Math (mostly Linear Algebra and Calculus) Refresher

Javier Martinez Hernandez

Western University, Economics Department

fmarti23@uwo.ca

Roadmap

Matrix Algebra

Optimization

Before we begin

- I will be following Appendix A from Greene.
- Equation numbering follows Greene's appendix.
- Doubts and questions are welcome and encouraged.

Terminology

- A matrix is a rectangular array of numbers, denoted

$$\mathbf{A} = [a_{ik}] = [\mathbf{A}]_{ik} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{bmatrix} \tag{A-1}$$

- The typical element is used to denote the matrix. A subscripted element of a matrix is always read as $a_{\text{row, column}}$

- A vector is an ordered set of numbers arranged either in a row or a column.
- A row vector is also a matrix with one row, whereas a column vector is a matrix with one column.
- A matrix can also be viewed as a set of column vectors or as a set of row vectors.
- The dimensions of a matrix are the numbers of rows and columns it contains. " $\mathbf{A}$ is an $n \times K$ matrix" (read " n by K ") will always mean that $\mathbf{A}$ has n rows and K columns.
- If n equals K, then $\mathbf{A}$ is a square matrix.

Matrix commonly employed in economics

- A symmetric matrix is one in which $a_{ik} = a_{ki}$ for all i and k.
- A diagonal matrix is a square matrix whose only nonzero elements appear on the main diagonal, that is, moving from upper left to lower right.
- A scalar matrix is a diagonal matrix with the same value in all diagonal elements.
- An identity matrix is a scalar matrix with ones on the diagonal. This matrix is always denoted $\mathbf{I}$. A subscript is sometimes included to indicate its size, or order. For example, $\mathbf{I}_4$ indicates a 4×4 identity matrix.
- A triangular matrix is one that has only zeros either above or below the main diagonal. If the zeros are above the diagonal, the matrix is lower triangular.

Algebraic manipulation of matrices

- Matrices (or vectors) **A** and **B** are equal if and only if they have the same dimensions and each element of **A** equals the corresponding element of **B**.

$$\mathbf{A} = \mathbf{B} \quad \text{if and only if } a_{ik} = b_{ik} \quad \text{for all } i \text{ and } k \tag{A-2}$$

- The transpose of a matrix **A**, denoted $\mathbf{A}'$, is obtained by creating the matrix whose *kth* row is the *kt* column of the original matrix.

- If $\mathbf{B} = \mathbf{A}'$, then each column of **A** will appear as the corresponding row of **B**. If **A** is $n \times K$, then $\mathbf{A}'$ is $K \times n$.

$$\mathbf{B} = \mathbf{A}' \Leftrightarrow b_{ik} = a_{ki} \quad \text{for all } i \text{ and } k \tag{A-3}$$

- The definition of a symmetric matrix implies that

$$\text{if (and only if) } \mathbf{A} \text{ is symmetric, then } \mathbf{A} = \mathbf{A}' \tag{A-4}$$

- It also follows from the definition that for any $\mathbf{A}$,

$$(\mathbf{A}')' = \mathbf{A} \tag{A-5}$$

- Finally, the transpose of a column vector, a, is a row vector:

$$\mathbf{a}' = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

Vectorization

- The matrix function Vec($\mathbf{A}$) takes the columns of an $n \times K$ matrix and rearranges them in a long $nK \times 1$ vector.

$$\text{Vec}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = [1, 2, 2, 4]'$$

Matrix Addition

- The operations of addition and subtraction are extended to matrices by defining

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = [a_{ik} + b_{ik}] \tag{A-6}$$
$$\mathbf{A} - \mathbf{B} = [a_{ik} - b_{ik}] \tag{A-7}$$

- Matrices cannot be added unless they have the same dimensions, in which case they are said to be conformable for addition.
- A zero matrix or null matrix is one whose elements are all zero.
- In the addition of matrices, the zero matrix plays the same role as the scalar 0 in scalar addition.

$$\mathbf{A} + \mathbf{0} = \mathbf{A} \tag{A-8}$$

Matrix Addition

- Matrix addition is commutative,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{A-9}$$

- and associative,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{A-10}$$

- and it follows that

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \tag{A-11}$$

Vector Multiplication

- The inner product, or dot product, of two vectors, **a** and **b**, is a scalar and is written

$$\mathbf{a}'\mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{j=1}^{n} a_j b_j \tag{A-12}$$

- Throughout this refresher (and in general in Math books), an untransposed vector will always be a column vector.
- We will often require a notation for the column vector that is the transpose of a row of a matrix.
- $\mathbf{a}_k$, or $\mathbf{a}_l$ or $\mathbf{a}_m$ will denote column k, l, or m of the matrix $\mathbf{A}$,
- $\mathbf{a}_i$, or $\mathbf{a}_j$ or $\mathbf{a}_t$ or $\mathbf{a}_s$ will denote the column vector formed by the transpose of row i, j, t, or s of matrix $\mathbf{A}$.
- Thus, $\mathbf{a}_i'$ is row i of $\mathbf{A}$.

Matrix Multiplication and Scalar Multiplication

- For an $n \times K$ matrix $\mathbf{A}$ and a $K \times M$ matrix $\mathbf{B}$, the product matrix, $\mathbf{C} = \mathbf{AB}$, is an $n \times M$ matrix whose ik th element is the inner product of row i of $\mathbf{A}$ and column k of $\mathbf{B}$. Thus, the product matrix $\mathbf{C}$ is

$$\mathbf{C} = \mathbf{AB} \Rightarrow c_{ik} = \mathbf{a}_i'\mathbf{b}_k \tag{A-15}$$

- To multiply two matrices, the number of columns in the first must be the same as the number of rows in the second, in which case they are conformable for multiplication.
- Multiplication of matrices is generally not commutative. In some cases, $\mathbf{AB}$ may exist, but $\mathbf{BA}$ may be undefined or, if it does exist, may have different dimensions.

