Lecture Notes - Econometrics: Some Facts about Matrices

Paul Söderlind[1]

8 July 2013

[1] University of St. Gallen and CEPR. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: EcmXMat.TeX.

Contents

22 Some Facts about Matrices **2**

22.1 Rank 2

22.2 Vector Norms 2

22.3 Systems of Linear Equations and Matrix Inverses 3

22.4 Complex matrices 5

22.5 Eigenvalues and Eigenvectors 6

22.6 Special Forms of Matrices 7

22.6.1 Triangular Matrices 7

22.6.2 Orthogonal Vector and Matrices 7

22.6.3 Positive Definite Matrices 8

22.6.4 Symmetric Matrices 8

22.6.5 Idempotent Matrices 9

22.7 Matrix Decompositions 9

22.8 Matrix Calculus 14

22.9 Miscellaneous 17

Chapter 22

Some Facts about Matrices

Some references: Greene (2000), Golub and van Loan (1989), Björk (1996), Anton (1987), Greenberg (1988).

22.1 Rank

Fact 22.1 *(Submatrix) Any matrix obtained from the $m \times n$ matrix A by deleting at most $m - 1$ rows and at most $n - 1$ columns is a submatrix of A.*

Fact 22.2 *(Rank) The rank of the $m \times n$ matrix A is ρ if the largest submatrix with non-zero determinant is $\rho \times \rho$. The number of linearly independent row vectors (and column vectors) of A is then ρ.*

22.2 Vector Norms

Fact 22.3 *(Vector p-norm) Let x be an $n \times 1$ matrix. The p-norm is defined as/*

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

The Euclidian norm corresponds to $p = 2$

$$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \sqrt{x'x}.$$

22.3 Systems of Linear Equations and Matrix Inverses

Fact 22.4 *(Linear systems of equations) Consider the linear system $Ax = c$ where A is $m \times n$, x is $n \times 1$, and c is $m \times 1$. A solution is a vector x such that $Ax = c$. It has a unique solution if and only if* $\text{rank}(A) = \text{rank}(\begin{bmatrix} A & c \end{bmatrix}) = n$*; an infinite number of solutions if and only if* $\text{rank}(A) = \text{rank}(\begin{bmatrix} A & c \end{bmatrix}) < n$*; and no solution if and only if* $\text{rank}(A) \neq \text{rank}(\begin{bmatrix} A & c \end{bmatrix})$.

Example 22.5 *(Linear systems of equations, unique solution when $m = n$) Let x be 2×1, and consider the linear system*

$$Ax = c \text{ with } A = \begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix} \text{ and } c = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Here $\text{rank}(A) = 2$ *and* $\text{rank}(\begin{bmatrix} A & c \end{bmatrix}) = 2$*. The unique solution is* $x = \begin{bmatrix} 3 & 0 \end{bmatrix}'$.

Example 22.6 *(Linear systems of equations, no solution when $m > n$) Let x be a scalar, and consider the linear system*

$$Ax = c \text{ with } A = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } c = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

Here $\text{rank}(A) = 1$ *and* $\text{rank}(\begin{bmatrix} A & c \end{bmatrix}) = 2$*. There is then no solution.*

Example 22.7 *(Inverse of 2×2 matrices). For a 2×2 matrix we have*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

In particular, for a triangular matrix we have

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} 1/a & 0 \\ -c/(ad) & 1/d \end{bmatrix}.$$

Fact 22.8 *(Least squares) Suppose that no solution exists to $Ax = c$. The best approximate solution, in the sense of minimizing (the square root of) the sum of squared errors, $\left[(c - A\hat{x})' (c - A\hat{x})\right]^{1/2} = \|c - A\hat{x}\|_2$, is $\hat{x} = (A'A)^{-1} A'c$, provided the inverse exist.*

