Mathematics for Machine Learning

Adam Prügel-Bennett

January 23, 2020

1 Matrix Algebra

Let $\mathbf{A}$ be a matrix and $[\mathbf{A}]_{ij}$ be the element in the i^{th} row and j^{th} column. That is, if $\mathbf{A}$ is the $n \times m$ matrix

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1,m} \\ A_{21} & A_{22} & \dots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{n,m} \end{pmatrix}$$

then $[\mathbf{A}]_{ij} = A_{ij}$.

In matrix notations an n-component vector $\boldsymbol{x}$ can be treated as a matrix with n rows and 1 column—note that we write normal (column) vectors as $\boldsymbol{x}$ and row vectors as $\boldsymbol{x}^\mathsf{T}$. The components of a vector are written $[\boldsymbol{x}]_{i1} = x_i$.

Transpose

We denote the transpose of the matrix $\mathbf{A}$ by $\mathbf{A}^\mathsf{T}$. Then

$$[\mathbf{A}^\mathsf{T}]_{ij} = [\mathbf{A}]_{ji} = A_{ji}.$$

A matrix is *symmetric* if $\mathbf{A} = \mathbf{A}^\mathsf{T}$ or $[A]_{ij} = [A]_{ji}$ for all i and j.

Matrix Multiplication

We write the product of a matrices $l \times m$ matrix, $\mathbf{A}$, and a $m \times n$ matrix, $\mathbf{B}$, by $\mathbf{AB}$ which is a $l \times n$ matrix with elements

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{m} [\mathbf{A}]_{ik} [\mathbf{B}]_{kj}$$

for all i and j.

Note that $[\mathbf{A}]_{ik}$ with i fixed is the i^{th} row and $[\mathbf{B}]_{kj}$ with j fixed is the j^{th} column so that $[\mathbf{AB}]_{ij}$ is the i^{th} row of matrix $\mathbf{A}$ times the j^{th} column of matrix $\mathbf{B}$.

For two n components vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ the dot product can be written down in matrix notation

$$\boldsymbol{x}^\mathsf{T}\boldsymbol{y} = \sum_{k=1}^{n} [\boldsymbol{x}^\mathsf{T}]_{1k} [\boldsymbol{y}]_{k1} = \sum_{k=1}^{n} [\boldsymbol{x}]_{k1} [\boldsymbol{y}]_{k1} = \sum_{k=1}^{n} x_k y_k.$$

This should not be confused with the outer product $\boldsymbol{x}\boldsymbol{y}^{\mathsf{T}}$ which is an $n \times n$ matrix with components $[\boldsymbol{x}\boldsymbol{y}^{\mathsf{T}}]_{ij} = x_i y_j$.

Exercise 1 *Let*

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}$$

Compute

$$\mathbf{AB} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}$$

If you want to you can use matlab (or octave an open source matlab). You can write

```
> A = [1, 3; 4, 2]
> B = [3,-1; 2 ,1]
> A*B
```

Exercise 2 *Compute* $\mathbf{A}^T$, $\mathbf{B}^T$ *and* $\mathbf{B}^T\mathbf{A}^T$

$$\mathbf{A}^T = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix} \qquad \mathbf{B}^T = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix} \qquad \mathbf{B}^T\mathbf{A}^T = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}$$

Hence verify that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.
In Matlab

```
> A'
> B'
> (A*B)'
> A'*B'
```

Exercise 3 *From the definition of transpose and matrix multiplication write down the following terms*

$$[(\mathbf{AB})^T]_{ij} = $$
$$[\mathbf{B}^T\mathbf{A}^T]_{ij} = $$

Hence show that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ *in general.*

Exercise 4 *Let*

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

Compute

$$\mathbf{AB} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}$$

and thus show that $\mathbf{BA}$ *does not generally equal* $\mathbf{AB}$.

```
> A=[1,2;2,3]
> B=[3,1;1,2]
> A*B
> B*A
```

Exercise 5 *Let* $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}$ *and* $\mathbf{C} = \begin{pmatrix} 3 & -2 \\ 2 & 2 \end{pmatrix}$ *compute*

$$(\mathbf{AB})\mathbf{C} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}$$

$$\mathbf{A}(\mathbf{BC}) = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix} = \begin{pmatrix} \quad & \quad \\ \quad & \quad \end{pmatrix}$$

In matlab

```
> A = [1, 3; 4, 2]
> B = [3,-1; 2 ,1]
> C = [2, -2; 3, 2]
> AB = A*B
> AB*C
> BC = B*C
> A*BC
```

