List of Concepts Chatper 5

Section 5.1: The scalar product in $\mathbb{R}^n$

- If $\mathbf{x}=[x_1,\ x_2,\ldots,x_n]^T$ and $\mathbf{y}=[y_1,\ y_2,\ldots,y_n]^T$, the ***scalar product*** of **x** and **y** (also called the *dot product* or *inner product*) is denoted by $\mathbf{x}\cdot\mathbf{y}$ and defined by $\mathbf{x}\cdot\mathbf{y}=\mathbf{x}^T\mathbf{y}=x_1 y_1+x_2 y_2+\ldots+x_n y_n$.
- If $\mathbf{x}=[x_1,\ x_2,\ldots,x_n]^T$, the ***norm*** of **x** or its euclidean *length* is defined as $\|\mathbf{x}\|=\sqrt{x_1^2+x_2^2+\ldots+x_n^2}$
- Theorem: $\mathbf{x}^T\mathbf{y}=\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ where θ is the smallest angle between the vectors **x** and **y**.
- The vectors **x** and **y** are ***orthogonal*** if their dot product is zero, i.e. $\mathbf{x}^T\mathbf{y}=0$
- The ***scalar orthogonal projection*** of the vector **x** onto the vector **y** is given by $\alpha=\dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$
- The ***vector orthogonal projection*** of the vector **x** onto the vector **y** is given by $\mathbf{p}=\alpha\dfrac{\mathbf{y}}{\|\mathbf{y}\|}=\dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}$
- The distance from the vector **y** to the line spanned by the vector **v** is given by $d=\|\mathbf{y}-\mathbf{p}\|$ where **p** is the orthogonal projection of **y** onto **v**.

Section 5.2. Orthogonal complements.

- Let Y be a subspace of $\mathbb{R}^n$. The set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in Y is denoted $Y^\perp$ and it is called the *orthogonal complement* of *Y.* The orthogonal complement is a subspace of $\mathbb{R}^n$ with the properties:
  - $Y\cap Y^\perp=\mathbf{0}$
  - $(Y^\perp)^\perp=Y$.
  - $\dim(Y)+\dim(Y^\perp)=n$

- **Fundamental Spaces Theorem**: Let **A** be an $m\times n$ matrix, then
  (a) $N(\mathbf{A})=R(\mathbf{A}^T)^\perp$ (i.e. all vectors in NullSpace of **A** are orthogonal to all vectors in the row space of **A**)
  (b) $N(\mathbf{A}^T)=R(\mathbf{A})^\perp$ (i.e. all vectors in NullSpace of $\mathbf{A}^T$ are orthogonal to all vectors in the column space of **A**)

- Let *S* be the subspace spanned by the vectors $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$. To find a basis for the orthogonal complement, $S^\perp$, we can create the matrix **A** with columns the vectors $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k$. Then S = R(**A**) and $R(\mathbf{A})^\perp=N(\mathbf{A}^T)$.
  Thus, all we have to do is find a basis for the nullspace of $\mathbf{A}^T$ by reducing $\mathbf{A}^T$ to row echelon form.

Section 5.3 – Least Squares

Least squares solutions to overdetermined systems:

Given an $m\times n$ system $A\mathbf{x}=\mathbf{b}$ with $m>n$ (more equations than unknowns), the system is inconsistent if the vector **b** is not in the column space of *A*.

For each $\mathbf{x}\in\mathbb{R}^n$ we can form a residual $r(\mathbf{x})=\mathbf{b}-A\mathbf{x}$. The distance between **b** and $A\mathbf{x}$ is given by $\|\mathbf{b}-A\mathbf{x}\|=\|\mathbf{r}(\mathbf{x})\|$ and we want to find **x** for which $\|r(\mathbf{x})\|$ is minimum.

- The vector $\hat{\mathbf{x}}$ that minimizes $\|r(\mathbf{x})\|$ is the solution to the ***Normal Equations***:
  $$A^T A\mathbf{x}=A^T\mathbf{b}$$
  If *A* has rank *n*, the normal equations have a unique solution which is called the ***least squares solution*** to the system $A\mathbf{x}=\mathbf{b}$.
- The ***projection vector*** $\mathbf{p}=A\hat{\mathbf{x}}$ is the element of R(*A*) (column space of *A*) that is closest to **b** in the least square sense, that is, such that $\|\mathbf{b}-\mathbf{p}\|$ is minimum.
- The ***projection matrix*** that projects any vector in $\mathbb{R}^m$ onto the column space of *A* is: $P=A(A^TA)^{-1}A^T$

Least square fit to data points

Suppose we are given a collection of data points of the form $\{(x_i, y_i)\}_{i=1}^{n}$ and we want to determine the polynomial of degree p that "best" fits the data.