- even if $\mathbf{AB}$ and $\mathbf{BA}$ do have the same dimensions, they will not be equal.
- In view of this, we define premultiplication and postmultiplication of matrices.
- In the product $\mathbf{AB}, \mathbf{B}$ is premultiplied by $\mathbf{A}$, whereas $\mathbf{A}$ is postmultiplied by $\mathbf{B}$.
- Scalar multiplication of a matrix is the operation of multiplying every element of the matrix by a given scalar. For scalar c and matrix $\mathbf{A}$,

$$c\mathbf{A} = [ca_{ik}] \tag{A-16}$$

- If two matrices **A** and **B** have the same number of rows and columns, then we can compute the direct product (also called the Hadamard product or the Schur product or the entrywise product).
- This product is a new matrix (or vector) whose ij element is the product of the corresponding elements of **A** and **B**.
- The usual symbol for this operation is " $\circ$."

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \circ \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} 1a & 2b \\ 2b & 3c \end{bmatrix} \text{ and } \begin{pmatrix} 3 \\ 5 \end{pmatrix} \circ \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 20 \end{pmatrix}$$

- The product of a matrix and a vector is written

$$\mathbf{c} = \mathbf{A}\mathbf{b}$$

- A simple way to check the conformability of two matrices for multiplication is to write down the dimensions of the operation, for example, $(n \times K)$ times $(K \times M)$.
- The inner dimensions must be equal; the result has dimensions equal to the outer values.

- The number of elements in **b** must equal the number of columns in **A**; the result is a vector with number of elements equal to the number of rows in **A**.

$$\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

  We can interpret this in two ways. First, it is a compact way of writing the three equations

$$\begin{aligned} 5 &= 4a + 2b + 1c \\ 4 &= 2a + 6b + 1c \\ 1 &= 1a + 1b + 0c \end{aligned}$$

- Second, by writing the set of equations as

$$\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} = a \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} + c \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

  we see that the right-hand side is a linear combination of the columns of the matrix where the coefficients are the elements of the vector.

- For the general case,

$$\mathbf{c} = \mathbf{Ab} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_K\mathbf{a}_K \tag{A-17}$$

- In the calculation of a matrix product $\mathbf{C} = \mathbf{AB}$, each column of $\mathbf{C}$ is a linear combination of the columns of $\mathbf{A}$, where the coefficients are the elements in the corresponding column of B. That is,

$$\mathbf{C} = \mathbf{AB} \Leftrightarrow \mathbf{c}_k = \mathbf{A}\mathbf{b}_k \tag{A-18}$$

A way to define the identity matrix

- Let $\mathbf{e}_k$ be a column vector that has zeros everywhere except for a one in the *kth* position.
- Then $\mathbf{Ae}_k$ is a linear combination of the columns of $\mathbf{A}$ in which the coefficient on every column but the k th is zero, whereas that on the k th is one.
- The result is

$$\mathbf{a}_k = \mathbf{Ae}_k \tag{A-19}$$

- Combining this result with (A-17) produces

$$(\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_n) = \mathbf{A}(\mathbf{e}_1\mathbf{e}_2 \cdots \mathbf{e}_n) = \mathbf{AI} = \mathbf{A} \tag{A-20}$$

- In matrix multiplication, the identity matrix is analogous to the scalar 1.
- For any matrix or vector $\mathbf{A}$, $\mathbf{AI} = \mathbf{A}$.
- In addition, $\mathbf{IA} = \mathbf{A}$, although if $\mathbf{A}$ is not a square matrix, the two identity matrices are of different orders.
- A conformable matrix of zeros produces the expected result: $\mathbf{A0} = \mathbf{0}$.

General rules for matrix multiplication

- Associative law: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.
- Distributive law: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.
- Transpose of a product: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.
- Transpose of an extended product: $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$

Sum of values

- Denote by $\mathbf{i}$ a vector that contains a column of ones. Then,

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n = \mathbf{i}'\mathbf{x} \tag{A-25}$$

- If all elements in $\mathbf{x}$ are equal to the same constant a, then $\mathbf{x} = a\mathbf{i}$ and

$$\sum_{i=1}^{n} x_i = \mathbf{i}'(\mathbf{ai}) = a\left(\mathbf{i}'\mathbf{i}\right) = na \tag{A-26}$$

- For any constant a and vector $\mathbf{x}$,

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i = a\mathbf{i}'\mathbf{x} \tag{A-27}$$

- If $a = 1/n$, then we obtain the arithmetic mean,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \mathbf{i}' \mathbf{x} \tag{A-28}$$

- from which it follows that

$$\sum_{i=1}^{n} x_i = \mathbf{i}' \mathbf{x} = n\bar{x}$$

- The sum of squares of the elements in a vector $\mathbf{x}$ is

$$\sum_{i=1}^{n} x_i^2 = \mathbf{x}' \mathbf{x} \tag{A-29}$$

  while the sum of the products of the n elements in vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\sum_{i=1}^{n} x_i y_i = \mathbf{x} \mathbf{y} \tag{A-30}$$

- By the definition of matrix multiplication,

$$\left[\mathbf{X}'\mathbf{X}\right]_{kl} = \left[\mathbf{x}_k'\mathbf{x}_l\right] \tag{A-31}$$

- is the inner product of the k th and l th columns of $\mathbf{X}$.
- If $\mathbf{X}$ is $n \times K$, then [again using (A-14)]

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'$$

- This form shows that the $K \times K$ matrix $\mathbf{X}'\mathbf{X}$ is the sum of n $(K \times K)$ matrices, each formed from a single row (year) of $\mathbf{X}$.