Exercise 6 *For arbitrary matrices* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *write down* $[(\mathbf{AB})\mathbf{C}]_{ij}$ *and* $[\mathbf{A}(\mathbf{BC})]_{ij}$ *and thus show that matrix multiplication is associative.*

$$[(\mathbf{AB})\mathbf{C}]_{ij} = \qquad$$

$$[\mathbf{A}(\mathbf{BC})]_{ij} = \qquad$$

Exercise 7 *Let* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *be vectors with components* $[\boldsymbol{x}]_{i1} = x_i$ *and* $[\boldsymbol{y}]_{i,1} = y_i$. *Write down* $\boldsymbol{x}^\mathsf{T}\boldsymbol{M}\boldsymbol{y}$ *and* $\boldsymbol{y}^\mathsf{T}\boldsymbol{M}^\mathsf{T}\boldsymbol{x}$.

$$\boldsymbol{x}^\mathsf{T}\boldsymbol{M}\boldsymbol{y} = \qquad\qquad$$
$$\boldsymbol{y}^\mathsf{T}\boldsymbol{M}^\mathsf{T}\boldsymbol{x} = \qquad\qquad$$

2 Vector Calculus

The gradient of a function defined over an n dimensional space is defined as

$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

where to calculate the partial derivative $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ we differentiate with respect to x_i assuming x_j is a constant if $j \neq i$. When it is obvious what set of variables we are differentiating with respect to we write $\boldsymbol{\nabla}$ rather than $\boldsymbol{\nabla}_{\boldsymbol{x}}$

We note that since partial derivatives behave the same as ordinary derivatives for a function of one variable that they obey the chain rule

$$\frac{\partial g(f(\boldsymbol{x}))}{\partial x_i} = g'(f(\boldsymbol{x}))\frac{\partial f(\boldsymbol{x})}{\partial x_i}.$$

Thus by the definition of gradients

$$\boldsymbol{\nabla} g(f(\boldsymbol{x})) = g'(f(\boldsymbol{x}))\boldsymbol{\nabla} f(\boldsymbol{x}).$$

Exercise 8 *Compute the gradient of* $f(\boldsymbol{x}) = 3x_1x_2^2 + 4x_1 + 3x_2 - x_3 + 2$

$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{pmatrix} \\ \\ \\ \end{pmatrix}$$

Exercise 9 *For* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ *and* $\boldsymbol{a} = \begin{pmatrix} 1 \\ 5 \\ -6 \end{pmatrix}$ *write down* $\boldsymbol{x}^\mathsf{T}\boldsymbol{a}$ *and compute* $\boldsymbol{\nabla}\boldsymbol{x}^\mathsf{T}\boldsymbol{a}$

$$\boldsymbol{x}^\mathsf{T}\boldsymbol{a} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ -6 \end{pmatrix} = \underline{\qquad\qquad\qquad\qquad\qquad\qquad}$$

$$\boldsymbol{\nabla}(\boldsymbol{x}^\mathsf{T}\boldsymbol{a}) = \begin{pmatrix} \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \end{pmatrix}$$

Exercise 10 *Let* $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *and* $\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$ *write down* $f(\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}\mathbf{M}\boldsymbol{x}$ *and compute* $\boldsymbol{\nabla}f(\boldsymbol{x})$

$$f(\boldsymbol{x}) = \underline{\qquad\qquad\qquad\qquad\qquad\qquad}$$

$$\boldsymbol{\nabla}f(\boldsymbol{x}) = \begin{pmatrix} \underline{\qquad\qquad} \\ \underline{\qquad\qquad} \end{pmatrix}$$

Compute $(\mathbf{M} + \mathbf{M}^\mathsf{T})\boldsymbol{x}$ *and show this is the same as* $\boldsymbol{\nabla}f(\boldsymbol{x})$.

$$\begin{aligned}(\mathbf{M} + \mathbf{M}^\mathsf{T})\boldsymbol{x} &= \left(\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} + \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{pmatrix} \right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} \underline{\qquad} & \underline{\qquad} \\ \underline{\qquad} & \underline{\qquad} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \underline{\qquad\qquad} \\ \underline{\qquad\qquad} \end{pmatrix}\end{aligned}$$

Exercise 11 *Write out* $f(\boldsymbol{x}) = \left(\boldsymbol{a}^\mathsf{T}\boldsymbol{x} - c\right)^2$ *in component form. Compute the partial derivative and show the gradient is what you would compute from using the chain rule.*

$$\begin{aligned} f(\boldsymbol{x}) &= \underline{\qquad\qquad\qquad\qquad\qquad\qquad} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_i} &= \underline{\qquad\qquad\qquad\qquad\qquad\qquad} \\ \boldsymbol{\nabla}f(\boldsymbol{x}) &= \underline{\qquad\qquad\qquad\qquad\qquad\qquad} \end{aligned}$$

3 Taylor Expansion

The Taylor expansion of a function of more than one variable $f(\boldsymbol{x})$ around a point $\boldsymbol{y}$ is given by

$$
\begin{aligned}
f(\boldsymbol{x}) &= f(\boldsymbol{y}) + \sum_i (x_i - y_i)\frac{\partial f(\boldsymbol{y})}{\partial y_i} + \frac{1}{2}\sum_{ij}(x_i - y_i)(x_j - y_j)\frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j} + \cdots \\
&= f(\boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})^{\mathsf{T}} \boldsymbol{\nabla}_{\boldsymbol{y}} f(\boldsymbol{y}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^{\mathsf{T}} \mathbf{H}(\boldsymbol{x} - \boldsymbol{y}) + O(\|\boldsymbol{x} - \boldsymbol{y}\|^3)
\end{aligned}
$$

where $\mathbf{H}$ is the matrix of second derivatives, $[\mathbf{H}]_{ij} = \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j}$ known as the *Hessian* matrix. This is well defined provided $f(\boldsymbol{x})$ is doubly differentiable at $\boldsymbol{y}$.

