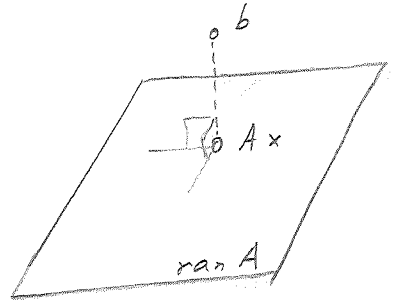
<u>Application</u>. Least squares problems

Suppose we want a solution to $Ax = b$.

In applications, it might be that $b \notin \operatorname{ran} A$.

Then we know that there is no x s.t. $Ax = b$.

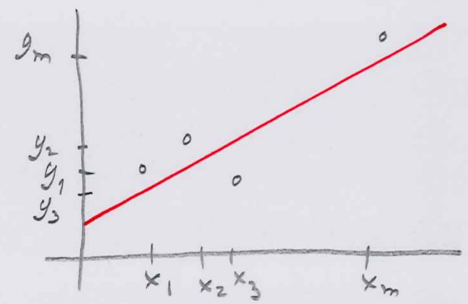
But often one is interested in finding an x which is "closest" in solving $Ax = b$. i.e, an x minimizing $\|Ax - b\|^2$.



$$\begin{cases} \text{Minimize } \frac{1}{2}\|Ax-b\|^2 = \frac{1}{2}(Ax-b)^T(Ax-b) = \frac{1}{2}x^TA^TAx - \frac{1}{2}b^TAx - \frac{1}{2}x^TA^Tb + \frac{1}{2}b^Tb \\ \qquad = \frac{1}{2}x^T\underbrace{(A^TA)}_{H}x \underbrace{- b^TA}_{c^T}x + \underbrace{\frac{1}{2}b^Tb}_{c_0} \\ \text{Subject to } x \in \mathbb{R}^n \end{cases}$$

<u>Example</u>. Experimental data:

$(x_1, y_1), \ldots, (x_m, y_m)$



Want to fit a straight line $y = \sigma x + \kappa$.

Want ideally:

$$\begin{cases} \sigma x_1 + \kappa = y_1 \\ \sigma x_2 + \kappa = y_2 \\ \vdots \\ \sigma x_m + \kappa = y_m . \end{cases}$$

But settle for σ, κ which is the optimal solution to:

$$\begin{cases} \text{minimize} & (\sigma x_1 + \kappa - y_1)^2 + \ldots + (\sigma x_m + \kappa - y_m)^2 \\ \text{s.t.} & \sigma, \kappa \in \mathbb{R} . \end{cases}$$

i.e.,

$$\begin{cases} \text{minimize} & \left\| \begin{bmatrix} \sigma x_1 + \kappa - y_1 \\ \vdots \\ \sigma x_m + \kappa - y_m \end{bmatrix} \right\|^2 \\ \text{s.t.} & \sigma, \kappa \in \mathbb{R} \end{cases}$$

i.e.,

$$\begin{cases} \text{minimize} & \left\| \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \\ x_m & 1 \end{bmatrix}}_{A} \begin{bmatrix} \sigma \\ \kappa \end{bmatrix} - \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{b} \right\|^2 \\ \text{s.t.} & \begin{bmatrix} \sigma \\ \kappa \end{bmatrix} \in \mathbb{R}^2 \end{cases}$$

How one solves the least squares problem.

$$\begin{cases} \text{minimize} & \frac{1}{2}\|Ax-b\|^2 = \frac{1}{2}x^T\underbrace{(A^TA)}_{H}x + \underbrace{(-A^Tb)}_{c}{}^T x + \underbrace{\frac{1}{2}b^Tb}_{c_0} \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

For a solution to exist H should be p.s.d.

Q. Is A^TA p.s.d?

A. Yes! $x^THx = x^TA^TAx = (Ax)^T\underbrace{Ax}_{\substack{\parallel \\ y=\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}} = [y_1 \dots y_m]\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = y_1^2+\dots+y_m^2 \geq 0.$

Moreover, $-c$ should belong to $\text{ran } H$.

