Linear Least-Squares Problem (LSP)
Method of Normal Equation

Rafikul Alam
Department of Mathematics
Indian Institute of Technology Guwahati
Guwahati - 781039, INDIA

Outline

- Least squares problem
- Method of Normal Equation

Least-squares problem (LSP)

Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Usually $m \gg n$. Find $x \in \mathbb{C}^n$ that minimizes

$$\|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big| \Big(\sum_{j=1}^{n} a_{ij}x_j - b_i \Big) \Big|^2 .$$

This is called least-squares problem because we minimize the sum of the squares of the errors

$$|r_1|^2 + \cdots + |r_m|^2 \text{ where } r := Ax - b.$$

Least-squares problem (LSP)

Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Usually $m \gg n$. Find $x \in \mathbb{C}^n$ that minimizes

$$\|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big| \Big(\sum_{j=1}^{n} a_{ij}x_j - b_i \Big) \Big|^2.$$

This is called least-squares problem because we minimize the sum of the squares of the errors

$$|r_1|^2 + \cdots + |r_m|^2 \text{ where } r := Ax - b.$$

The vector $r := Ax - b$ is called residual vector and $\|r\|_2$ is called residual error of the least squares problem. We write a solution x of the LSP as

$$x = \arg\min_{y \in \mathbb{C}^n} \|Ay - b\|_2.$$

Least-squares problem (LSP)

Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Usually $m \gg n$. Find $x \in \mathbb{C}^n$ that minimizes

$$\|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big| \Big(\sum_{j=1}^{n} a_{ij} x_j - b_i \Big) \Big|^2.$$

This is called least-squares problem because we minimize the sum of the squares of the errors

$$|r_1|^2 + \cdots + |r_m|^2 \text{ where } r := Ax - b.$$

The vector $r := Ax - b$ is called residual vector and $\|r\|_2$ is called residual error of the least squares problem. We write a solution x of the LSP as

$$x = \arg\min_{y \in \mathbb{C}^n} \|Ay - b\|_2.$$

=> x belongs to the set containing y's, s.t. norm (Ay - b) is minimum.

The LSP is called a linear least squares problem and is written as

$$\text{solve } Ax \approx b \text{ or } \text{LSP } Ax \approx b.$$

Remark: If x is a solution of the LSP $Ax \approx b$ then so is $x + z$ for any $z \in N(A)$.

Normal equation

If $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, then define

$$f(x) := \|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} x_j - b_i \Big)^2.$$

Then gradient $\nabla f(x) = 2A^\top (Ax - b)$ and Hessian $H_f(x) = A^\top A$. For a minimum $\nabla f(x) = 0$ yields the normal equation $A^\top A x = A^\top b$.

Normal equation

If $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then define

$$f(x) := \|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}x_j - b_i\Big)^2.$$

Then gradient $\nabla f(x) = 2A^\top(Ax - b)$ and Hessian $H_f(x) = A^\top A$. For a minimum $\nabla f(x) = 0$ yields the normal equation $A^\top Ax = A^\top b$. Since Hessian is symmetric positive semidefinite, f has a minimum at x.

Normal equation

If $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then define

$$f(x) := \|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}x_j - b_i \Big)^2.$$

Then gradient $\nabla f(x) = 2A^\top(Ax - b)$ and Hessian $H_f(x) = A^\top A$. For a minimum $\nabla f(x) = 0$ yields the normal equation $A^\top Ax = A^\top b$. Since Hessian is symmetric positive semidefinite, f has a minimum at x.

Note that $A^\top A$ is positive semidefinite and the normal equation

How?

$$A^\top Ax = A^\top b$$

is always consistent and has a solution. If $\mathrm{rank}(A) = n$ then $A^\top A$ is positive definite and $x = (A^\top A)^{-1}A^\top b = A^+ b$ is a unique solution of the LSP $Ax \approx b$.

Normal equation

If $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then define

$$f(x) := \|Ax - b\|_2^2 = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij}x_j - b_i\Big)^2.$$

Then gradient $\nabla f(x) = 2A^\top(Ax - b)$ and Hessian $H_f(x) = A^\top A$. For a minimum $\nabla f(x) = 0$ yields the normal equation $A^\top Ax = A^\top b$. Since Hessian is symmetric positive semidefinite, f has a minimum at x.

