Econometrics Lecture 1: Review of Matrix Algebra

R. G. Pierse

1 Introduction

A *matrix* is a rectangular array of numbers. If the matrix has n rows and m columns it is said to be an $n \times m$ matrix. This is called the *dimension* of the matrix. A matrix with a single column ($n \times 1$) is called a *column vector*. A matrix with a single row ($1 \times m$) is called a *row vector*. A matrix with only one row and one column (a single number) is called a *scalar*.

The standard convention for denoting a matrix is to use a capital letter in bold typeface as in **A**, **B**, **C**. A column vector is denoted with a lowercase letter in bold typeface as in **a**, **b**, **c**. A row vector is denoted with a lowercase letter in bold typeface, followed by a prime, as in $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$. A scalar is generally denoted with a lowercase letter is normal typeface as in a, b, c.

An $n \times m$ matrix **A** can be written out explicitly in terms of its elements as in:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2j} & a_{2m} \\ \vdots & \vdots & \ddots & & \\ a_{i1} & a_{i2} & & a_{ij} & a_{im} \\ & & & & \\ a_{n1} & a_{n2} & & a_{nj} & a_{nm} \end{bmatrix}.$$

Each element has two subscripts: the first is the row index and the second the column index so that a_{ij} refers to the element in the ith row and jth column of **A**.

2 Matrix Operations

The standard operations of addition, subtraction and multiplication can be defined for two matrices as long as the dimensions of the matrices satisfy appropriate conditions to ensure that the operation makes sense. If so, then the two matrices are said to be *conformable* for the operation. If not, then the operation is not defined for these matrices.

2.1 Matrix Addition and Subtraction

$$\mathbf{C} = \mathbf{A} + \mathbf{B}; \quad c_{ij} = a_{ij} + b_{ij}$$

$$\mathbf{C} = \mathbf{A} - \mathbf{B}; \quad c_{ij} = a_{ij} - b_{ij}$$

For $\mathbf{A}$ and $\mathbf{B}$ to be conformable for addition or subtraction, they must be of the same dimension $(n \times m)$. Then the resultant matrix $\mathbf{C}$ will also be $n \times m$ with each element equal to the sum (difference) of the corresponding elements of $\mathbf{A}$ and $\mathbf{B}$. Matrix addition obeys the rules that

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

and

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = \mathbf{A} + \mathbf{B} + \mathbf{C}.$$

2.2 Matrix Multiplication

$$\mathbf{C} = \mathbf{AB}; \quad c_{ij} = \sum_k a_{ik} b_{kj}$$

For $\mathbf{A}$ and $\mathbf{B}$ to be conformable for matrix multiplication, the number of columns of $\mathbf{A}$ must be equal to the number of rows of $\mathbf{B}$. If $\mathbf{A}$ is of dimension $n \times m$ and $\mathbf{B}$ is of dimension $m \times p$, then the resultant matrix $\mathbf{C}$ will be of dimension $n \times p$. The ijth element of $\mathbf{C}$ is the sum of the product of the elements of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$.

Note that, except under very special conditions,

$$\mathbf{AB} \neq \mathbf{BA}$$

and in fact both products will only be defined in the special case that $p = n$. Because of the fact that the order of multiplication matters, it is important to distinguish between *pre-multiplying* and *post-multiplying* a matrix.

Matrix products obey the rules of

$$\mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$$

and distribution

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}.$$

2.3 Matrix Transposition

The *transpose* of an $n \times m$ matrix $\mathbf{A}$, denoted as $\mathbf{A}'$, is the $m \times n$ matrix defined by

$$\mathbf{C} = \mathbf{A}'; \quad c_{ij} = a_{ji}$$

so that the ith row of $\mathbf{C}$ is the ith column of $\mathbf{A}$. The transpose operator obeys the rules that

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$$

and

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

3 Square Matrices

A matrix with the same number of columns as rows is called a *square matrix*. The number of rows (columns) is called the *order* of the matrix. The elements with row index equal to column index as in a_{11}, a_{22}, *etc.* are called the *diagonal elements* and the elements a_{ij}, $i \neq j$ are called the *off-diagonal elements*.