A Useful Idempotent Matrix

- The "centering matrix" that is used to transform data to deviations from their mean.

$$\mathbf{i}\bar{x} = \mathbf{i}\frac{1}{n}\mathbf{i}'\mathbf{x} = \begin{bmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} = \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x} \tag{A-32}$$

- The matrix $(1/n)\mathbf{i}\mathbf{i}'$ is an $n \times n$ matrix with every element equal to $1/n$. The set of values in deviations form is

$$\begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \cdots \\ x_n - \bar{x} \end{bmatrix} = [\mathbf{x} - \mathbf{i}\bar{x}] = \left[\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right] \tag{A-33}$$

- Because $\mathbf{x} = \mathbf{I}\mathbf{x}$

$$\left[\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right] = \left[\mathbf{I}\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right] = \left[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right]\mathbf{x} = \mathbf{M}^0\mathbf{x} \tag{A-34}$$

- The symbol $\mathbf{M}^0$ will be used only for this matrix.
- Its diagonal elements are all $(1 - 1/n)$, and its off-diagonal elements are $-1/n$.
- The matrix $\mathbf{M}^0$ is primarily useful in computing sums of squared deviations.

- Some computations are simplified by the result

$$\mathbf{M}^0\mathbf{i} = \left[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right]\mathbf{i} = \mathbf{i} - \frac{1}{n}\mathbf{i}\left(\mathbf{i}'\mathbf{i}\right) = \mathbf{0}$$

- which implies that $\mathbf{i}'\mathbf{M}^0 = \mathbf{0}'$.

- The sum of deviations about the mean is then

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \mathbf{i}'\left[\mathbf{M}^0\mathbf{x}\right] = \mathbf{0}'\mathbf{x} = 0 \tag{A-35}$$

- For a single variable $\mathbf{x}$, the sum of squared deviations about the mean is

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2 \tag{A-36}$$

- In matrix terms,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (\mathbf{x} - \bar{x}\mathbf{i})'(\mathbf{x} - \bar{x}\mathbf{x}) = \left(\mathbf{M}^0\mathbf{x}\right)'\left(\mathbf{M}^0\mathbf{x}\right) = \mathbf{x}'\mathbf{M}^0\mathbf{M}^0\mathbf{x}$$

- Because all off-diagonal elements of $\mathbf{M}^0$ equal $-1/n, \mathbf{M}^0$ is symmetric.
- $\mathbf{M}^0$ is equal to its square; $\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$.

- An idempotent matrix, $\mathbf{M}$, is one that is equal to its square, that is, $\mathbf{M}^2 = \mathbf{MM} = \mathbf{M}$. If $\mathbf{M}$ is a symmetric idempotent matrix (all of the idempotent matrices we shall encounter are symmetric), then $\mathbf{M}'\mathbf{M} = \mathbf{M}$ as well.

- Thus, $\mathbf{M}^0$ is a symmetric idempotent matrix. Combining results, we obtain

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mathbf{x}'\mathbf{M}^0\mathbf{x} \tag{A-37}$$

- Consider constructing a matrix of sums of squares and cross products in deviations from the column means. For two vectors $\mathbf{x}$ and $\mathbf{y}$

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\left(\mathbf{M}^0\mathbf{x}\right)'\left(\mathbf{M}^0\mathbf{y}\right) \tag{A-38}$$

so

$$\left[\begin{array}{cc} \sum_{i=1}^{n}(x_i-\bar{x})^2 & \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\ \sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) & \sum_{i=1}^{n}(y_i-\bar{y})^2 \end{array}\right]=\left[\begin{array}{cc} \mathbf{x}'\mathbf{M}^0\mathbf{x} & \mathbf{x}'\mathbf{M}^0\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{x} & \mathbf{y}'\mathbf{M}^0\mathbf{y} \end{array}\right] \tag{A-39}$$

- If we put the two column vectors $\mathbf{x}$ and $\mathbf{y}$ in an $n\times 2$ matrix $\mathbf{Z}=[\mathbf{x},\mathbf{y}]$, then $\mathbf{M}^0\mathbf{Z}$ is the $n\times 2$ matrix in which the two columns of data are in mean deviation form. Then

$$\left(\mathbf{M}^0\mathbf{Z}\right)'\left(\mathbf{M}^0\mathbf{Z}\right)=\mathbf{Z}'\mathbf{M}^0\mathbf{M}^0\mathbf{Z}=\mathbf{Z}'\mathbf{M}^0 Z$$

Vector Spaces

- The K elements of a column vector

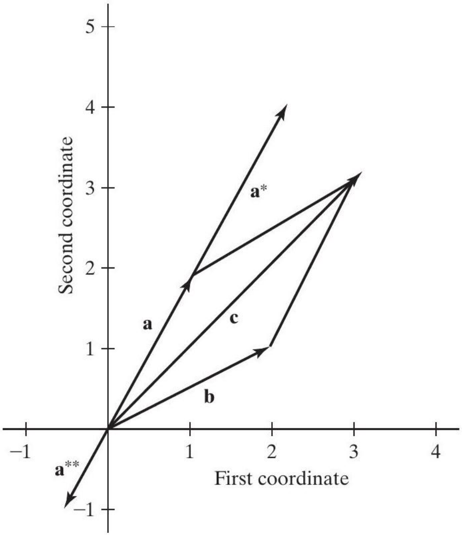
$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_K \end{bmatrix}$$

can be viewed as the coordinates of a point in a K-dimensional space.

