Rob 501 Handout: Grizzle
Orthogonal Projection and Normal Equations

Projection Theorem (Continued)

Orthogonal Projection Operator
Let $\mathcal{X}$ be a finite dimensional (real) inner product space and M a subspace of $\mathcal{X}$. For $x \in \mathcal{X}$ and $m_0 \in M$. The Projection Theorem shows the TFAE:

(a) $x - m_0 \perp M$.

(b) $\exists \tilde{m} = M^{\perp}$ such that $x = m_0 + \tilde{m}$.

(c) $\|x - m_0\| = d(x, M) = \inf_{m \in M} \|x - m\|$.

Def. P: $\mathcal{X} \to M$ by $P(x) = m_0$, where m_0 satisfies any of (a),(b) or (c), is called the orthogonal projection of $\mathcal{X}$ onto M.

Exercise1: P: $\mathcal{X} \to M$ is a linear operator.

Exercise2: P: Let $\{v^1, \cdots, v^k\}$ be an orthonormal basis for M.Then

$$P(x) = \sum_{i=1}^{k} \langle x, v^i \rangle v^i.$$

Normal Equations

Let $\mathcal{X}$ be a finite dimensional (real) inner product space and $M = \text{span}\{y^1, \cdots, y^k\}$, with $\{y^1, \cdots, y^k\}$ linearly independent. Given $x \in \mathcal{X}$, seek $\hat{x} \in M$ such that

$$\|x - \hat{x}\| = d(x, M) = \inf_{m \in M} \|x - m\| = \min_{m \in M} \|x - m\|$$

where we can write "min" because the Projection Theorem assures the existence of a minimizing vector $\hat{x} \in M$.

Notation: $\hat{x} = \text{argmin}\, d(x, M)$

Remark: One solution is Gram Schmidt and the orthogonal projection operator. We provide an alternative way to compute the answer.

By the Projection Theorem, $\hat{x}$ exists and is characterized by $x - \hat{x} \perp M$. Write

$$\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k$$

and impose $(x - \hat{x}) \perp M \Leftrightarrow (x - \hat{x}) \perp y^i,\ 1 \le i \le k$.

Then, $\langle x - \hat{x}, y^i \rangle = 0, \quad \forall 1 \le i \le k$ yields

$$\begin{aligned} &\langle \hat{x}, y^i \rangle = \langle x, y^i \rangle \quad i = 1, 2, \cdots, k \\ \Leftrightarrow &\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = \langle x, y^i \rangle \quad i = 1, 2, \cdots, k. \end{aligned}$$

We now write this out in matrix form.

$\underline{i = 1}$

$$\alpha_1 \langle y^1, y^1 \rangle + \alpha_2 \langle y^2, y^1 \rangle + \cdots + \alpha_k \langle y^k, y^1 \rangle = \langle x, y^1 \rangle$$

$\underline{i = 2}$

$$\alpha_1 \langle y^1, y^2 \rangle + \alpha_2 \langle y^2, y^2 \rangle + \cdots + \alpha_k \langle y^k, y^2 \rangle = \langle x, y^2 \rangle$$

$$\vdots$$

$\underline{i = k}$

$$\alpha_1 \langle y^1, y^k \rangle + \alpha_2 \langle y^2, y^k \rangle + \cdots + \alpha_k \langle y^k, y^k \rangle = \langle x, y^k \rangle$$

These are called the Normal Equations.

Def. $G = G(y^1, \cdots, y^k) := \begin{bmatrix} \langle y^1, y^1 \rangle & \langle y^1, y^2 \rangle & \cdots & \langle y^1, y^k \rangle \\ \langle y^2, y^1 \rangle & \langle y^2, y^2 \rangle & \cdots & \langle y^2, y^k \rangle \\ \vdots & \vdots & & \vdots \\ \langle y^k, y^1 \rangle & \langle y^k, y^2 \rangle & \cdots & \langle y^k, y^k \rangle \end{bmatrix}$

$G_{ij} = \langle y^i, y^j \rangle$ is called the Gram matrix.

Remark: Because we are assuming $\mathcal{F} = \mathbb{R}$, $\langle y^i, y^j \rangle = \langle y^j, y^i \rangle$, and we therefore have $G = G^T$.

Let $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}$, we have

$$G^T \alpha = \beta \text{ (normal equations in matrix form)}$$

where

$$\beta_i = \langle x, y^i \rangle, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

Def. $g(y^1, y^2, \cdots, y^k)$ =det $G(y^1, \cdots, y^k)$ is the determinant of the Gram Matrix.

