Homework Solutions

ME 555

Papalambros

Mathematics Review

This material reviews basic mathematical definitions and results that are used throughout the course. The reader is assumed to have some familiarity with the concepts presented; however, the intent is to provide a relatively self-contained mathematics review. The emphasis of this review is twofold: to provide a concise review of concepts from calculus, matrix algebra, and topology that are used throughout the course, and to familiarize the reader with various notations that will be encountered in the course notes.

1. Vectors

A vector x is an ordered collection of scalars x_i, where i=1,2,..,n. Here, n is the dimension of the vector **x**, and x_i is called the ith component or element of the vector **x**. A conventional notation for **x** is to represent it as a column vector. For example,

$$\mathbf{x} = \begin{pmatrix} x1 \\ x2 \\ \cdot \\ \cdot \\ \cdot \\ xn \end{pmatrix}$$

Often, it is useful to write **x** as a row vector:

$$\mathbf{x}^T = (x_1, x_2, \ldots, x_n)$$

Where $\mathbf{x}^T$ denotes the transpose of **x**. Roughly speaking, the transpose of a quantity can be obtained by interchanging the rows and columns of the quantity. We will revisit this idea when discussing matrices.

Two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are equal if their corresponding elements are equal, that is, if $x_i = y_i$ for all i.

Vector Addition and Multiplication by a scalar

Let x and y be two n-dimensional vectors. The sum of **x** and **y** is denoted by **x+y**, and the product of a vector **x** and a scalar t is denoted by **tx**, where

$$\mathbf{x} + \mathbf{y} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n)$$

and

$$t\mathbf{x} = (tx_1, tx_2, \ldots, tx_n)$$

Vector Scalar Product (Dot Product)

Let **x** and **y** be two n-dimensional vectors. The scalar product of **x** and **y** is denoted by $\mathbf{x}^T\mathbf{y}$ or $\mathbf{x}\cdot\mathbf{y}$, where

$$\mathbf{x}^T\mathbf{y} = \mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2+...+x_ny_n$$

If $\mathbf{x}^T\mathbf{y} = 0$, the vectors **x** and **y** are said to be <u>orthogonal</u>. Note that it is possible for the scalar product of two non-zero vectors to be zero. An example of this is the dot product of the vectors which make up the cartesian coordinate axes in the plane.

Euclidean Norm

The Euclidean norm of a vector x is denoted by $\|x\|$, where

$$\|x\| = (x_1^2 + x_2^2+...+x_n^2)^{1/2}$$

The Euclidean norm of a vector provides a measure of the "length" of the vector. Note that from the definition of the dot product, the norm can also be written as

$$\|x\|^2 = (x_1^2 + x_2^2+...+x_n^2) = \mathbf{x}^T\mathbf{x}$$

These definitions just formalize the familiar concept of distance determination in the plane.

2. Matrices

A matrix **B** can be expressed as a square array of scalar elements. Often, it is desirable to express a matrix as an ordered collection of vectors $\mathbf{b}_j$, where j=1,...m, and

$$\mathbf{b}_j = \begin{pmatrix} b_{1j} \\ b_{2j} \\ \cdot \\ \cdot \\ \cdot \\ b_{nj} \end{pmatrix}$$

The matrix **B** can then be expressed as

$$\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_m] = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix}$$

The size or dimension of the matrix **B** is n by m, or n x m. The element of the matrix **B** in the ith row and the jth column is denoted b_{ij}. We can also express the matrix **B** as the ordered collection of row vectors $\mathbf{b}_i^T$, where i=1,...,n, and $\mathbf{b}_i^T = (b_{i1}, b_{i2}, ..., b_{im})$. In this case, the matrix **B** becomes

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \dots \\ \mathbf{b}_m^T \end{bmatrix}$$

The transpose of a matrix is the matrix that results when the roles of the row and column indices are interchanged. The transpose **B** is denoted $\mathbf{B}^T$, and is defined by

$$(b^T)_{ij} = b_{ji} \qquad \text{for all i and j}$$

For example, if

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{then} \quad \mathbf{B}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Matrix Addition and Multiplication by a scalar.