- Two basic arithmetic operations are defined for vectors, scalar multiplication and addition.

- A scalar multiple of a vector, $\mathbf{a}$, is another vector, say $\mathbf{a}^*$, whose coordinates are the scalar multiple of a's coordinates.

Figure A1. Vector Space



- Consider the following vectors:

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{a}^* = 2\mathbf{a} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad \mathbf{a}^{**} = -\frac{1}{2}\mathbf{a} = \begin{bmatrix} -\frac{1}{2} \\ -1 \end{bmatrix}$$

- The set of all possible scalar multiples of $\mathbf{a}$ is the line through the origin, $\mathbf{0}$ and $\mathbf{a}$.
- Any scalar multiple of $\mathbf{a}$ is a segment of this line.
- The sum of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a third vector whose coordinates are the sums of the corresponding coordinates of $\mathbf{a}$ and $\mathbf{b}$.

- Consider:

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

- Geometrically, **c** is obtained by moving in the distance and direction defined by **b** from the tip of **a** or, because addition is commutative, from the tip of **b** in the distance and direction of **a**.

- Note that scalar multiplication and addition of vectors are special cases of (A-16) and (A-6) for matrices.

- The two-dimensional plane is the set of all vectors with two real-valued coordinates. We label this set $\mathbb{R}^2$ (" R two," not " R squared"). It has two important properties.
  - $\mathbb{R}^2$ is closed under scalar multiplication; every scalar multiple of a vector in $\mathbb{R}^2$ is also in $\mathbb{R}^2$.
  - $\mathbb{R}^2$ is closed under addition; the sum of any two vectors in the plane is always a vector in $\mathbb{R}^2$.
- A vector space is any set of vectors that is closed under scalar multiplication and addition.

Linear Combinations and Basis Vectors

- A set of vectors in a vector space is a basis for that vector space if they are linearly independent and any vector in the vector space can be written as a linear combination of that set of vectors.

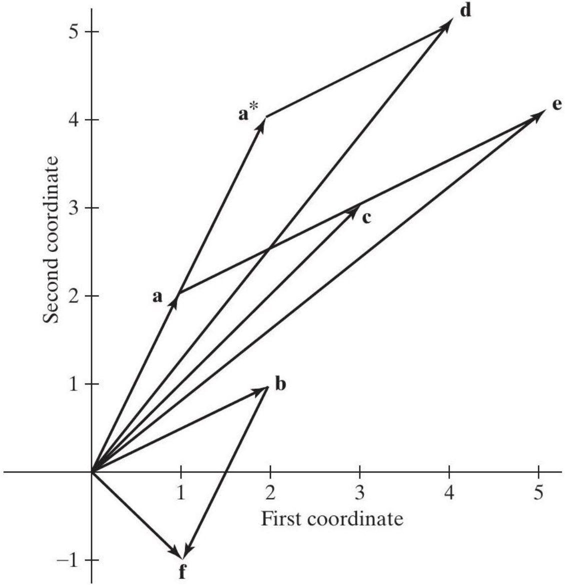
- Vectors have both magnitude and direction, but not position.

- The magnitude of a vector can be measured with a norm. For example, the $\mathcal{L}^2$ norm, also called Euclidean distance.

$$||\mathbf{x}||^2 = \sqrt{\sum_{k=1}^{n} (x_k)^2}$$

- A closed vector space: $\alpha, \beta \in \mathbb{R}$ and $x, y \in V$ (where V is our vector space), then $\alpha x + \beta y \in V$.

Figure A2. Linear combination of vectors



Subspace and Span

- A subspace of $\mathbb{R}^N$ (or any vectors space V), is a nonempty set, $\mathcal{S}$, of vectors satisfying:
  - Closed under addition: $\forall v_1, v_2 \in \mathcal{S}$ and $\alpha, \beta \in \mathbb{R} \Rightarrow \alpha v_1 + \beta v_2 \in \mathcal{S}$.
  - Closed under scalar multiplication: $v_1 \in \mathcal{S}$, $\alpha \in \mathbb{R} \Rightarrow \alpha v_1 \in \mathcal{S}$.
- Row subspace of a matrix: all possible linear combinations of the matrix rows.
- Column subspace of a matrix: all linear combinations of the columns of a matrix.

- Consider a set of vectors $V = \{v_1, v_2, ..., v_n\}$.
- The span of a vector set is the subspace composed of all the possible linear combinations of the vectors in that set.

$$span(V) = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n$$

such that $\forall \alpha_i \in \mathbb{R}, i = 1, ..., n; v_i \in V, i = 1, ..., n$.

- A set of vectors in a vector space is a basis for that vector space if they are linearly independent and any vector in the vector space can be written as a linear combination of that set of vectors.
- Any pair of two-element vectors that point in different directions will form a basis for $\mathbb{R}^2$.

- Consider an arbitrary set of three vectors in $\mathbb{R}^2$, $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$.

- If $\mathbf{a}$ and $\mathbf{b}$ are a basis, then we can find numbers α_1 and α_2 such that $\mathbf{c} = \alpha_1 \mathbf{a} + \alpha_2 \mathbf{b}$.

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\begin{aligned} c_1 &= \alpha_1 a_1 + \alpha_2 b_1 \\ c_2 &= \alpha_1 a_2 + \alpha_2 b_2 \end{aligned} \tag{A-40}$$

- The solutions (α_1, α_2) to this pair of equations are

$$\alpha_1 = \frac{b_2 c_1 - b_1 c_2}{a_1 b_2 - b_1 a_2}, \quad \alpha_2 = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - b_1 a_2} \tag{A-41}$$

- This result gives a unique solution unless $(a_1 b_2 - b_1 a_2) = 0$.
- If $(a_1 b_2 - b_1 a_2) = 0$, then $a_1/a_2 = b_1/b_2$, which means that **b** is just a multiple of **a**.
- This returns us to our original condition, that **a** and **b** must point in different directions.
- The implication is that if **a** and **b** are any pair of vectors for which the denominator in (A-41) is not zero, then any other vector **c** can be formed as a unique linear combination of **a** and **b**.