Q. Does A^Tb belong to $\text{ran } A^TA$?

A Yes! Take $S = \text{ran } A$. Then $S^\perp = (\text{ran } A)^\perp = \ker A^T$.

$b = y + z$ where $y \in \text{ran } A$ and $z \in \ker A^T$.

So $b = Ax + z$ for some x.

So $A^Tb = A^TAx + \underbrace{A^Tz}_{=0} = A^TAx \in \text{ran } A^TA$.

So an optimal solution exists. Any $\hat{x}$ satisfying $H\hat{x} = -c$ is optimal,

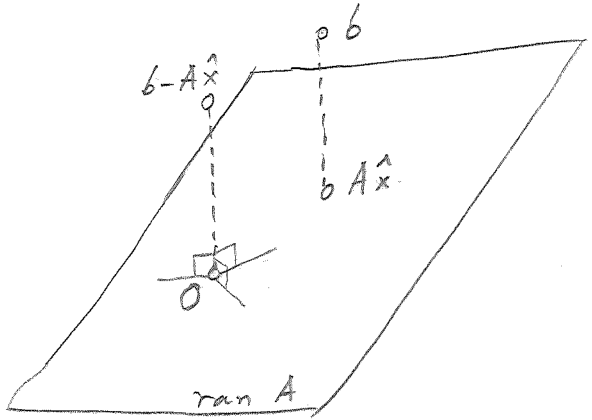
i.e., $\boxed{A^TA\hat{x} = A^Tb}$ Normal equations.

So the problem (P): $\left\{\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax-b\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n \end{array}\right\}$ always has an optimal solution.

Moreover, $\boxed{\hat{x} \text{ is optimal for (P)}} \iff \boxed{A^TA\hat{x} = A^Tb}$.

<u>Geometric interpretation of optimal solution:</u>

$$
\begin{aligned}
& A^TA\hat{x} = A^Tb \\
\iff & A^T(A\hat{x}-b) = 0 \\
\iff & A\hat{x}-b \in \ker A^T \\
& \phantom{A\hat{x}-b \in {}} \parallel \\
\iff & A\hat{x}-b \in (\operatorname{ran} A)^\perp \iff A\hat{x}-b \perp \text{ every } y \in \operatorname{ran} A.
\end{aligned}
$$



We have so far seen quadratic optimization without constraints.

$$\begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

Now we study quadratic optimization with equality constraints.

$$\begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & Ax = b \end{cases}$$

Here $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and as before $H = H^T \in \mathbb{R}^{n \times n}$, $c \in \mathbb{R}^n$, $c_0 \in \mathbb{R}$.

Feasible set $\mathcal{F} = \{ x \in \mathbb{R}^n : Ax = b \}$.

$b \notin \operatorname{ran} A \Rightarrow \mathcal{F} = \emptyset$. So we assume $b \in \operatorname{ran} A$.

$\ker A = \{0\} \Rightarrow \mathcal{F}$ has only one element. So we assume $\ker A \neq \{0\}$.

$\left(\begin{matrix} Ax_1 = b \\ Ax_2 = b \end{matrix} \Rightarrow A(x_1 - x_2) = 0 \text{ and so } x_1 = x_2 \right)$

$$\boxed{\text{Assumption: } \quad b \in \operatorname{ran} A \quad \text{and} \quad \ker A \neq \{0\}}$$

A characterization of feasible elements.

(1) Let $k = \dim(\ker A)$ and let $\{z_1, \dots, z_k\}$ be a basis for $\ker A$.

Define $Z = \begin{bmatrix} | & & | \\ z_1 & \cdots & z_k \\ | & & | \end{bmatrix}$

Then $z \in \ker A \Rightarrow \exists\, v_1, \dots, v_k$ s.t. $z = v_1 z_1 + \dots + v_k z_k = \begin{bmatrix} | & & | \\ z_1 & \cdots & z_k \\ | & & | \end{bmatrix} \underbrace{\begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}}_{=v} = Zv$.