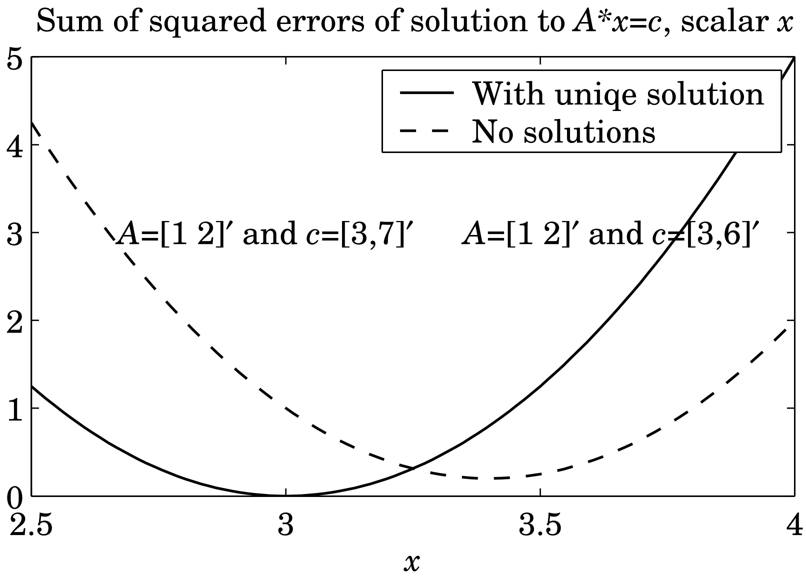

Figure 22.1: Value of qudratic loss function.

This is obviously the least squares solution. In the example with $c = [\ 3 \quad 7\ ]'$*, it is*

$$\hat{x} = \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}' \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix}' \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$
$$= \frac{17}{5} \text{ or } 3.4.$$

This is illustrated in Figure 22.1. *(Translation to OLS notation:* c *is the vector of dependent variables for* m *observations,* A *is the matrix with explanatory variables with the* t^{th} *observation in row* t*, and* x *is the vector of parameters to estimate).*

Fact 22.9 *(Pseudo inverse or generalized inverse) Suppose that no solution exists to* $Ax = c$*, and that* $A'A$ *is not invertible. There are then several approximations,* $\hat{x}$*, which all minimize* $\|c - A\hat{x}\|_2$*. The one with the smallest* $\|\hat{x}\|_2$ *is given by* $\hat{x} = A^{+}c$*, where* A^{+} *is the Moore-Penrose pseudo (generalized) inverse of* A*. See Fact 22.56.*

Example 22.10 *(Linear systems of equations, unique solution when* $m > n$*) Change* c *in Example 22.6 to* $c = [\ 3 \quad 6\ ]'$*. Then* $\operatorname{rank}(A) = 1$ *and* $\operatorname{rank}([\ A \quad c\ ]) = 1$*, and the unique solution is* $x = 3$*.*

Example 22.11 *(Linear systems of equations, infinite number of solutions, $m < n$) Let x be 2×1, and consider the linear system*

$$Ax = c \text{ with } A = \begin{bmatrix} 1 & 2 \end{bmatrix} \text{ and } c = 5.$$

Here $\operatorname{rank}(A) = 1$ *and* $\operatorname{rank}(\begin{bmatrix} A & c \end{bmatrix}) = 1$. *Any value of x_1 on the line $5 - 2x_2$ is a solution.*

Example 22.12 *(Pseudo inverses again) In the previous example, there is an infinite number of solutions along the line $x_1 = 5 - 2x_2$. Which one has the smallest norm $\|\hat{x}\|_2 = [(5-2x_2)^2 + x_2^2]^{1/2}$? The first order condition gives $x_2 = 2$, and therefore $x_1 = 1$. This is the same value as given by $\hat{x} = A^+c$, since $A^+ = [0.2, 0.4]$ in this case.*

Fact 22.13 *(Rank and computers) Numerical calculations of the determinant are poor indicators of whether a matrix is singular or not. For instance,* $\det(0.1 \times I_{20}) = 10^{-20}$. *Use the condition number instead (see Fact 22.53).*

Fact 22.14 *(Some properties of inverses) If A, B, and C are invertible, then $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$; $(A^{-1})' = (A')^{-1}$; if A is symmetric, then A^{-1} is symmetric; $(A^n)^{-1} = \left(A^{-1}\right)^n$.*

Fact 22.15 *(Changing sign of column and inverting) Suppose the square matrix A_2 is the same as A_1 except that the i^{th} and j^{th} columns have the reverse signs. Then A_2^{-1} is the same as A_1^{-1} except that the i^{th} and j^{th} rows have the reverse sign.*

22.4 Complex matrices

Fact 22.16 *(Modulus of complex number) If $\lambda = a + bi$, where $i = \sqrt{-1}$, then $|\lambda| = |a + bi| = \sqrt{a^2 + b^2}$.*

Fact 22.17 *(Complex matrices) Let A^H denote the transpose of the complex conjugate of A, so that if*

$$A = \begin{bmatrix} 1 & 2+3i \end{bmatrix} \text{ then } A^H = \begin{bmatrix} 1 \\ 2-3i \end{bmatrix}.$$

