Métodos Numéricos modo virtual
(pandemia COVID-19)
Material Complementario

Cuadrados Mínimos Lineales - versión 1.0

Este es material complementario de las diapos de la clase de cuadrados mínimos usadas durante el dictado virtual (pandemia COVID-19).
Este documento incluye formulación del problema, existencia de solución y análisis de la unicidad.

Dados pares ordenados (x_i, y_i) para $i = 1, \ldots, m$, buscamos una función $f(x)$ perteneciente a una familia $\mathcal{F}$ de funciones tal que *mejor aproxime* a los datos.
¿Cómo *traducimos* en términos matemáticos *mejor aproxime*? Hay diferentes alternativas:

- $\min\limits_{f \in \mathcal{F}} \max\limits_{i=1,\ldots,m} |f(x_i) - y_i|$
- $\min\limits_{f \in \mathcal{F}} \sum\limits_{i=1}^{m} |f(x_i) - y_i|$
- $\min\limits_{f \in \mathcal{F}} \sum\limits_{i=1}^{m} (f(x_i) - y_i)^2$

El primer criterio, al considerar el máximo error, es sensible a valores atípicos que puedan aparecer en la muestra. En el segundo y tercer criterio, los valores atípicos ya dejan de dominar a la muestra y reflejan en mejor manera el concepto de ajuste. El tercer criterio es conocido como el criterio de cuadrados mínimos y es el más usado en el contexto de aproximación ya que bajo ciertos escenarios tiene propiedades teóricas y prácticas que facilitan obtener la solución.
Nos vamos a concentrar particularmente en familias de funciones que se describen como combinación lineal de un conjunto finito de funciones linealmente independientes.
Si $\phi_1(x), \phi_2(x), \ldots, \phi_n(x)$ son funciones linealmente independientes, la familia $\mathcal{F}$ está definida por

$$\mathcal{F} = \{f(x) : f(x) = c_1\phi_1(x) + c_2\phi_2(x) + \ldots + c_n\phi_n(x) \text{ con } c_i \in R \; \forall i = 1, \ldots, n\}$$

Entonces

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (f(x_i) - y_i)^2 = \min_{c_1,\ldots,c_n} \sum_{i=1}^{m} (\sum_{k=1}^{n} c_k\phi_k(x_i) - y_i)^2$$

Veamos cómo podemos expresar este problema en forma matricial. Consideremos lo siguiente:

$$A = \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_k(x_1) & \dots & \phi_n(x_1) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \phi_1(x_i) & \phi_2(x_i) & \dots & \phi_k(x_i) & \dots & \phi_n(x_i) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \phi_1(x_m) & \phi_2(x_m) & \dots & \phi_k(x_m) & \dots & \phi_n(x_m) \end{bmatrix} \qquad x = \begin{bmatrix} c_1 \\ \vdots \\ c_k \\ \vdots \\ c_n \end{bmatrix} \qquad b = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{bmatrix}$$

$$Ax - b = \begin{bmatrix} \sum_{k=1}^n \phi_k(x_1)c_k \\ \sum_{k=1}^n \phi_k(x_2)c_k \\ \vdots \\ \sum_{k=1}^n \phi_k(x_i)c_k \\ \vdots \\ \sum_{k=1}^n \phi_k(x_m)c_k \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}$$

Entonces

$$\min_{c_1,\dots,c_n} \sum_{i=1}^m (\sum_{k=1}^n c_k\phi_k(x_i) - y_i)^2 = \min_{x\in R^n} ||Ax-b||_2^2$$

A partir de ahora nuestro problema será:

Dados $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^m$, buscamos $x \in \mathbb{R}^n$ tal que minimice $||Ax - b||_2^2$.

Veamos en primer lugar si este problema tiene solución y qué características tiene.

