p. 472

SECTION VI E:

LEAST-SQUARES SOLUTIONS

Recall that the linear system

$$A\vec{x} = \vec{b}$$

is <u>consistent</u> if it has a solution.

This section will address

WHAT TO DO if a

linear system is inconsistent?

p. 473

Example 6.46

Consider the linear system

$$x_1 + x_2 = 1.01$$
$$2x_1 + 2x_2 = 2$$

$$\left(A \equiv \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}, \ \vec{x} \equiv \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \ \vec{b} \equiv \begin{bmatrix} 1.01 \\ 2 \end{bmatrix} \right).$$

This is an inconsistent linear system. Our human intelligences might speculate a small error in the appearance of "1.01" instead of "1", but none of this matters to computers that are fed linear systems, to

perform Gauss-Jordan on.
Mr. Computer will say "inconsistent" and nothing more. For those who have seen the movie "2001" (made in 1968) (see YouTube "2001 a Space Odyssey" or "HAL 9000: 'I'm sorry Dave, I'm afraid I can't do that'"), the computer, when given an inconsistent linear system, will tell you only "I'm sorry Dave, I'm afraid I can't solve that."

This is not a constructive response. Any scientist or engineer must deal with the inevitability of errors, in measurements or models; even if you are clean and sober, mistakes will occur. From the point of view of "consistent" versus "inconsistent," these mistakes, no matter how small, might be catastrophic. As in Example 6.46, for any $\epsilon > 0$ (ϵ stands for "error"),

$$x_1 + x_2 = (1 + \epsilon)$$

$$2x_1 + 2x_2 = 2$$

is inconsistent.

HOW TO RESPOND?

Do we give up, like the metallic intolerance of a computer?

BETTER!

Get an "approximate solution."

Notice that $A\vec{x} = \vec{b}$ if & only if $(A\vec{x} - \vec{b}) = \vec{0}$ if & only if $\|A\vec{x} - \vec{b}\| = 0$.

DEFINITION 6.47

p. 477

A (best) least-squares solution of

$$A\vec{x} = \vec{b}$$

is x^* that minimizes

$$\| A\vec{x} - \vec{b} \|;$$ that is

$$\| A x^* - \vec{b} \| \leq \| A\vec{x} - \vec{b} \|,$$

for all $\vec{x}$.

Before we characterize the least-squares solution (Theorem 6.55), we need a few more dot product ideas.

Recall that A^T means the transpose of a matrix A. The following result, whose proof is a tedious calculation, gives a better picture of transpose than we've seen up to now.

PROPOSITION 6.48

If A is an $(m \times n)$ matrix, then

$$\vec{x} \cdot (A^T \vec{y}) = (A\vec{x}) \cdot \vec{y}$$

for all $\vec{x}$ in $\mathbb{R}^n$, $\vec{y}$ in $\mathbb{R}^m$.

Here is an idea that opens up possibilities for decomposing vector spaces.

DEFINITION 6.49

If W is a subset of $\mathbb{R}^n$, then

$W^\perp \equiv$ the **orthogonal complement** of W

p. 480

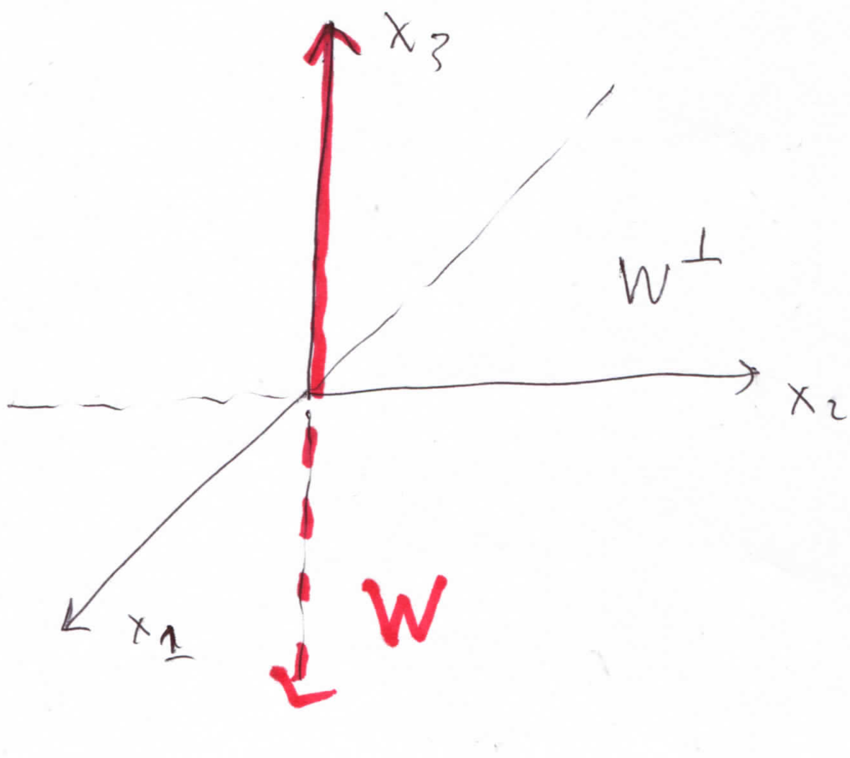
is $\{\vec{x} \text{ in } \mathbb{R}^n \mid \vec{x} \bullet \vec{w} = 0$
for all $\vec{w}$ in $W\}$

Example 6.50

If $W = \{(0, 0, x_3) \mid x_3 \text{ is real}\}$,
the x_3 axis in $\mathbb{R}^3$, then

$W^\perp = \{(x_1, x_2, 0) \mid x_1, x_2 \text{ is real}\}$,

the $x_1 x_2$ plane.