A square matrix A is unitary *(similar to orthogonal) if $A^H = A^{-1}$, for instance,*

$$A = \begin{bmatrix} \frac{1+i}{2} & \frac{1+i}{2} \\ \frac{1-i}{2} & \frac{-1+i}{2} \end{bmatrix} \text{ gives } A^H = A^{-1} = \begin{bmatrix} \frac{1-i}{2} & \frac{1+i}{2} \\ \frac{1-i}{2} & \frac{-1-i}{2} \end{bmatrix}.$$

and it Hermitian *(similar to symmetric) if $A = A^H$, for instance*

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1+i}{2} \\ \frac{1-i}{2} & \frac{-1}{2} \end{bmatrix}.$$

A Hermitian matrix has real elements along the principal diagonal and A_{ji} is the complex conjugate of A_{ij}. Moreover, the quadratic form $x^H Ax$ is always a real number.

22.5 Eigenvalues and Eigenvectors

Fact 22.18 *(Homogeneous linear system). Consider the linear system in Fact 22.4 with $c = \mathbf{0}$: $A_{m\times n}x_{n\times 1} = \mathbf{0}_{m\times 1}$. Then* $\text{rank}(A) = \text{rank}([\ A \quad c\ ])$, *so it has a unique solution if and only if* $\text{rank}(A) = n$; *and an infinite number of solutions if and only if* $\text{rank}(A) < n$. *Note that $x = \mathbf{0}$ is always a solution, and it is the unique solution if* $\text{rank}(A) = n$. *We can thus only get a nontrivial solution (not all elements are zero), only if* $\text{rank}\,(A) < n$.

Fact 22.19 *(Eigenvalues) The n eigenvalues, λ_i, $i = 1, \ldots, n$, and associated eigenvectors, z_i, of the $n \times n$ matrix A satisfy*

$$(A - \lambda_i I)\, z_i = \mathbf{0}_{n\times 1}.$$

We require the eigenvectors to be non-trivial (not all elements are zero). From Fact 22.18, an eigenvalue must therefore satisfy

$$\det(A - \lambda_i I) = 0.$$

Fact 22.20 *(Right and left eigenvectors) A "right eigenvector" z (the most common) satisfies $Az = \lambda z$, and a "left eigenvector" v (seldom used) satisfies $v'A = \lambda v'$, that is, $A'v = \lambda v$.*

Fact 22.21 *(Rank and eigenvalues) For any $m \times n$ matrix A,* $\text{rank}\,(A) = \text{rank}\,(A') = \text{rank}\,(A'A) = \text{rank}\,(AA')$ *and equals the number of non-zero eigenvalues of $A'A$ or AA'.*

Example 22.22 *Let x be an $n \times 1$ vector, so* $\text{rank}\,(x) = 1$. *We then have that the outer product, xx' also has rank 1.*

Fact 22.23 *(Determinant and eigenvalues) For any $n \times n$ matrix A,* $\det(A) = \Pi_{i=1}^{n} \lambda_i$.

22.6 Special Forms of Matrices

22.6.1 Triangular Matrices

Fact 22.24 *(Triangular matrix) A lower (upper) triangular matrix has zero elements above (below) the main diagonal.*

Fact 22.25 *(Eigenvalues of triangular matrix) For a triangular matrix A, the eigenvalues equal the diagonal elements of A. This follows from that*

$$det(A - \lambda I) = (A_{11} - \lambda)(A_{22} - \lambda) \ldots (A_{nn} - \lambda).$$

Fact 22.26 *(Squares of triangular matrices) If T is lower (upper) triangular, then TT is as well.*

22.6.2 Orthogonal Vector and Matrices

Fact 22.27 *(Orthogonal vector) The $n \times 1$ vectors x and y are orthogonal if $x'y = 0$.*

Fact 22.28 *(Orthogonal matrix) The $n \times n$ matrix A is orthogonal if $A'A = I$. Properties: If A is orthogonal, then* $\det(A) = \pm 1$*; if A and B are orthogonal, then AB is orthogonal.*

Example 22.29 *(Rotation of vectors ("Givens rotations").) Consider the matrix $G = I_n$ except that $G_{ik} = c$, $G_{ik} = s$, $G_{ki} = -s$, and $G_{kk} = c$. If we let $c = \cos\theta$ and $s = \sin\theta$ for some angle θ, then $G'G = I$. To see this, consider the simple example where $i = 2$ and $k = 3$*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & s \\ 0 & -s & c \end{bmatrix}' \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & s \\ 0 & -s & c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c^2 + s^2 & 0 \\ 0 & 0 & c^2 + s^2 \end{bmatrix},$$

which is an identity matrix since $\cos^2\theta + \sin^2\theta = 1$. *$G$ is thus an orthogonal matrix. It is often used to "rotate" an* $n \times 1$ *vector* ε *as in* $u = G'\varepsilon$, *where we get*

$$
\begin{aligned}
u_t &= \varepsilon_t \text{ for } t \neq i, k \\
u_i &= \varepsilon_i c - \varepsilon_k s \\
u_k &= \varepsilon_i s + \varepsilon_k c.
\end{aligned}
$$

The effect of this transformation is to rotate the i^{th} *and* k^{th} *vectors counterclockwise through an angle of* θ.