If we suppose that

$$y_i = c_0 + c_1 x_i + c_2 x_i^2 + \ldots + c_p x_i^p, \qquad (1)$$

then we have to estimate the p unknowns $c_0, c_{1,} \ldots, c_p$.

The system (1) can be written in the following form:

$$X\mathbf{c} = \mathbf{y}, \qquad (2)$$

where $\mathbf{y} = [y_1,\ y_2, \ldots,\ y_n]^T$ is an n-vector of observations, $\mathbf{c} = [c_0,\ c_1, \ldots,\ c_p]^T$ is the p+1-vector of unknown parameters, and X is the $n \times (p+1)$ matrix:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ & & \vdots & & \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{bmatrix}$$

System (2) is usually overdetermined, thus we are looking for a least square solution which is given by the solution $\hat{\mathbf{c}}$ to the normal equations

$$X^T X\,\mathbf{c} = X^T \mathbf{y}$$

This solution $\hat{\mathbf{c}}$ minimizes the quantity $\|\mathbf{y} - X\mathbf{c}\|$ that is, minimizes the sum of squares of the differences between the "predicted" value of **y** ($= X\mathbf{c}$) from the "observed" value of **y**, where the sum is taken over every point in the data set.

REMARK: The above idea can be generalized to any function. For instance, if we want to find the trigonometric function of the form $f(t) = c_0 + c_1 \sin(t) + c_2 \cos(t)$ that best fits the data, then we can obtain the overdetermined system by requiring the function to go through the points. Once we have the system we can find the least squares solution by solving the corresponding normal equations.

Section 5.4 – Inner Product Spaces

An inner product on a vector space V is an operation on V that assigns to each pair of vectors **x** and **y** in V a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ such that:

I. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$

II. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all **x** and **y**.

III. $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$ for all **x**, **y**, and **z** in V and all scalars α and β.

A vector space with an inner product is called an *Inner Product space.*

Examples of inner product spaces:

- $\mathbb{R}^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$
- $\mathbb{R}^{m \times n}$: $\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$
- $C([a, b])$: $\langle f, g \rangle = \int_a^b f(x) g(x)\, dx$
- P_n: $\langle p, q \rangle = \sum_{i=1}^{n} p(x_i) q(x_i)$ where $x_1, x_2, \ldots, x_n$ are given.

Properties:

- Length or norm of **v**: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
- **u** and **v** are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$
- Pythagorean Law: If **u** and **v** are orthogonal then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.
- The <u>scalar projection</u> of **u** onto **v** is given by $\alpha = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|}$

  and the <u>vector projection</u> is given by $\mathbf{p} = \alpha \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2}\, \mathbf{v}$

- The angle between **u** and **v** is given by $\theta = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{v}\|\|\mathbf{u}\|}\right)$

A vector space V is said to be a *normed* linear space if to each vector **v** in V is associated a real number ||**v**||, called the *norm* of V, such that:
I. $\|\mathbf{v}\| \geq 0$ with equality if and only if **v** = **0**.
II. $\|\alpha \mathbf{v}\| \geq |\alpha|\|\mathbf{v}\|$ for any scalar α.
III. $\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all **v**, **w** in V.

Examples of norms in $\mathbb{R}^n$*:*

- $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ (this is the norm derived from the inner product)
- More generally: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$ with *p* any positive integer.
- $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$

Section 5.5 – Orthonormal Sets

- The set of non-zero vectors $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ form an *orthogonal* set if $\langle \mathbf{v_i}, \mathbf{v_j}\rangle = 0$ whenever $i \neq j$.

- Orthogonal sets of vectors are linearly independent.

- An ***orthonormal*** set of vectors is an orthogonal set of **unit** vectors.

- Let $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ be an orthonormal basis for an inner product space V. The following Theorem hold:

  Theorem 5.5.2:. If $\mathbf{v} = \sum_{i=i}^{n} c_i \mathbf{u_i}$ then $c_i = \langle \mathbf{v}, \mathbf{u_i}\rangle$.

  Corollary 5.5.3: If $\mathbf{u} = \sum_{i=i}^{n} a_i \mathbf{u_i}$ and $\mathbf{v} = \sum_{i=i}^{n} b_i \mathbf{u_i}$ then $\langle \mathbf{u}, \mathbf{v}\rangle = \sum_{i=1}^{n} a_i b_i$.

  Theorem 5.5.4 (Parseval's Formula): If $\mathbf{v} = \sum_{i=i}^{n} c_i \mathbf{u_i}$ then $\|\mathbf{v}\|^2 = \sum_{i=i}^{n} c_i^2$

- **Theorem 5.5.6. Orthogonality and Least Squares:** If the column vectors of A form an orthonormal set of vectors in $\mathbb{R}^m$, then $A^T A = I$ and the solution to the least squares problem $A\mathbf{x} = \mathbf{b}$ is $\hat{\mathbf{x}} = A^T \mathbf{b}$.
  The projection matrix is $P = A A^T$.

