7.8 Best approximation and least squares

We wish to find the best approximation to the solutions.

Minimum distance problem

The Minimum Distance Problem in R^n Given a subspace W and a vector $\mathbf{b}$ in R^n, find a vector $\hat{\mathbf{w}}$ in W that is closest to $\mathbf{b}$ in the sense that $\|\mathbf{b} - \hat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$ for every vector $\mathbf{w}$ in W that is distinct from $\hat{\mathbf{w}}$. Such a vector $\hat{\mathbf{w}}$, if it exists, is called a ***best approximation to* b *from W*** (Figure 7.8.1).

- Answer:

Theorem 7.8.1 (***Best Approximation Theorem***) *If W is a subspace of R^n, and $\mathbf{b}$ is a point in R^n, then there is a unique best approximation to $\mathbf{b}$ from W, namely $\hat{\mathbf{w}} = \text{proj}_W \mathbf{b}$.*

- Distance from b to a subspace W.
- d=||b-proj_W(b)||=||proj_W^c(b)||.

Least square solutions of the linear system.

- Ax=b.
- If this is inconsistent, minimize ||b-Ax||.

Definition 7.8.2 If A is an $m \times n$ matrix and $\mathbf{b}$ is a vector in R^m, then a vector $\hat{\mathbf{x}}$ in R^n is called a ***best approximate solution*** or a ***least squares solution*** of $A\mathbf{x} = \mathbf{b}$ if

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \tag{5}$$

for all $\mathbf{x}$ in R^n. The vector $\mathbf{b} - A\hat{\mathbf{x}}$ is called the ***least squares error vector***, and the scalar $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the ***least squares error***.

Finding the least squares solutions of linear systems.

- Ax=b.
- Ax is in col(A).
- ||b-Ax|| is minimized when Ax=proj_col(A)b.
- This is consistent. Every system has the least squares solution.
- b-Ax=b-proj_col(A)b.
- A^T(b-Ax)=A^T(b-proj_col(A)b).
- proj_null(A^T)b=b-proj_col(A)b.
- Thus, A^T(b-Ax)=O or A^TAx=A^Tb.
- This is called the normal equation associated with Ax=b.

Theorem 7.8.3

(*a*) *The least squares solutions of a linear system* $A\mathbf{x} = \mathbf{b}$ *are the exact solutions of the normal equation*

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{11}$$

(*b*) *If A has full column rank, the normal equation has a unique solution, namely*

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b} \tag{12}$$

(*c*) *If A does not have full column rank, then the normal equation has infinitely many solutions, but there is a unique solution in the row space of A. Moreover, among all solutions of the normal equation, the solution in the row space of A has the smallest norm.*

- Proof: (a) done
- (b). Theorem 7.5.10 implies $A^T A$ is invertible.
- (c) omit.
- Example 3.

Error vector

- b=proj_col(A)b + proj_null(A)^T(b).
- b-Ax = (proj_col(A)b – Ax)+ proj_null(A^T)b.
- By (7) proj_col(A)b-Ax=0 if x is lss.
- b-Ax'=proj_null(A^T)b.
- Least squares error=||b-Ax'|| =||proj_null(A^T)b||.

Theorem 7.8.4 *A vector* $\hat{\mathbf{x}}$ *is a least squares solution of* $A\mathbf{x} = \mathbf{b}$ *if and only if the error vector* $\mathbf{b} - A\hat{\mathbf{x}}$ *is orthogonal to the column space of* A.

- Example 4.

Fitting a curve to experimental data

- Data (x_1,y_1), (x_2, y_2),…,(x_n, y_n).
- y=a+bx. Find the best a, b.
- If the line passes through the data, then
- Mv=y where

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, v = \begin{bmatrix} a \\ b \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- $M^T M v = M^t y$. Normal system
- If the x-coordinates are all distinct, then M has column rank 2 which is full,
- $v = (M^T M)^{-1} M^T y$.
- v gives us y=ax+b, the least squares line of best fit or regression line.
- What is minimizes is $S=(y_1-(a+bx_1))^2+\ldots+(y_n-(a+bx_n))^2$, the squares of residuals.
- Example 5.

Least squares by higher-degree polynomials

- Data (x_1,y_1), (x_2, y_2),…,(x_n, y_n).
- y=a_0+a_1x+…+a_mx^m. (m < n-1)
- Again, we can write this as:

$$Mv = y,$$

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}$$

- If $m \geq n-1$, then exact solutions exist. (Lagrange interpolation formula)
- If $m < n-1$, we need to find the best solution.
- $v=(M^T M)^{-1} M^T y$.
- Example 7 to find the gravitational constant. (Read yourselves.)