Linear Independence

- A set of vectors is linearly independent if and only if the only solution $(\alpha_1, \ldots, \alpha_K)$ to

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \cdots + \alpha_K \mathbf{a}_K = \mathbf{0}$$

  is

$$\alpha_1 = \alpha_2 = \cdots = \alpha_K = 0$$

- How can we relate it to multicolinearity in Econometrics?

- The column space of a matrix is the vector space that is spanned by its column vectors.
- The column rank of a matrix is the dimension of the vector space that is spanned by its column vectors.
- Consider:

$$\mathbf{A} = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 6 & 8 \\ 7 & 1 & 8 \end{bmatrix}$$

- It contains three vectors from $\mathbb{R}^3$, but the third is the sum of the first two, so the column space of this matrix cannot have three dimensions.
- If the column rank of a matrix happens to equal the number of columns it contains, then the matrix is said to have full column rank.
- Use the term full rank to describe a matrix whose rank is equal to the number of columns it contains.

Determinant of a Matrix

- Determinants are not defined for nonsquare matrices.
- The determinant of a matrix is nonzero if and only if it has full rank.
- For 2×2 matrices, the computation of the determinant is

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc \tag{A-50}$$

- For more than two dimensions, the determinant can be obtained by using an expansion by cofactors. Using any row, say, i, we obtain

$$|\mathbf{A}| = \sum_{k=1}^{K} a_{ik}(-1)^{i+k} \left|\mathbf{A}_{(ik)}\right|, \quad k = 1, \ldots, K \tag{A-51}$$

  where $\mathbf{A}_{(ik)}$ is the matrix obtained from $\mathbf{A}$ by deleting row i and column k.

- The determinant of $\mathbf{A}_{(ik)}$ is called a minor of $\mathbf{A}$.[5] When the correct sign, $(-1)^{i+k}$, is added, it becomes a cofactor.

Orthogonality

- Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal, written $\mathbf{a} \perp \mathbf{b}$, if and only if

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = \mathbf{0}$$

Systems of Linear Equations

- Consider the set of n linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{A-56}$$

  in which the K elements of $\mathbf{x}$ constitute the unknowns.

- $\mathbf{A}$ is a known matrix of coefficients, and $\mathbf{b}$ is a specified vector of values.
- We are interested in knowing whether a solution exists.
- If so, then how to obtain it; and finally, if it does exist, then whether it is unique.

- We will consider only square systems of equations. That is, **A** is a square matrix: n=K.

- The number of rows in **A** is the number of equations, whereas the number of columns in **A** is the number of variables, this case is the familiar one of " n equations in n unknowns."

- Two types of systems of equations: Homogeneous and Nonhomogeneous.

- A **homogeneous** system is of the form $\mathbf{Ax} = \mathbf{0}$.
- By definition, a nonzero solution to such a system will exist if and only if $\mathbf{A}$ does not have full rank. If so, then for at least one column of $\mathbf{A}$, we can write the preceding as

$$\mathbf{a}_k = -\sum_{m \neq k} \frac{x_m}{x_k} \mathbf{a}_m$$

- This means, as we know, that the columns of $\mathbf{A}$ are linearly dependent and that $|\mathbf{A}| = \mathbf{0}$.

- A **nonhomogeneous** system of equations is of the form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{b}$ is a nonzero vector.
- The vector $\mathbf{b}$ is chosen arbitrarily and is to be expressed as a linear combination of the columns of $\mathbf{A}$.
- Because $\mathbf{b}$ has $\boldsymbol{K}$ elements, this solution will exist only if the columns of A span the entire K-dimensional space, $\mathbb{R}^K$
- Equivalently, we shall require that the columns of $\mathbf{A}$ be linearly independent or that $|\mathbf{A}|$ not be equal to zero.
- If $\mathbf{A}$ does not have full rank, then the nonhomogeneous system will have solutions for some vectors $\mathbf{b}$, namely, any $\mathbf{b}$ in the column space of $\mathbf{A}$.
- But we are interested in the case in which there are solutions for all nonzero vectors $\mathbf{b}$, which requires $\mathbf{A}$ to have full rank.

- To solve the system $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$, something akin to division by a matrix is needed. Suppose that we could find a square matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}$.
- If the equation system is premultiplied by this $\mathbf{B}$, then the following would be obtained:

$$\mathbf{BAx} = \mathbf{Ix} = \mathbf{x} = \mathbf{Bb} \tag{A-57}$$

- If the matrix $\mathbf{B}$ exists, then it is the inverse of $\mathbf{A}$, denoted

$$\mathbf{B} = \mathbf{A}^{-1}$$

- From the definition,

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

- In addition, by premultiplying by $\mathbf{A}$, postmultiplying by $\mathbf{A}^{-1}$, and then canceling terms, we find

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

as well.

- If the inverse exists, then it must be unique. Suppose that it is not and that $\mathbf{C}$ is a different inverse of $\mathbf{A}$.
- Then $\mathbf{CAB} = \mathbf{CAB}$, but $(\mathbf{CA})\mathbf{B} = \mathbf{IB} = \mathbf{B}$ and $\mathbf{C}(\mathbf{AB}) = \mathbf{C}$, which would be a contradiction if $\mathbf{C}$ did not equal $\mathbf{B}$.
- Because, by (A-57), the solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, the solution to the equation system is unique as well.

