Ch.12 Solving Linear Equations

12.1 Least-Squares Analysis

- Consider a system of linear equations

$$Ax = b$$

where $A \in \Re^{m\times n},\ b \in \Re^{m},\ m > n$, and rank $A = n$.

- If b does not belong to the range of A; that is, if b is not in $R(A)$, no solution exists.
- Our goal is to find the vector x minimizing $\|Ax - b\|^2$.
- Vector x^* is a least-squares solution to $Ax = b$.

$$\|Ax - b\|^2 \geq \|Ax^* - b\|^2$$

- **Lemma 12.1** *Let $A \in \Re^{m\times n}, m \geq n$. Then, rank $A = n$ iff rank $A^TA = n$(i.e., the square matrix A^TA is nonsingular).*
- **Theorem 12.1** *The unique vector x^* that minimizes $\|Ax - b\|^2$ is given by the solution to the equation $A^TAx = A^Tb$; that is, $x^* = (A^TA)^{-1}A^Tb$.*
  Proof. Let $x^* = (A^TA)^{-1}A^Tb$.

$$\begin{aligned}\|Ax - b\|^2 &= \|(A(x - x^*) + (Ax^* - b)\|^2 \\ &= (A(x - x^*) + (Ax^* - b))^T\end{aligned}$$

$$
\begin{aligned}
& \times(A(x-x^*)+(Ax^*-b)) \\
= \; & \|A(x-x^*)\|^2+\|Ax^*-b\|^2 \\
& +2[A(x-x^*)]^T(Ax^*-b).
\end{aligned}
$$

For the last term,

$$
\begin{aligned}
[A(x-x^*)]^T(Ax^*-b) \; = \; & (x-x^*)^T A^T \\
& \times[A(A^TA)^{-1}A^T-I_n]b=0
\end{aligned}
$$

Hence,

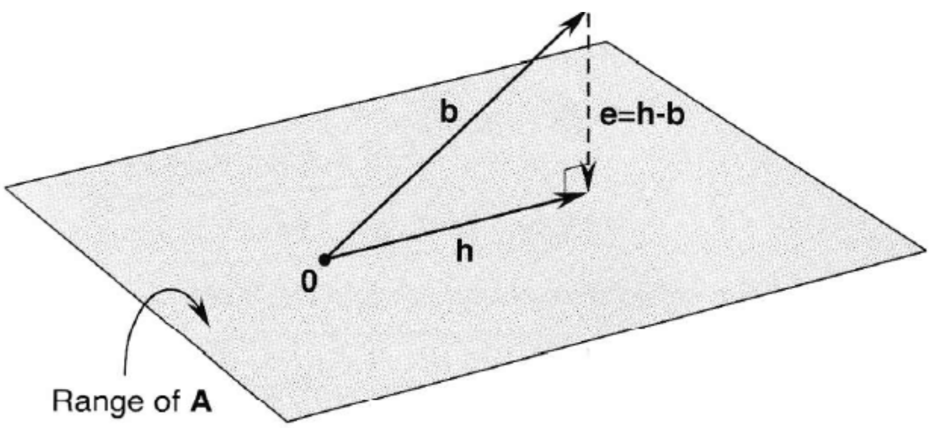
$$\|Ax-b\|^2=\|A(x-x^*)\|^2+\|Ax^*-b\|^2$$

If $x \neq x^*$, then $\|A(x-x^*)\|^2>0$,

$$\|Ax-b\|^2>\|Ax^*-b\|^2$$

Thus, $x^*=(A^TA)^{-1}A^Tb$ is the unique minimizer of $\|Ax-b\|^2$.

- **Proposition 12.1** *Let $h \in R(A)$ such that $h-b$ is orthogonal to $R(A)$. Then, $h=Ax^*=A(A^TA)^{-1}A^Tb$.*
  ♣ Fig. 12.1 on page 189.



- The vector $h \in R(A)$ minimizing $\|b - h\|$ is exactly the orthogonal projection of b onto $R(A)$. In other words, the vector x^* minimizing $\|Ax - b\|$ is exactly the vector that makes $Ax - b$ orthogonal to $R(A)$.
- Gram matrix(or Grammian)

$$A^T A = \begin{bmatrix} < a_1, a_1 > & \cdots & < a_n, a_1 > \\ \vdots & \vdots & \vdots \\ < a_1, a_n > & \cdots & < a_n, a_n > \end{bmatrix}$$