Let **A** and **B** be two n x m matrices. The sum of **A** and **B** is denoted by **A** + **B**, and the product of a matrix **A** and a scalar t is denoted by t**A**, where

$$\mathbf{A+B} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2m}+b_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \dots & a_{nm}+b_{nm} \end{bmatrix}$$

and

$$t\mathbf{A} = \begin{bmatrix} ta_{11} & ta_{12} & \dots & ta_{1m} \\ ta_{21} & ta_{22} & \dots & ta_{2m} \\ \dots & \dots & \dots & \dots \\ ta_{n1} & ta_{n2} & \dots & ta_{nm} \end{bmatrix}$$

Matrix Product

Let **A** be a n x m matrix, **B** be a m x p matrix. The product of **A** and **B** is a n x p matrix **AB** whose ijth element is the scalar product of the ith row vector of **A** and the jth column vector of **B**. If $\mathbf{a}_i^T$ is the ith row vector of **A**, and $\mathbf{b}_j^T$ is the jth column vector of **B**, then the product of **A** and **B** is

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \dots \\ \mathbf{a}_m^T \end{bmatrix} (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p) = \begin{bmatrix} \mathbf{a}_1^T\mathbf{b}_1 & \mathbf{a}_1^T\mathbf{b}_2 & \dots & \mathbf{a}_1^T\mathbf{b}_p \\ \mathbf{a}_2^T\mathbf{b}_1 & \mathbf{a}_2^T\mathbf{b}_2 & \dots & \mathbf{a}_2^T\mathbf{b}_p \\ \dots & \dots & \dots & \dots \\ \mathbf{a}_n^T\mathbf{b}_1 & \mathbf{a}_n^T\mathbf{b}_2 & \dots & \mathbf{a}_n^T\mathbf{b}_p \end{bmatrix}$$

Notice that when the sizes of A and B are written side by side in the same order as the product, that is, (n x m) (m x p), the inner dimensions must be equal, and the outer dimensions give the size of the product AB. Symbolically,

(n x m) (m x p) = (n x p)

Matrix Inverse

The inverse of a matrix **A**, denoted A^{-1}, is the matrix which, when multiplied by A, yields the identity matrix, i.e.

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

The inverse is very useful when solving matrix equations. For example to solve the system

$$Ax = b$$

we can multiply both sides by A^{-1}, and find

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

Matrix Partition

On certain occasions it is useful to partition a matrix into several smaller matrices called "submatrices." For example, one possible way of partitioning a 3 x 4 matrix would be

$$A = \begin{bmatrix} a_{11} & \vdots & a_{12} & a_{13} & a_{14} \\ \cdots & \cdot\vdots & \cdots & \cdots & \cdots \\ a_{21} & \vdots & a_{22} & a_{23} & a_{24} \\ a_{31} & \vdots & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where

$$A_{12} = [a_{12}, a_{13}, a_{14}],\ A_{21} = \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix},\ A_{22} = \begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \end{bmatrix}$$

Rather than perform operations element by element on such a partitioned matrices, they can instead be done in terms of the submatrices, provided the partitioning are such that the operations are defined. For example, if B is a partitioned 4 x 1 matrix such that

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ B_2 \end{bmatrix}, \quad \text{(MR-21)}$$

then

$$AB = \begin{bmatrix} a_{11}b_1 + A_{12}B_2 \\ \cdots\cdots\cdots\cdots \\ A_{21}b_1 + A_{22}B_2 \end{bmatrix}$$

3. Vector Spaces

In this section, we consider any n-dimensional vector as a particular element of a vector space. The space itself, n-dimensional Euclidean space, is denoted by R^n.