- A matrix is nonsingular if and only if its inverse exists.
- We shall use a^{ik} to indicate the *ik* th element of $\mathbf{A}^{-1}$. The general formula for computing an inverse matrix is

$$a^{ik} = \frac{|\mathbf{C}_{ki}|}{|\mathbf{A}|} \tag{A-59}$$

  where $|\mathbf{C}_{ki}|$ is the *ki* th cofactor of $\mathbf{A}$.
- It follows, therefore, that for $\mathbf{A}$ to be nonsingular, $|\mathbf{A}|$ must be nonzero. Notice the reversal of the subscripts

- Some computational results involving inverses are

$$\left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$$
$$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$$
$$\left(\mathbf{A}^{-1}\right)' = (\mathbf{A}')^{-1}$$

- If $\mathbf{A}$ is symmetric, then $\mathbf{A}^{-1}$ is symmetric.

- When both inverse matrices exist,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{A-64}$$

- For the nonhomogeneous system

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

if $\mathbf{A}$ is nonsingular, then the unique solution is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Eigenvalues and Eigenvectors

- Let A be an $n \times n$ matrix, and $\lambda \in \mathbb{R}$.
- If the following system has nontrivial solutions:

  $$Ax = \lambda x$$

  we say that λ is an eigenvalue of A.
- If λ is an eigenvalue, a vector satisfying $Ax = \lambda x$ is called eigenvector.
- The eigenvalue is a scale factor that changes the magnitude of thevector x.
- Think of the eigenvector x as a vector that has its direction unchanged by a linear transformation; rather it is only scaled (by the eigenvalue) when the linear transformation is applied to it.

Eigenvalues and Eigenvectors

- By definition, eigenvectors are nonzero, eigenvalues may be equal to zero.
- We do not ocnsider the zero vector to be an eigenvector: since $A0 = 0 = \lambda 0$ for every scalar λ. the associated eigenvalue would be undefined.
- The eigenvectors with eigenvalue λ, if any, are the nonzero solutions of the equation $Ax = \lambda x$. We can rewrite this equation as follows:

$$
\begin{aligned}
&Ax = \lambda x \\
&Ax - \lambda I_n x \\
&(A - \lambda I_n)x = 0
\end{aligned}
$$

- The eigenvectors of A with eigenvalue λ, if any, are the nontrivial solutions of the matrix equation $(A - \lambda I_n)x = 0$. If this equation has no nontrivial solutions, then λ is not an eigenvector of A.

Strategy for the eigenvalue problem

- (1) The eigenvalues are scalars such that

$$\det(A - \lambda I) = 0$$

- If the eigenvalues $\lambda_1, \ldots, \lambda_k$ are given by the previous step, then the eigenvectors solve the systems $(i = 1, \ldots, k)$

$$(A - \lambda_i)\,\mathbf{v}_i = \mathbf{0}$$

Matrix:

$$A = \begin{bmatrix} 5 & -4 \\ 8 & -7 \end{bmatrix}$$

Characteristic equation:

$$\begin{vmatrix} 5-\lambda & -4 \\ 8 & -7-\lambda \end{vmatrix} = 0 \quad \Longleftrightarrow \quad \lambda^2 + 2\lambda - 3 = 0$$

Eigenvalues:

$$\lambda_1 = -3, \quad \lambda_2 = 1$$

Computation of $A - \lambda_1 I$: We get

$$A + 3I = \begin{bmatrix} 8 & -4 \\ 8 & -4 \end{bmatrix}$$

Reduced row-echelon form:

$$A + 3I \sim \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

Eigenvectors for $\lambda_1 = -3$:

$$\{r\mathbf{v}_1; r \in \mathbb{R}\}, \quad \text{where} \quad \mathbf{v}_1 = (1, 2)$$

Eigenvectors for $\lambda_2 = 1$:

$$\{r\mathbf{v}_2; r \in \mathbb{R}\}, \quad \text{where} \quad \mathbf{v}_2 = (1, 1)$$

Quadratic Forms and Definite Matrices

- Many optimization problems involve double sums of the form

$$q = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij} \tag{A-109}$$

- This quadratic form can be written

$$q = \mathbf{x}' \mathbf{A} \mathbf{x}$$

  where $\mathbf{A}$ is a symmetric matrix.

- In general, q may be positive, negative, or zero; it depends on $\mathbf{A}$ and $\mathbf{x}$.

- There are some matrices, however, for which q will be positive regardless of $\mathbf{x}$, and others for which q will always be negative (or nonnegative or nonpositive).

- For a given matrix $\mathbf{A}$,
  1. If $\mathbf{x}'\mathbf{A}\mathbf{x} > (<)0$ for all nonzero $\mathbf{x}$, then $\mathbf{A}$ is positive (negative) definite.
  2. If $\mathbf{x}'\mathbf{A}\mathbf{x} \geq (\leq)0$ for all nonzero $\mathbf{x}$, then $\mathbf{A}$ is nonnegative definite or positive semidefinite (nonpositive definite).

Theorem

- Let $\boldsymbol{A}$ be a symmetric matrix. If all the eigenvalues of $\mathbf{A}$ are positive (negative), then $\mathbf{A}$ is positive definite (negative definite). If some of the eigenvalues are zero, then $\mathbf{A}$ is nonnegative (nonpositive) definite if the remainder are positive (negative). If $\mathbf{A}$ has both negative and positive eigenvalues, then $\mathbf{A}$ is indefinite.