Vice versa, if $z = Zv$ for some $v \in \mathbb{R}^k$, then $z \in \ker A$.

So $\boxed{z \in \ker A} \Longleftrightarrow \boxed{z = Zv \text{ for some } v \in \mathbb{R}^k}$

(2) Let $\bar{x}$ be (any) one solution to $Ax = b$.

Then $\boxed{x \in \mathcal{F}} \Longleftrightarrow \boxed{x = \bar{x} + Zv \text{ for some } v \in \mathbb{R}^k}$

$\boxed{\Rightarrow}$: $x \in \mathcal{F} \Rightarrow Ax = b$. But $A\bar{x} = b$. So $A(x - \bar{x}) = b - b = 0$. So $x - \bar{x} \in \ker A$

$\Rightarrow x - \bar{x} = Zv$ for some $v \in \mathbb{R}^k$

i.e., $x = \bar{x} + Zv$.

$\boxed{\Leftarrow}$: $x = \bar{x} + Zv \Rightarrow Ax = A\bar{x} + A\underbrace{Zv}_{\in \ker A} = A\bar{x} + 0 = A\bar{x} = b$. So $x \in \mathcal{F}$.

$$f(\bar{x}+Zv) = \frac{1}{2}(\bar{x}+Zv)^T H (\bar{x}+Zv) + c^T(\bar{x}+Zv) + c_0$$

$$= \frac{1}{2} v^T Z^T H Z v + \frac{1}{2}\bar{x}^T H Z v + \frac{1}{2} v^T Z^T H \bar{x} + c^T Z v + \frac{1}{2}\bar{x}^T H \bar{x} + c^T \bar{x} + c_0 .$$

$$= \frac{1}{2} v^T (Z^T H Z) v + \bar{x}^T H Z v + c^T Z v + f(\bar{x})$$

$$= \frac{1}{2} v^T (Z^T H Z) v + (Z^T H \bar{x} + Z^T c)^T v + f(\bar{x})$$

$$= \frac{1}{2} v^T \underbrace{(Z^T H Z)}_{\tilde{H}} v + \underbrace{\left(Z^T (H\bar{x} + c)\right)}_{\tilde{c}}^T v + \underbrace{f(\bar{x})}_{\tilde{c}_0}$$

$$=: \varphi(v)$$

Old problem

$$(Q): \begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & Ax = b \end{cases}$$

$x = \bar{x} + Zv$
$v \in \mathbb{R}^k$

New problem

$$(Q'): \begin{cases} \text{minimize} & \varphi(v) = \frac{1}{2} v^T \tilde{H} v + \tilde{c}^T v + \tilde{c}_0 \\ \text{s.t.} & v \in \mathbb{R}^k \end{cases}$$

Constrained $\longrightarrow$ unconstrained !

9.8

<u>Claim 1</u> If $\hat{v}$ is optimal for (Q'), then $\hat{x} := \bar{x} + Z\hat{v}$ is optimal for (Q).

<u>Proof</u> $\hat{x}$ is feasible for (Q).

Let x be feasible for (Q). Then $x = \bar{x} + Zv$ for some $v \in \mathbb{R}^k$

As $\hat{v}$ is optimal for (Q'), we have

$$\begin{array}{ccc} \varphi(\hat{v}) & \leq & \varphi(v). \\ \| & & \| \\ f(\bar{x} + Z\hat{v}) & & f(\bar{x} + Zv). \\ \| & & \| \\ f(\hat{x}) & & f(x). \end{array}$$

So $f(\hat{x}) \leq f(x)$ $\forall$ x feasible for (Q). Hence $\hat{x}$ is optimal for (Q). □

<u>Claim 2</u>. If $\hat{x}$ is optimal for (Q), then $\hat{v}$ is optimal for (Q'), where $\hat{v}$ is such that $\hat{x} = \bar{x} + Z\hat{v}$.

