Mathematics for Computer Science
Linear Algebra (Part 2)
Linear Regression

Karl Southern

Durham University

March 3rd, 2025

Thanks to Andrei Krokhin and William Moses for use of slides.

Outline

Roadmap for Lectures 5 - 8

- **End Goal:** Application - linear regression.
- **Using:** QR decomposition.
- Requires knowledge of some basics: Inner product spaces.

Roadmap for Lectures 5 - 8

- **End Goal:** Application - linear regression.
- **Using:** QR decomposition.
- Requires knowledge of some basics: Inner product spaces.

Now we recap last lecture & look at what we'll cover today.

Last Lecture Reminder

Definition (Least Squares Problem)

Given a linear system $A\mathbf{x} = \mathbf{b}$ with m equations and n variables, find a vector $\mathbf{x}$ that minimises $||\mathbf{b} - A\mathbf{x}||$ (w.r.t. the Euclidean inner product on $\mathbb{R}^m$).

- We call such a vector $\mathbf{x}$ a least squares solution to the system, the vector $\mathbf{b} - A\mathbf{x}$ is the least squares error vector, and the number $||\mathbf{b} - A\mathbf{x}||$ is the least squares error.
- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- $A^T A\mathbf{x} = A^T\mathbf{b}$ is the normal equation (or normal system) associated with $A\mathbf{x} = \mathbf{b}$.
- The solution to $A^T A\mathbf{x} = A^T\mathbf{b}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$.

Today's Lecture Contents

- **Application:** Least squares fitting to data (aka linear regression)
- The benefit of a QR decomposition

Outline

Application: Least Squares Fitting to Data

- Assume that you want to determine, possibly approximately, the (quantitative) dependency between two parameters x and y in some process.

Application: Least Squares Fitting to Data

- Assume that you want to determine, possibly approximately, the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).

Application: Least Squares Fitting to Data

- Assume that you want to determine, possibly approximately, the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first.

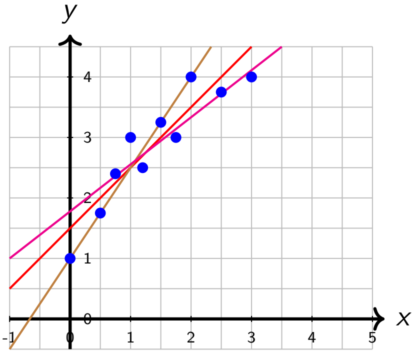
  (The method extends to more than 2 variables and to more complex curves)

Application: Least Squares Fitting to Data

- **Assume:** you want the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first.

Application: Least Squares Fitting to Data

- **Assume:** you want the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first.



Application: Least Squares Fitting to Data

- **Assume:** you want the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first.
- This can then be written as a series of equations

$$\begin{aligned} a + bx_1 &= y_1 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

- Or in matrix form $\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

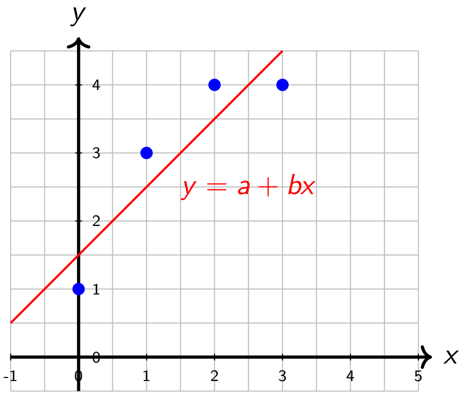
Application: Least Squares Fitting to Data

- **Assume:** you want the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first.
- This can then be written as a series of equations

$$\begin{aligned} a + bx_1 &= y_1 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

- Or in matrix form $\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Example



$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \\ a + bx_3 &= y_3 \\ a + bx_4 &= y_4 \end{aligned}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

Example 18.1

Need to find a least squares solution to the system $A\mathbf{v} = \mathbf{y}$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

Matrix A has linearly independent columns, so least squares solution is unique.

Example 18.1

Need to find a least squares solution to the system $A\mathbf{v} = \mathbf{y}$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

Matrix A has linearly independent columns, so least squares solution is unique.

From the associated normal system $A^T A\mathbf{v} = A^T\mathbf{y}$, we get

$$\begin{pmatrix} a \\ b \end{pmatrix} = \left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

Example 18.1

Need to find a least squares solution to the system $A\mathbf{v} = \mathbf{y}$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

Matrix A has linearly independent columns, so least squares solution is unique.