Notice how W and $W^\perp$ break up $\mathbb{R}^3$ into two orthogonal pieces

$$(x_1, x_2, x_3) = (0, 0, x_3) + (x_1, x_2, 0)$$

$$\qquad\qquad W \qquad\qquad\qquad W^\perp$$

that are easily taped back together to give $\mathbb{R}^3$.

In general, W in Definition 6.49 need not be a subspace; e.g.,

$$\{(0,0,1)\}^\perp = \{(x_1, x_2, 0) \mid x_1, x_2 \text{ real}\}.$$

In the following list of properties of orthogonal complements, keep in mind that

$$(\text{span of } W) = W$$

when W is a subspace.

Recall also that

$\mathcal{N}(A) \equiv$ <u>null</u> <u>space</u> of a matrix A

$\mathcal{R}(A) \equiv$ <u>range</u> <u>space</u> of A.

THEOREM 6.51

Suppose W is contained in $\mathbb{R}^n$, W is nonempty, and A is an $(m \times n)$ matrix.

(a) $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

(b) $(\text{span of } W)^{\perp} = W^{\perp}$.

(c) $(W^{\perp})^{\perp} = \text{span of } W$.

(d) $\vec{0}$ is the only vector in both $(\text{span of } W)$ and $W^{\perp}$.

(e) $\mathbb{R}^n = (\text{span of } W) \oplus W^\perp \equiv$

$\left\{ (\vec{x} + \vec{y}) \;\middle|\; \begin{array}{l} \vec{x} \text{ is in (span of } W), \\ \vec{y} \text{ is in } W^\perp \end{array} \right\}$

(f) $W = \{\vec{0}\}$ if and only if

$W^\perp = \mathbb{R}^n$

(g) $(\text{span of } W) = \mathbb{R}^n$ if and only if

$W^\perp = \{\vec{0}\}$

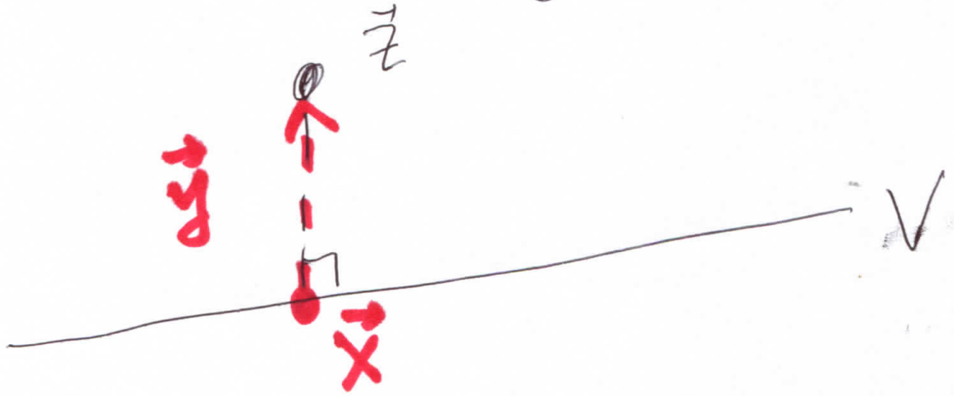
(h) $\left[\mathcal{N}(A)\right]^\perp = \mathcal{R}(A^T)$

(see pictures and discussion preceding 4.51)

Proof of more interesting parts:

(e) Let $V \equiv$ (span of W). For any $\vec{z}$ in $\mathbb{R}^n$, define

$\vec{x} \equiv P_V(\vec{z})$, $\vec{y} \equiv (\vec{z} - P_V(\vec{z}))$



(f) If $W = \{\vec{0}\}$, then, for any $\vec{x}$ in $\mathbb{R}^n$, $\vec{x} \cdot \vec{0} = 0$, thus $\vec{x}$ is in $W^\perp$. Conversely, if $W^\perp = \mathbb{R}^n$, then for any $\vec{w}$ in W,

$0 = \vec{w} \bullet \vec{w} = \|\vec{w}\|^2$, thus

$\vec{w} = \vec{0}$.

(g) If (span of W) $= \mathbb{R}^n$

and $\vec{x}$ is in $W^\perp$, then

$$\vec{x} = c_1 \vec{w}_1 + c_2 \vec{w}_2 + \dots$$

for some real $c_1, c_2, \dots$, thus

$\|\vec{x}\|^2 = \vec{x} \bullet \vec{x} = c_1(\vec{x} \bullet \vec{w}_1) + c_2(\vec{x} \bullet \vec{w}_2)$

$+ \dots = c_1(0) + c_2(0) + \dots = 0$,

thus $\vec{x} = \vec{0}$.