<u>Proof</u>. Let $v \in \mathbb{R}^k$. Define $x = \bar{x} + Zv$. Then x is feasible for (Q)

As $\hat{x}$ is optimal for (Q), we have $f(\hat{x}) \leq f(x)$.

i.e.,
$$\begin{array}{ccc} f(\bar{x} + Z\hat{v}) & \leq & f(\bar{x} + Zv) \\ \| & & \| \\ \varphi(\hat{v}) & & \varphi(v). \end{array}$$

So $\varphi(\hat{v}) \leq \varphi(v)$ $\forall$ $v \in \mathbb{R}^k$. Hence $\hat{v}$ is optimal for (Q'). □

Look at (Q'): $\begin{cases} \text{minimize} & \varphi(v) := \frac{1}{2} v^T (Z^T H Z) v + \left(Z^T (H\bar{x} + c)\right)^T v + f(\bar{x}) \\ \text{s.t.} & v \in \mathbb{R}^k . \end{cases}$

$Z^T H Z$ not p.s.d. $\Rightarrow$ (Q') has no optimal solution $\Rightarrow$ (Q) has no optimal solution

$Z^T H Z$ p.s.d. Then (Q') has an optimal solution $\hat{v}$ if and only if

$$Z^T H Z \hat{v} = - Z^T (H\bar{x} + c) .$$

So (Q) has an optimal solution $\hat{x}$ if and only if

$\hat{x} = \bar{x} + Z\hat{v}$, where $\hat{v}$ satisfies $Z^T H Z \hat{v} = - Z^T (H\bar{x} + c)$.

So we have the following result.

Theorem. Consider $(Q): \begin{cases} \text{minimize} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & Ax = b. \end{cases}$

Let $\bar{x}$ be any fixed vector s.t. $A\bar{x} = b$ and let the columns of $Z \in \mathbb{R}^{n \times k}$ form a basis for $\ker A$.

If $Z^T H Z$ is not p.s.d., then (Q) has no optimal solution.

If $Z^T H Z$ is p.s.d., then the following are equivalent:

(1) $\hat{x}$ is an optimal solution for (P).

(2) $\exists \hat{v}$ s.t. $Z^T H Z \hat{v} = -Z^T(H\bar{x} + c)$ and $\hat{x} = \bar{x} + Z\hat{v}$.

(3) $\exists u$ s.t. $H\hat{x} + c = A^T u$ and $A\hat{x} = b$.

Proof. We have already seen that $(1) \Leftrightarrow (2)$.

Now we show $(2) \Leftrightarrow (3)$.

$\underline{(2) \Rightarrow (3)}$: $\exists \hat{v}$ s.t. $\underbrace{\hat{x} = \bar{x} + Z\hat{v}}_{}$ and $Z^T H Z \hat{v} = -Z^T(H\bar{x} + c)$

$\downarrow$ premultiply by $Z^T H$

$$Z^T H \hat{x} = Z^T H \bar{x} + Z^T H Z \hat{v}$$

$$= \cancel{Z^T H \bar{x}} - Z^T(\cancel{H\bar{x}} + c) = -Z^T c.$$

So $Z^T(H\hat{x} + c) = 0$ i.e., $(H\hat{x} + c)^T Z = 0$.

So $\forall\, v \in \mathbb{R}^k$, $(H\hat{x} + c)^T Z v = 0$. So $\forall\, z \in \ker A$, $(H\hat{x} + c)^T z = 0$.

Hence $H\hat{x} + c \in (\ker A)^\perp$. But for $M \in \mathbb{R}^{m \times n}$, $(\operatorname{ran} M)^\perp = \ker M^T$.

So with $M = A^T$, $\ker A = (\operatorname{ran} A^T)^\perp$.