3.1 The trace operator

The *trace* of a square matrix, denoted tr, is the sum of its diagonal elements

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

The trace operator obeys the rules that

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$$

and

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$$

if both $\mathbf{AB}$ and $\mathbf{BA}$ exist.

3.2 Special matrices

3.2.1 Symmetric matrices

A square matrix $\mathbf{A}$ that satisfies the property $\mathbf{A} = \mathbf{A}'$ is said to be *symmetric*. It has the property that $a_{ij} = a_{ji}$ for all values of the indices i and j.

3.2.2 Diagonal matrices

A square matrix with all off-diagonal elements equal to zero is called a *diagonal matrix*. A diagonal matrix is symmetric.

3.2.3 Triangular matrices

A square matrix with all elements below the diagonal equal to zero is called an *upper triangular matrix*. Similarly a matrix with all elements above the diagonal equal to zero is called a *lower triangular matrix*.

3.2.4 The Identity matrix

The square matrix of order n with all diagonal elements equal to one, and all off-diagonal elements equal to zero is called the *identity matrix* of order n and is denoted as $\mathbf{I}_n$. The identity matrix is symmetric aand diagonal. It has the property that, for *any* $n \times m$ matrix $\mathbf{A}$,

$$\mathbf{I}_n\mathbf{A} = \mathbf{A} \quad \text{and} \quad \mathbf{A}\mathbf{I}_m = \mathbf{A}$$

so that any matrix when pre- or post-multiplied by the identity matrix is unchanged. The identity matrix is the equivalent of the number one in standard (scalar) algebra.

4 Matrix Inversion

If $\mathbf{A}$ is a square $n \times n$ matrix, then it may or may not be possible to find a square $n \times n$ matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{I}_n.$$

If $\mathbf{B}$ does exist then it is called the *inverse* of $\mathbf{A}$ and is written $\mathbf{A}^{-1}$. Where the matrix inverse exists, it satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n \quad \text{and} \quad \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

To state the conditions for the existence of the matrix inverse $\mathbf{A}^{-1}$ we need to consider the concept of *linear independence* of a set of vectors and the concept of the *rank* of a matrix.

4.1 Linear independence

Let $\mathbf{a}_1$, $\mathbf{a}_2$, $\cdots$, $\mathbf{a}_m$ be a set of column vectors of dimension $n \times 1$, and let λ_1, λ_2, $\cdots$, λ_m be a set of scalar weights. Then the vector $\mathbf{c}$ defined by

$$\mathbf{c} = \sum_{i=1}^{m} \lambda_i \mathbf{a}_i$$

is called a *linear combination* of the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\cdots$, $\mathbf{a}_m$.

Under what conditions on the weights λ_i will this linear combination be equal to the $n \times 1$ zero column vector $\mathbf{0}_n$? Clearly this will be the case if all the weights are zero, $\lambda_i = 0$, $\forall i$. If this is the *only* condition under which $\mathbf{c} = \mathbf{0}$ then the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\cdots$, $\mathbf{a}_m$ are called *linearly independent*. However, if there are values for λ_i such that $\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}$ where at least one $\lambda_i \neq 0$, then the vectors $\mathbf{a}_i$ are said to be *linearly dependent*.

If a set of vectors are linearly dependent, then it is possible to write one of the vectors as a linear combination of the others. For example if $\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}$ with $\lambda_j \neq 0$ then

$$\mathbf{a}_j = -\frac{1}{\lambda_j} \sum_{\substack{i=1 \\ i \neq j}}^{m} \lambda_i \mathbf{a}_i.$$

Note that if $m > n$, then the set of m column vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\cdots$, $\mathbf{a}_m$ must be linearly dependent. Similarly, if any vector is equal to $\mathbf{0}_n$, then the set of vectors must be linearly dependent.