- An alternative method of arriving at the least-squares solution

$$\begin{aligned} f(x) &= \|Ax - b\|^2 \\ &= (Ax - b)^T (Ax - b) \\ &= \frac{1}{2} x^T (2A^T A) x - x^T (2A^T b) + b^T b. \\ \nabla f(x) &= 2A^T Ax - 2A^T b = 0 \\ x^* &= (A^T A)^{-1} A^T b \end{aligned}$$

- **Ex.12.2** Single input $t \in \Re$ and a single output $y \in \Re$ are given as follows.
  ♣ Table 12.1 on page 191.

i	0	1	2
t_i	2	3	4
y_i	3	4	15

Input labeled t_i and the output labeled y_i. We would like to find a straight line given by

$$y = mt + c$$

that fits the experimental data. There is no straight line that passes through all three points simultaneously. Therefore, we would like to find the values of m and c that best fit the data.
Three linear equations

$$\begin{aligned} 2m + c &= 3 \\ 3m + c &= 4 \\ 4m + c &= 15 \end{aligned}$$

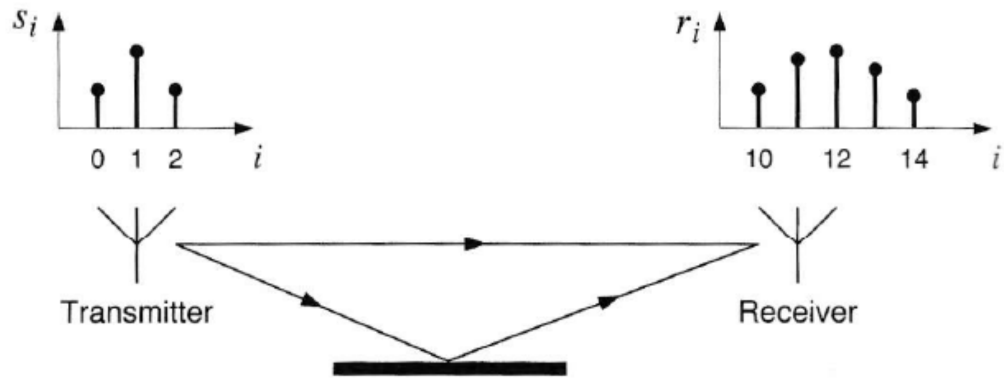
$$\begin{aligned} Ax &= b \\ A &= \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}. \quad b = \begin{bmatrix} 3 \\ 4 \\ 15 \end{bmatrix}, \quad x = \begin{bmatrix} m \\ c \end{bmatrix} \end{aligned}$$

Since rank $A <$ rank $[A,\ b]$, the vector b does not belong to the range of A. The solution to this least-squares problem is

$$x^* = \begin{bmatrix} m^* \\ c^* \end{bmatrix} = (A^T A)^{-1} A^T b = \begin{bmatrix} 6 \\ -32/3 \end{bmatrix}$$

Note that the error vector $e = Ax^* - b$ is orthogonal to each column of A.

- **Ex. 12.3** A wireless transmitter sends a discrete-time signal $\{s_0, s_1, s_2\}$ to a receiver.
  ♣ Fig. 12.3 on page 193.



The transmitted signal takes two paths to the receiver: a direct path, with delay 10 and attenuation factor a_1, and an indirect (reflected) path, with delay 12 and attenuation factor a_2.

s_0	s_1	s_2	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}
1	2	1	4	7	8	6	3

That is

$$\begin{aligned} s_0 a_1 &= r_{10} \quad s_1 a_1 = r_{11} \quad s_2 a_1 + s_0 a_2 = r_{12} \\ s_1 a_2 &= r_{13} \quad s_2 a_2 = r_{14} \end{aligned}$$

In matrix form,

$$A = \begin{bmatrix} s_0 & 0 \\ s_1 & 0 \\ s_2 & s_0 \\ 0 & s_1 \\ 0 & s_2 \end{bmatrix}, \quad x = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad b = \begin{bmatrix} r_{10} \\ r_{11} \\ r_{12} \\ r_{13} \\ r_{14} \end{bmatrix}$$

The least-squares estimate is given by

$$\begin{aligned} \begin{bmatrix} a_1^* \\ a_2^* \end{bmatrix} &= (A^T A)^{-1} A^T b \\ &= \frac{1}{35} \begin{bmatrix} 133 \\ 112 \end{bmatrix} \end{aligned}$$

12.2 Recursive Least-Squares Algorithm

- We are given experimental points (t_0, y_0), (t_1, y_1), and (t_2, y_2), and we find the parameters m^* and c^* of the straight line that best fits these data in the least-squares sense.
- We are now given an extra measurement point (t_3, y_3).
- We simply update our values of m^* and c^* to accommodate the new data point. $\Rightarrow$ Recursive least-squares (RLS) algorithm