22.6.3 Positive Definite Matrices

Fact 22.30 *(Positive definite matrix) The* $n \times n$ *matrix* A *is positive definite if for any non-zero* $n \times 1$ *vector* x, $x'Ax > 0$. *(It is positive semidefinite if* $x'Ax \geq 0$.*)*

Fact 22.31 *(Some properties of positive definite matrices) If* A *is positive definite, then all eigenvalues are positive and real. (To see why, note that an eigenvalue satisfies* $Ax = \lambda x$. *Premultiply by* x' *to get* $x'Ax = \lambda x'x$. *Since both* $x'Ax$ *and* $x'x$ *are positive real numbers,* λ *must also be.)*

Fact 22.32 *(More properties of positive definite matrices) If* B *is a* $r \times n$ *matrix of rank* r *and* A *is a* $n \times n$ *positive definite matrix, then* BAB' *is also positive definite and has rank* r. *For instance,* B *could be an invertible* $n \times n$ *matrix. If* $A = I_n$, *then we have that* BB' *is positive definite.*

Fact 22.33 *(More properties of positive definite matrices) If* A *is positive definite, then* $\det(A) > 0$ *and all diagonal elements are positive; if* A *is positive definite, then* A^{-1} *is too.*

Fact 22.34 *(Cholesky decomposition) See Fact 22.42.*

22.6.4 Symmetric Matrices

Fact 22.35 *(Symmetric matrix)* A *is symmetric if* $A = A'$.

Fact 22.36 *(Properties of symmetric matrices) If* A *is symmetric, then all eigenvalues are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Fact 22.37 *If* A *is symmetric, then* A^{-1} *is symmetric.*

22.6.5 Idempotent Matrices

Fact 22.38 *(Idempotent matrix) A is idempotent if $A = AA$. If A is also symmetric, then $A = A'A$.*

22.7 Matrix Decompositions

Fact 22.39 *(Diagonal decomposition) An $n \times n$ matrix A is diagonalizable if there exists a matrix C such that $C^{-1}AC = \Lambda$ is diagonal. We can thus write $A = C\Lambda C^{-1}$. The $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. We can then take C to be the matrix of the eigenvectors (in columns), and Λ the diagonal matrix with the corresponding eigenvalues along the diagonal.*

Fact 22.40 *(Spectral decomposition.) If the eigenvectors are linearly independent, then we can decompose A as*

$$A = Z\Lambda Z^{-1}, \text{ where } \Lambda = diag(\lambda_1, ..., \lambda_1) \text{ and } Z = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix},$$

where Λ is a diagonal matrix with the eigenvalues along the principal diagonal, and Z is a matrix with the corresponding eigenvalues in the columns.

Fact 22.41 *(Diagonal decomposition of symmetric matrices) If A is symmetric (and possibly singular) then the eigenvectors are orthogonal, $C'C = I$, so $C^{-1} = C'$. In this case, we can diagonalize A as $C'AC = \Lambda$, or $A = C\Lambda C'$. If A is $n \times n$ but has rank $r \leq n$, then we can write*

$$A = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} C_1 & C_2 \end{bmatrix}' = C_1 \Lambda_1 C_1',$$

where the $n \times r$ matrix C_1 contains the r eigenvectors associated with the r non-zero eigenvalues in the $r \times r$ matrix Λ_1.

Fact 22.42 *(Cholesky decomposition) Let Ω be an $n \times n$ symmetric positive definite matrix. The Cholesky decomposition gives the unique lower triangular P such that $\Omega = PP'$ (some software returns an upper triangular matrix, that is, Q in $\Omega = Q'Q$ instead). Note that each column of P is only identified up to a sign transformation; they can be reversed at will.*

Example 22.43 *(2×2 matrix) For a 2×2 matrix we have the following Cholesky decomposition*

$$chol\left(\begin{bmatrix} a & b \\ b & d \end{bmatrix}\right) = \begin{bmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{d - b^2/a} \end{bmatrix}.$$