Hence $(\ker A)^\perp = ((\operatorname{ran} A^T)^\perp)^\perp = \operatorname{ran} A^T$.

So $H\hat{x} + c \in \operatorname{ran} A^T$, i.e., $H\hat{x} + c = A^T u$ for some u.

Also, from $\hat{x} = \bar{x} + Z\hat{v}$,

we have $A\hat{x} = b$.

(3) $\Rightarrow$ (2): Suppose $\exists\, u$ s.t. $H\hat{x} + c = A^T u$ where $A\hat{x} = b$.

Then: from $A\hat{x} = b$, it follows that $\hat{x} = \bar{x} + Z\hat{v}$ for some $\hat{v}$.

As $H\hat{x} + c \in \operatorname{ran} A^T = (\ker A)^\perp$,

$$(H\hat{x} + c)^T z = 0 \quad \forall\, z \in \ker A.$$

So $(H\hat{x}+c)^T z_1 = 0$, $(H\hat{x}+c)^T z_2 = 0$, $\ldots$, $(H\hat{x}+c)^T z_k = 0$.

Hence $(H\hat{x}+c)^T \underbrace{\begin{bmatrix} z_1 & \cdots & z_k \end{bmatrix}}_{Z} = 0$ i.e., $(H\hat{x}+c)^T Z = 0$.

i.e., $(H(\bar{x} + Z\hat{v}) + c)^T Z = 0$

i.e., $Z^T(H(\bar{x} + Z\hat{v}) + c) = 0$

i.e., $Z^T H Z \hat{v} = -Z^T(H\bar{x} + c)$ $\square$.

Summary. $\left\{\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & Ax = b. \end{array}\right\}$

If $Z^T H Z$ not p.s.d., then no optimal solution exists.

If $Z^T H Z$ is p.s.d, then

$\boxed{\hat{x} \text{ is an optimal solution}} \iff \underbrace{\boxed{\exists\, u \text{ s.t. } \begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \hat{x} \\ u \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}}}_{\text{Lagrange method}} \iff \underbrace{\boxed{\exists\, \hat{v} \text{ s.t. } Z^T H Z \hat{v} = -Z^T(H\bar{x}+c) \text{ and } \hat{x} = \bar{x} + Z\hat{v}}}_{\text{Null space method}}$

Lagrange method

Null space method.

Least squares problem revisited.

Consider the equation system $\underbrace{\begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix} 3 \\ 2 \end{bmatrix}}_{b}$

This has no solution: If $\left\{ \begin{array}{l} 2x_1 + 6x_2 = 3 \\ x_1 + 3x_2 = 2 \end{array} \right\}$, then $0 = 2x_1 + 6x_2 - 2(x_1 + 3x_2)$
$= 3 - 2.2 = -1$, a contradiction

So consider the problem: $\left\{ \begin{array}{ll} \text{minimize} & \frac{1}{2}\| Ax - b \|^2 \\ \text{s.t.} & x \in \mathbb{R}^2 \end{array} \right.$

Any solution to the normal equations $A^T A x = A^T b$ is an optimal solution.

$$A^T A = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 45 \end{bmatrix} = 5 \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}.$$

$$A^T b = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 24 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

$5 \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} x = 8 \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ i.e., $\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} x = \frac{8}{5} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ i.e., $\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} x = \frac{8}{5} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

i.e., $\begin{bmatrix} 1 & 3 \end{bmatrix} x = \frac{8}{5}$.

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{8}{5} - 3\alpha \\ \alpha \end{bmatrix}, \quad \alpha \in \mathbb{R}$$

Which x should one choose?

It is reasonable to choose an x which is the smallest

So we consider the auxiliary problem

$$\begin{cases} \text{minimize} & \frac{1}{2}\|x\|^2 = \frac{1}{2} x^T I x \\ \text{s.t.} & A^T A x = A^T b. \end{cases}$$

Suppose $\bar{x}$ satisfies $A^T A \bar{x} = A^T b$.