Conversely, if $W^\perp = \{\vec{0}\}$, then,

by (c) and (f),

(span of W) $= (W^\perp)^\perp = \{\vec{0}\}^\perp$

$= \mathbb{R}^n$.

(h) By Proposition 6.48 and (f) of this theorem,

$\vec{x}$ is in $[\mathcal{R}(A^T)]^{\perp}$ if and only if

$(A^T \vec{y}) \cdot \vec{x} = 0$ for all $\vec{y}$ in $\mathbb{R}^m$

if and only if

$\vec{y} \cdot (A\vec{x}) = 0$ for all $\vec{y}$ in $\mathbb{R}^m$

if and only if

$A\vec{x} = \vec{0}$ if and only if

$\vec{x}$ is in $\mathcal{N}(A)$.

Thus $\mathcal{N}(A) = [\mathcal{R}(A^T)]^{\perp}$; (c) of this theorem concludes the proof.

The square matrix (A^TA) will soon be seen to be of interest for getting least-squares solutions (see 6.54 and 6.55).

COROLLARY 6.52

For any matrix A, $\mathcal{N}(A^TA) = \mathcal{N}(A)$.

Proof: If $A\vec{x} = \vec{0}$, then $(A^TA)\vec{x} = A^T(A\vec{x}) = A^T\vec{0} = \vec{0}$. Thus $\mathcal{N}(A)$ is contained in $\mathcal{N}(A^TA)$.

Conversely, if $(A^T A)\vec{x} = \vec{0}$, then $A\vec{x}$ is in both $\mathcal{N}(A^T)$ and $\mathcal{R}(A)$, thus by (d) and (h) of Theorem 6.51, with A^T replacing A, we conclude that $A\vec{x} = \vec{0}$. Thus $\mathcal{N}(A^T A)$ is contained in $\mathcal{N}(A)$, so that $\mathcal{N}(A^T A) = \mathcal{N}(A)$.

COROLLARY 6.53

A is nonsingular if and only if $(A^T A)$ is invertible.

Proof: Recall that nonsingular means trivial nullspace. This corollary thus follows from Corollary 6.52 and Theorem 4.58.

6.54 NORMAL EQUATIONS

for $A\vec{x} = \vec{b}$ are

$$A^T A \vec{x} = A^T \vec{b}$$

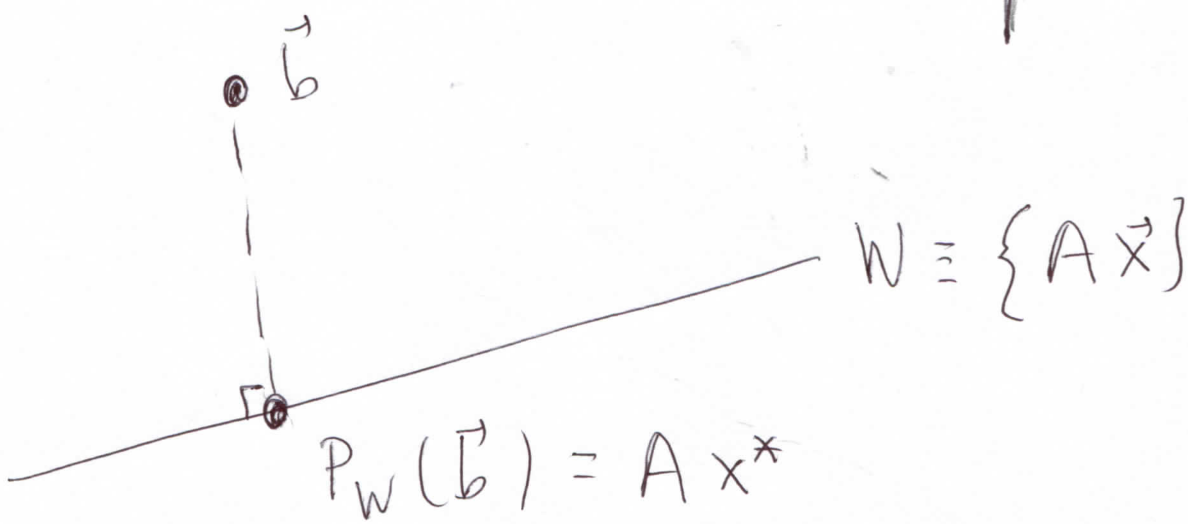
THEOREM 6.55

(1) Every linear system has a least-squares solution.

(2) x^* is a least-squares solution of $A\vec{x} = \vec{b}$
if and only if
x^* is a solution of the Normal Equations

$$A^T A \vec{x} = A^T \vec{b}$$

Proof: We are experts at minimizing distances to subspaces with projections (see Theorem 6.14). Our subspace of interest here is $W \equiv R(A) \equiv \{A\vec{x} \mid \vec{x} \text{ is in } \mathbb{R}^n\}$. ($n \equiv$ number of columns of A)



Since $P_W(\vec{b})$ is in W, it equals Ax^*, for some x^* in $\mathbb{R}^n$.

By Theorem 6.14,

$$\|Ax^* - \vec{b}\| \leq \|A\vec{x} - \vec{b}\|$$

for all $\vec{x}$ in $\mathbb{R}^n$, thus x^* is a least-squares solution of $A\vec{x} = \vec{b}$, proving (1).