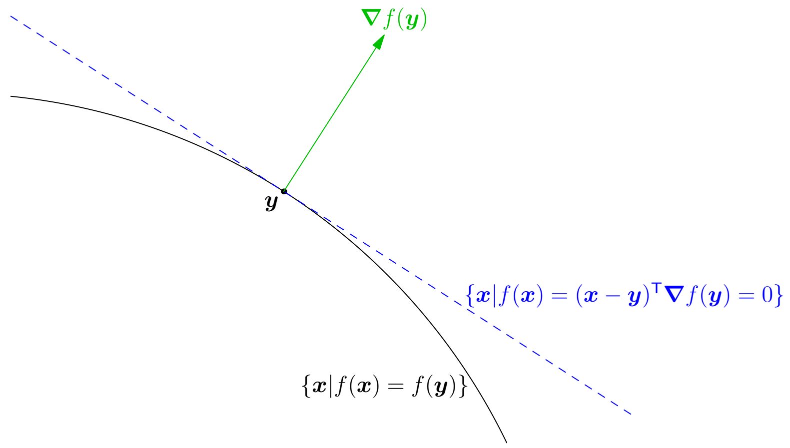
We can write the set of points orthogonal to the gradient at $\boldsymbol{y}$ by

$$\left\{\boldsymbol{x} \middle| (\boldsymbol{x} - \boldsymbol{y})^{\mathsf{T}} \boldsymbol{\nabla}_{\boldsymbol{y}} f(\boldsymbol{y}) = 0\right\}.$$

By the Taylor expansion this set is equal to

$$\left\{\boldsymbol{x} \middle| f(\boldsymbol{x}) = f(\boldsymbol{y}) + O(\|\boldsymbol{x} - \boldsymbol{y}\|^2)\right\}.$$

For small $|\boldsymbol{x} - \boldsymbol{y}|$ these points will be (almost) coincident with the contour surface $\left\{\boldsymbol{x} \middle| f(\boldsymbol{x}) = f(\boldsymbol{y})\right\}$. Thus, the gradient is orthogonal to the contour surface.



Exercise 12 *Show that when $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{\delta}^{(i)}$ with $\delta_j^{(i)} = 0$ for $j \neq i$ and $\delta_i^{(i)} = \delta$ that the Taylor expansion reduces to the one dimensional version of the Taylor expansion.*

$$f(\boldsymbol{y} + \boldsymbol{\delta}^{(i)}) = $$

4 Eigen-Systems

A vector $\boldsymbol{v}$ is an *eigenvector* of matrix $\boldsymbol{M}$ if there exists some scalar (number) λ such that

$$\mathbf{M}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

In this case λ is known as the *eigenvalue*. If $\boldsymbol{v}$ is an eigenvector then so is $a\,\boldsymbol{v}$ for any scalar $a \neq 0$. As a consequence we are always free to normalise an eigenvector so that $\|\boldsymbol{v}\| = 1$.

An $n \times n$ symmetric matrix will have n orthonormal eigenvectors $\boldsymbol{v}^{(i)}$ such that

$$\mathbf{M}\boldsymbol{v}^{(i)} = \lambda^{(i)}\boldsymbol{v}^{(i)}$$

and $\boldsymbol{v}^{(i)}\boldsymbol{v}^{(j)} = 0$ (the eigenvectors are orthogonal) and $\|\boldsymbol{v}^{(i)}\| = 1$ (the eigenvectors of normalised). Note that in general the eigenvectors don't need to be normalised or even always orthogonal, but we can choose a set of them which are.

If we construct a matrix of orthonormal eigenvectors $\mathsf{V} = (\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(n)})$ (i.e. each column of the matrix is an eigenvector so $[\mathsf{V}]_{ij} = [\boldsymbol{v}^{(y)}]_{1j} = v_i^{(j)}$) then $\mathsf{V}^{\mathsf{T}}\mathsf{V} = \mathbf{I}$ the identity element. Such matrices are known as *orthogonal matrices* and describe a rotation.