Fact 22.44 *(Triangular Decomposition) Let Ω be an $n \times n$ symmetric positive definite matrix. There is a unique decomposition $\Omega = ADA'$, where A is lower triangular with ones along the principal diagonal, and D is diagonal with positive diagonal elements. This decomposition is usually not included in econometric software, but it can easily be calculated from the commonly available Cholesky decomposition since P in the Cholesky decomposition is of the form*

$$P = \begin{bmatrix} \sqrt{D_{11}} & 0 & \cdots & 0 \\ \sqrt{D_{11}} A_{21} & \sqrt{D_{22}} & & 0 \\ \vdots & & \ddots & \vdots \\ \sqrt{D_{11}} A_{n1} & \sqrt{D_{22}} A_{n2} & \cdots & \sqrt{D_{nn}} \end{bmatrix}.$$

Fact 22.45 *(Schur decomposition) The decomposition of the $n \times n$ matrix A gives the $n \times n$ matrices T and Z such that*

$$A = ZTZ^H$$

where Z is a unitary $n \times n$ matrix and T is an $n \times n$ upper triangular Schur form with the eigenvalues along the diagonal. Note that premultiplying by $Z^{-1} = Z^H$ and postmultiplying by Z gives

$$T = Z^H AZ,$$

which is upper triangular. The ordering of the eigenvalues in T can be reshuffled, although this requires that Z is reshuffled conformably to keep $A = ZTZ^H$, which involves a bit of tricky "book keeping."

Fact 22.46 *(Generalized Schur Decomposition) The decomposition of the $n \times n$ matrices G and D gives the $n \times n$ matrices Q, S, T, and Z such that Q and Z are unitary and S and T upper triangular. They satisfy*

$$G = QSZ^H \text{ and } D = QTZ^H.$$

The generalized Schur decomposition solves the generalized eigenvalue problem $Dx = \lambda Gx$, where λ are the generalized eigenvalues (which will equal the diagonal elements in T divided by the corresponding diagonal element in S). Note that we can write

$$Q^H GZ = S \text{ and } Q^H DZ = T.$$

Example 22.47 *If $G = I$ in the generalized eigenvalue problem $Dx = \lambda Gx$, then we are back to the standard eigenvalue problem. Clearly, we can pick $S = I$ and $Q = Z$ in this case, so $G = I$ and $D = ZTZ^H$, as in the standard Schur decomposition.*

Fact 22.48 *(QR decomposition) Let A be $m \times n$ with $m \geq n$. The QR decomposition is*

$$\begin{aligned} A_{m\times n} &= Q_{m\times m} R_{m\times n} \\ &= \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ \mathbf{0} \end{bmatrix} \\ &= Q_1 R_1. \end{aligned}$$

where Q is orthogonal ($Q'Q = I$) and R upper triangular. The last line is the "thin QR decomposition," where Q_1 is an $m \times n$ orthogonal matrix and R_1 an $n \times n$ upper triangular matrix.

Fact 22.49 *(Inverting by using the QR decomposition) Solving $Ax = c$ by inversion of A can be very numerically inaccurate (no kidding, this is a real problem). Instead, the problem can be solved with QR decomposition. First, calculate Q_1 and R_1 such that $A = Q_1 R_1$. Note that we can write the system of equations as*

$$Q_1 Rx = c.$$

Premultply by Q_1' to get (since $Q_1' Q_1 = I$)

$$Rx = Q_1' c.$$

This is an upper triangular system which can be solved very easily (first solve the first equation, then use the solution is the second, and so forth.)

Fact 22.50 *(Singular value decomposition) Let A be an $m \times n$ matrix of rank ρ. The singular value decomposition is*

$$A = U_{m\times m} S_{m\times n} V'_{n\times n}$$

where U and V are orthogonal and S is diagonal with the first ρ elements being non-zero, that is,

$$S = \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ where } S_1 = \begin{bmatrix} s_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_{\rho\rho} \end{bmatrix}.$$

Fact 22.51 *(Singular values and eigenvalues) The singular values of A are the nonnegative square roots of AA^H if $m \leq n$ and of $A^H A$ if $m \geq n$.*

Remark 22.52 *If the square matrix A is symmetric and idempotent ($A = A'A$), then the singular values are the same as the eigevalues. From Fact (22.41) we know that a symmetric A can be decomposed as $A = C\Lambda C'$. It follows that this is the same as the singular value decomposition.*

Fact 22.53 *(Condition number) The condition number of a matrix is the ratio of the largest (in magnitude) of the singular values to the smallest*

$$c = |s_{ii}|_{\max} / |s_{ii}|_{\min}.$$

For a square matrix, we can calculate the condition value from the eigenvalues of AA^H or $A^H A$ (see Fact 22.51). In particular, for a square matrix we have

$$c = \left|\sqrt{\lambda_i}\right|_{\max} / \left|\sqrt{\lambda_i}\right|_{\min},$$

where λ_i are the eigenvalues of AA^H and A is square.

