Westminster Business School
School of Finance and Accounting
7FNCE044 Predictive Analysis for Decision Making
Semester 1
Dr. Issam Malki

MATRIX ALGEBRA REVIEW[1]

Vectors and matrices

An $m \times n$ matrix is a rectangular array of numbers (or scalars) with m rows and n columns. Thus, a matrix $\mathbf{A}$ contains the elements a_{ij}, $i = 1,...,m$, $j = 1,...,n$, as follows:

$$\mathbf{A} = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

A vector is a matrix with only one row or one column. Throughout these notes, all vectors will be assumed to be column vectors by default (these can easily be converted to row vectors using the transpose operator, introduced below).

By convention, matrices are denoted by bold uppercase letters and vectors are denoted by bold lowercase letters. Thus, the matrix $\mathbf{A}$ is made up of n column vectors, $\mathbf{a}_1,...,\mathbf{a}_n$, and can be written:

$$\mathbf{A} = \left[\mathbf{a}_1,...,\mathbf{a}_n\right].$$

We could also interpret $\mathbf{A}$ as the ordered set of m row vectors, stacked one upon another.

Matrix operations

Two matrices are said to be conformable if they have the appropriate dimensions for some operation. If two matrices are not conformable, it is not possible to complete the operation.

In the case of matrix addition/subtraction, two matrices are conformable if they have the same number of rows and columns. If $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices, $\mathbf{A} + \mathbf{B}$ is an $m \times n$ matrix containing the elements $a_{ij} + b_{ij}$.

[1] **Last revised:** 23/01/2021

Any matrix can be multiplied by a scalar, γ, in which case each element of the matrix is multiplied by the scalar, so that $\gamma\mathbf{A}$ contains the elements γa_{ij} .

In the case of matrix multiplication, conformability requires that the number of columns of the first matrix equals the number of rows of the second matrix, *i.e*. if **A** is an $m\times n$ matrix, then **B** must be an $n\times r$ matrix in order for the matrix product **AB** to exist. In this case, **AB** contains the elements $\sum_{k=1}^{n} a_{ik}b_{kj}$, $i=1,\ldots,m$, $j=1,\ldots,r$. We say that **A** is "post-multiplied" by **B** or that **B** is "pre-multiplied" by **A**.

Matrix addition and matrix multiplication satisfy the following properties:

1. $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$;

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

3. $(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})=\mathbf{ABC}$;

4. $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$ and $(\mathbf{B}+\mathbf{C})\mathbf{A}=\mathbf{BA}+\mathbf{CA}$.

Note, however, the following facts (often the source of silly mistakes):

1. **AB** is not necessarily equal to **BA** (pre-multiplication is not the same as post-multiplication).

2. $\mathbf{AB}=\mathbf{0}$ is possible even if $\mathbf{A}\neq\mathbf{0}$ and $\mathbf{B}\neq\mathbf{0}$.

3. $\mathbf{CD}=\mathbf{CE}$ is possible even if $\mathbf{C}\neq\mathbf{0}$ and $\mathbf{D}\neq\mathbf{E}$.

Transpose

The transpose of an $m\times n$ matrix **A** is the $n\times m$ matrix **B** such that:

$b_{ij}=a_{ji}$, for $i=1,\ldots,n$, $j=1,\ldots,m$,

The transpose satisfies the following properties:

1. $(\mathbf{A}')'=\mathbf{A}$;

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

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

Some special matrices

1. Square matrix: an $m \times n$ matrix where $m=n$.

2. Symmetric matrix: a square matrix **A** where $\mathbf{A} = \mathbf{A}'$.

3. Diagonal matrix: a symmetric matrix **A** where all off-diagonal terms are zero, *i.e.*:

$a_{ij} = 0$, for $i \neq j$.

4. Identity matrix (denoted **I**): a diagonal matrix where all elements on the principal diagonal (*i.e.* the diagonal running from top-left to bottom-right) are equal to one, *i.e.*:

$a_{ii} = 1$, for $i = 1,...,n$;

$a_{ij} = 0$, for $i \neq j$.

Therefore:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Sometimes it is convenient to denote identity matrices $\mathbf{I}_n$, where n denotes the number of rows (or columns).