For (2), note that

x^* is a least-squares solution

if and only if

$$(Ax^* - \vec{b}) \perp \mathcal{R}(A);$$

by Theorem 6.51 (h), this is

equivalent to

$(Ax^* - \vec{b})$ being in $\mathcal{N}(A^T)$;

that is,

$$A^T(Ax^* - \vec{b}) = \vec{0}, \text{ or}$$

$$A^TAx^* = A^T\vec{b},$$

a solution of the Normal

Equations.

COROLLARY 6.56

If A is nonsingular, then the least-squares solution of $A\vec{x} = \vec{b}$ is unique, given by

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

Example 6.57 Find the least-squares solution of the linear system in Example 6.46.

This is $A\vec{x} = \vec{b}$, with

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1.01 \\ 2 \end{bmatrix};$$

$$A^T = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix},$$

$$A^TA = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}, \quad A^T\vec{b} = \begin{bmatrix} 5.01 \\ 5.01 \end{bmatrix},$$

thus the Normal Equations are

$$5x_1 + 5x_2 = 5.01$$
$$5x_1 + 5x_2 = 5.01,$$

with solution

$$\boxed{x^* = \begin{bmatrix} 1.002 - x_2 \\ x_2 \end{bmatrix}, \quad x_2 \text{ arbitrary},}$$

also known as least-squares solutions of

$$x_1 + x_2 = 1.01$$
$$2x_1 + 2x_2 = 2$$

Note that x^* is not a solution of $A\vec{x} = \vec{b}$; it minimizes

$$\| A\vec{x} - \vec{b} \|,$$

but does not make it equal to 0 (that would be a solution).

DEFINITION 6.58

The **least-squares error** for $A\vec{x} = \vec{b}$ is

$$\| Ax^* - \vec{b} \|,$$

where x^* is the least-squares solution of $A\vec{x} = \vec{b}$.

The least-squares error is the smallest that

$$\| A\vec{x} - \vec{b} \|$$

can be; it measures how far from consistency $A\vec{x} = \vec{b}$ is.

Examples 6.59

(a) Find the least-squares error for the linear system in Example 6.46.

SOLUTION:

See Example 6.57.

$\| A x^* - \vec{b} \| =$

$$\left\| \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1.002 - x_2 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1.01 \\ 2 \end{bmatrix} \right\| =$$

$$\left\| \begin{bmatrix} 1.002 \\ 2.004 \end{bmatrix} - \begin{bmatrix} 1.01 \\ 2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -0.008 \\ 0.004 \end{bmatrix} \right\|$$

$$= 0.004 \left\| \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\| = \boxed{0.004\sqrt{5}}$$

Note that the least-squares error is independent of which least-squares solution x^* we choose.

(b) Get the least-squares solution and least-squares error of

$$x_1 + x_2 = 1$$
$$x_1 + x_2 = 0$$
$$2x_1 - x_2 = 2$$

SOLUTION: $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, thus

$A^T = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix}$, $A^T\vec{b} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$,

and $A^T A = \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix}$, so that the Normal Eq'ns are

$$\begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

or

$$\begin{aligned} 6x_1 \quad &= 5 \\ 3x_2 &= -1, \end{aligned}$$

easily solved:

$$x^* = \begin{bmatrix} 5/6 \\ -1/3 \end{bmatrix}$$

Notice that here we could have used Corollary 6.56:

$$x^* = \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ -1 \end{bmatrix}.$$

In (a), $A^T A = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$

is not invertible, so we couldn't have used Corollary 6.56.

For least-squares error in (b):

$$\|A x^* - \vec{b}\| = \left\| \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 5/6 \\ -1/3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right\|$$

$$= \frac{1}{6} \left\| \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \end{bmatrix} - \begin{bmatrix} 6 \\ 0 \\ 12 \end{bmatrix} \right\|$$

$$= \cdots \frac{1}{6} \left\| \begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}}$$

(c) I weigh myself,
in pounds:

My first measurement is 183. This doesn't fit my self-image, so I weigh again and get 180. Encouraged by the trend, I weigh again: another 183. This is not so encouraging, so I stop making measurements.

WHAT TO TELL

an authority figure hovering nearby?

The demand is for a single number that is allegedly my weight. This is not so clear from the data. Denoting by x my weight, my experiment produces the following linear system.

$$\begin{aligned} x &= 183 \\ x &= 180 \qquad (*) \\ x &= 183 \end{aligned}$$

This linear system has no solution. I conclude that I have no weight; weight is an illusion.

Authority figures hate that kind of answer.

Let's get a least-squares solution of (*). Here

$A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 183 \\ 180 \\ 183 \end{bmatrix}$, thus

$A^T = [1 \quad 1 \quad 1]$, $A^T A = [3]$

and $A^T \vec{b} = (183 + 180 + 183)$, so

the Normal Equations are

$$3x = (183 + 180 + 183)$$

$$\boxed{x^* = \frac{1}{3}(183 + 180 + 183) = 182}$$

There's a name for what we just did: the <u>average</u> or <u>mean</u> of the measurements.

Averaging is somewhat reflexive, but it wasn't always. In the early days (late 18^{th} century) of statistical inference, other summaries of contradictory data, such as the median, were initially used.