Linear Combination, Linearly Dependent and Linearly Independent

Let $a_1, a_2, \ldots a_k$ be a set of vectors in R^n and $\lambda_1, \lambda_2, \ldots \lambda_k$ be a set of scalars. A linear combination of $a_1, a_2, \ldots a_k$ is a vector **b** such that

$$b = \lambda_1 a_1 + \lambda_2 a_2 + \ldots \lambda_n a_n$$

A set of vectors $a_1, a_2, \ldots a_k$ is said to be linearly dependent if there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$, not all zero, such that

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k = 0$$

If no such set of scalars exist, then the vectors are said to be linearly independent.

Span, Basis, Subspace, and Null Space

The collection of all linear combinations of $a_1, a_2, \ldots a_k$ is the set S which is spanned by the vectors. In other works, $\{a_1, a_2, \ldots a_k\}$ is the span of S. Notice that $a_1, a_2, \ldots a_k$ might not be linearly independent.

A subspace S of R^n is a subset that is closed under the operations of vector addition and scalar multiplication; that is, if **a** and **b** are vectors in S, then $\alpha 1 \mathbf{a} + \alpha 2 \mathbf{b}$ is also in S for any pair of scalars α_1, α_2.

A linearly independent set of vectors that span R^n is said to be a basis for R^n. Similarly, a linearly independent set of vectors that span a subspace S is said to be a basis for S

Let $a_1, a_2, \ldots a_k$ be the k basis for a subspace L in Rn. A set N is said to be the null space corresponding to $a_1, a_2, \ldots a_k$ if any vector in N is orthogonal to every vector in L, i.e.,

$$y^T x = 0$$

Note that N is also a subspace of R^n

Eigenvalues and Eigenvectors

Let A be an n x n square matrix. A scalar λ and a nonzero vector **x** satisfying the equation

$$Ax = \lambda x$$

are said to be, respectively, an eigenvalue and eigenvector of **A**. This equation can also be written as

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

which implies that the matrix $A - \lambda I$ is singular if λ is an eigenvalue.

4. Topological Concepts

A sequence of vectors $x_1, x_2, \ldots, x_k$, denoted $\{x_k\}$, said to converge to the limit x_* if

$$\| x_k - x_* \| \to 0 \text{ as } k \to \infty$$

that is, if given an $\varepsilon > 0$, there is an N such that for all $k \geq N$ implies $\| x_k - x_* \| < \varepsilon$. If $\{x_k\}$ converges to the limit x_*, we write $x_k \to x_*$.

A point x_* is a limit point of a sequence of $\{x_k\}$ if there is a subsequence of $\{x_k\}$ convergent to x_*. For example, let $x_k = (-1)^k \frac{k}{k+1}$. Then $\{1\}$ and $\{-1\}$ are both limit points.

A sequence $\{x_k\}$ is said to be bounded if there is a finite value M such that $\| x_k \| \leq M$ for all k. Similarly, a set of vectors S is said to be bounded if there is a finite value M such that $\| x \| \leq M$ for all $x \in S$.

Neighborhood, openness, closeness and compactness

A neighborhood around x^0 is defined as, for some $\varepsilon > 0$

$$N(x^0, \varepsilon) = \{x : \| x_k - x_* \| < \varepsilon\}$$

A subset S of R^n is said to be open if for any $x \in S$, there is an $\varepsilon > 0$ such that $N(x^0, \varepsilon)$ is contained entirely in S.

A subset S of R^n is said to be closed if for any convergent sequence $x_k \to x_*$ where $x_k \in S$, then $x_* \in S$.

A subset S of R^n is said to be compact if S is both closed and bounded.

Interior and Boundary

A point x is on the boundary of a set S in R^n if every ε neighborhood of x contains points in S and points not in S.

A point x is in the interior of a set S in R^n if there is an ε neighborhood of x contains entirely in S.

5.Functions

A real-valued function $f(\mathbf{x})$ defined on a subset S of R^n calculates a function value for each set of $\mathbf{x}$ which is an n-dimensional vector $(x_1, x_2, ..., x_n)^T$.