Note that $A^\top A$ is positive semidefinite and the normal equation

$$A^\top Ax = A^\top b$$

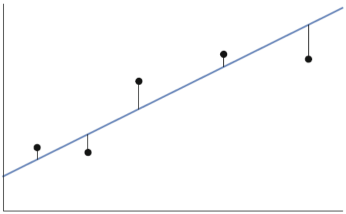
is always consistent and has a solution. If $\mathrm{rank}(A) = n$ then $A^\top A$ is positive definite and $x = (A^\top A)^{-1}A^\top b = A^+ b$ is a unique solution of the LSP $Ax \approx b$.

then A(T)*A is SPD

Remark: If $\mathrm{rank}(A) = n$ then $A^\top Ax = A^\top b$ can be solved by Cholesky factorization. However, $A^\top A$ may be highly ill-conditioned.

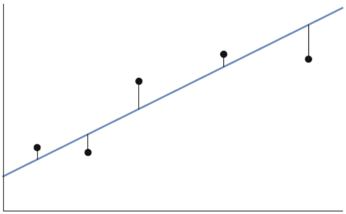
Linear regression

Given data points $(t_1, b_1), \ldots, (t_m, b_m)$ in $\mathbb{R}^2$, find a straight line $f(t) := x_1 + x_2 t$ that best fit the data. The task is to minimize the error $\sum_{j=1}^{m}(f(t_j) - b_j)^2$ for all $x_1, x_2 \in \mathbb{R}$.



Linear regression

Given data points $(t_1, b_1), \ldots, (t_m, b_m)$ in $\mathbb{R}^2$, find a straight line $f(t) := x_1 + x_2 t$ that best fit the data. The task is to minimize the error $\sum_{j=1}^{m}(f(t_j) - b_j)^2$ for all $x_1, x_2 \in \mathbb{R}$.



Setting $r_i := f(t_i) - b_i \implies f(t_i) = b_i + r_i \implies x_1 + x_2 t_i = b_i + r_i$ for $i = 1 : m$. This yields the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Linear regression

Now consider the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Linear regression

Now consider the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Set $\mu_t := (t_1 + \cdots + t_m)/m, \sigma_t^2 := (t_1^2 + \cdots + t_m^2)/m, \mu_b := (b_1 + \cdots + b_m)/m$ and $\sigma_{tb} := (t_1 b_1 + \cdots + t_m b_m)/m.$ Then the normal equation $A^\top Ax = A^\top b$ gives

Linear regression

Now consider the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Set $\mu_t := (t_1 + \cdots + t_m)/m, \sigma_t^2 := (t_1^2 + \cdots + t_m^2)/m, \mu_b := (b_1 + \cdots + b_m)/m$ and $\sigma_{tb} := (t_1 b_1 + \cdots + t_m b_m)/m.$ Then the normal equation $A^\top Ax = A^\top b$ gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} b \Longrightarrow \begin{bmatrix} 1 & \mu_t \\ \mu_t & \sigma_t^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_b \\ \sigma_{tb} \end{bmatrix}.$$

Linear regression

Now consider the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Set $\mu_t := (t_1 + \cdots + t_m)/m, \sigma_t^2 := (t_1^2 + \cdots + t_m^2)/m, \mu_b := (b_1 + \cdots + b_m)/m$ and $\sigma_{tb} := (t_1 b_1 + \cdots + t_m b_m)/m.$ Then the normal equation $A^\top Ax = A^\top b$ gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} b \Longrightarrow \begin{bmatrix} 1 & \mu_t \\ \mu_t & \sigma_t^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_b \\ \sigma_{tb} \end{bmatrix}.$$

Hence $x_1 = (\mu_b \sigma_t^2 - \mu_t \sigma_{tb})/(\sigma_t^2 - \mu_t^2)$ and $x_2 = (\sigma_{tb} - \mu_t \mu_b)/(\sigma_t^2 - \mu_t^2).$