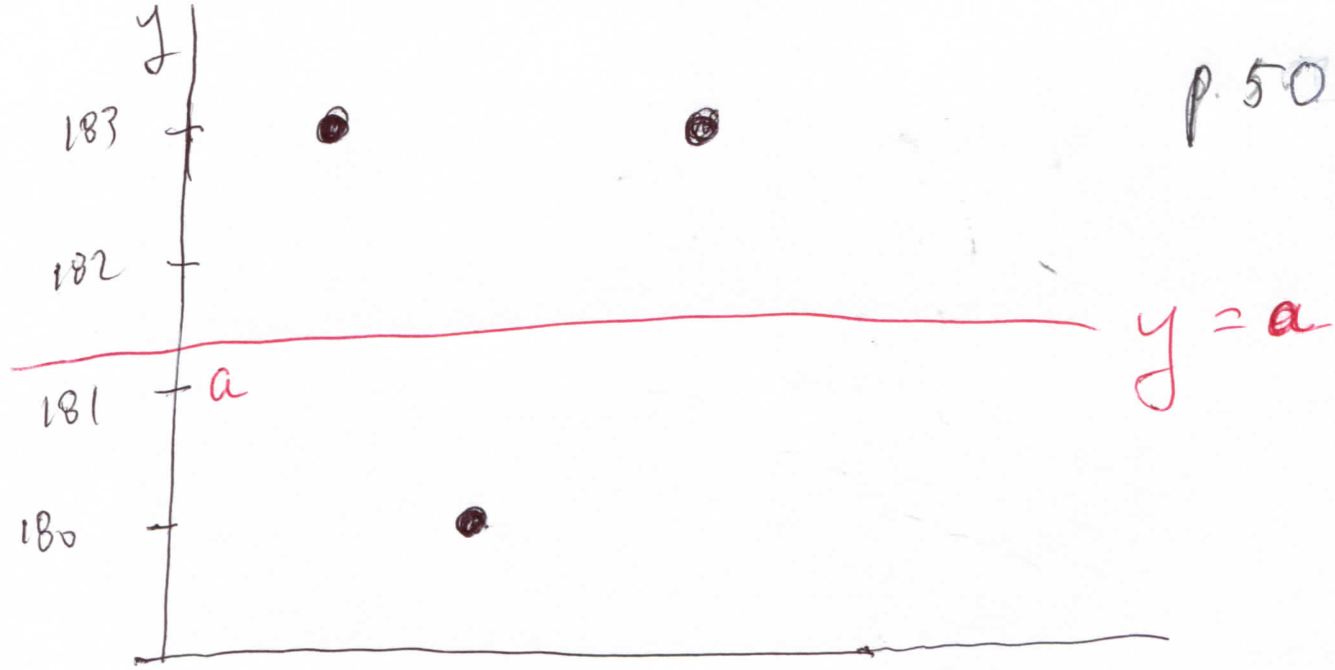
What turns out to make minimizing error possible and not so difficult is to look at squares of individual errors.

Getting back to (*):
we may think of our least-squares approach as the search for a horizontal line

$$y = a$$

that is closest to our data:

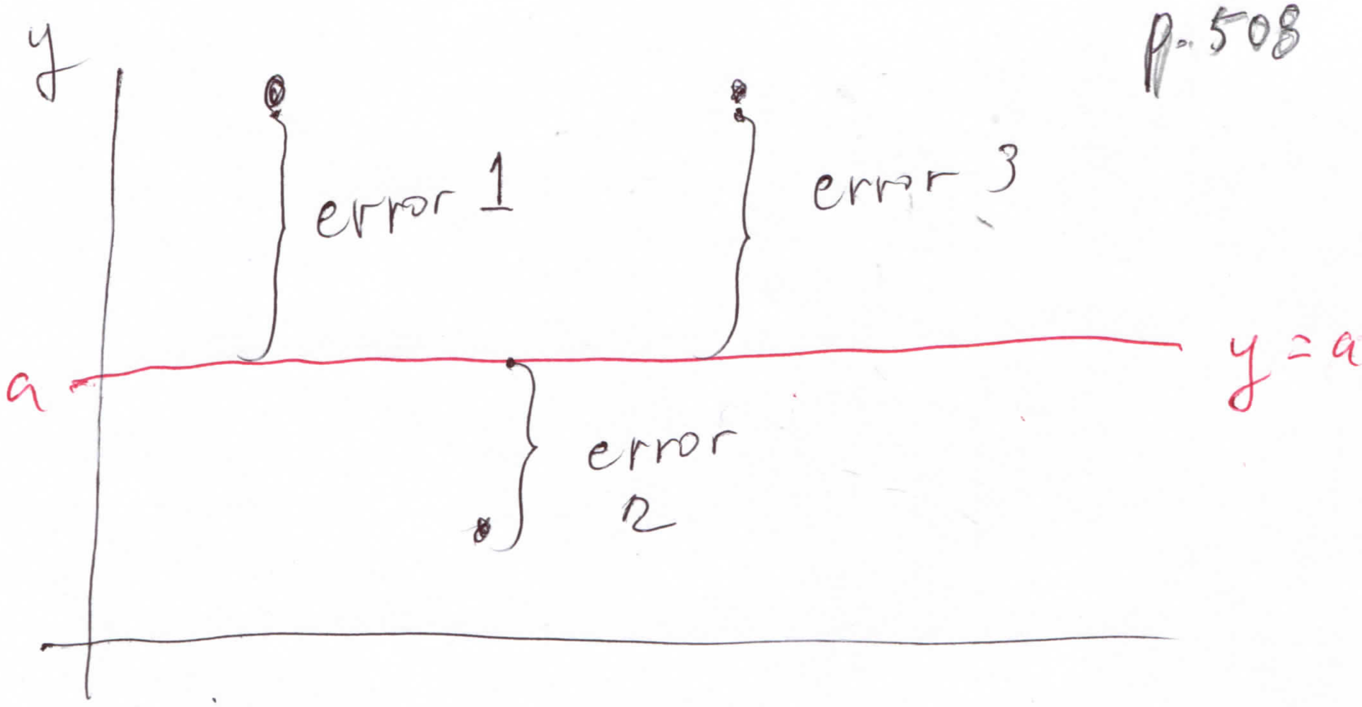
p. 507



Which choice of a number a is "closest" to the data?

Quotation marks are necessary, because we must define "distance from {ordered pairs} to a line."

We choose sum of squares of vertical errors:



That is, we are choosing a number a that minimizes

$$(183-a)^2+(180-a)^2+(183-a)^2;$$

Normal Equations gave us

$$a = 182$$

p. 509

BIVARIATE DATA 6.60

It is often the case that a pair of variables are related.

For example, if

$x \equiv$ temperature

and

$y \equiv$ pounds per hour of ice cream sold, an entrepreneur might expect a relationship between x and y; namely, that y increases as x increases.

To get a more specific relationship between x and y, we need to form ordered pairs $\{(x_1, y_1), (x_2, y_2), \ldots\}$ of measurements of x and y made simultaneously; this is what we mean by **bivariate data.**

p. 511

DEFINITION 6.61

A least-squares approximating line

to data

$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots)$

is $y = a + bx$ (a, b = real numbers)