Prop. $g(y^1, y^2, \cdots, y^k) \neq 0 \Leftrightarrow \{y^1, \cdots, y^k\}$ is linearly independent.

The proof is given at the end of the handout.

Summary: Here is the solution of our best approximation problem by the normal equations. Assume the set $\{y^1, \cdots, y^k\}$ is linearly independent and $M := \text{span}\{y^1, \cdots, y^k\}$. Then $\hat{x} = \arg\min d(x, M)$ if, and only if,

$$\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k$$

$$G^T \alpha = \beta$$

$$G_{ij} = \langle y^i, y^j \rangle$$

$$\beta_i = \langle x, y^i \rangle.$$

Application: Over determined system of linear equations in $\mathbb{R}^n$

$$A\alpha = b,$$

where $A = n \times m$ real matrix, $n \geq m$, rank$(A) = m$ (columns of A are linearly independent). From the dimension of A, we have that $\alpha \in \mathbb{R}^m, b \in \mathbb{R}^n$.

<u>Original Problem Formulation:</u>

Seek $\hat{\alpha}$ such that

$$\|A\hat{\alpha} - b\| = \min_{\alpha \in \mathbb{R}^m} \|A\alpha - b\|,$$

where

$$\|x\|^2 = \sum_{i=1}^{n} (x_i)^2.$$

Better Formulation via the Normal Equations

$$\mathcal{X} = \mathbb{R}^n, \quad \mathcal{F} = \mathbb{R}, \quad \langle x, y \rangle = x^T y = y^T x = \sum_{i=1}^{n} x_i y_i$$

Therefore,

$$\|x\|^2 = \langle x, x \rangle = \sum_{i=1}^{n} |x_i|^2.$$

Write

$$A = [A_1 | A_2 | \cdots | A_m] \text{ and } \alpha = [\alpha_1, \alpha_2, \cdots, \alpha_m]^\top$$

and we note that

$$A\alpha = \alpha_1 A_1 + \alpha_2 A_2 + \cdots \alpha_m A_m.$$

<u>New Problem Formulation:</u>
Seek

$$\hat{x} = A\hat{\alpha} \in \text{span}\{A_1, A_2, \cdots, A_m\} =: M$$

such that

$$\|\hat{x} - b\| = d(b, M) \Leftrightarrow \hat{x} - b \perp M.$$

From the Projection Theorem and the Normal Equations,

$$\hat{x} = \hat{\alpha}_1 A_1 + \hat{\alpha}_2 A_2 + \cdots \hat{\alpha}_m A_m$$

and $G^\top \hat{\alpha} = \beta$, with

$$G_{ij} = \langle A_i, A_j \rangle = A_i^\top A_j$$
$$\beta_i = \langle b, A_i \rangle = b^\top A_i = A_i^\top b.$$

<u>Aside</u>

$$A^\top = \begin{bmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_m^\top \end{bmatrix} \qquad A = [A_1 | \cdots | A_m]$$

$$(A^\top A)_{ij} = A_i^\top A_j$$
$$G = G^\top = A^\top A$$
$$(A^\top b)_i = A_i^\top b$$

Normal Equations are

$$A^\top A \hat{\alpha} = A^\top b.$$

From the Proposition, $G^\top = A^\top A$ is invertible $\Leftrightarrow$ columns of A are linearly independent. Hence,

$$\hat{\alpha} = (A^\top A)^{-1} A^\top b.$$

Prop. $g(y^1, y^2, \cdots, y^k) \neq 0 \Leftrightarrow \{y^1, \cdots, y^k\}$ is linearly independent.

Proof: $g(y^1, y^2, \cdots, y^k) = 0 \leftrightarrow \exists \alpha \neq 0$ such that $G^\top \alpha = 0$.

From our construction of the normal equations, $G^\top \alpha = 0$ if, and only if

$$\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = 0 \quad i = 1, 2, \cdots, k.$$

This is equivalent to

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \perp y^i = 0 \quad i = 1, 2, \cdots, k$$

which is equivalent to

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \perp \text{span}\{y^1, \cdots, y^k\} =: M$$

and thus

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \in M^\perp.$$

Because $\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k \in M$, we have that

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \in M \cap M^\perp$$

and therefore

$$\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k = 0.$$

By the linear independence of $\{y^1, \cdots, y^k\}$, we deduce that

$$\alpha_1 = \alpha_2 = \cdots \alpha_k = 0. \; \square$$