Linear regression

Now consider the LSP

$$Ax = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = b.$$

Set $\mu_t := (t_1 + \cdots + t_m)/m, \sigma_t^2 := (t_1^2 + \cdots + t_m^2)/m, \mu_b := (b_1 + \cdots + b_m)/m$ and $\sigma_{tb} := (t_1 b_1 + \cdots + t_m b_m)/m$. Then the normal equation $A^\top Ax = A^\top b$ gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} b \Longrightarrow \begin{bmatrix} 1 & \mu_t \\ \mu_t & \sigma_t^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_b \\ \sigma_{tb} \end{bmatrix}.$$

Hence $x_1 = (\mu_b \sigma_t^2 - \mu_t \sigma_{tb})/(\sigma_t^2 - \mu_t^2)$ and $x_2 = (\sigma_{tb} - \mu_t \mu_b)/(\sigma_t^2 - \mu_t^2)$.

The best fit is given by the line $y = \beta(t - \mu_t) + \mu_b$, where $\beta = (\sigma_{tb} - \mu_t \mu_b)/(\sigma_t^2 - \mu_t^2)$.

Polynomial data fitting problem

For $(n-1)$ degree polynomial $p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ fitting the data $(t_1, b_1), \ldots, (t_m, b_m)$, we have $p(t_i) = b_i + r_i$ for $i = 1 : m$. This yields the LSP

Polynomial data fitting problem

For $(n-1)$ degree polynomial $p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ fitting the data $(t_1, b_1), \ldots, (t_m, b_m)$, we have $p(t_i) = b_i + r_i$ for $i = 1 : m$. This yields the LSP

$$\begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Polynomial data fitting problem

For $(n-1)$ degree polynomial $p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ fitting the data $(t_1, b_1), \ldots, (t_m, b_m)$, we have $p(t_i) = b_i + r_i$ for $i = 1 : m$. This yields the LSP

$$\begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

In practice, one considers $n = 2, 3, 4$ which correspond to straightline, quadratic and cubic polynomial fit, respectively.

Polynomial data fitting problem

For $(n-1)$ degree polynomial $p(t) = x_1 + x_2 t + \cdots + x_n t^{n-1}$ fitting the data $(t_1, b_1), \ldots, (t_m, b_m)$, we have $p(t_i) = b_i + r_i$ for $i = 1 : m$. This yields the LSP

$$\begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \approx \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

In practice, one considers $n = 2, 3, 4$ which correspond to straightline, quadratic and cubic polynomial fit, respectively.

The matrix in the LSP has full rank and is solved by normal equation method. However, for large n the matrix becomes highly ill-conditioned.

Example

Consider the data points

t	-1.0	-0.5	0.0	0.5	0.1
b	1.0	0.5	0.0	1.5	2.0

Example

Consider the data points

t	-1.0	-0.5	0.0	0.5	0.1
b	1.0	0.5	0.0	1.5	2.0

For the quadratic polynomial fit, we have the LSP

$$\begin{bmatrix} 1 & -1.0 & 1.0 \\ 1 & -0.5 & 0.25 \\ 1 & 0.0 & 0.0 \\ 1 & 0.5 & 0.25 \\ 1 & 0.1 & 1.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \approx \begin{bmatrix} 1.0 \\ 0.5 \\ 0.0 \\ 1.5 \\ 2.0 \end{bmatrix}.$$

Example

Consider the data points

t	-1.0	-0.5	0.0	0.5	0.1
b	1.0	0.5	0.0	1.5	2.0

For the quadratic polynomial fit, we have the LSP

$$\begin{bmatrix} 1 & -1.0 & 1.0 \\ 1 & -0.5 & 0.25 \\ 1 & 0.0 & 0.0 \\ 1 & 0.5 & 0.25 \\ 1 & 0.1 & 1.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \approx \begin{bmatrix} 1.0 \\ 0.5 \\ 0.0 \\ 1.5 \\ 2.0 \end{bmatrix}.$$

Solving the LSP we have $x = \begin{bmatrix} 0.086 & 0.40 & 1.4 \end{bmatrix}^\top$ which yields the polynomial $p(t) = 0.086 + 0.4t + 1.4t^2$.