From the associated normal system $A^T A\mathbf{v} = A^T\mathbf{y}$, we get

$$\begin{pmatrix} a \\ b \end{pmatrix} = \left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

From this, $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$ so $y = 1.5 + x$ is the least squares straight line fit

(aka the regression line)

Outline

When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

When are least squares solutions unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

When are least squares solutions unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^T A\mathbf{x} = \mathbf{0}$, i.e., $A^T A\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.
- Thus $A\mathbf{x} = \mathbf{0}$ has only the trivial solution iff the same is true for $A^T A\mathbf{x} = \mathbf{0}$.

□

Finding a Unique Least Squares Solution

Assume that A is an $m \times n$ matrix with linearly independent column vectors.

Finding a Unique Least Squares Solution

Assume that A is an $m \times n$ matrix with linearly independent column vectors.

- From the previous theorem, for every column $\mathbf{b} \in \mathbb{R}^m$, the least squares solution to $A\mathbf{x} = \mathbf{b}$ (i.e. a solution to $A^T A\mathbf{x} = A^T\mathbf{b}$) is unique:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}.$$

Finding a Unique Least Squares Solution

Assume that A is an $m \times n$ matrix with linearly independent column vectors.

- From the previous theorem, for every column $\mathbf{b} \in \mathbb{R}^m$, the least squares solution to $A\mathbf{x} = \mathbf{b}$ (i.e. a solution to $A^T A\mathbf{x} = A^T\mathbf{b}$) is unique:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}.$$

For the last item, substitute $A = QR$ into the first equation above and simplify:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b} = ((QR)^T (QR))^{-1} (QR)^T \mathbf{b} = (R^T Q^T QR)^{-1} R^T Q^T \mathbf{b}$$
$$= (R^T IR)^{-1} R^T Q^T \mathbf{b} = (R^T R)^{-1} R^T Q^T \mathbf{b} = R^{-1} (R^T)^{-1} R^T Q^T \mathbf{b} = R^{-1} Q^T \mathbf{b}.$$

The simplification uses $Q^T Q = I$ — true because Q has orthonormal columns.

Finding a Unique Least Squares Solution

Assume that A is an $m \times n$ matrix with linearly independent column vectors.

- From the previous theorem, for every column $\mathbf{b} \in \mathbb{R}^m$, the least squares solution to $A\mathbf{x} = \mathbf{b}$ (i.e. a solution to $A^TA\mathbf{x} = A^T\mathbf{b}$) is unique:

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}.$$

  For the last item, substitute $A = QR$ into the first equation above and simplify:

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} = ((QR)^T(QR))^{-1}(QR)^T\mathbf{b} = (R^TQ^TQR)^{-1}R^TQ^T\mathbf{b}$$
$$= (R^TIR)^{-1}R^TQ^T\mathbf{b} = (R^TR)^{-1}R^TQ^T\mathbf{b} = R^{-1}(R^T)^{-1}R^TQ^T\mathbf{b} = R^{-1}Q^T\mathbf{b}.$$

  The simplification uses $Q^TQ = I$ — true because Q has orthonormal columns.

- If $A = QR$ is a QR decomposition (which exists under our assumption), then

$$\mathbf{x} = R^{-1}Q^T\mathbf{b} \quad \text{(or, equivalently, } R\mathbf{x} = Q^T\mathbf{b}\text{)}$$

Projection Matrices

We showed before: for any column vector $\mathbf{b}$, if $W = \mathcal{C}(A)$ is the column space of a matrix A and $\mathbf{x}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$, then $\mathrm{proj}_W \mathbf{b} = A\mathbf{x}$.

Projection Matrices

We showed before: for any column vector $\mathbf{b}$, if $W = \mathcal{C}(A)$ is the column space of a matrix A and $\mathbf{x}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$, then $\mathrm{proj}_W \mathbf{b} = A\mathbf{x}$.

If A has linearly independent column vectors, then we can use the first formula from the previous slide:

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

Projection Matrices

We showed before: for any column vector $\mathbf{b}$, if $W = \mathcal{C}(A)$ is the column space of a matrix A and $\mathbf{x}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$, then $\mathrm{proj}_W \mathbf{b} = A\mathbf{x}$.

If A has linearly independent column vectors, then we can use the first formula from the previous slide:

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

- A matrix of the form $P = A(A^T A)^{-1} A^T$ is called a projection matrix.

Projection Matrices

We showed before: for any column vector $\mathbf{b}$, if $W = \mathcal{C}(A)$ is the column space of a matrix A and $\mathbf{x}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$, then $\mathrm{proj}_W \mathbf{b} = A\mathbf{x}$.