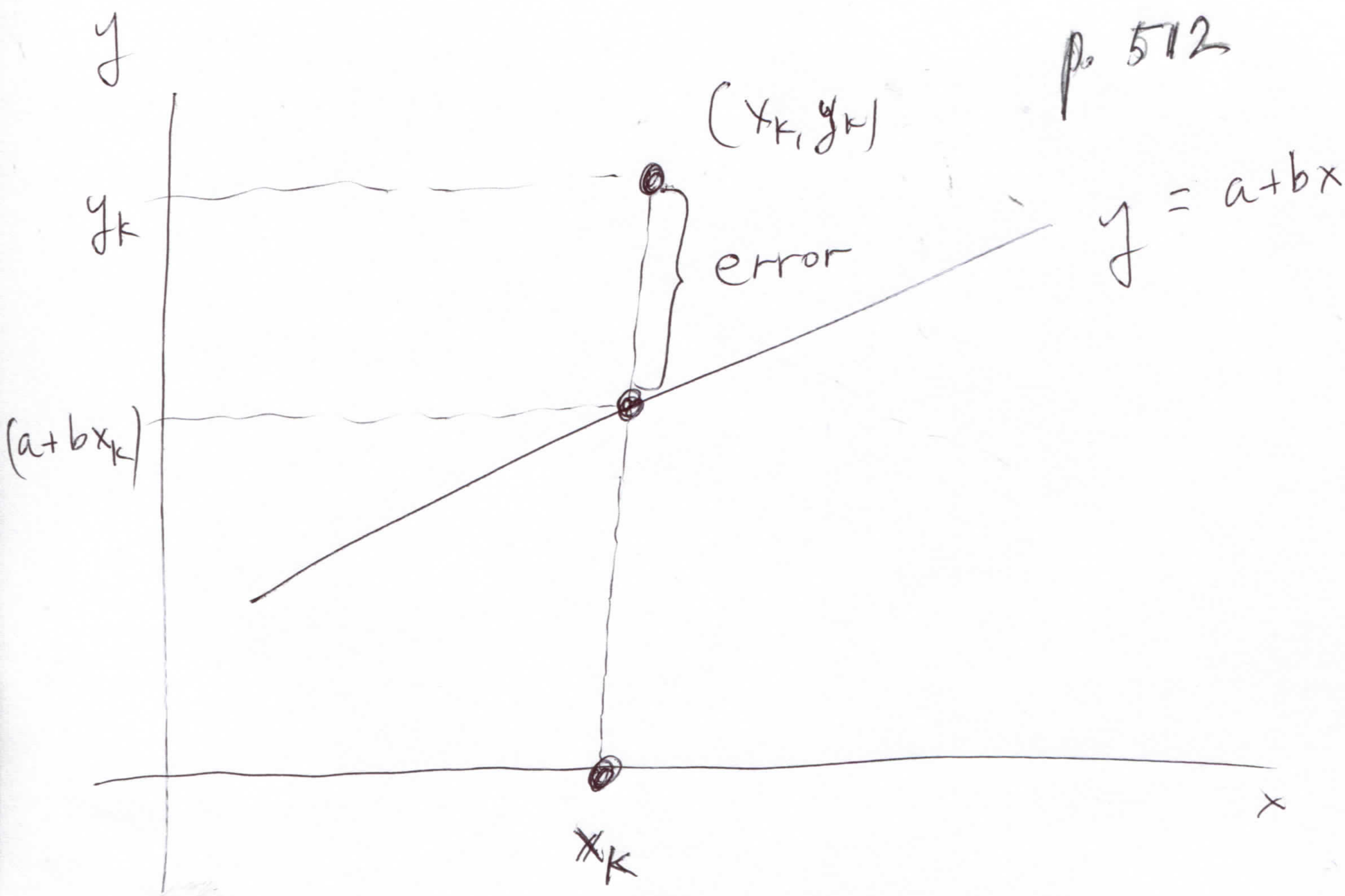
that minimizes

$[(a+bx_1 - y_1)^2 + (a+bx_2 - y_2)^2 +$

$(a+bx_3 - y_3)^2 + \ldots]$

~ sum of squares of vertical errors

p. 512



<u>DISCUSSION</u> <u>6.62</u>

If the line $y = a+bx$ goes through all the data points $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots\}$, then

$$a + bx_1 = y_1$$
$$a + bx_2 = y_2$$
$$a + bx_3 = y_3$$
$$\vdots \qquad \vdots$$

in matrix form,

$$(*) \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}$$

Notice now that what we are minimizing

$$\left[(a+bx_1-y_1)^2 + (a+bx_2-y_2)^2 + (a+bx_3-y_3)^2 + \cdots \right]$$

equals

$$\left\| \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix} \right\|^2 ;$$

thus a and b in our least-squares approximating line $y = a+bx$ are the least-squares solutions of (*)

THEOREM 6.63 p. 515

The numbers a and b in the least-squares approximating line $y = a + bx$ to some data are the least-squares solutions of the linear system one gets by requiring that the line go through all the data points.

Examples 6.64

In each part, find the least-squares approximating line to the data.

(a) Data

$\{(-1,3), (0,2), (1,0)\}$

$\boxed{\text{STEP ONE:}}$ Write down a perfect fit; that is, a linear system in a and b whose solution would produce a line $y = a + bx$ that goes through all the data points.

$$(x,y) = (-1,3) \longrightarrow a + b(-1) = 3$$

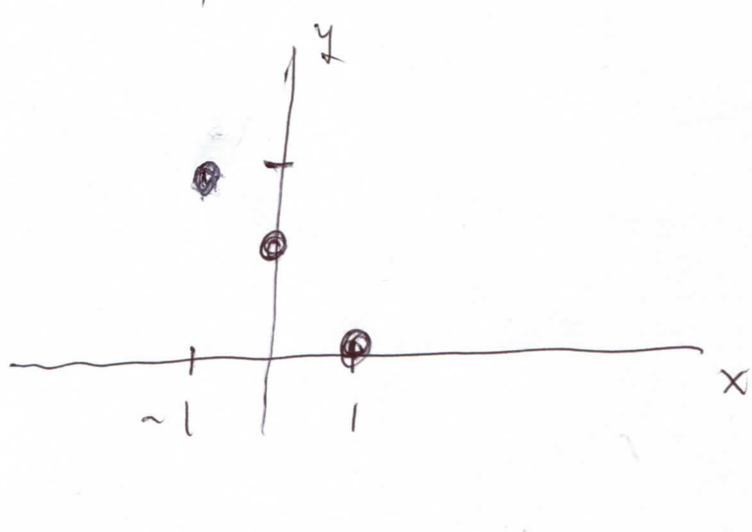
$$(x,y) = (0,2) \longrightarrow a + b(0) = 2$$

$$(x,y) = (1,0) \longrightarrow a + b(1) = 0$$

Linear system for perfect fit:

$$\begin{aligned} a - b &= 3 \\ a \phantom{{}+b} &= 2 \\ a + b &= 0 \end{aligned} \quad \rightarrow \quad \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}}_{\vec{b}}$$

(IF a solution (a,b) exists, then $y = a + bx$ goes through $(-1,3)$, $(0,2)$, $(1,0)$)



$\boxed{\text{STEP TWO:}}$ Get a least-squares solution of the linear system created by STEP ONE,

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

$$A^T A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad A^T \vec{b} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

NORMAL EQUATIONS:

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$\rightarrow x^* = \begin{bmatrix} 5/3 \\ -3/2 \end{bmatrix}; \text{ least-squares}$$

line is $\boxed{y = \frac{5}{3} - \frac{3}{2}x}$

(b) Data

$\{(-1,0), (0,2), (1,1), (2,0)\}$.

$\boxed{\text{STEP ONE:}}$ Perfect fit by

$y = a + bx$

$(x,y) = (-1,0) \longrightarrow a - b = 0$

$(x,y) = (0,2) \longrightarrow a = 2$

$(x,y) = (1,1) \longrightarrow a + b = 1$

$(x,y) = (2,0) \longrightarrow a + 2b = 0$

$$\longrightarrow \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}}_{\vec{b}}$$

p 520

$\boxed{\text{STEP TWO:}}$

Least-squares solution:

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \to A^T\vec{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$; NORMAL EQUATIONS

$$\left[\begin{array}{cc|c} 4 & 2 & 3 \\ 2 & 6 & 1 \end{array}\right] \to$$ (Gauss-Jordan, left to reader)

$a = 4/5$, $b = -\frac{1}{10}$ $\to$

$$\boxed{y = \frac{4}{5} - \frac{1}{10}x}$$

Other models besides linear ones are possible.