Calculus

- A variable y is a function of another variable x written

$$y = f(x), \quad y = g(x), \quad y = y(x)$$

- In this relationship, y and x are sometimes labeled the dependent variable and the independent variable, respectively.
- We can find the rate of change at a point of a function by finding the derivative at that point.
- The rate of change tells us how the dependent variable (function output) changes for an infinitely small change in the independent variable (input).
- Notation: $f : X \to Y$ (function from domain set X to co-domain Y)

- Assuming that the function $f(x)$ is continuous and differentiable, we obtain the following derivatives:

$$f'(x) = \frac{dy}{dx}, f''(x) = \frac{d^2y}{dx^2},$$

and so on.

- A frequent use of the derivatives of $f(x)$ is in the Taylor series approximation. A Taylor series is a polynomial approximation to $f(x)$. Letting x^0 be an arbitrarily chosen expansion point

$$f(x) \approx f\left(x^0\right) + \sum_{i=1}^{P} \frac{1}{i!} \frac{d^i f\left(x^0\right)}{d\left(x^0\right)^i} \left(x - x^0\right)^i \tag{A-121}$$

The choice of P, the number of terms, is arbitrary; the more that are used, the more accurate the approximation will be.

- The approximation used most frequently in econometrics is the linear approximation,

$$f(x) \approx \alpha + \beta x, \tag{A-122}$$

where, by collecting terms in (A-121), $\alpha = \left[f\left(x^0\right) - f'\left(x^0\right)x^0\right]$ and $\beta = f'\left(x^0\right)$. The superscript " 0 " indicates that the function is evaluated at x^0.

- The quadratic approximation is

$$f(x) \approx \alpha + \beta x + \gamma x^2 \tag{A-123}$$

where $\alpha = \left[f^0 - f'^0 x^0 + \frac{1}{2} f''^0 \left(x^0\right)^2\right], \beta = \left[f'^0 - f''^0 x^0\right]$ and $\gamma = \frac{1}{2} f''^0$

- We can regard a function $y = f(x_1, x_2, \ldots, x_n)$ as a scalar-valued function of a vector; that is, $y = f(\mathbf{x})$. The vector of partial derivatives, or gradient vector, or simply gradient, is

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \cdots \\ \partial y/\partial x_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{bmatrix} \tag{A-124}$$

- The vector $\mathbf{g}(\mathbf{x})$ or $\mathbf{g}$ is used to represent the gradient. Notice that it is a column vector The shape of the derivative is determined by the denominator of the derivative.

- A second derivatives matrix or Hessian is computed as

$$\mathbf{H} = \begin{bmatrix} \partial^2 y/\partial x_1 \partial x_1 & \partial^2 y/\partial x_1 \partial x_2 & \cdots & \partial^2 y/\partial x_1 \partial x_n \\ \partial^2 y/\partial x_2 \partial x_1 & \partial^2 y/\partial x_2 \partial x_2 & \cdots & \partial^2 y/\partial x_2 \partial x_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial^2 y/\partial x_n \partial x_1 & \partial^2 y/\partial x_n \partial x_2 & \cdots & \partial^2 y/\partial x_n \partial x_n \end{bmatrix} = [f_{ij}] \qquad \text{(A-125)}$$

  In general, $\mathbf{H}$ is a square, symmetric matrix. (The symmetry is obtained for continuous and continuously differentiable functions from Young's theorem.). Each column of $\mathbf{H}$ is the derivative of $\mathbf{g}$ with respect to the corresponding variable in $\mathbf{x}'$.

Optimization

- Consider finding the x where $f(x)$ is maximized or minimized.
- Because $f'(x)$ is the slope of $f(x)$, either optimum must occur where $f'(x) = 0$.
- Otherwise, the function will be increasing or decreasing at x.
- This result implies the first-order or necessary condition for an optimum (maximum or minimum):