4.2 The rank of a matrix

The *column rank* of an $n \times m$ matrix $\mathbf{A}$ is defined to be the maximum number of linearly independent columns of $\mathbf{A}$. The *row rank* is defined to be the maximum number of linearly independent rows of $\mathbf{A}$. Since it can be shown that the column rank and row rank of a matrix are always equal, we can simply refer to the rank of $\mathbf{A}$, denoted $\operatorname{rank}(\mathbf{A})$. The following results hold for the rank of a matrix:

$$0 \leq \operatorname{rank}(\mathbf{A}) \leq \min(n, m)$$

and

$$\operatorname{rank}(\mathbf{A}') = \operatorname{rank}(\mathbf{A}).$$

If $\operatorname{rank}(\mathbf{A}) = \min(n, m)$ then the matrix is said to be of *full rank*.

4.3 The matrix inverse

The inverse of a square $n \times n$ matrix $\mathbf{A}$ exists *if and only if*

$$\text{rank}(\mathbf{A}) = n$$

so that $\mathbf{A}$ is of full rank. The matrix inverse has the following properties:

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

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

4.4 Example: solving linear equations

Consider the set of n linear equations in the n variables $x_1, x_2, \cdots, x_n$ defined by

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= c_n \end{aligned}$$

or, in matrix form,

$$\mathbf{Ax} = \mathbf{c}.$$

If the matrix $\mathbf{A}$ is nonsingular, then these equations have a solution which is given by pre-multiplying the set of equations by the matrix inverse $\mathbf{A}^{-1}$ to give

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{c}.$$

If the matrix is singular, then no solution to these equations will exist. In particular, this will be the case if $\mathbf{A}$ is not square, with either too few or too many equations to uniquely determine $\mathbf{x}$. More generally, linear dependence of the equations will mean that no solution exists, corresponding to the singularity of the matrix $\mathbf{A}$.

5 Determinants

The determinant of a square $n \times n$ matrix $\mathbf{A}$ is defined by the expression

$$\det(\mathbf{A}) = |\mathbf{A}| = \sum (\pm)\, a_{1i}a_{2j} \cdots a_{nr}$$

where the summation is taken over all permutations of the second subscripts. Each term has a plus sign for even permutations and a minus sign for odd permutations. For example, for the second order matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

the determinant is given by the expression

$$\det(\mathbf{A}) = |\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}.$$

A singular matrix has determinant equal to zero while a nonsingular matrix has a non-zero determinant.

6 Quadratic Forms

Let $\mathbf{A}$ be an $n \times n$ square, symmetric matrix, and $\mathbf{x}$ be an $n \times 1$ column vector. Then the scalar expression

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

is called a *quadratic form.* If $\mathbf{A}$ is a nonsingular matrix then the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ can only be equal to zero if $\mathbf{x} = \mathbf{0}$.

A *positive definite (pd) matrix* is one for which all quadratic forms are greater than zero for all values of $\mathbf{x} \neq \mathbf{0}$. Formally

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}.$$

A *negative definite (nd) matrix* is one for which all quadratic forms are less than zero for all values of $\mathbf{x} \neq \mathbf{0}$. Formally

$$\mathbf{x}'\mathbf{A}\mathbf{x} < 0, \quad \forall \mathbf{x} \neq \mathbf{0}$$

Similarly, a *positive semi-definite (psd) matrix* is one for which

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0, \quad \forall \mathbf{x}$$

and a *negative semi-definite (nsd) matrix* is one for which

$$\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0, \quad \forall \mathbf{x}$$

7 Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be an $n \times n$ square matrix. Consider the equation system

$$\mathbf{Ax} = \lambda \mathbf{x}$$

where $\mathbf{x}$ is an $n \times 1$ vector with $\mathbf{x} \neq \mathbf{0}$ and λ is a scalar. A value of $\mathbf{x}$ that solves this system of equations is called an *eigenvector* (or *characteristic vector* or *latent vector*) of the matrix $\mathbf{A}$. λ is the corresponding *eigenvalue* (or *characteristic value* or *latent root*). In general there will be n solutions to this system of equations although these need not be distinct. If the matrix is not symmetric, then the eigenvalues λ_i may include complex numbers.