Exercise 13 *From the definition of matrix multiplication and the fact that the eigenvectors are orthonormal show that* $\mathsf{V}^{\mathsf{T}}\mathsf{V} = \mathbf{I}$.

$$[\mathsf{V}^{\mathsf{T}}\mathsf{V}]_{ij} =$$

Exercise 14 *Show that for an orthogonal matrix* V *and an arbitrary vector* $\boldsymbol{x}$ *that* $\|\mathsf{V}\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$, *i.e. operating on a space by the matrix* V *does not change the length of any vector*

$$\|\mathsf{V}\boldsymbol{x}\|^2 =$$

Exercise 15 *Show that for an orthogonal matrix* V *and two arbitrary vectors* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *that the angle between* $\boldsymbol{x}' = \mathsf{V}\boldsymbol{x}$ *and* $\boldsymbol{y}' = \mathsf{V}\boldsymbol{y}$ *is the same as that between* $\boldsymbol{x}$ *and* $\boldsymbol{y}$. *(Hint* $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y} = \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|\cos(\theta)$.*) Thus show that operating on a space by the matrix* V *does not change the angle between any pair of vectors.*

$$\boldsymbol{x}'^{\mathsf{T}}\boldsymbol{y}' =$$

Exercise 16 *For any vector* $\boldsymbol{x}$ *we can find a vector* $\boldsymbol{y}$ *such that* $\boldsymbol{x} = \mathsf{V}\boldsymbol{y}$ *since multiplying both sides by* V^{T} *we have* $\mathsf{V}^{\mathsf{T}}\boldsymbol{x} = \mathsf{V}^{\mathsf{T}}\mathsf{V}\boldsymbol{y} = \boldsymbol{y}$. *Starting from* $\mathsf{V}^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{y}$ *show that* $\mathsf{V}\mathsf{V}^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{x}$. *Hence argue that* $\mathsf{V}\mathsf{V}^{\mathsf{T}} = \mathbf{I}$.

Exercise 17 *Show if $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$ share a common eigenvalue, λ, then so does $a\boldsymbol{v}^{(i)} + b\boldsymbol{v}^{(j)}$.*

$$\mathbf{M}(a\boldsymbol{v}^{(i)} + b\boldsymbol{v}^{(j)}) = \qquad\qquad$$
$$= \qquad\qquad$$

Matrix Decomposition

Any symmetric matrix, $\mathbf{M}$, can be decomposed as a product of matrices

$$\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}$$

where $\mathbf{V}$ is the orthogonal matrix formed from the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with components $[\boldsymbol{\Lambda}]_{ii} = \lambda^{(i)}$.

To find the eigenvalue decomposition of a matrix in MATLAB (octave) write

```
> [V,Lambda] = eig(M)
```

Exercise 18 *This question is a series of exercises to perform in MATLAB or (octave) to verify properties of matrices.*
Generate a random 5×5 matrix, $\mathbf{X}$

```
> X = rand(5,5)
```

Generate a symmetric matrix $\mathbf{M} = \mathbf{X}^T\boldsymbol{X}$

```
> M = X'*X
```

Note that in MATLAB the apostrophy is a shorthand for taking the transpose of a matrix. Verify that $\mathbf{M}$ is symmetric

```
> M-M'
```

Compute the eigenvalues and eigenvectors of $\mathbf{M}$

```
> [V,Lambda] = eig(M)
```

Check the matrix decomposition $\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$

```
> V*Lambda*V' - M
```

Note that there will be rounding errors of the size of the machine precision (approximately 10^{-15}). Eigenvector, $\boldsymbol{v}^{(i)}$, is the i^{th} column of $\mathbf{V}$ which in MATLAB is written `V(:,i)`*. Verify that $\mathbf{M}\boldsymbol{v}^{(2)} = \lambda^{(2)}\boldsymbol{v}^{(2)}$*

```
> M*V(:,2) - Lambda(2,2)*V(:,2)
```

Verify that $\mathbf{V}$ is an orthogonal matrix

```
> V'*V
> V*V'
```

Exercise 19 *If a symmetric matrix* $\mathbf{M}$ *has a decomposition* $\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ *show that the inverse is given by* $\mathbf{M}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T$ *where* $[\boldsymbol{\Lambda}^{-1}]_{ii} = 1/[[\boldsymbol{\Lambda}]_{ii} = \lambda^{(i)}$

$$\mathbf{M}\mathbf{M}^{-1} =$$
$$\mathbf{M}^{-1}\mathbf{M} =$$

Verify that this in MATLAB

```
> V*inv(Lambda)*V' - inv(M)
```

Exercise 20 *Show that in general for any invertible matrices* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *that* $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$ *by computing*

$$(\mathbf{ABC})^{-1}(\mathbf{ABC}) =$$
$$(\mathbf{ABC})(\mathbf{ABC})^{-1} =$$

Hence compute the inverse of $\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ *directly*

$$\mathbf{M}^{-1} =$$

5 Singular Value Decomposition

Any $n \times m$ matrix, $\mathbf{X}$ can be decomposed into a product of three matrices

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$$

where $\mathbf{U}$ is a $n \times n$ orthogonal matrix, $\mathbf{V}$ is a $m \times m$ orthogonal matrix, and $\mathbf{S}$ is a matrix whose only non-zero elements are the diagonals $[\mathbf{S}]_{ii} = s_i$ which are known as the singular values.