Example

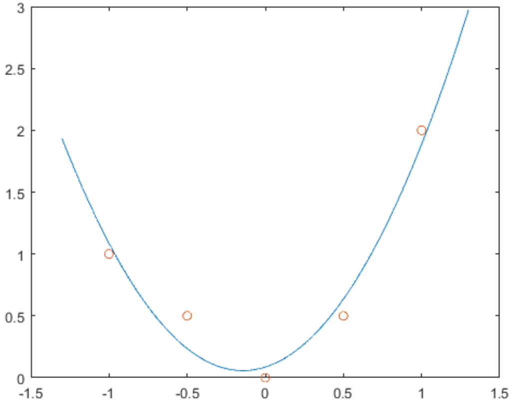


Figure : The plot of $p(t) = 0.086 + 0.4t + 1.4t^2$ and the data points.

Least squares data fitting problem

Task: Given data points $(\mathbf{t}_1, b_1), \ldots, (\mathbf{t}_m, b_m)$ in $\mathbb{C}^p \times \mathbb{C}$ and model functions $\phi_1, \ldots, \phi_n$, determine a function $f \in \mathrm{span}(\phi_1, \cdots, \phi_n) =: S$ that "best fit" the data:

Least squares data fitting problem

Task: Given data points $(\mathbf{t}_1, b_1), \ldots, (\mathbf{t}_m, b_m)$ in $\mathbb{C}^p \times \mathbb{C}$ and model functions $\phi_1, \ldots, \phi_n$, determine a function $f \in \mathrm{span}(\phi_1, \cdots, \phi_n) =: S$ that "best fit" the data:

$$f = \arg\min_{\phi \in S} \sum_{j=1}^{m} |\phi(\mathbf{t}_j) - b_j|^2.$$

Least squares data fitting problem

Task: Given data points $(\mathbf{t}_1, b_1), \ldots, (\mathbf{t}_m, b_m)$ in $\mathbb{C}^p \times \mathbb{C}$ and model functions $\phi_1, \ldots, \phi_n$, determine a function $f \in \mathrm{span}(\phi_1, \cdots, \phi_n) =: S$ that "best fit" the data:

$$f = \arg\min_{\phi \in S} \sum_{j=1}^{m} |\phi(\mathbf{t}_j) - b_j|^2.$$

We have $f = x_1\phi_1 + \cdots + x_n\phi_n$ for some $x := [x_1, \ldots, x_n]^\top \in \mathbb{C}^n$. Now, forcing f to pass through the data $(\mathbf{t}_i, b_i)$, we have $f(\mathbf{t}_i) = b_i + r_i$, where r_i is the error for $i = 1 : m$.

Least squares data fitting problem

Task: Given data points $(\mathbf{t}_1, b_1), \ldots, (\mathbf{t}_m, b_m)$ in $\mathbb{C}^p \times \mathbb{C}$ and model functions $\phi_1, \ldots, \phi_n$, determine a function $f \in \mathrm{span}(\phi_1, \cdots, \phi_n) =: S$ that "best fit" the data:

$$f = \arg\min_{\phi \in S} \sum_{j=1}^{m} |\phi(\mathbf{t}_j) - b_j|^2.$$

We have $f = x_1\phi_1 + \cdots + x_n\phi_n$ for some $x := [x_1, \ldots, x_n]^\top \in \mathbb{C}^n$. Now, forcing f to pass through the data $(\mathbf{t}_i, b_i)$, we have $f(\mathbf{t}_i) = b_i + r_i$, where r_i is the error for $i = 1 : m$.

Now $f(\mathbf{t}_i) = b_i + r_i \Longrightarrow x_1\phi_1(\mathbf{t}_i) + \cdots + x_n\phi_n(\mathbf{t}_i) = b_i + r_i$ for $i = 1 : m$. This yields

$$Ax = \begin{bmatrix} \phi_1(\mathbf{t}_1) & \cdots & \phi_n(\mathbf{t}_1) \\ \phi_1(\mathbf{t}_2) & \cdots & \phi_n(\mathbf{t}_2) \\ \vdots & \cdots & \vdots \\ \phi_1(\mathbf{t}_m) & \cdots & \phi_n(\mathbf{t}_m) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = b + r.$$