- First consider the problem of minimizing $\|A_0 x - b^{(0)}\|^2$.
- The solution to this is given by $x^{(0)} = G_0^{-1} A_0^T b^{(0)}$, where $G_0 = A_0^T A_0$.
- Consider now the problem of minimizing

$$\left\| \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} x - \begin{bmatrix} b^{(0)} \\ b^{(1)} \end{bmatrix} \right\|^2$$

The solution is given by

$$\begin{aligned} x^{(1)} &= G_1^{-1} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}^T \begin{bmatrix} b^{(0)} \\ b^{(1)} \end{bmatrix} \\ G_1 &= \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}^T \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} \end{aligned}$$

Our goal is to write $x^{(1)}$ as a function of $x^{(0)}$, G_0, and the new data A_1 and $b^{(1)}$.

$$\begin{aligned} G_1 &= \begin{bmatrix} A_0^T & A_1^T \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} \\ &= A_0^T A_0 + A_1^T A_1 \\ &= G_0 + A_1^T A_1 \end{aligned}$$

$$\begin{aligned}
\begin{bmatrix} A_0 \\ A_1 \end{bmatrix}^T \begin{bmatrix} b^{(0)} \\ b^{(1)} \end{bmatrix} &= \begin{bmatrix} A_0^T & A_1^T \end{bmatrix} \begin{bmatrix} b^{(0)} \\ b^{(1)} \end{bmatrix} \\
&= A_0^T b^{(0)} + A_1^T b^{(1)}
\end{aligned}$$

We write $A_0^T b^{(0)}$ as

$$\begin{aligned}
A_0^T b^{(0)} &= G_0 G_0^{-1} A_0^T b^{(0)} \\
&= G_0 x^{(0)} \\
&= (G_1 - A_1^T A_1) x^{(0)} \\
&= G_1 x^{(0)} - A_1^T A_1 x^{(0)}
\end{aligned}$$

We can write $x^{(1)}$ as

$$\begin{aligned}
x^{(1)} &= G_1^{-1} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}^T \begin{bmatrix} b^{(0)} \\ b^{(1)} \end{bmatrix} \\
&= G_1^{-1}(G_1 x^{(0)} - A_1^T A_1 x^{(0)} + A_1^T b^{(1)}) \\
&= x^{(0)} + G_1^{-1} A_1^T (b^{(1)} - A_1 x^{(0)}) \\
G_1 &= G_0 + A_1^T A_1
\end{aligned}$$

- At the $(k+1)$st iteration

$$\begin{aligned}
G_{k+1} &= G_k + A_{k+1}^T A_{k+1} \\
x^{(k+1)} &= x^{(k)} + G_{k+1}^{-1} A_{k+1}^T (b^{(k+1)} - A_{k+1} x^{(k)})
\end{aligned}$$

- **Lemma 12.2** *Let A be a nonsingular matrix. Let U and V be matrices such that $I + VA^{-1}U$ is nonsingular.*

Then, $A+UV$ is nonsingular, and

$$(A+UV)^{-1} = A^{-1} - (A^{-1}U)(I+VA^{-1}U)^{-1}(VA^{-1})$$

- Using the result of the above lemma,

$$\begin{aligned} G_{k+1}^{-1} &= (G_k + A_{k+1}^T A_{k+1})^{-1} \\ &= G_k^{-1} - G_k^{-1} A_{k+1}^T (I + A_{k+1} G_k^{-1} A_{k+1}^T)^{-1} A_{k+1} G_k^{-1} \end{aligned}$$

- We rewrite G_K^{-1} as P_k.

$$\begin{aligned} P_{k+1} &= P_k - P_k A_{k+1}^T (I + A_{k+1} P_k A_{k+1}^T)^{-1} A_{k+1} P_k \\ x^{(k+1)} &= x^{(k)} + P_{k+1} A_{k+1}^T (b^{(k+1)} - A_{k+1} x^{(k)}) \end{aligned}$$

- In the special case where A_{k+1} is a matrix consisting of a single row, $A_{k+1} = a_{k+1}^T$,

$$\begin{aligned} P_{k+1} &= P_k - \frac{P_k a_{k+1} a_{k+1}^T P_k}{1 + a_{k+1}^T P_k a_{k+1}} \\ x^{(k+1)} &= x^{(k)} + P_{k+1} a_{k+1} (b_{k+1} - a_{k+1}^T x^{(k)}) \end{aligned}$$