Fact 22.54 *(Condition number and computers) The determinant is not a good indicator of the realibility of numerical inversion algorithms. Instead, let c be the condition number of a square matrix. If $1/c$ is close to the a computer's floating-point precision (10^{-13} or so), then numerical routines for a matrix inverse become unreliable. For instance, while $\det(0.1 \times I_{20}) = 10^{-20}$, the condition number of $0.1 \times I_{20}$ is unity and the matrix is indeed easy to invert to get $10 \times I_{20}$.*

Fact 22.55 *(Inverting by using the SVD decomposition) The inverse of the square matrix A is found by noting that if A is square, then from Fact 22.50 we have*

$$\begin{aligned} AA^{-1} &= I \text{ or} \\ USV'A^{-1} &= I\text{, so} \\ A^{-1} &= VS^{-1}U', \end{aligned}$$

provided S is invertible (otherwise A will not be). Since S is diagonal, S^{-1} is also diagonal with the inverses of the diagonal elements in S, so it is very easy to compute.

Fact 22.56 *(Pseudo inverse or generalized inverse) The Moore-Penrose pseudo (generalized) inverse of an $m \times n$ matrix A is defined as*

$$A^{+} = VS^{+}U', \text{ where } \underset{nxm}{S^{+}} = \begin{bmatrix} S_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where V and U are from Fact 22.50. The submatrix S_{11}^{-1} contains the reciprocals of the non-zero singular values along the principal diagonal. A^{+} satisfies the A^{+} satisfies the Moore-Penrose conditions

$$AA^{+}A = A,\ A^{+}AA^{+} = A^{+},\ \left(AA^{+}\right)' = AA^{+}, \text{ and } \left(A^{+}A\right)' = A^{+}A.$$

See Fact 22.9 for the idea behind the generalized inverse.

Fact 22.57 *(Some properties of generalized inverses) If A has full rank, then $A^{+} = A^{-1}$; $(BC)^{+} = C^{+}B^{+}$; if B, and C are invertible, then $(BAC)^{-1} = C^{-1}A^{+}B^{-1}$; $(A^{+})' = (A')^{+}$; if A is symmetric, then A^{+} is symmetric.*

Example 22.58 *(Pseudo inverse of a square matrix) For the matrix*

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}, \text{ we have } A^{+} = \begin{bmatrix} 0.02 & 0.06 \\ 0.04 & 0.12 \end{bmatrix}.$$

Fact 22.59 *(Pseudo inverse of symmetric matrix) If A is symmetric, then the SVD is identical to the spectral decomposition $A = Z\Lambda Z'$ where Z is a matrix of the orthogonal eigenvectors ($Z'Z = I$) and Λ is a diagonal matrix with the eigenvalues along the main* *diagonal. By Fact 22.56) we then have $A^+ = Z\Lambda^+ Z'$, where*

$$\Lambda^+ = \begin{bmatrix} \Lambda_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

with the reciprocals of the non-zero eigen values along the principal diagonal of Λ_{11}^{-1}.

22.8 Matrix Calculus

Fact 22.60 *(Matrix differentiation of non-linear functions, $\partial y/\partial x'$) Let the vector $y_{n\times 1}$ be a function of the vector $x_{m\times 1}$*

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}.$$

Then, let $\partial y/\partial x'$ be the $n \times m$ matrix

$$\frac{\partial y}{\partial x'} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x'} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x'} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{bmatrix}.$$

This matrix is often called the Jacobian of the f functions. (Note that the notation implies that the derivatives of the first element in y, denoted y_1, with respect to each of the elements in x' are found in the first row of $\partial y/\partial x'$. A rule to help memorizing the format of $\partial y/\partial x'$: y is a column vector and x' is a row vector.)

Fact 22.61 *($\partial y'/\partial x$ instead of $\partial y/\partial x'$) With the notation in the previous Fact, we get*

$$\frac{\partial y'}{\partial x} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} & \cdots & \frac{\partial f_n(x)}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(x)}{\partial x_m} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{bmatrix} = \left(\frac{\partial y}{\partial x'}\right)'.$$

Fact 22.62 *(Matrix differentiation of linear systems) When $y_{n\times 1} = A_{n\times m} x_{m\times 1}$, then*

$f(x)$ is a linear function

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

In this case $\partial y/\partial x' = A$ and $\partial y'/\partial x = A'$.