Least squares data fitting problem

Task: Given data points $(\mathbf{t}_1, b_1), \ldots, (\mathbf{t}_m, b_m)$ in $\mathbb{C}^p \times \mathbb{C}$ and model functions $\phi_1, \ldots, \phi_n$, determine a function $f \in \mathrm{span}(\phi_1, \cdots, \phi_n) =: S$ that "best fit" the data:

$$f = \arg\min_{\phi \in S} \sum_{j=1}^{m} |\phi(\mathbf{t}_j) - b_j|^2.$$

We have $f = x_1\phi_1 + \cdots + x_n\phi_n$ for some $x := [x_1, \ldots, x_n]^\top \in \mathbb{C}^n$. Now, forcing f to pass through the data $(\mathbf{t}_i, b_i)$, we have $f(\mathbf{t}_i) = b_i + r_i$, where r_i is the error for $i = 1 : m$.

Now $f(\mathbf{t}_i) = b_i + r_i \Longrightarrow x_1\phi_1(\mathbf{t}_i) + \cdots + x_n\phi_n(\mathbf{t}_i) = b_i + r_i$ for $i = 1 : m$. This yields

$$Ax = \begin{bmatrix} \phi_1(\mathbf{t}_1) & \cdots & \phi_n(\mathbf{t}_1) \\ \phi_1(\mathbf{t}_2) & \cdots & \phi_n(\mathbf{t}_2) \\ \vdots & \cdots & \vdots \\ \phi_1(\mathbf{t}_m) & \cdots & \phi_n(\mathbf{t}_m) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = b + r.$$

For the best fit, we have to choose $x \in \mathbb{C}^n$ for which $\|r\|_2 = \|Ax - b\|_2$ is minimized. This yields the LSP $Ax \approx b$.

Geometry of Least-squares problem

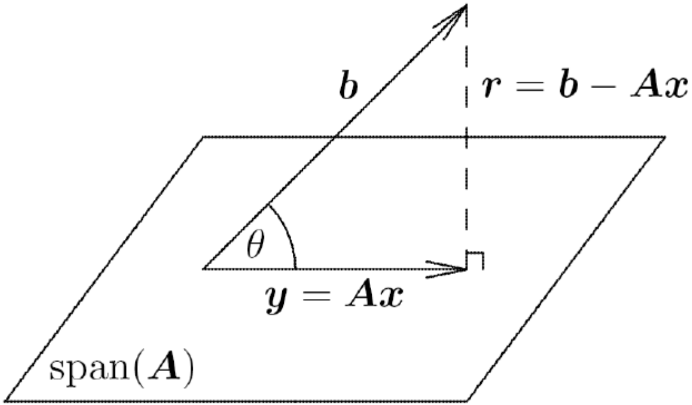


Figure : Relationships among b; r and $R(A)$:

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v,\ v_1 \rangle v_1 + \cdots + \langle v,\ v_n \rangle v_n$.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x, y \rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x, y \rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$.

Orthogonal direct sum: $\mathcal{M} \oplus \mathcal{N}$ and $\mathcal{M} \perp \mathcal{N}$.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v,\, v_1\rangle v_1 + \cdots + \langle v,\, v_n\rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x,\, y\rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$.

Orthogonal direct sum: $\mathcal{M} \oplus \mathcal{N}$ and $\mathcal{M} \perp \mathcal{N}$.

Let $S \subset \mathbb{C}^n$. Then $S^\perp := \{v \in \mathbb{C}^n : v \perp S\}$ is called the orthogonal complement of S. Note that $S^\perp$ is a subspace.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_n\rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x, y\rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$.

Orthogonal direct sum: $\mathcal{M} \oplus \mathcal{N}$ and $\mathcal{M} \perp \mathcal{N}$.

Let $S \subset \mathbb{C}^n$. Then $S^\perp := \{v \in \mathbb{C}^n : v \perp S\}$ is called the orthogonal complement of S. Note that $S^\perp$ is a subspace.