Then x satisfies $A^T A x = A^T b$ <u>if and only if</u> $A x = A \bar{x}$.

<u>Reason</u>: If $Ax = A\bar{x}$, then $A^T A x = A^T A \bar{x} = A^T b$. Done!

If $A^T A x = A^T b$, then

$$\begin{aligned} & A^T A x = A^T A \bar{x} \\ \Rightarrow\ & A^T A (x - \bar{x}) = 0 \\ \Rightarrow\ & (x - \bar{x})^T A^T \underbrace{A(x - \bar{x})}_{=: y} = 0 \\ \Rightarrow\ & y^T y = 0 \\ \Rightarrow\ & y = 0 \\ \Rightarrow\ & A(x - \bar{x}) = 0 \\ \Rightarrow\ & A x = A \bar{x}. \end{aligned}$$

9.15

So the problem becomes:

$$(P): \begin{cases} \text{minimize} & \frac{1}{2}\|x\|^2 = \frac{1}{2}x^T I x \\ \text{s.t.} & Ax = \underbrace{A\bar{x}}_{b} . \end{cases}$$

(Note that since $H = I$ is p.d., $Z^T H Z = Z^T Z$ is p.s.d.)

Take $S = \operatorname{ran} A^T$ Then $S^\perp = (\operatorname{ran} A^T)^\perp = \ker((A^T)^T) = \ker A$.

$\bar{x} = \underset{\underset{S = \operatorname{ran} A^T}{\cap}}{\xi} + z \in S^\perp = \ker A$. So $\bar{x} = A^T u + z$ for some u.

Hence $A\bar{x} = AA^T u + \underbrace{Az}_{=0}$ $(\because z \in \ker A)$

$= AA^T u$

So there is at least one u s.t. $AA^T u = A\bar{x}$. Take any such u.

Now define $\hat{x} = A^T u$.

We have $A\hat{x} = AA^T u = A\bar{x}$.

Also $H\hat{x} + c = I\hat{x} + 0 = \hat{x} = A^T u$.

So $\hat{x}$ is optimal for (P).

All u's give the same $\hat{x}$: Suppose $AA^T u = A\bar{x} = AA^T u'$.

Then $AA^T(u-u') = 0$. And so $(u-u')AA^T(u-u') = 0$

i.e., $A^T(u-u') = 0$

i.e., $\underbrace{A^T u}_{\hat{x}} = A^T u'$.

<u>Summary</u> Minimum norm solution to $\begin{cases} \text{minimize} & \frac{1}{2}\|Ax-b\|^2 \\ \text{s.t.} & Ax=b \end{cases}$.

(1) Find any solution $\bar{x}$ to $A^TA\bar{x} = A^Tb$.

(2) Find any solution to $AA^Tu = A\bar{x}$.

(3) Take $\hat{x} = A^Tu$.

<u>Example</u> (continued). $\underset{(\alpha=0)}{\bar{x}} = \begin{bmatrix} \frac{8}{5} \\ 0 \end{bmatrix}$. $AA^T = \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} 40 & 20 \\ 20 & 10 \end{bmatrix} = 10\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

$A\bar{x} = \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 8/5 \\ 0 \end{bmatrix} = \begin{bmatrix} 16/5 \\ 8/5 \end{bmatrix}$.

$AA^Tu = A\bar{x}$ becomes

$10\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}u = \frac{8}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ i.e., $\begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix}u = \frac{4}{25}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

i.e., $\begin{bmatrix} 2 & 1 \end{bmatrix}u = \frac{4}{25}$. $u = \begin{bmatrix} 0 \\ 4/25 \end{bmatrix}$.

$$\hat{x} = A^Tu = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix}\begin{bmatrix} 0 \\ 4/25 \end{bmatrix} = \frac{4}{25}\begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

is ~~the~~ minimum norm solution to the least squares problem.