An $n \times n$ identity matrix satisfies the following properties for any $n \times n$ matrix A:

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

5. Idempotent matrix: a square matrix **A** where $\mathbf{AA} = \mathbf{A}$.

Rank

A set of m-vectors, $\mathbf{x}_1,...,\mathbf{x}_n$, is linearly dependent if there exist numbers $\gamma_1,...,\gamma_n$, not all zero, such that:

$$\sum_{i=1}^{n} \gamma_i \mathbf{x}_i = 0.$$

If the vectors are not linearly dependent, they are linearly independent.

The rank of a set of vectors is the maximum number of linearly independent vectors that can be chosen from the set of vectors.

Any set of m+1 m-vectors is linearly dependent. This means that the rank of a set of m-vectors is at most m.

The above means that if one vector in a set can be expressed as a linear combination of the others, the set of vectors will be linearly dependent. In this case, the rank of the set of vectors will be less than the number of vectors.

It can be shown that for any matrix the rank of the set of row vectors that comprise the matrix is equal to the rank of the set of column vectors.

If **A** is an $m \times m$ matrix, then **A** is called non-singular if the rank of **A** is m (called "full rank") and singular if the rank of **A** is less than m.

Inverse

An $m \times m$ matrix **A** is called invertible if there exists an $m \times m$ matrix **B** such that:

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

In this case, **B** is called the inverse of **A**, denoted $\mathbf{A}^{-1}$. Inverse matrices satisfy the following properties (here we assume that **A** and **B** are both invertible matrices):

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

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

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

It can be shown (using the Basis Theorem) that **A** is invertible if and only if **A** is non-singular. Hence, it is equivalent to say that a square matrix has full rank, is non-singular or is invertible.

Calculating the inverse of a matrix is difficult by hand, except in the case of a 2×2 matrix, when there is a useful rule:

If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\mathbf{A}^{-1} = \dfrac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Trace

The trace, denoted tr(), is the sum of the elements on the principal diagonal of a square $n \times n$ matrix:

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

The trace satisfies the following properties, where **A**, **B** and **C** are $n \times n$ matrices and γ is a scalar:

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

2. $\mathrm{tr}(\gamma\mathbf{A}) = \gamma\,\mathrm{tr}(\mathbf{A})$;

3. $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA}) = \mathrm{tr}(\mathbf{CAB})$.

Matrix differentiation

When we differentiate a scalar with respect to another scalar the result is a scalar. When matrices or vectors are involved, things are more complicated. We can differentiate an $m \times n$ matrix with respect to a scalar, in which case the result is an $m \times n$ matrix. We can differentiate a scalar with respect to an m column vector, in which case the result is an m column vector. Finally, we can differentiate an n row vector with respect to an m column vector, in which case the result is an $m \times n$ matrix.

The following results are useful:

1. $\dfrac{\partial \mathbf{a}'\mathbf{b}}{\partial \mathbf{b}} = \dfrac{\partial \mathbf{b}'\mathbf{a}}{\partial \mathbf{b}} = \mathbf{a}$;

2. $\dfrac{\partial \mathbf{Ab}}{\partial \mathbf{b}'} = \mathbf{A}$ and $\dfrac{\partial \mathbf{b}'\mathbf{A}}{\partial \mathbf{b}} = \mathbf{A}$;

3. $\dfrac{\partial \mathbf{b}'\mathbf{Ab}}{\partial \mathbf{b}} = (\mathbf{A}+\mathbf{A}')\mathbf{b}$.

A special case of property 3 is when **A** is symmetric (so that $\mathbf{A} = \mathbf{A}'$), in which case:

$$\frac{\partial \mathbf{b}'\mathbf{Ab}}{\partial \mathbf{b}} = 2\mathbf{Ab}.$$

Ordinary least squares with matrices

Our primary goal in econometrics is often to analyse multivariate relationships, that is, we want to explain one variable, y, (the "dependent variable" or "regressand") in terms of a set of k other variables, $x_2,\ldots,x_k$, (the "independent variables" or "regressors"). Usually, we do this by specifying a linear relationship between the variables (known as a "model"):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + u_i,\; i = 1,\ldots,n.$$

Here, the i subscripts denote the observation. u is the error term (or disturbance) and is a random variable. It allows for the fact that, as well as being influenced by the values of the *x* variables, *y* contains a random component. The model contains *k*+1 parameters: the *k* slope parameters, $\beta_1,...,\beta_k$, and the variance of *u*, σ^2.