Fact 22.63 *(Matrix differentiation of inner product) The inner product of two column vectors, $y = z'x$, is a special case of a linear system with $A = z'$. In this case we get $\partial (z'x)/\partial x' = z'$ and $\partial (z'x)/\partial x = z$. Clearly, the derivatives of $x'z$ are the same (a transpose of a scalar).*

Example 22.64 *($\partial (z'x)/\partial x = z$ when x and z are 2×1 vectors)*

$$\frac{\partial}{\partial x}\left(\begin{bmatrix} z_1 & z_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

Fact 22.65 *(First order Taylor series) For each element $f_i(x)$ in the $n\times$ vector $f(x)$, we can apply the mean-value theorem*

$$f_i(x) = f_i(c) + \frac{\partial f_i(b_i)}{\partial x'}(x - c),$$

for some vector b_i between c and x. Stacking these expressions gives

$$\begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} = \begin{bmatrix} f_1(c) \\ \vdots \\ f_n(c) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1(b_1)}{\partial x_1} & \cdots & \frac{\partial f_1(b_1)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial f_n(b_n)}{\partial x_1} & \cdots & \frac{\partial f_n(b_n)}{\partial x_m} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \text{ or}$$

$$f(x) = f(c) + \frac{\partial f(b)}{\partial x'}(x - c),$$

where the notation $f(b)$ is a bit sloppy. It should be interpreted as that we have to evaluate the derivatives at different points for the different elements in $f(x)$.

Fact 22.66 *(Matrix differentiation of quadratic forms) Let $x_{m\times 1}$ be a vector, $A_{m\times m}$ a* *matrix, and* $f(x)_{n\times 1}$ *a vector of functions. Then,*

$$\begin{aligned}\frac{\partial f(x)' A f(x)}{\partial x} &= \left(\frac{\partial f(x)}{\partial x'}\right)' \left(A + A'\right) f(x) \\ &= 2\left(\frac{\partial f(x)}{\partial x'}\right)' A f(x) \text{ if } A \text{ is symmetric.}\end{aligned}$$

If $f(x) = x$*, then* $\partial f(x)/\partial x' = I$*, so* $\partial(x'Ax)/\partial x = 2Ax$ *if* A *is symmetric.*

Example 22.67 *(*$\partial(x'Ax)/\partial x = 2Ax$ *when* x *is* 2×1 *and* A *is* 2×2*)*

$$\begin{aligned}\frac{\partial}{\partial x}\left(\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) &= \left(\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \\ &= 2\begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ if } A_{21} = A_{12}.\end{aligned}$$

Example 22.68 *(Least squares) Consider the linear model* $Y_{m\times 1} = X_{m\times n}\beta_{n\times 1} + u_{m\times 1}$*. We want to minimize the sum of squared fitted errors by choosing the* $n\times 1$ *vector* β*. The fitted errors depend on the chosen* β*:* $u(\beta) = Y - X\beta$*, so quadratic loss function is*

$$\begin{aligned}L &= u(\beta)'u(\beta) \\ &= (Y - X\beta)'(Y - X\beta).\end{aligned}$$

In thus case, $f(\beta) = u(\beta) = Y - X\beta$*, so* $\partial f(\beta)/\partial\beta' = -X$*. The first order condition for* $u'u$ *is thus*

$$-2X'\left(Y - X\hat{\beta}\right) = \mathbf{0}_{n\times 1} \text{ or } X'Y = X'X\hat{\beta},$$

which can be solved as

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y.$$

Fact 22.69 *(Matrix of 2nd order derivatives of of a non-linear function,* $\partial^2 y/\partial x\partial x'$*) Let the scalar* y *be a function of the vector* $x_{m\times 1}$

$$y = f(x).$$

Then, let $\partial^2 y/\partial x \partial x'$ be the $m \times m$ matrix with $\partial^2 y/\partial x_i \partial x_j$ in cell (i, j)

$$\frac{\partial^2 y}{\partial x \partial x'} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_m \partial xm} \end{bmatrix}.$$

This matrix is often called the Hessian of the f function. This is clearly a symmetric matrix.

22.9 Miscellaneous

Fact 22.70 *(Some properties of transposes) $(A + B)' = A' + B'$; $(ABC)' = C'B'A'$ (if conformable).*

Fact 22.71 *(Kronecker product) If A and B are matrices, then*

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

Some properties: $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (if conformable); $(A \otimes B)(C \otimes D) = AC \otimes BD$ (if conformable); $(A \otimes B)' = A' \otimes B'$; if a is $m \times 1$ and b is $n \times 1$, then $a \otimes b = (a \otimes I_n)b$; if A is symmetric and positive definite, then chol$(A \otimes I)$ =chol$(A) \otimes I$ and chol$(I \otimes A) = I \otimes$chol$(A)$.