If A has linearly independent column vectors, then we can use the first formula from the previous slide:

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

- A matrix of the form $P = A(A^T A)^{-1} A^T$ is called a projection matrix.
- For any column vector $\mathbf{b}$, the vector $P\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ onto $W = \mathcal{C}(A)$, the column space of A.

Projection Matrices

We showed before: for any column vector $\mathbf{b}$, if $W = \mathcal{C}(A)$ is the column space of a matrix A and $\mathbf{x}$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$, then $\mathrm{proj}_W \mathbf{b} = A\mathbf{x}$.

If A has linearly independent column vectors, then we can use the first formula from the previous slide:

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

- A matrix of the form $P = A(A^T A)^{-1} A^T$ is called a projection matrix.
- For any column vector $\mathbf{b}$, the vector $P\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ onto $W = \mathcal{C}(A)$, the column space of A.
- Such matrices P are used in ML and Data Science to map vectors $\mathbf{b}$ from a high-dimensional space to a suitably chosen small-dimensional space W.

Example 18.2

Given the data points: $(1,-1),(2,2),(0,-6),(3,6),(1,0),(2,3)$, where we weight the experiments as $(4,4,3,3,1,1)$, fit the data to a curve of the form $y = a + bx + cx^2$.

This is equivalent to finding the least squares solution to the system $A\mathbf{x} = \mathbf{y}$:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 0 & 0 \\ 1 & 3 & 9 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -6 \\ 6 \\ 0 \\ 3 \end{pmatrix}$$

Example 18.2 ... Gram-Schmidt

N.B inner product is weighted Euclidean dot product with weights $(4, 4, 3, 3, 1, 1)$.

- $\mathbf{v}_1 = (1, 1, 1, 1, 1, 1)$
- $\mathbf{v}_2 = \frac{1}{2}(-1, 1, -3, 3, -1, 1)$
- $\mathbf{v}_3 = \frac{1}{4}(-3, -3, 5, 5, 3, -3)$

Normalising to

- $\mathbf{q}_1 = \frac{1}{4}(1, 1, 1, 1, 1, 1)$
- $\mathbf{q}_2 = \frac{1}{8}(-1, 1, -3, 3, -1, 1)$
- $\mathbf{q}_3 = \frac{1}{4\sqrt{15}}(-3, -3, 5, 5, -3, -3)$

Example 18.2 ... QR decomp

N.B inner product is weighted Euclidean dot product with weights $(4, 4, 3, 3, 1, 1)$.

$$Q = \begin{pmatrix} \frac{1}{4} & \frac{-1}{8} & \frac{-3}{4\sqrt{15}} \\ \frac{1}{4} & \frac{1}{8} & \frac{-3}{4\sqrt{15}} \\ \frac{1}{4} & \frac{-3}{8} & \frac{5}{4\sqrt{15}} \\ \frac{1}{4} & \frac{3}{8} & \frac{5}{4\sqrt{15}} \\ \frac{1}{4} & \frac{-1}{8} & \frac{-3}{4\sqrt{15}} \\ \frac{1}{4} & \frac{1}{8} & \frac{-3}{4\sqrt{15}} \end{pmatrix} R = \begin{pmatrix} 4 & 6 & 13 \\ 0 & 4 & 12 \\ 0 & 0 & \sqrt{15} \end{pmatrix}$$

$$\mathbf{x} = R^{-1} Q^T \mathbf{y} = \frac{1}{160} \begin{pmatrix} -315 \\ 306 \\ -32 \end{pmatrix} \approx \begin{pmatrix} 1.97 \\ 1.91 \\ -0.2 \end{pmatrix}$$

Outline

Exam Question

(d) Given the following points, use the least squares method to find the straight line and quadratic polynomial that best approximate them.

$$P = (0,-1),(0,0),(1,-1),(1,-2),(2,0),(2,2).$$

(e) Which of the two polynomials is the best approximation? Justify your answer. **[3 Marks]**

Wrapping Things Up

Today:

- Least squares solutions for inconsistent linear systems
- Least squares fitting to data (aka linear regression)
- Benefit of a QR decomposition

Next time:

- Orthogonal matrices and spectral decomposition

The End

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^T A\mathbf{x} = \mathbf{0}$, i.e., $A^T A\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^T A\mathbf{x} = \mathbf{0}$, i.e., $A^T A\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.
- Thus $A\mathbf{x} = \mathbf{0}$ has only the trivial solution iff the same is true for $A^T A\mathbf{x} = \mathbf{0}$.

$\square$