- **Ex.12.6** Let

$$\begin{aligned} A_0 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \qquad b^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ A_1 &= a_1^T = [2 \;\; 1] \qquad b^{(1)} = b_1 = [3] \\ A_2 &= a_2^T = [3 \;\; 1] \qquad b^{(2)} = b_2 = [4] \end{aligned}$$

(1) First compute the vector $x^{(0)}$ minimizing $\|A_0 x - b^{(0)}\|^2$. (2) Use the RLS algorithm to find $x^{(2)}$ minimizing

$$\left\| \begin{bmatrix} A_0 \\ A_1 \\ A_2 \end{bmatrix} x - \begin{bmatrix} b^{(0)} \\ b^{(1)} \\ b^{(2)} \end{bmatrix} \right\|^2$$

$$\begin{aligned} P_0 &= (A_0^T A_0)^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \\ x^{(0)} &= P_0 A_0^T b^{(0)} = \begin{bmatrix} 2/3 \\ 2/3 \end{bmatrix} \end{aligned}$$

Applying the RLS algorithm twice,

$$\begin{aligned} P_1 &= P_0 - \frac{P_0 a_1 a_1^T P_0}{1 + a_1^T P_0 a_1} = \begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \\ x^{(1)} &= x^{(0)} + P_1 a_1 (b_1 - a_1^T x^{(0)}) = \begin{bmatrix} 1 \\ 2/3 \end{bmatrix} \\ P_2 &= P_1 - \frac{P_1 a_2 a_2^T P_1}{1 + a_2^T P_1 a_2} = \begin{bmatrix} 1/6 & -1/4 \\ -1/4 & 5/8 \end{bmatrix} \end{aligned}$$

$$x^{(2)} = x^{(1)} + P_2 a_2 (b_2 - a_2^T x^{(1)}) = \begin{bmatrix} 13/12 \\ 5/8 \end{bmatrix}$$

12.3 Solution to a Linear Equation with Minimum Norm

($Ax = b$ minimizing $\|x\|$)

- Consider a system of linear equation

$$Ax = b$$

where $A \in \Re^{m \times n}$, $b \in \Re^m$, $m \leq n$, and rank $A = m$.

- There may exist an infinite number of solutions to this system of equations. However, there is only one solution that is closest to the origin: the solution to $Ax = b$ whose norm $\|x\|$ is minimal.

- x^* is the solution to the problem

$$\begin{array}{rl} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- **Theorem 12.2** *The unique solution x^* to $Ax = b$ that minimizes the norm $\|x\|$ is given by*

$$x^* = A^T (AA^T)^{-1} b$$

Proof, Let $x^* = A^T (AA^T)^{-1} b$.

$$\|x\|^2 = \|(x - x^*) + x^*\|^2$$

$$
\begin{aligned}
&= ((x-x^*)+x^*)^T((x-x^*)+x^*) \\
&= \|x-x^*\|^2 + \|x^*\|^2 + 2x^{*T}(x-x^*)
\end{aligned}
$$

$$
\begin{aligned}
x^{*T}(x-x^*) &= [A^T(AA^T)^{-1}b]^T[x - A^T(AA^T)^{-1}b] \\
&= b^T(AA^T)^{-1}[Ax - (AA^T)(AA^T)^{-1}b] \\
&= b^T(AA^T)^{-1}[b-b] = 0 \\
\|x\|^2 &= \|x^*\|^2 + \|x-x^*\|^2 \\
\|x\|^2 &> \|x^*\|^2 \\
\|x\| &> \|x^*\|
\end{aligned}
$$

- **Ex. 12.7** Find the point closest to the origin of $\Re^3$ on the line of intersection of the two planes defined by the following two equations:

$$
\begin{aligned}
x_1 + 2x_2 - x_3 &= 1 \\
4x_1 + x_2 + 3x_3 &= 0
\end{aligned}
$$

$$
\begin{aligned}
\text{minimize} \quad & \|x\| \\
\text{subject to} \quad & Ax = b
\end{aligned}
$$

$$
A = \begin{bmatrix} 1 & 2 & -1 \\ 4 & 1 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
$$

$$
x^* = A^T(AA^T)^{-1}b = \begin{bmatrix} 0.09 \\ 0.33 \\ -0.23 \end{bmatrix}
$$