As econometricians, we typically have a sample of values of *y* and $x_2,...,x_k$, but not $\beta_1,...,\beta_k$ or *u*. Our aim is to find the best possible estimates of $\beta_1,...,\beta_k$ and σ^2. This task is made immensely easier if we "stack" the *n* equations and create a set of vectors:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + ... + \beta_k x_{k1} + u_1 \\ \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + ... + \beta_k x_{k2} + u_2 \\ \vdots \\ \beta_1 + \beta_2 x_{2n} + \beta_3 x_{3n} + ... + \beta_k x_{kn} + u_n \end{bmatrix}$$

$$= \beta_1 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_2 \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} + \beta_3 \begin{bmatrix} x_{31} \\ x_{32} \\ \vdots \\ x_{3n} \end{bmatrix} + ... + \beta_k \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kn} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$\mathbf{y} = \beta_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3 + ... + \beta_k \mathbf{x}_k + \mathbf{u}.$$

In the above, we have expressed the **y** vector as a linear combination of the **x** vectors and the error vector. Note that one of the independent vectors (implicitly, $\mathbf{x}_1$) is a column of ones.

We can now collect all the **x** vectors into an $n \times k$ matrix **X** and all the β parameters into a *k*-vector **β**, defined as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & & x_{k2} \\ \vdots & & \ddots & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix};$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

In this case, the multivariate linear model can be written as simply:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

We will return to the issue of estimating σ^2 in the next lecture, but for now consider the simpler problem of finding a good estimate of $\boldsymbol{\beta}$, which we will denote $\hat{\boldsymbol{\beta}}$.

For any choice of $\hat{\boldsymbol{\beta}}$, a vector of residuals, $\hat{\mathbf{u}}$, can be defined as follows:

$$\hat{\mathbf{u}} \equiv \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Each element in $\hat{\mathbf{u}}$ is the residual for a different observation (*i.e.* row). The least squares method involves choosing a set of parameter estimates, $\hat{\boldsymbol{\beta}}$, in order to minimise the sum of squared residuals, *i.e.* $\sum_{i=1}^{n} \hat{u}_i^2$, where $\hat{u}_i$ is the individual residual for observation *i*, defined (analogously to $\hat{\mathbf{u}}$) as

$$\hat{u}_i \equiv y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \ldots - \hat{\beta}_k x_{ki}$$

But now note the following:

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_n \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \ldots + \hat{u}_n^2 = \sum_{i=1}^{n} \hat{u}_i^2 .$$

So, minimising the sum of squared residuals is equivalent to minimising the scalar $\hat{\mathbf{u}}'\hat{\mathbf{u}}$. Using the definition of $\hat{\mathbf{u}}$ and the properties of transpose and matrix multiplication, we know that:

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}. \qquad \text{(why?)}$$

We know that to find the minimum of a differentiable function, a necessary condition is that its first derivative is equal to zero. Using the rules of matrix differentiation, this means:

$$\frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\boldsymbol{\beta}}} = \frac{\partial}{\partial \hat{\boldsymbol{\beta}}}\mathbf{y}'\mathbf{y} - \frac{\partial}{\partial \hat{\boldsymbol{\beta}}}2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \frac{\partial}{\partial \hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \qquad \text{(why?)}$$

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

This equation is known as the normal equation. As long as $\mathbf{X}'\mathbf{X}$ is invertible, we can pre-multiply both sides of the normal equation by $(\mathbf{X}'\mathbf{X})^{-1}$ as follows:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\,.$$

This equation is the basic equation defining the ordinary least squares (OLS) estimate of $\boldsymbol{\beta}$.

To finish, recall that in usual (scalar) calculus, to confirm that we have a minimum not a maximum, we should calculate the second derivative and confirm that it is positive. The matrix equivalent of the second derivative is the following:

$$\frac{\partial^2 \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'} = \frac{\partial}{\partial\hat{\boldsymbol{\beta}}'}(-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= 2\mathbf{X}'\mathbf{X}.$$

The matrix equivalent of checking that this is positive is to check that it is a positive definite matrix. It turns out that this is true whenever $\mathbf{X}'\mathbf{X}$ is invertible, which we required to calculate $\hat{\boldsymbol{\beta}}$ in the first place. Hence, as long as we can calculate the least squares coefficients, we know we are minimising the sum of squared residuals, not maximising them.

The big question is when $\mathbf{X}'\mathbf{X}$ is invertible. We will discuss this in the next lecture.

References

J&D: Appendix A.1-A.2.7; Section 3.1.1.

Greene: Appendix A.1-A.4; Section 3.2.