Exercise 21 *Generate a random* 5×4 *matrix* $\boldsymbol{X}$ *and compute its singular value decomposition.*

```
> X = rand(5,4)
> [U,S,V]= svd(X)
```

Verify $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ *and* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthogonal*

```
> U*S*V'-X
> U*U'
> U'*U
> V*V'
> V'*V
```

Exercise 22 *Show that* $\mathbf{X}^T = (\mathbf{USV}^T)^T = \mathbf{VSU}^T$

```
> V*S'*U' - X'
```

Show that $\mathbf{XX}^T = \mathbf{USS}^T\mathbf{U}^T$

```
> U*S*S'*U' - X*X'
```

Compute the eigenvalues and eigenvectors of $\mathbf{XX}^T$ *and compare this with* $\mathbf{SS}^T$ *and* $\mathbf{U}$

```
> [Vxxt, Lxxt] = eig(X*X')
> [U, S*S']
```

Note that these are not necessarily in the same order, i.e. they can be made identical by a permutation of columns. Also the eigenvectors are defined only up to a sign.

6 Trace and Determinant of Matrices

Trace

The trace of a square matrix, $\mathbf{M}$, is written $\mathrm{Tr}\,\mathbf{M}$ and is equal to the sum of the diagonal elements

$$\mathrm{Tr}\,\mathbf{M} = \sum_i [\mathbf{M}]_{ii}$$

Exercise 23 *Using the definition of matrix multiplication and trace expand out the following*

$$\mathrm{Tr}\,\mathbf{AB} = \qquad$$
$$\mathrm{Tr}\,\mathbf{BA} = \qquad$$

Thus show that $\mathrm{Tr}\,\mathbf{AB} = \mathrm{Tr}\,\mathbf{BA}$.

Exercise 24 *Test* $\mathrm{Tr}\,\mathbf{AB} = \mathrm{Tr}\,\mathbf{BA}$ *in MATLAB for a pair of random matrices*

```
> A = rand(4,4)
> B = rand(4,4)
> AB = A*B
> BA = B*A
> trace(AB)
> trace(BA)
```

Note that in MATLAB `diag(M)` *returns a vector containing the diagonal elements of a matrix* $\mathbf{M}$ *so that* `trace(M)` *is the same as* `sum(diag(M))`.

Exercise 25 *Show that for a symmetric matrix* $\mathbf{M}$ *that* $\operatorname{Tr}\mathbf{M} = \sum_i \lambda^{(i)}$. *(Hint use the eigenvalue decomposition* $\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}}$

$$\begin{aligned}\operatorname{Tr}\mathbf{M} &= \operatorname{Tr}(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}})\\ &= \end{aligned}$$

Use MATLAB to verify this for a random symmetric matrix

```
> X = rand(4,4)
> M = X*X'
> lambda = eig(M)
> sum(lambda)
> trace(M)
```

Determinant

The determinant of a square matrix, $\mathbf{M}$, written either $|\mathbf{M}|$ or $\det(\mathbf{M})$ measures the change in volume produced by the mapping $\mathbf{M}$. In addition, the determinant is negative if the volume is reflected. That is, given any closed shape with volume, V, in the original space the points in that volume will be mapped to a new closed shape with volume V' such that $\det(\mathbf{M}) = \pm V'/V$. For a diagonal matrix the determinant is just the product of the diagonal elements. Here it is obvious that the determinant will measure the change in volume.

For general matrices the determinants are

$$\det\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}\end{pmatrix} = M_{11}M_{22} - M_{12}M_{21}$$

$$\begin{aligned}\det\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33}\end{pmatrix} &= M_{11}\det\begin{pmatrix} M_{22} & M_{23} \\ M_{32} & M_{33}\end{pmatrix} - M_{12}\det\begin{pmatrix} M_{21} & M_{23} \\ M_{31} & M_{33}\end{pmatrix} + M_{13}\det\begin{pmatrix} M_{21} & M_{22} \\ M_{31} & M_{22}\end{pmatrix}\\ &= M_{11}M_{22}M_{33} - M_{11}M_{32}M_{23} - M_{12}M_{21}M_{33} + M_{12}M_{31}M_{23}\\ &\quad + M_{13}M_{21}M_{22} - M_{13}M_{31}M_{22}\end{aligned}$$

For an $n \times n$ matrix the determinant is equal to

$$\det(\mathbf{M}) = \sum_{i=1}^{n} (-1)^n [\mathbf{M}]_{1i} \det(\mathbf{M}_{/1i})$$

where $\mathbf{M}_{/1i}$ is an $(n-1)\times(n-1)$ matrix obtained by removing the first row and the i^{th} column of $\mathbf{M}$. By construction $\det M = \det M^{\mathsf{T}}$.