The eigenvalues of a matrix have many useful properties. In particular, the trace of a matrix is the sum of its eigenvalues

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$$

and the determinant of a matrix is the product of its eigenvalues

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i.$$

A positive definite matrix has eigenvalues that are all positive and a negative definite matrix has eigenvalues that are all negative. In addition, if $\mathbf{A}$ is a symmetric matrix, then its rank is equal to the number of its non-zero eigenvalues.

If $\mathbf{A}$ is a symmetric matrix, then the n eigenvectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$ have the property of *orthogonality* that

$$\mathbf{x}_i'\mathbf{x}_j = 0 \quad i \neq j \quad \text{and} \quad \mathbf{x}_i'\mathbf{x}_i = 1 \quad , i, j = 1, \cdots, n.$$

Stacking these eigenvectors into a matrix

$$\mathbf{X} = [\mathbf{x}_1 : \mathbf{x}_2 : \cdots : \mathbf{x}_n]$$

with the property that $\mathbf{X}^{-1} = \mathbf{X}'$, it follows that

$$\mathbf{A} = \mathbf{X \Lambda X}'$$

where $\mathbf{\Lambda}$ is an $n \times n$ diagonal matrix with the eigenvalues $\lambda_1, \cdots, \lambda_n$ along the diagonal. This result is called the *eigenvalue decomposition* of the symmetric matrix $\mathbf{A}$.

8 Cholesky Decomposition

Let $\mathbf{A}$ be an $n \times n$ *symmetric positive-definite* matrix. Then it can be shown that

$$\mathbf{A} = \mathbf{H}\mathbf{H}'$$

where $\mathbf{H}$ is a lower triangular matrix of order $n \times n$. This is known as the *Cholesky decomposition* of the symmetric positive-definite matrix $\mathbf{A}$.

It follows that

$$\mathbf{A}^{-1} = \mathbf{H}^{-1\prime}\mathbf{H}^{-1} \; .$$

9 Idempotent Matrices

A square $n \times n$ matrix $\mathbf{A}$ is idempotent if

$$\mathbf{A}\mathbf{A} = \mathbf{A}.$$

If the matrix is symmetric it also follows that

$$\mathbf{A}'\mathbf{A} = \mathbf{A}.$$

Idempotent matrices are also called *projection matrices*. The eigenvalues of an idempotent matrix are all either zero or one. It follows that most idempotent matrices are singular. The exception is the identity matrix $\mathbf{I}_n$. If $\mathbf{A}$ is idempotent, then so also is $\mathbf{I}_n - \mathbf{A}$.

Idempotent matrices have the property that their rank is equal to their trace, or,

$$\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}).$$

Idempotent matrices are very important in econometrics. Let $\mathbf{X}$ be an $n \times k$ matrix of data of rank k. Then the matrix

$$\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

is a symmetric idempotent matrix since

$$\begin{aligned} \mathbf{M}\mathbf{M} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M}. \end{aligned}$$

The rank of $\mathbf{M}$ can be determined using the results above since

$$\begin{aligned} \text{rank}(\mathbf{M}) &= \text{tr}(\mathbf{M}) = \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= \text{tr}(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = \text{tr}(\mathbf{I}_k) = k. \end{aligned}$$

10 The Kronecker Product

The *Kronecker product* (or *tensor product*) of the $n \times m$ matrix $\mathbf{A}$ and the $p \times q$ matrix $\mathbf{B}$, which is denoted $\mathbf{A} \otimes \mathbf{B}$, is defined by the $np \times mq$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & & a_{nm}\mathbf{B} \end{bmatrix}.$$

