^{-1} = a^{-1}\sum_{i=0}^{\infty}(-1)^i(ba^{-1})^i \tag{190}$$

*Proof.*

$$\begin{aligned}(a+b)a^{-1}\sum_{i=0}^{\infty}(-1)^i(ba^{-1})^i &= aa^{-1}(I - ba^{-1} + (ba^{-1})^2 - (ba^{-1})^3 + \cdots) \\ &+ ba^{-1}(I - ba^{-1} + (ba^{-1})^2 - (ba^{-1})^3 + \cdots) = I \end{aligned} \tag{191}$$

□

**Lemma 2.** *Let $a, u, b, v$ matrices.*

$$(a+ubv)^{-1} = a^{-1} - a^{-1}u(b^{-1} + va^{-1}u)^{-1}va^{-1} \tag{192}$$

*Proof.* First note

$$(a+ubv)(a^{-1}(I + ubva^{-1})^{-1}) = I \tag{193}$$

Thus

$$(a+ubv)^{-1} = a^{-1}(I+ubva^{-1})^{-1} = a^{-1}\sum_{k=0}^{\infty}(-1)^k(ubva^{-1})^k \tag{194}$$

$$= a^{-1}(I + \sum_{k=1}^{\infty}(-1)^k(ubva^{-1})^k) \tag{195}$$

$$\tag{196}$$

Note

$$(ubva^{-1})^k = ub(va^{-1}ub)(va^{-1}ub)(va^{-1}ub)\cdots(va^{-1}ub)va^{-1} \tag{197}$$

$$= ub(va^{-1}ub)^{k-1}va^{-1} \tag{198}$$

Thus

$$(a+ubv)^{-1} = a^{-1}(I - ub(\sum_{k=0}^{\infty}(-1)^k(ubva^{-1})^k)va^{-1} \tag{199}$$

$$= a^{-1}(I - ub(I + va^{-1}ub)^{-1})va^{-1} \tag{200}$$

$$= a^{-1} - a^{-1}u(b^{-1} + va^{-1}u)^{-1}va^{-1} \tag{201}$$

where we used the fact that

$$(b^{-1} + va^{-1}u)^{-1} = (I - va^{-1}ub) \tag{202}$$

□