# matrix identities

sam roweis

---

(revised June 1999)

**note** that $\mathbf{a}$,$\mathbf{b}$,$\mathbf{c}$ and $\mathbf{A}$,$\mathbf{B}$,$\mathbf{C}$ do not depend on $\mathbf{X}$,$\mathbf{Y}$,$\mathbf{x}$,$\mathbf{y}$ or $z$

## 0.1 basic formulae

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{A}\mathbf{B}+\mathbf{A}\mathbf{C} \tag{1a}$$

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T+\mathbf{B}^T \tag{1b}$$

$$(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T \tag{1c}$$

$$\text{if individual inverses exist} \quad (\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{1d}$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \tag{1e}$$

## 0.2 trace, determinant and rank

$$|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}| \tag{2a}$$

$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|} \tag{2b}$$

$$|\mathbf{A}| = \prod \text{evals} \tag{2c}$$

$$\mathrm{Tr}\left[\mathbf{A}\right] = \sum \text{evals} \tag{2d}$$

if the cyclic products are well defined,

$$\mathrm{Tr}\left[\mathbf{A}\mathbf{B}\mathbf{C}\ldots\right] = \mathrm{Tr}\left[\mathbf{B}\mathbf{C}\ldots\mathbf{A}\right] = \mathrm{Tr}\left[\mathbf{C}\ldots\mathbf{A}\mathbf{B}\right] = \ldots \tag{2e}$$

$$\mathrm{rank}\left[\mathbf{A}\right] = \mathrm{rank}\left[\mathbf{A}^T\mathbf{A}\right] = \mathrm{rank}\left[\mathbf{A}\mathbf{A}^T\right] \tag{2f}$$

$$\text{condition number} = \gamma = \sqrt{\frac{\text{biggest eval}}{\text{smallest eval}}} \tag{2g}$$

**derivatives** of scalar forms with respect to scalars, vectors, or matricies are indexed in the obvious way. similarly, the indexing for derivatives of vectors and matrices with respect to scalars is straightforward.

## 0.3 derivatives of traces

$$\frac{\partial \mathrm{Tr}\left[\mathbf{X}\right]}{\partial \mathbf{X}} = \mathbf{I} \tag{3a}$$

$$\frac{\partial \mathrm{Tr}\left[\mathbf{X}\mathbf{A}\right]}{\partial \mathbf{X}} = \frac{\partial \mathrm{Tr}\left[\mathbf{A}\mathbf{X}\right]}{\partial \mathbf{X}} = \mathbf{A}^T \tag{3b}$$

$$\frac{\partial \mathrm{Tr}\left[\mathbf{X}^T\mathbf{A}\right]}{\partial \mathbf{X}} = \frac{\partial \mathrm{Tr}\left[\mathbf{A}\mathbf{X}^T\right]}{\partial \mathbf{X}} = \mathbf{A} \tag{3c}$$

$$\frac{\partial \mathrm{Tr}\left[\mathbf{X}^T\mathbf{A}\mathbf{X}\right]}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{X} \tag{3d}$$

$$\frac{\partial \mathrm{Tr}\left[\mathbf{X}^{-1}\mathbf{A}\right]}{\partial \mathbf{X}} = -\mathbf{X}^{-1}\mathbf{A}^T\mathbf{X}^{-1} \tag{3e}$$

## 0.4 derivatives of determinants

$$\frac{\partial |\mathbf{A}\mathbf{X}\mathbf{B}|}{\partial \mathbf{X}} = |\mathbf{A}\mathbf{X}\mathbf{B}|(\mathbf{X}^{-1})^T = |\mathbf{A}\mathbf{X}\mathbf{B}|(\mathbf{X}^T)^{-1} \tag{4a}$$

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \tag{4b}$$

$$\frac{\partial \ln |\mathbf{X}(z)|}{\partial z} = \mathrm{Tr}\left[\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial z}\right] \tag{4c}$$

$$\frac{\partial |\mathbf{X}^T\mathbf{A}\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}^T\mathbf{A}\mathbf{X}|(\mathbf{A}\mathbf{X}(\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \tag{4d}$$

## 0.5 derivatives of scalar forms

$$\frac{\partial (\mathbf{a}^T\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial (\mathbf{x}^T\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a} \tag{5a}$$

$$\frac{\partial (\mathbf{x}^T\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} \tag{5b}$$

$$\frac{\partial (\mathbf{a}^T\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{5c}$$

$$\frac{\partial (\mathbf{a}^T\mathbf{X}^T\mathbf{b})}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{5d}$$

$$\frac{\partial (\mathbf{a}^T\mathbf{X}\mathbf{a})}{\partial \mathbf{X}} = \frac{\partial (\mathbf{a}^T\mathbf{X}^T\mathbf{a})}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{5e}$$

$$\frac{\partial (\mathbf{a}^T\mathbf{X}^T\mathbf{C}\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{C}^T\mathbf{X}\mathbf{a}\mathbf{b}^T + \mathbf{C}\mathbf{X}\mathbf{b}\mathbf{a}^T \tag{5f}$$

$$\frac{\partial \left((\mathbf{X}\mathbf{a} + \mathbf{b})^T\mathbf{C}(\mathbf{X}\mathbf{a} + \mathbf{b})\right)}{\partial \mathbf{X}} = (\mathbf{C} + \mathbf{C}^T)(\mathbf{X}\mathbf{a} + \mathbf{b})\mathbf{a}^T \tag{5g}$$