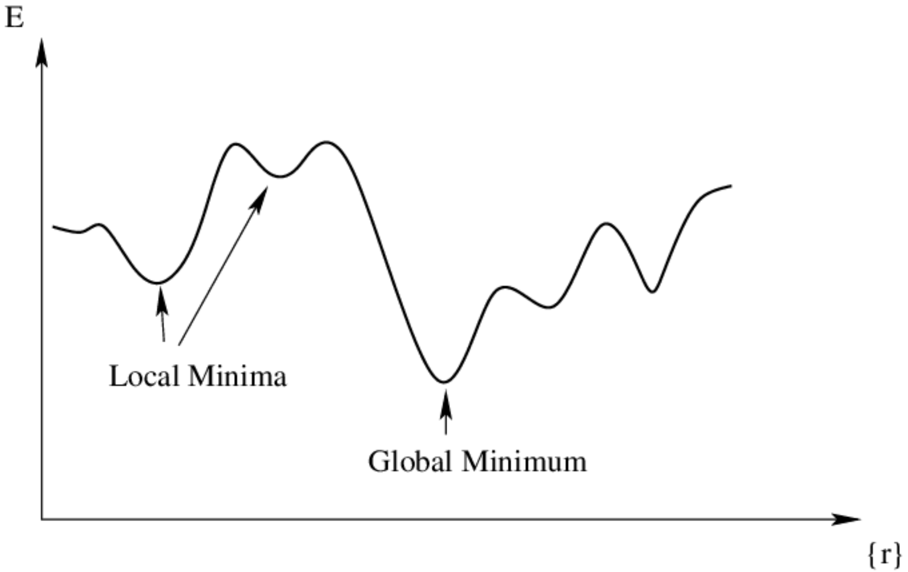
$$\frac{dy}{dx} = 0 \tag{A-134}$$

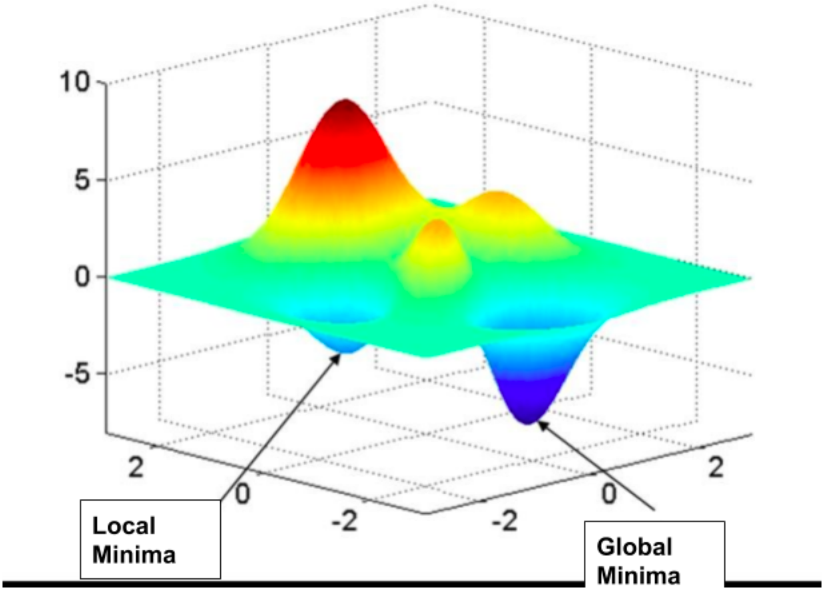
- For a maximum, the function must be concave; for a minimum, it must be convex. The sufficient condition for an optimum is.

$$\begin{aligned} &\text{For a maximum, } \frac{d^2y}{dx^2} < 0 \\ &\text{for a minimum, } \frac{d^2y}{dx^2} > 0 \end{aligned} \tag{A-135}$$

- Some functions have many local optima, that is, many minima and maxima.

- Certain functions, such as a quadratic, have only a single optimum. These functions are globally concave if the optimum is a maximum and globally convex if it is a minimum.


E
Local Minima
Global Minimum
{r}


10
5
0
-5
2
0
-2
-2
0
2
Local Minima
Global Minima

- For maximizing or minimizing a function of several variables, the first-order conditions are

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0} \tag{A-136}$$

- This result is interpreted in the same manner as the necessary condition in the univariate case.

- At the optimum, it must be true that no small change in any variable leads to an improvement in the function value.

- In the single-variable case, d^2y/dx^2 must be positive for a minimum and negative for a maximum.

- The second-order condition for an optimum in the multivariate case is that, at the optimizing value,

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \tag{A-137}$$

must be positive definite for a minimum and negative definite for a maximum.

- In a single-variable problem, the second-order condition can usually be verified by inspection.

- This situation will not generally be true in the multivariate case.

- For most of the problems encountered in econometrics, however, the second-order condition will be implied by the structure of the problem. That is, the matrix **H** will usually be of such a form that it is always definite.

Constrained Optimization

- We seek to

$$\begin{aligned}
&\text{maximize}_{\mathbf{x}}\, f(\mathbf{x}) \text{ subject to } c_1(\mathbf{x}) = 0\\
&c_2(\mathbf{x}) = 0\\
&\cdots\\
&c_J(\mathbf{x}) = 0
\end{aligned} \tag{A-140}$$

- The Lagrangean approach to this problem is to find the stationary points-that is, the points at which the derivatives are zero - of

$$L^*(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j c_j(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x}) \tag{A-141}$$

- The solutions satisfy the equations

$$\begin{aligned}\frac{\partial L^*}{\partial \mathbf{x}} &= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}(n \times 1), \\ \frac{\partial L^*}{\partial \boldsymbol{\lambda}} &= \mathbf{c}(\mathbf{x}) = \mathbf{0}(J \times 1)\end{aligned} \tag{A-142}$$

- The second term in $\partial L^*/\partial \mathbf{x}$ is

$$\frac{\partial \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{c}(\mathbf{x})'\boldsymbol{\lambda}}{\partial \mathbf{x}} = \left[\frac{\partial \mathbf{c}(\mathbf{x})'}{\partial \mathbf{x}}\right]\boldsymbol{\lambda} = \mathbf{C}'\boldsymbol{\lambda} \tag{A-143}$$

  where $\mathbf{C}$ is the matrix of derivatives of the constraints with respect to $\mathbf{x}$.

- The j th row of the $J \times n$ matrix $\mathbf{C}$ is the vector of derivatives of the j th constraint, $c_j(\mathbf{x})$, with respect to $\mathbf{x}'$.

- Upon collecting terms, the first-order conditions are

$$\begin{aligned}\frac{\partial L^*}{\partial \mathbf{x}} &= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{C}'\boldsymbol{\lambda} = \mathbf{0} \\ \frac{\partial L^*}{\partial \lambda} &= \mathbf{c}(\mathbf{x}) = \mathbf{0}\end{aligned} \tag{A-144}$$

- In the unconstrained solution, we have $\partial f(\mathbf{x})/\partial \mathbf{x} = \mathbf{0}$. From (A-144), we obtain, for a constrained solution,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{C}'\boldsymbol{\lambda} \tag{A-145}$$

which will not equal $\mathbf{0}$ unless $\boldsymbol{\lambda} = \mathbf{0}$.

This result has two important implications:

- The constrained solution cannot be superior to the unconstrained solution. This is implied by the nonzero gradient at the constrained solution. (That is, unless $\mathbf{C} = \mathbf{0}$ which could happen if the constraints were nonlinear. But, even if so, the solution is still not better than the unconstrained optimum.)

- If the Lagrange multipliers are zero, then the constrained solution will equal the unconstrained solution.