12.4 Kaczmarz's Algorithm

- Iterative algorithm for solving $Ax = b$
- $A \in \Re^{m\times n},\ b \in \Re^{m},\ m \le n$ and rank A=m, $0 < \mu < 2$.
- Kaczmarz's algorithm:
  1. Set $i := 0$, initial condition $x^{(0)}$.
  2. For $j = 1, \cdots, m$, set
     $x^{(im+j)} = x^{(im+j-1)} + \mu(b_j - a_j^T x^{(im+j-1)}) \frac{a_j}{a_j^T a_j}$
  3. Set $i := i + 1$; go to step 2.
- **Theorem 12.3** *In Kaczmarz's algorithm, if* $x^{(0)} = 0$*, then* $x^{(k)} \to x^* = A^T(AA^T)^{-1}b$ *as* $k \to \infty$*.*
- For the case where $x^{(0)} \neq 0$, Kaczmarz's algorithm converges to the unique point on $\{x : Ax = b\}$ minimizing the distance $\|x - x^{(0)}\|$.
- **Ex. 12.8**

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

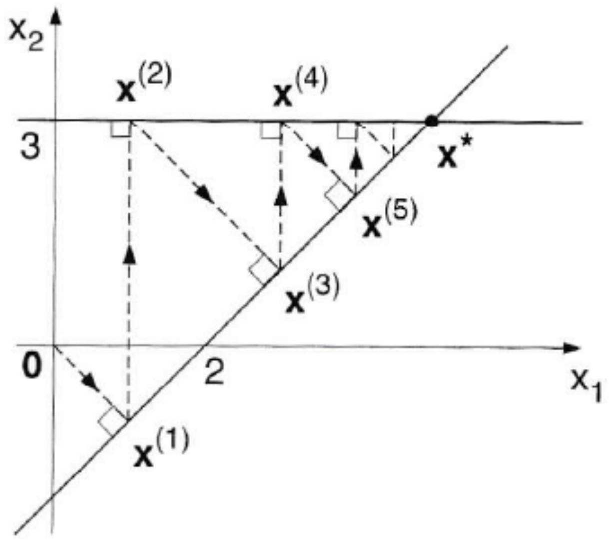
Begin with $\mu = 1$ and $x^{(0)} = 0$.

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + (2-0)\frac{1}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$x^{(2)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + (3-(-1)) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$x^{(3)} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} + (2-(-2))\frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

♣ Fig. 12.4 on page 205.



12.5 Solving Linear Equations ($Ax = b$) in General

- Solving a system of linear equations

$$Ax = b$$

where $A \in \Re^{m\times n}$, and rank A=r.

- **Lemma 12.3 Full-rank factorization** *Let* $A \in \Re^{m\times n}$, *rank* $A = r \le min(m, n)$. *Then, there exist* *matrices* $B \in \Re^{m\times r}$ *and* $C \in \Re^{r\times n}$ *such that*

$$\begin{aligned} A &= BC \\ rank\ A &= rank\ B = rank\ C = r \end{aligned}$$

- **Ex. 12.9**

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}$$

rank A=2.

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & 4 & 1 \end{bmatrix} = BC$$

- **Def. 12.1** *Given* $A \in \Re^{m\times n}$, *a matrix* $A^\dagger \in \Re^{n\times m}$ *is called a pseudoinverse of the matrix* A *if*

$$AA^\dagger A = A$$

and there exist matrices $U \in \Re^{n\times n}$, $V \in \Re^{m\times m}$ *such that*

$$A^\dagger = UA^T, \quad A^\dagger = A^T V$$

- For the case in which a matrix $A \in \Re^{m\times n}$ with $m \geq n$ and rank $A = n$,

$$A^\dagger = (A^T A)^{-1} A^T \quad \rightarrow \quad \text{left pseudoinverse of } A$$

Proof: $A(A^TA)^{-1}A^TA = A$.

- For the case in which a matrix $A \in \Re^{m \times n}$ with $m \leq n$ and rank $A = m$,

$$A^{\dagger} = A^T(AA^T)^{-1} \quad \rightarrow \quad \text{right pseudoinverse of } A$$

- **Theorem 12.5** *Let a matrix $A \in \Re^{m \times n}$ has a full-rank factorization $A = BC$, with rank A=rank B=rank C=r, $B \in \Re^{m \times r}$, $C \in \Re^{r \times n}$. Then,*

$$\begin{aligned} A^{\dagger} &= C^{\dagger}B^{\dagger} \\ B^{\dagger} &= (B^TB)^{-1}B^T \\ C^{\dagger} &= C^T(CC^T)^{-1} \end{aligned}$$

- **Theorem 12.6** *Consider a system of linear equations $Ax = b$, $A \in \Re^{m \times n}$, rank $A = r$. The vector $x^* = A^{\dagger}b$ minimizes $\|Ax - b\|^2$ on $\Re^n$. Furthermore, among all vectors in $\Re^n$ that minimize $\|Ax - b\|^2$, the vector $x^* = A^{\dagger}b$ is the unique vector with minimal norm.*