Consider a vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, where $n$ is the number of input variables and $m$ is the number of output variables. The Jacobian matrix of $\mathbf{f}$, denoted as $J(\mathbf{f})$ or $\mathbf{J}$, is an $m \times n$ matrix defined as follows:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

the **derivative** of one vector $\mathbf{y}$ with respect to another vector $\mathbf{x}$ is a matrix whose $(i,j)^{th}$ element is $\partial y(j)/\partial x(i)$. such a derivative should be written as $\partial \mathbf{y}^T/\partial \mathbf{x}$ in which case it is the *Jacobian* matrix of $\mathbf{y}$ wrt $\mathbf{x}$. its determinant represents the ratio of the hypervolume $d\mathbf{y}$ to that of $d\mathbf{x}$ so that $\int f(\mathbf{y})d\mathbf{y} = \int f(\mathbf{y}(\mathbf{x}))|\partial \mathbf{y}^T/\partial \mathbf{x}|d\mathbf{x}$. however, the sloppy forms $\partial \mathbf{y}/\partial \mathbf{x}$, $\partial \mathbf{y}^T/\partial \mathbf{x}^T$ and $\partial \mathbf{y}/\partial \mathbf{x}^T$ are often used for this Jacobain matrix.

## 0.6 derivatives of vector/matrix forms

$$\frac{\partial(\mathbf{X}^{-1})}{\partial z} = -\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial z}\mathbf{X}^{-1} \tag{6a}$$

$$\frac{\partial(\mathbf{A}\mathbf{x})}{\partial z} = \mathbf{A}\frac{\partial \mathbf{x}}{\partial z} \tag{6b}$$

$$\frac{\partial(\mathbf{X}\mathbf{Y})}{\partial z} = \mathbf{X}\frac{\partial \mathbf{Y}}{\partial z} + \frac{\partial \mathbf{X}}{\partial z}\mathbf{Y} \tag{6c}$$

$$\frac{\partial(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial z} = \mathbf{A}\frac{\partial \mathbf{X}}{\partial z}\mathbf{B} \tag{6d}$$

$$\frac{\partial(\mathbf{x}^T\mathbf{A})}{\partial \mathbf{x}} = \mathbf{A} \tag{6e}$$

$$\frac{\partial(\mathbf{x}^T)}{\partial \mathbf{x}} = \mathbf{I} \tag{6f}$$

$$\frac{\partial(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T)}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}\mathbf{x}^T + \mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{I} \tag{6g}$$

## 0.7 constrained maximization

the maximum over $\mathbf{x}$ of the quadratic form:

$$\boldsymbol{\mu}^T\mathbf{x} - \frac{1}{2}\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x} \tag{7a}$$

subject to the $J$ conditions $c_j(\mathbf{x}) = 0$ is given by:

$$\mathbf{A}\boldsymbol{\mu} + \mathbf{A}\mathbf{C}\boldsymbol{\Lambda}, \qquad \boldsymbol{\Lambda} = -4(\mathbf{C}^T\mathbf{A}\mathbf{C})\mathbf{C}^T\mathbf{A}\boldsymbol{\mu} \tag{7b}$$

where the $j$th column of $\mathbf{C}$ is $\partial c_j(\mathbf{x})/\partial \mathbf{x}$

## 0.8 symmetric matrices

have real eigenvalues, though perhaps not distinct and can always be diagonalized to the form:

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^T \tag{8}$$

where the columns of $\mathbf{C}$ are (orthonormal) eigenvectors (i.e. $\mathbf{C}\mathbf{C}^T = \mathbf{I}$) and the diagonal of $\mathbf{\Lambda}$ has the eigenvalues

### 0.9 block matrices

for conformably partitioned block matrices, addition and multiplication is performed by adding and multiplying blocks in exactly the same way as scalar elements of regular matrices
however, determinants and inverses of block matrices are very tricky; for 2 blocks by 2 blocks the results are:

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{22}| \cdot |\mathbf{F}_{11}| = |\mathbf{A}_{11}| \cdot |\mathbf{F}_{22}| \tag{9a}$$

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{F}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_{22}^{-1} \\ -\mathbf{F}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_{22}^{-1} \end{bmatrix} \tag{9b}$$

$$= \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{F}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{F}_{11}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{F}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix}$$

where

$$\mathbf{F}_{11} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \qquad\qquad \mathbf{F}_{22} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

for block *diagonal* matrices things are much easier:

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}||\mathbf{A}_{22}| \tag{9d}$$

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix} \tag{9e}$$

### 0.10 matrix inversion lemma (sherman-morrison-woodbury)

using the above results for block matrices we can make some substitutions and get the following important results:

$$(\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{B}^{-1} + \mathbf{X}^T\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{A}^{-1} \tag{10}$$

$$|\mathbf{A} + \mathbf{X}\mathbf{B}\mathbf{X}^T| = |\mathbf{B}||\mathbf{A}||\mathbf{B}^{-1} + \mathbf{X}^T\mathbf{A}^{-1}\mathbf{X}| \tag{11}$$

where $\mathbf{A}$ and $\mathbf{B}$ are *square* and *invertible* matrices but need not be of the same dimension. this lemma often allows a really hard inverse to be converted into an easy inverse. the most typical example of this is when $\mathbf{A}$ is large but diagonal, and $\mathbf{X}$ has many rows but few columns