Appendix: Details of matrix differentiation

Details of the matrix differentiation rules are given below. In each case, let: $\mathbf{a}=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, $\mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ and $\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix}$.

1. $\mathbf{a}'\mathbf{b}=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$= a_1b_1 + a_2b_2 + ... + a_nb_n$$

$$= \mathbf{b}'\mathbf{a}$$

$$\frac{\partial \mathbf{a}'\mathbf{b}}{\partial \mathbf{b}} = \frac{\partial \mathbf{b}'\mathbf{a}}{\partial \mathbf{b}} = \begin{bmatrix} \dfrac{\partial \mathbf{a}'\mathbf{b}}{\partial b_1} \\ \dfrac{\partial \mathbf{a}'\mathbf{b}}{\partial b_2} \\ \vdots \\ \dfrac{\partial \mathbf{a}'\mathbf{b}}{\partial b_n} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\partial (a_1b_1 + a_2b_2 + ... + a_nb_n)}{\partial b_1} \\ \dfrac{\partial (a_1b_1 + a_2b_2 + ... + a_nb_n)}{\partial b_2} \\ \vdots \\ \dfrac{\partial (a_1b_1 + a_2b_2 + ... + a_nb_n)}{\partial b_n} \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}\,.$$

2. $\mathbf{Ab} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$= \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + ... + a_{1n}b_n \\ a_{21}b_1 + a_{22}b_2 + ... + a_{2n}b_n \\ \vdots \\ a_{n1}b_1 + a_{n2}b_2 + ... + a_{nn}b_n \end{bmatrix}$$

$$\frac{\partial \mathbf{Ab}}{\partial \mathbf{b}'} = \begin{bmatrix} \frac{\partial \mathbf{Ab}}{\partial b_1} & \frac{\partial \mathbf{Ab}}{\partial b_2} & \cdots & \frac{\partial \mathbf{Ab}}{\partial b_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial(a_{11}b_1 + a_{12}b_2 + ... + a_{1n}b_n)}{\partial b_1} & \frac{\partial(a_{11}b_1 + a_{12}b_2 + ... + a_{1n}b_n)}{\partial b_2} & \cdots & \frac{\partial(a_{11}b_1 + a_{12}b_2 + ... + a_{1n}b_n)}{\partial b_n} \\ \frac{\partial(a_{21}b_1 + a_{22}b_2 + ... + a_{2n}b_n)}{\partial b_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial(a_{n1}b_1 + a_{n2}b_2 + ... + a_{nn}b_n)}{\partial b_1} & \cdots & \cdots & \frac{\partial(a_{n1}b_1 + a_{n2}b_2 + ... + a_{nn}b_n)}{\partial b_n} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix} = \mathbf{A}\,.$$

$\frac{\partial \mathbf{b}'\mathbf{A}}{\partial \mathbf{b}} = \mathbf{A}$ is similar, except here we are taking the derivative of a row vector with respect to a column vector, thereby creating a matrix.

3. $\mathbf{b}'\mathbf{Ab} = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$= a_{11}b_1^2 + a_{21}b_1b_2 + ... + a_{n1}b_1b_n + a_{12}b_1b_2 + a_{22}b_2^2 + ... + a_{n2}b_n + ...$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_ib_j$$

$$\frac{\partial \mathbf{b}'\mathbf{A}\mathbf{b}}{\partial \mathbf{b}} = \begin{bmatrix} \dfrac{\partial \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_i b_j}{\partial b_1} \\ \dfrac{\partial \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_i b_j}{\partial b_2} \\ \vdots \\ \dfrac{\partial \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_i b_j}{\partial b_n} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{i1}b_i + \sum_{j=1}^{n} a_{1j}b_j \\ \sum_{i=1}^{n} a_{i2}b_i + \sum_{j=1}^{n} a_{2j}b_j \\ \vdots \\ \sum_{i=1}^{n} a_{in}b_i + \sum_{j=1}^{n} a_{nj}b_j \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{i1}b_i + \sum_{j=1}^{n} a_{1j}b_j \\ \sum_{i=1}^{n} a_{i2}b_i + \sum_{j=1}^{n} a_{2j}b_j \\ \vdots \\ \sum_{i=1}^{n} a_{in}b_i + \sum_{j=1}^{n} a_{nj}b_j \end{bmatrix} = \mathbf{A}'\mathbf{b} + \mathbf{A}\mathbf{b} = (\mathbf{A} + \mathbf{A}')\mathbf{b}\,.$$