We can interpret the matrix $\mathbf{AB}$ as the application of two mappings: firstly a mapping defined by $\mathbf{B}$ followed by a mapping $\mathbf{A}$. As a consequence a unit volume will first be mapped to a volume of size $|\det(\mathbf{B})|$ and this will be mapped to a volume of size $|\det(\mathbf{A})| \times |\det(\mathbf{B})|$. Similarly the points are reflected by the first mapping if $\det(\mathbf{B}) < 0$ and are reflected by the combined mapping if $\det(\mathbf{A}) \times \det(\mathbf{B}) < 0$. In consequence we have the important property of determinant that

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Exercise 26 *Use MATLAB or octave to verify* $\det(\mathbf{AB}) = \det(A)\det(B)$ *for some particular matrices*

```
> A = rand(4,4)
> B = rand(4,4)
> AB = A*B
> det(AB) - det(A)*det(B)
```

Remember this will only be accurate up to machine precision ($\approx 10^{-15}$)*. Show that* $\det(\mathbf{A}) = \det(\mathbf{A}^T)$

```
> det(A) - det(A')
```

Exercise 27 *Use the property of orthogonal matrices that* $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ *and properties of the determinant to show that* $\det(\mathbf{V}) = \pm 1$

$$\det(\mathbf{V}\mathbf{V}^T) = \quad$$
$$= \quad$$

In fact, these are rotation matrices which don't reflect any coordinates so that $\det V = 1$*.*

Exercise 28 *Using the property that any symmetric matrices,* $\mathbf{M}$ *can be decomposed as* $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}$ *where* $\mathbf{V}$ *is an orthogonal matrix and* $\boldsymbol{\Lambda}$ *is a diagonal matrix of eigenvalues to show that* $\det(\mathbf{M})$ *can be written as a product of its eigenvalues*

Answers

1

$$\mathbf{AB} = \begin{pmatrix} 9 & 2 \\ 16 & -2 \end{pmatrix}$$

2

$$\mathbf{A}^{\mathsf{T}} = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \qquad \mathbf{B}^{\mathsf{T}} = \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix} \qquad \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}} = \begin{pmatrix} 9 & 16 \\ 2 & -2 \end{pmatrix}$$

Note that this is the transpose of the answer to question 1.

3

$$[(\mathbf{AB})^{\mathsf{T}}]_{ij} = [\mathbf{AB}]_{ji} = \sum_k [\mathbf{A}]_{jk}[\mathbf{B}]_{ki}$$
$$[\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}]_{ij} = \sum_k [\mathbf{B}^{\mathsf{T}}]_{ik}[\mathbf{A}^{\mathsf{T}}]_{kj} = \sum_k [\mathbf{B}]_{ki}[\mathbf{A}]_{jk}$$

These are the same.

4

$$\mathbf{AB} = \begin{pmatrix} 5 & 5 \\ 9 & 8 \end{pmatrix} \qquad \mathbf{BA} = \begin{pmatrix} 5 & 9 \\ 5 & 8 \end{pmatrix}$$

Note that even symmetric matrices are do not generally commute. Also the product of two symmetric matrices will not in general be symmetric. Although for symmetric matrices $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top = \mathbf{BA}$. Thus $\mathbf{BA}$ will be the transpose of $\mathbf{AB}$ if $\mathbf{A}$ and $\mathbf{B}$ are symmetric.

5

$$\begin{aligned}(\mathbf{AB})\mathbf{C} &= \begin{pmatrix} 9 & 2 \\ 16 & -2 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 24 & -14 \\ 26 & -36 \end{pmatrix} \\ \mathbf{A}(\mathbf{BC}) &= \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 3 & -8 \\ 7 & -2 \end{pmatrix} = \begin{pmatrix} 24 & -14 \\ 26 & -36 \end{pmatrix}\end{aligned}$$

6

$$\begin{aligned}[(\mathbf{AB})\mathbf{C}]_{ij} &= \sum_k [\mathbf{AB}]_{ik} [\mathbf{C}]_{kj} = \sum_{k,l} [\mathbf{A}]_{il} [\mathbf{B}]_{lk} [\mathbf{C}]_{kj} \\ [\mathbf{A}(\mathbf{BC})]_{ij} &= \sum_k [\mathbf{A}]_{ik} [\mathbf{BC}]_{kj} = \sum_{k,l} [\mathbf{A}]_{ik} [\mathbf{B}]_{kl} [\mathbf{C}]_{lj}\end{aligned}$$

These are the same up to a relabelling of the summation indices.

7

$$\begin{aligned}\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{y} &= \sum_{ij} [\boldsymbol{x}^\top]_{1i} [\mathbf{M}]_{ij} [\boldsymbol{y}]_{j1} = \sum_{ij} [\boldsymbol{x}]_{i1} [\mathbf{M}]_{ij} [\boldsymbol{y}]_{j1} = \sum_{ij} x_i \mathbf{M}_{ij} y_j \\ \boldsymbol{y}^\top \boldsymbol{M}^\top \boldsymbol{x} &= \sum_{ij} [\boldsymbol{y}^\top]_{1i} [\mathbf{M}^\top]_{ij} [\boldsymbol{x}]_{j1} = \sum_{ij} [\boldsymbol{y}]_{i1} [\mathbf{M}]_{ji} [\boldsymbol{x}]_{j1} = \sum_{ij} y_i \mathbf{M}_{ji} x_j\end{aligned}$$

These are the same!