The Kronecker product has the following properties:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$$

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$$

and

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}.$$

11 Vectorisation

Let $\mathbf{A}$ be an $n \times m$ matrix with columns

$$\mathbf{A} = [\mathbf{a}_1 : \mathbf{a}_2 : \cdots : \mathbf{a}_m].$$

Then the column vectorisation of $\mathbf{A}$, denoted by $\text{vec}(\mathbf{A})$, is defined by the $nm \times 1$ vector

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

constructed by stacking the columns of $\mathbf{A}$.

The vec operator has the property that

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\text{vec}\,(\mathbf{B}).$$

12 Matrix Derivatives

The rules of differential calculus carry over to matrices in a straightforward way. The only issue is that of adopting a convention for ordering the derivatives.

12.1 Derivatives of a scalar wrt a matrix

The derivatives of a scalar function f with respect to a matrix argument $\mathbf{X}$ of dimension $n \times m$ is defined by the $n \times m$ dimensional matrix

$$\frac{\partial f}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1m}} \\ \vdots & \ddots & \\ \frac{\partial f}{\partial x_{n1}} & & \frac{\partial f}{\partial x_{nm}} \end{bmatrix}.$$

12.2 Derivatives of a vector wrt a vector

The derivatives of an $n \times 1$ vector $\mathbf{y}$ with respect to an $m \times 1$ vector $\mathbf{x}$ is defined by the $n \times m$ dimensional matrix

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & \ddots & \\ \frac{\partial y_n}{\partial x_1} & & \frac{\partial y_n}{\partial x_m} \end{bmatrix}.$$

12.3 Derivatives of a matrix wrt a matrix

There is no obvious way to order the derivatives of one matrix with respect to another matrix. In this case the most sensible procedure is to vectorise both matrices and look at the matrix of derivatives

$$\frac{\partial \operatorname{vec}(\mathbf{Y})}{\partial \operatorname{vec}(\mathbf{X})'}.$$

If $\mathbf{Y}$ is of order $p \times q$ and $\mathbf{X}$ is of order $n \times m$, then this matrix of derivatives is of order $pq \times nm$.

12.4 Some useful results

Two general rules allow the calculation of derivatives of complicated functions. These are followed by some useful derivatives of commonly used matrix functions.

12.4.1 Function of a function rule

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \frac{\partial \mathbf{y}}{\partial \mathbf{z}'}\frac{\partial \mathbf{z}}{\partial \mathbf{x}'}$$

12.4.2 Product rule

$$\frac{\partial \operatorname{vec}(\mathbf{AB})}{\partial \mathbf{x}'} = (\mathbf{B}' \otimes \mathbf{I})\,\frac{\partial \operatorname{vec}(\mathbf{A})}{\partial \mathbf{x}'} + (\mathbf{I} \otimes \mathbf{A})\,\frac{\partial \operatorname{vec}(\mathbf{B})}{\partial \mathbf{x}'}$$

12.4.3 Derivative of an inner product

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

12.4.4 Derivative of a quadratic form

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

12.4.5 Derivative of the trace of a matrix

$$\frac{\partial \operatorname{tr}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{I}$$

12.4.6 Derivative of the determinant of a matrix

$$\frac{\partial \det(\mathbf{A})}{\partial \mathbf{A}} = \det(\mathbf{A})(\mathbf{A}')^{-1}$$

12.4.7 Derivative of the log determinant of a matrix

$$\frac{\partial \ln \det(\mathbf{A})}{\partial \mathbf{A}} = (\mathbf{A}')^{-1}$$

12.4.8 Derivative of a matrix inverse

$$\frac{\partial \operatorname{vec}(\mathbf{A}^{-1})}{\partial \operatorname{vec}(\mathbf{A})'} = -\left((\mathbf{A}')^{-1} \otimes \mathbf{A}^{-1}\right)$$