Fact 22.72 *(Cyclical permutation of trace) Trace(ABC) =Trace(BCA) =Trace(CAB), if the dimensions allow the products.*

Fact 22.73 *(The* vec *operator).* vec *A where A is $m \times n$ gives an $mn \times 1$ vector with the columns in A stacked on top of each other. For instance,* $\operatorname{vec} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}$.

Properties: $\operatorname{vec}(A + B) = \operatorname{vec} A + \operatorname{vec} B$; $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$; *if a and b are column vectors, then* $\operatorname{vec}(ab') = b \otimes a$.

Fact 22.74 *(The vech operator) vech$A$ where A is $m \times m$ gives an $m(m+1)/2 \times 1$ vector with the elements on and below the principal diagonal A stacked on top of each other (columnwise). For instance, vech* $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix}$*, that is, like vec, but uses only the elements on and below the principal diagonal.*

Fact 22.75 *(Duplication matrix) The duplication matrix D_m is defined such that for any symmetric $m \times m$ matrix A we have* $\operatorname{vec} A = D_m \text{vech} A$. *The duplication matrix is therefore useful for "inverting" the vech operator (the step from* $\operatorname{vec} A$ *to A is trivial). For instance, to continue the example of the vech operator*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{21} \\ a_{22} \end{bmatrix} \text{ or } D_2 \text{vech} A = \operatorname{vec} A.$$

Fact 22.76 *(OLS notation) Let x_t be $k \times 1$ and y_t be $m \times 1$. Suppose we have T such vectors. The sum of the outer product (a $k \times m$ matrix) is*

$$S = \sum_{t=1}^{T} x_t y_t'.$$

Create matrices $X_{T \times k}$ and $Y_{T \times m}$ by letting x_t' and y_t' be the t^{th} rows

$$X_{T \times k} = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix} \text{ and } Y_{T \times m} = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}.$$

We can then calculate the same sum of outer product, S, as

$$S = X'Y.$$

(To see this, let $X(i,:)$ be the ith row of X, and similarly for Y, so

$$X'Y = \sum_{t=1}^{T} X(t,:)'Y(t,:),$$

which is precisely $\Sigma_{t=1}^{T} x_t y_t'$.) For instance, with

$$x_t = \begin{bmatrix} a_t \\ b_t \end{bmatrix} \text{ and } y_t = \begin{bmatrix} p_t \\ q_t \\ r_t \end{bmatrix},$$

and $T = 2$ we have

$$X'Y = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 \end{bmatrix} = \sum_{t=1}^{T} \begin{bmatrix} a_t \\ b_t \end{bmatrix} \begin{bmatrix} p_t & q_t & r_t \end{bmatrix}.$$

Fact 22.77 *(Matrix geometric series) Suppose the eigenvalues to the square matrix A are all less than one in modulus. Then,*

$$I + A + A^2 + \cdots = (1 - A)^{-1}.$$

To see why this makes sense, consider $(1 - A)\, \Sigma_{t=1}^{T} A^t$ (with the convention that $A^0 = I$). It can be written as

$$(1 - A)\, \Sigma_{t=1}^{T} A^t = \left(I + A + A^2 + \cdots\right) - A\left(I + A + A^2 + \cdots\right) = I - A^{T+1}.$$

If all the eigenvalues are stable, then $\lim_{T\to\infty} A^{T+1} = \mathbf{0}$, so taking the limit of the previous equation gives

$$(1 - A) \lim_{T\to\infty} \Sigma_{t=1}^{T} A = I.$$

Fact 22.78 *(Matrix exponential) The matrix exponential of an $n \times n$ matrix A is defined as*

$$\exp(At) = \sum_{s=0}^{\infty} \frac{(At)^s}{s!}.$$

Bibliography

Anton, H., 1987, *Elementary linear algebra*, John Wiley and Sons, New York, 5th edn.

Björk, Å., 1996, *Numerical methods for least squares problems*, SIAM, Philadelphia.

Golub, G. H., and C. F. van Loan, 1989, *Matrix computations*, The John Hopkins University Press, Baltimore, 2nd edn.

Greenberg, M. D., 1988, *Advanced engineering mathematics*, Prentice Hall, Englewood Cliffs, New Jersey.

Greene, W. H., 2000, *Econometric analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 4th edn.