8

$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{pmatrix} 3x_2^2 + 4 \\ 6x_2 x_2 + 3 \\ -1 \end{pmatrix}$$

9

$$\begin{aligned}\boldsymbol{x}^\top \boldsymbol{a} &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ -6 \end{pmatrix} = x_1 + 5x_2 - 6x_3 \\ \boldsymbol{\nabla}(\boldsymbol{x}^\top \boldsymbol{a}) &= \begin{pmatrix} 1 \\ 5 \\ -6 \end{pmatrix}\end{aligned}$$

We note that this is equal to $\boldsymbol{a}$. Since $\boldsymbol{x}^\top \boldsymbol{a} = \boldsymbol{a}^\top \boldsymbol{x}$ we have that $\boldsymbol{\nabla}(\boldsymbol{x}^\top \boldsymbol{a}) = \boldsymbol{\nabla}(\boldsymbol{a}^\top \boldsymbol{x}) = \boldsymbol{a}$.

10

$$
\begin{aligned}
f(\boldsymbol{x}) &= x_1^2 M_{11} + x_1 x_2 (M_{12} + M_{21}) + x_2^2 M_{22} \\
\boldsymbol{\nabla} f(\boldsymbol{x}) &= \begin{pmatrix} 2x_1 M_{11} + x_2(M_{12} + M_{21}) \\ 2x_2 M_{22} + x_1(M_{12} + M_{21}) \end{pmatrix} \\
(\mathbf{M} + \mathbf{M}^\mathsf{T})\boldsymbol{x} &= \left(\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} + \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{pmatrix} \right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \begin{pmatrix} 2M_{11} & M_{12} + M_{21} \\ M_{12} + M_{21} & 2M_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 M_{11} + x_2(M_{12} + M_{21}) \\ 2x_2 M_{22} + x_1(M_{12} + M_{21}) \end{pmatrix}
\end{aligned}
$$

This is the same as $\boldsymbol{\nabla} f(\boldsymbol{x})$.

11

$$
\begin{aligned}
f(\boldsymbol{x}) &= \left(\sum_j a_j x_j - c \right)^2 = \sum_{jk} a_j a_k x_j x_k - 2c \sum_j a_j x_j + c^2 \\
\frac{\partial f(\boldsymbol{x})}{\partial x_i} &= 2a_i \sum_j a_j x_j - 2c\, a_i = 2(\boldsymbol{a}^\mathsf{T} \boldsymbol{x} - c) a_i \\
\boldsymbol{\nabla} f(\boldsymbol{x}) &= 2(\boldsymbol{a}^\mathsf{T} \boldsymbol{x} - c) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = 2(\boldsymbol{a}^\mathsf{T} \boldsymbol{x} - c)\, \boldsymbol{a}
\end{aligned}
$$

This is what you get writing $f(\boldsymbol{x}) = g(\boldsymbol{a}^\mathsf{T}\boldsymbol{x} - c)$ where $g(z) = z^2$ so $g'(z) = 2z$ and $\boldsymbol{\nabla} \boldsymbol{a}^\mathsf{T}\boldsymbol{x} - c = \boldsymbol{a}$.

12

$$
f(\boldsymbol{y} + \boldsymbol{\delta}^{(i)}) = f(\boldsymbol{y}) + \delta \frac{\partial f(\boldsymbol{y})}{\partial y_i} + \frac{\delta^2}{2} \sum_i \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i^2} + \ldots
$$

which follows because $x_j - y_j = 0$ if $j \neq i$ and $x_i - y_i = \delta_i$ otherwise. This is the form of the one dimensional Taylor expansion in the i^{th} dimension.

13

$$
[\mathbf{V}^\mathsf{T}\mathbf{V}]_{ij} = \sum_k [\mathbf{V}^\mathsf{T}]_{ik}\, [\mathbf{V}]_{kj} = \sum_k [\mathbf{V}]_{ki} [\mathbf{V}]_{kj} = \sum_k v_k^{(i)} v_k^{(j)} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}
$$

14

$$
\|\mathbf{V}\boldsymbol{x}\|^2 = (\mathbf{V}\boldsymbol{x})^\mathsf{T}(\mathbf{V}\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}(\mathbf{V}^\mathsf{T}\mathbf{V})\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}\boldsymbol{x} = \|\boldsymbol{x}\|^2
$$

15

$$
\boldsymbol{x}'^\mathsf{T}\boldsymbol{y}' = (\mathbf{V}\boldsymbol{x})^\mathsf{T}(\mathbf{V}\boldsymbol{y}) = \boldsymbol{x}^\mathsf{T}(\mathbf{V}^\mathsf{T}\mathbf{V})\boldsymbol{y} = \boldsymbol{x}^\mathsf{T}\boldsymbol{y} = \|\boldsymbol{x}\|\, \|\boldsymbol{y}\| \cos(\theta)
$$

But $\boldsymbol{x}'^\mathsf{T}\boldsymbol{y}' = \|\boldsymbol{x}'\|\, \|\boldsymbol{y}'\| \cos(\theta')$ and we showed in the last question that $\|\boldsymbol{x}'\| = |\boldsymbol{x}|$ and $\|\boldsymbol{y}'\| = |\boldsymbol{y}|$ so $\cos(\theta') = \cos(\theta)$. Thus showing than angle between the vectors is unchanged. This is because orthogonal matrices just rotates the whole space. They are sometimes known as rotation matrices.