Continuity

A real-valued function $f(\mathbf{x})$ defined on a subset S of R^n is said to be continuous at $\mathbf{x}$ if $\mathbf{x}_k \rightarrow \mathbf{x}$ implies $f(\mathbf{x}_k) \rightarrow f(\mathbf{x})$. Equivalently, $f(\mathbf{x})$ is continuous at $\mathbf{x}$ if given an $\varepsilon > 0$, there is a $\mu > 0$ such that for all $\mathbf{y} \in S$ $\| \mathbf{y} - \mathbf{x} \| < \mu$ implies $| f(\mathbf{y}) - f(\mathbf{x}) | < \varepsilon$.

If $f(\mathbf{x})$ is continuous for every $\mathbf{x} \in S$, then $f(\mathbf{x})$ is said to be continuous on S.

Partial derivative, Gradient and Hessian

The partial derivative of $f(\mathbf{x})$ with respect to x_i at $\underline{x}$ is denoted by $\frac{\partial f(\underline{x})}{\partial x_i}$, and is defined as

$$\frac{\partial f(\underline{x})}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(\underline{x}_1, ..., \underline{x}_{i-1}, \underline{x}_i + h, \underline{x}_{i+1}, ..., \underline{x}_n) - f(\underline{x})}{h}$$

Note that each partial derivative may be regarded as a function of $\mathbf{x}$.

The gradient of $f(\mathbf{x})$ is the n-dimensional vector of those partial derivatives, and will be denoted by $\nabla f(\mathbf{x})$ where

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

If for all $\mathbf{x} \in S$, and all $i = 1, ..., n$, $\frac{\partial f(\mathbf{x})}{\partial x_i}$ exists and is continuous, then $f(\mathbf{x})$ is said to be differentiable in S, and the notation $f(\mathbf{x}) \in C^1$ on S is used.

The "second derivative" of a function $f(\mathbf{x})$ is defined by n^2 partial derivatives of the n first partial derivatives with respect to n elements of $\mathbf{x}$.

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right) \quad i = 1, ..., n; \ j = 1, ..., n$$

Conventional notations for the second partial derivative are:

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j},\ i \neq j;\ \frac{\partial^2 f(\mathbf{x})}{\partial x_i^2},\ i = j$$

The Hessian matrix of f(x) is denoted by **H(x)**, and is defined as

$$\mathbf{H(x)} = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

If $\frac{\partial f(\mathbf{x})}{\partial x_i}$, $\frac{\partial f(\mathbf{x})}{\partial x_j}$ and $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$ are continuous, then $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$ exists and $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$. This result also implies **H(x)** exists and is a square and symmetric matrix. Also, f(x) is said to be twice differentiable in S, and the notation $f(\mathbf{x}) \in C^2$ on S is used.

Taylor's Theorem

Suppose that $f(\mathbf{x}) \in C^1$ on an open set S. Let $\mathbf{x}_0$ and $\mathbf{x}_1$ be two points in S whose line segment is entirely contained in S. Then there is a point $\underline{\mathbf{x}}$ on the line segment such that

$$f(\mathbf{x}_1) = f(\mathbf{x}_0) + \nabla f(\underline{\mathbf{x}})^T(\mathbf{x}_1 - \mathbf{x}_0)$$

where $\underline{\mathbf{x}} = \mu \mathbf{x}_0 + (1-\mu)\, \mathbf{x}_1$, and $0 \leq \mu \leq 1$.
Furthermore, if $f(\mathbf{x}) \in C^2$, then there is a point $\underline{\mathbf{x}}$ such that

$$f(\mathbf{x}_1) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x}_1 - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}(\underline{\mathbf{x}})(\mathbf{x}_1 - \mathbf{x}_0)$$

where $\underline{\mathbf{x}} = \mu \mathbf{x}_0 + (1-\mu)\, \mathbf{x}_1$, and $0 \leq \mu \leq 1$.

These results are used often for approximating a function by giving the gradient or Hessian information at a particular point.