Here is a very general, but somewhat ambiguous, strategy that extends Theorem 6.63.

THEOREM 6.65

For least-squares approximating model to some data:

STEP ONE: Write down a perfect fit to the data, as a linear system in the model parameters.

STEP TWO : Find the least-squares solution of the linear system created by STEP ONE.

Example 6.66 Find the least-squares approximating parabola to the data from Examples 6.64 (b)

STEP ONE : $y = a + bx + cx^2 \rightarrow$

$$
\begin{aligned}
a - b + c &= 0 \qquad & (x, y) &= (-1, 0) \\
a &= 2 & (x, y) &= (0, 2) \\
a + b + c &= 1 & (x, y) &= (1, 1) \\
a + 2b + 4c &= 0 & (x, y) &= (2, 0)
\end{aligned}
$$

$\boxed{\text{STEP TWO:}}$

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

$$\rightarrow \quad A^T A = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}, \quad A^T \vec{b} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

NORMAL EQ'NS are

$$\left[\begin{array}{ccc|c} 4 & 2 & 6 & 3 \\ 2 & 6 & 8 & 1 \\ 6 & 8 & 18 & 1 \end{array}\right] \longrightarrow \text{(Gauss-Jordan, left to reader)}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 31 \\ 13 \\ -15 \end{bmatrix} \rightarrow \quad y = a + bx + cx^2 \Rightarrow$$

$$y = \frac{1}{20}\left(31 + 13x - 15x^2\right)$$

Example 6.67

Find the least-squares error, in Examples 6.64(b) and 6.66.

Note that the error in Example 6.66 cannot be larger than the error in Example 6.64(b), since 6.66 is the best approximation by parabolas $y = a + bx + cx^2$, a larger class than lines $y = a + bx$, as in 6.64(b)

In 6.64 (b), error is

$$\left\| \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 4/5 \\ -1/10 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\| =$$

$$\frac{1}{10} \left\| \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 8 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 20 \\ 10 \\ 0 \end{bmatrix} \right\| =$$

$$\frac{1}{10} \left\| \begin{bmatrix} 9 \\ 8 \\ 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 0 \\ 20 \\ 10 \\ 0 \end{bmatrix} \right\| = \frac{1}{10} \left\| \begin{bmatrix} 9 \\ -12 \\ -3 \\ 6 \end{bmatrix} \right\| =$$

$$\frac{3}{10} \left\| \begin{bmatrix} 3 \\ -4 \\ -1 \\ 2 \end{bmatrix} \right\| = \frac{3}{10}\sqrt{30} = \boxed{\frac{3\sqrt{3}}{\sqrt{10}} \simeq 1.64}$$

p. 526

In 6.66, error is

$$\left\| \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \cdot \frac{1}{20} \begin{bmatrix} 31 \\ 13 \\ -15 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\| =$$

$$\frac{1}{20} \left\| \begin{bmatrix} 3 \\ 31 \\ 29 \\ -3 \end{bmatrix} - \begin{bmatrix} 0 \\ 40 \\ 20 \\ 0 \end{bmatrix} \right\| = \frac{1}{20} \left\| \begin{bmatrix} 3 \\ -9 \\ 9 \\ -3 \end{bmatrix} \right\|$$

$$= \frac{3}{20} \left\| \begin{bmatrix} 1 \\ -3 \\ 3 \\ -1 \end{bmatrix} \right\| = \frac{3}{20}\sqrt{20} = \frac{3}{\sqrt{20}} \simeq 0.67$$

Definitely an improvement over the linear approximation, since the error is smaller.

p. 527

Example 6.68

Find the least-square approximating parabola through the origin, to the data $\{(-2,0), (0,1), (2,4)\}$.

$\boxed{\text{STEP ONE:}}$ Perfect fit, by $y = bx + cx^2$:

$$
\begin{aligned}
-2b + 4c^2 &= 0 \\
0 &= 1 \\
2b + 4c^2 &= 4
\end{aligned}
\qquad
\left(
\begin{aligned}
(x,y) &= (-2,0) \\
(x,y) &= (0,1) \\
(x,y) &= (2,4)
\end{aligned}
\right)
$$

$\boxed{\text{STEP TWO:}}$ Least-squares solution:

$$A \equiv \begin{bmatrix} -2 & 4 \\ 0 & 0 \\ 2 & 4 \end{bmatrix}, \quad \vec{b} \equiv \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} \to$$

$$A^T = \begin{bmatrix} -2 & 0 & 2 \\ 4 & 0 & 4 \end{bmatrix}, \quad A^T \vec{b} = \begin{bmatrix} 8 \\ 16 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 8 & 0 \\ 0 & 32 \end{bmatrix};$$ NORMAL EQ'NS are

$$\left[\begin{array}{cc|c} 8 & 0 & 8 \\ 0 & 32 & 16 \end{array}\right] \to b = 1, \quad c = \frac{1}{2},$$

$$\boxed{y = x + \frac{1}{2}x^2}$$