16 Multiplying $\mathbf{V}^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{y}$ by $\mathbf{V}$ we find

$$\mathbf{V}\mathbf{V}^{\mathsf{T}}\boldsymbol{x} = \mathbf{V}\boldsymbol{y} = \boldsymbol{x}$$

by the definition of $\boldsymbol{x}$. But this is true for any vector $\boldsymbol{x}$ thus $\mathbf{V}\mathbf{V}^{\mathsf{T}}$ must be the identity mapping $\mathbf{I}$. Note that this is non-trivial. We found that $\mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$, but it is far from obvious that $\mathbf{V}\mathbf{V}^{\mathsf{T}} = \mathbf{I}$.

17

$$\begin{aligned} \mathbf{M}(a\boldsymbol{v}^{(i)} + b\boldsymbol{v}^{(j)}) &= a\lambda\boldsymbol{v}^{(i)} + b\lambda\boldsymbol{v}^{(j)} \\ &= \lambda(a\boldsymbol{v}^{(i)} + b\boldsymbol{v}^{(j)}) \end{aligned}$$

Note that this means any vector in the subspace spanned by $\boldsymbol{v}^{(i)}$ and $\boldsymbol{v}^{(j)}$ is also an eigenvector (although it won't necessarily be normalised).

19

$$\begin{aligned} \mathbf{M}\mathbf{M}^{-1} &= \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}} = \mathbf{V}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}} = \mathbf{V}\mathbf{V}^{\mathsf{T}} = \mathbf{I} \\ \mathbf{M}^{-1}\mathbf{M} &= \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}} = \mathbf{V}\mathbf{V}^{\mathsf{T}} = \mathbf{I} \end{aligned}$$

20

$$\begin{aligned} (\mathbf{ABC})^{-1}(\mathbf{ABC}) &= \mathbf{ABC}\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{I} \\ (\mathbf{ABC})(\mathbf{ABC})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{ABC} = \mathbf{I} \end{aligned}$$

Thus

$$\mathbf{M}^{-1} = (\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}})^{1} \qquad = \mathbf{V}^{\mathsf{T}-1}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}}.$$

23

$$\begin{aligned} \operatorname{Tr}\mathbf{AB} &= \sum_i [\mathbf{AB}]_{ii} = \sum_{i,k} [\mathbf{A}]_{ik}[\mathbf{B}]_{ki} \\ \operatorname{Tr}\mathbf{BA} &= \sum_i [\mathbf{BA}]_{ii} = \sum_{i,k} [\mathbf{B}]_{ik}\mathbf{A}]_{ki}. \end{aligned}$$

These are the same we a renaming of the indices.

25

$$\begin{aligned} \operatorname{Tr}\mathbf{M} &= \operatorname{Tr}\left((\mathbf{V}\boldsymbol{\Lambda})\mathbf{V}^{\mathsf{T}}\right) = \operatorname{Tr}(\mathbf{V}^{\mathsf{T}}\mathbf{V}\boldsymbol{\Lambda}) \\ &= \operatorname{Tr}\lambda = \sum_i [\Lambda]_{ii} = \sum_i \lambda^{(i)} \end{aligned}$$

27

$$\begin{aligned} \det(\mathbf{V}\mathbf{V}^{\mathsf{T}}) &= \det(\mathbf{V})\det(\mathbf{V}^{\mathsf{T}}) = \det(\mathbf{V})^2 \\ &= \det(\mathbf{I}) = 1 \end{aligned}$$

Thus $\det(\mathbf{V}) = \pm 1$.

28

$$
\begin{aligned}
\det(\mathbf{M}) &= \det(\mathbf{V}\,\boldsymbol{\Lambda}\,\mathbf{V}^{\mathsf{T}}) \\
&= \det(\mathbf{V})\,\det(\boldsymbol{\Lambda})\,\det(\mathbf{V}^{\mathsf{T}}) \\
&= \det(\boldsymbol{\Lambda})
\end{aligned}
$$

since $\det(\mathbf{V}) = \det(\mathbf{V}^{\mathsf{T}}) = 1$. But $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements, $\Lambda_{ii} = \lambda_i$, therefore

$$
\det(\mathbf{M}) = \sum_{i=1}^{n} \lambda_i .
$$