- **Theorem 5.5.7:** Let S be a nonzero subspace of an inner product space V and let $\mathbf{b} \in V$.
  If $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ is an orthonormal basis for S, then the projection **p** of **b** onto S is given by
  $$\mathbf{p} = \langle \mathbf{b}, \mathbf{u}_1\rangle \mathbf{u}_1 + \langle \mathbf{b}, \mathbf{u}_2\rangle \mathbf{u}_2 + \ldots + \langle \mathbf{b}, \mathbf{u}_k\rangle \mathbf{u}_k$$
  p is the element of S that is closest to **b,** that is ||**y** – **b**|| > || **p** – **b**|| for any **y** ≠ **p** in S.

Section 5.6 - The Gram – Schmidt orthogonalization process.

In order to transform an ordinary basis $\{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ into an orthonormal basis $\{\mathbf{q_1}, \mathbf{q_2}, \ldots, \mathbf{q_n}\}$, we can use the Gram-Schmidt orthogonalization process:
We begin by normalizing $\mathbf{x}_1$:

$$\mathbf{q}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

Let $\mathbf{p}_1$ be the projection of $\mathbf{x}_2$ onto span($\mathbf{x}_1$) = span($\mathbf{q}_1$): $\mathbf{p}_1 = \langle \mathbf{x}_2, \mathbf{u}_1\rangle \mathbf{u}_1$
then $\mathbf{x}_2 - \mathbf{p}_1$ is orthogonal to $\mathbf{q}_1$ and we can set

$$\mathbf{q}_2 = \frac{\mathbf{x}_2 - \mathbf{p}_1}{\|\mathbf{x}_2 - \mathbf{p}_1\|}$$

Let $\mathbf{p}_2$ be the projection of $\mathbf{x}_3$ onto span($\mathbf{x}_1$, $\mathbf{x}_2$) = span($\mathbf{q}_1$, $\mathbf{q}_2$): $\quad \mathbf{p}_2 = \langle \mathbf{x}_3, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{x}_3, \mathbf{q}_2 \rangle \mathbf{q}_2$
then $\mathbf{x}_3 - \mathbf{p}_2$ is orthogonal to $\mathbf{q}_2$ and $\mathbf{q}_3$ and we can set

$$\mathbf{q}_3 = \frac{\mathbf{x}_3 - \mathbf{p}_2}{\|\mathbf{x}_3 - \mathbf{p}_2\|}$$

we proceed this way until we have created n orthonormal vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$.

Gram – Schmidt and QR factorization:
If A is an $m \times n$ matrix of rank n, then A can be factored into a product QR, where Q is an $m \times n$ matrix with orthonormal column vectors and R is an upper triangular $n \times n$ matrix whose diagonal entries are all positive.
If $\mathrm{A} = [\mathbf{a}_1 \ \ \mathbf{a}_2 \ \ \ldots \ \ \mathbf{a}_n]$, the procedure for finding the QR factorization is similar to the Gram-Schmidt orthogonalization process. We just have to record the values of the intermediate steps:

$r_{11} = \|\mathbf{a}_1\|$ $\qquad \mathbf{q}_1 = \dfrac{\mathbf{a}_1}{r_{11}}$

$r_{12} = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle$ $\qquad \mathbf{p}_2 = r_{12}\,\mathbf{q}_1$ $\qquad r_{22} = \|\mathbf{a}_2 - \mathbf{p}_2\|$ $\qquad \mathbf{q}_2 = \dfrac{\mathbf{a}_2 - \mathbf{p}_2}{r_{22}}$

$r_{13} = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle$ $\qquad r_{23} = \langle \mathbf{a}_3, \mathbf{q}_2 \rangle$ $\qquad \mathbf{p}_3 = r_{13}\,\mathbf{q}_1 + r_{23}\,\mathbf{q}_2$ $\qquad \mathrm{r}_{33} = \|\mathbf{p}_3 - \mathbf{a}_3\|$ $\qquad \mathbf{q}_3 = \dfrac{\mathbf{a}_3 - \mathbf{p}_3}{r_{33}}$

and so on ....

<u>**Theorem 5.6.3**</u>: Let A be an $m \times n$ matrix of rank n and let A = QR. The Normal Equations for the system A$\mathbf{x}$ = $\mathbf{b}$ reduce to

$$R\,\mathbf{x} = Q^T\,\mathbf{b}.$$

The solution
$\hat{\mathbf{x}}$ may be obtained <u>by using back substitution.</u>

Note that using the QR factorization to find the least squares solution is computationally much more efficient and accurate than using the normal equations.