Fact: Let $\mathcal{M}$ is a subspace of $\mathbb{C}^n$. Then $\mathbb{C}^n = \mathcal{M} \oplus \mathcal{M}^\perp$.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v,\, v_1\rangle v_1 + \cdots + \langle v,\, v_n\rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x,\, y\rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$.

Orthogonal direct sum: $\mathcal{M} \oplus \mathcal{N}$ and $\mathcal{M} \perp \mathcal{N}$.

Let $S \subset \mathbb{C}^n$. Then $S^\perp := \{v \in \mathbb{C}^n : v \perp S\}$ is called the orthogonal complement of S. Note that $S^\perp$ is a subspace.

Fact: Let $\mathcal{M}$ is a subspace of $\mathbb{C}^n$. Then $\mathbb{C}^n = \mathcal{M} \oplus \mathcal{M}^\perp$. Thus if $w \in \mathbb{C}^n$ then there are unique vectors $u \in \mathcal{M}$ and $v \in \mathcal{M}^\perp$ such that $w = u + v$ and $u \perp v$.

Orthogonal direct sum

Fact: Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbb{C}^n$ and $v \in \mathbb{C}^n$. Then $v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n$.

Let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $\mathbb{C}^n$. Then $\mathcal{M}$ is said to be orthogonal to $\mathcal{N}$ and written as $\mathcal{M} \perp \mathcal{N}$ if $\langle x, y \rangle = 0$ for all $x \in \mathcal{M}$ and all $y \in \mathcal{N}$. https://www.math.arizona.edu/~rsims/ma528b/orthogonality.pdf

Orthogonal direct sum: $\mathcal{M} \oplus \mathcal{N}$ and $\mathcal{M} \perp \mathcal{N}$.

Let $S \subset \mathbb{C}^n$. Then $S^\perp := \{v \in \mathbb{C}^n : v \perp S\}$ is called the orthogonal complement of S. Note that $S^\perp$ is a subspace.

Fact: Let $\mathcal{M}$ is a subspace of $\mathbb{C}^n$. Then $\mathbb{C}^n = \mathcal{M} \oplus \mathcal{M}^\perp$. Thus if $w \in \mathbb{C}^n$ then there are unique vectors $u \in \mathcal{M}$ and $v \in \mathcal{M}^\perp$ such that $w = u + v$ and $u \perp v$.

Fact: Let $A \in \mathbb{C}^{m \times n}$. Then the adjoint $A^* \in \mathbb{C}^{n \times \mathrm{m}}$ is the unique matrix such that

$$\langle Ax, y \rangle = \langle x, A^* y \rangle \text{ for all } x \in \mathbb{C}^n \text{ and all } y \in \mathbb{C}^m.$$

Normal equation

Consider the range space $R(A) := \{Ax : x \in \mathbb{C}^n\} \subset \mathbb{C}^m$ and the null space $N(A) := \{x \in \mathbb{C}^n : Ax = 0\} \subset \mathbb{C}^n$ of $A \in \mathbb{C}^{m \times n}$. Then

$$\begin{aligned} \mathbb{C}^m &= R(A) \oplus N(A^*) \text{ and } R(A) \perp N(A^*) \\ \mathbb{C}^n &= N(A) \oplus R(A^*) \text{ and } N(A) \perp R(A^*). \end{aligned}$$

Normal equation

Consider the range space $R(A) := \{Ax : x \in \mathbb{C}^n\} \subset \mathbb{C}^m$ and the null space $N(A) := \{x \in \mathbb{C}^n : Ax = 0\} \subset \mathbb{C}^n$ of $A \in \mathbb{C}^{m\times n}$. Then

$$\begin{aligned} \mathbb{C}^m &= R(A) \oplus N(A^*) \text{ and } R(A) \perp N(A^*) \\ \mathbb{C}^n &= N(A) \oplus R(A^*) \text{ and } N(A) \perp R(A^*). \end{aligned}$$

Theorem: The LSP $Ax = b$ has a solution and

$$x = \operatorname*{argmin}_{y\in\mathbb{C}^n} \|Ay - b\|_2 \iff (Ax - b) \perp R(A) \iff A^*Ax = A^*b.$$

Normal equation

Consider the range space $R(A) := \{Ax : x \in \mathbb{C}^n\} \subset \mathbb{C}^m$ and the null space $N(A) := \{x \in \mathbb{C}^n : Ax = 0\} \subset \mathbb{C}^n$ of $A \in \mathbb{C}^{m\times n}$. Then

$$\begin{aligned} \mathbb{C}^m &= R(A) \oplus N(A^*) \text{ and } R(A) \perp N(A^*) \\ \mathbb{C}^n &= N(A) \oplus R(A^*) \text{ and } N(A) \perp R(A^*). \end{aligned}$$

Theorem: The LSP $Ax = b$ has a solution and

$$x = \operatorname*{argmin}_{y \in \mathbb{C}^n} \|Ay - b\|_2 \iff (Ax - b) \perp R(A) \iff A^*Ax = A^*b.$$

Proof: $(Ax - b) \perp R(A) \iff b - Ax \in N(A^*) \iff A^*(Ax - b) = 0 \iff A^*Ax = A^*b.$

Normal equation

Consider the range space $R(A) := \{Ax : x \in \mathbb{C}^n\} \subset \mathbb{C}^m$ and the null space $N(A) := \{x \in \mathbb{C}^n : Ax = 0\} \subset \mathbb{C}^n$ of $A \in \mathbb{C}^{m\times n}$. Then

$$\begin{aligned}\mathbb{C}^m &= R(A) \oplus N(A^*) \text{ and } R(A) \perp N(A^*)\\ \mathbb{C}^n &= N(A) \oplus R(A^*) \text{ and } N(A) \perp R(A^*).\end{aligned}$$

Theorem: The LSP $Ax = b$ has a solution and

$$x = \operatorname*{argmin}_{y\in\mathbb{C}^n} \|Ay - b\|_2 \iff (Ax - b) \perp R(A) \iff A^*Ax = A^*b.$$

Proof: $(Ax - b) \perp R(A) \iff b - Ax \in N(A^*) \iff A^*(Ax - b) = 0 \iff A^*Ax = A^*b.$

Let $b = b_1 + b_2$ with $b_1 \in R(A)$ and $b_2 \in N(A^*)$. Then

$$\begin{aligned}\|Ax - b\|_2^2 &= \|Ax - b_1 - b_2\|_2^2 = \|Ax - b_1\|_2^2 + \|b_2\|_2^2 = \|b_2\|_2^2\\ &\Leftrightarrow Ax = b_1 \iff b_2 = b - Ax \iff (Ax - b) \perp R(A). \blacksquare\end{aligned}$$

Normal equation

Consider the range space $R(A) := \{Ax : x \in \mathbb{C}^n\} \subset \mathbb{C}^m$ and the null space $N(A) := \{x \in \mathbb{C}^n : Ax = 0\} \subset \mathbb{C}^n$ of $A \in \mathbb{C}^{m \times n}$. Then

$$\begin{aligned} \mathbb{C}^m &= R(A) \oplus N(A^*) \text{ and } R(A) \perp N(A^*) \\ \mathbb{C}^n &= N(A) \oplus R(A^*) \text{ and } N(A) \perp R(A^*). \end{aligned}$$

Theorem: The LSP $Ax = b$ has a solution and

$$x = \underset{y \in \mathbb{C}^n}{\operatorname{argmin}} \|Ay - b\|_2 \iff (Ax - b) \perp R(A) \iff A^*Ax = A^*b.$$

Proof: $(Ax - b) \perp R(A) \iff b - Ax \in N(A^*) \iff A^*(Ax - b) = 0 \iff A^*Ax = A^*b.$

Let $b = b_1 + b_2$ with $b_1 \in R(A)$ and $b_2 \in N(A^*)$. Then

$$\begin{aligned} \|Ax - b\|_2^2 &= \|Ax - b_1 - b_2\|_2^2 = \|Ax - b_1\|_2^2 + \|b_2\|_2^2 = \|b_2\|_2^2 \\ &\Leftrightarrow \quad Ax = b_1 \iff b_2 = b - Ax \iff (Ax - b) \perp R(A). \blacksquare \end{aligned}$$

The system $A^*Ax = A^*b$ is called the normal equation for $Ax \approx b$.