Proposición: *Existe $\bar{x}$ solución del problema de cuadrados mínimos lineales.*

Demostración:
Hay un resultado de álgebra lineal que establece que $Im(A) + Nu(A^t) = \mathbb{R}^m$. Es decir, todo vector en $\mathbb{R}^m$ se puede escribir como la suma de dos vectores, uno en el subespacio imagen de A y otro en el núcleo de A^t. Además $Nu(A^t) = Im(A)^\perp$.
Aplicando este resultado al vector b, tenemos que $b = b^{(1)} + b^{(2)}$ con $b^{(1)} \in Im(A)$ y $b^{(2)} \in Nu(A^t)$.
Entonces

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ||Ax - b^{(1)} - b^{(2)}||_2^2$$

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ((Ax - b^{(1)}) - b^{(2)})^t((Ax - b^{(1)}) - b^{(2)})$$

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} (Ax - b^{(1)})^t(Ax - b^{(1)}) - 2(Ax - b^{(1)})^t b^{(2)} + {b^{(2)}}^t b^{(2)}$$

Como $Ax - b^{(1)} \in Im(A)$ y $b^{(2)} \in Nu(A^t) = Im(A)^\perp$, entonces $(Ax - b^{(1)})^t b^{(2)} = 0$ y por lo tanto

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} (Ax - b^{(1)})^t(Ax - b^{(1)}) + {b^{(2)}}^t b^{(2)} = \min_{x\in\mathbb{R}^n} ||Ax - b^{(1)})||_2^2 + ||b^{(2)}||_2^2$$

Como el segundo término es constante, el mínimo será aquel que $\min_{x\in\mathbb{R}^n} ||Ax - b^{(1)})||_2^2$. Pero como $b^{(1)} \in Im(A)$, entonces existe $\bar{x} \in \mathbb{R}^n$ tal que $A\bar{x} = b^{(1)}$ y por lo tanto $||A\bar{x} - b^{(1)})||_2^2 = 0$ que es lo mínimo que puede valer una norma.
En conclusión, siempre existe $\bar{x} \in \mathbb{R}^n$ solución del problema de cuadrados mínimos lineales y verifica que $A\bar{x} = b^{(1)}$.

■

¿Cuándo podremos afirmar la unicidad de la solución? Por la caracterización a la que llegamos, sabemos que una solución del problema satisface que $A\bar{x} = b^{(1)}$. Es decir, se escribe a $b^{(1)}$ como combinación lineal de las columnas de la matriz A. Si las columnas de A son linealmente independientes, entonces hay una única manera de escribir a $b^{(1)}$ (y la inversa también es válida). En el caso que no lo fueran, entonces tendremos infinitas soluciones. Veamos esto más formalmente.

Proposición: *Existe única solución de cuadrados mínimos lineales sii las columnas de A son li.*

Demostración:
Sea $\bar{x}$ única solución del problema. Si las columnas de A fueran ld, entonces existe z, $z \neq 0$, tal que $Az = 0$. Entonces

$$A(\bar{x} + z) = A\bar{x} + Az = b^{(1)} + 0 = b^{(1)}$$

Por lo tanto $\bar{x} + z$ también es solución del problema, lo que contradice la unicidad.
Supongamos ahora que las columnas de A son li y existen al menos dos soluciones distintas $\bar{x}_1$ y $\bar{x}_2$. Entonces

$$A(\bar{x}_1 - \bar{x}_2) = A\bar{x}_1 - A\bar{x}_2 = b^{(1)} - b^{(1)} = 0$$

Es decir, $\bar{x}_1 - \bar{x}_2$ es un vector no nulo solución del sistema homogéneo. Esto contradice que las columnas de A son li.

■

En resumen:

Siempre existe solución del problema de cuadrados mínimos lineales. La solución es única sii las columnas de A son li.

DEPARTAMENTO DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Métodos Numéricos modo virtual
(pandemia COVID-19)
Material Complementario

CML - Ecuaciones normales - versión 1.0

Este es material complementario de las diapos de la clase de cuadrados mínimos usadas durante el dictado virtual (pandemia COVID-19). En este documento deducimos las ecuaciones normales y analizamos su utilidad.

La caracterización que hemos dado de la solución de cuadrados mínimos lineales es mediante el sistema

$$Ax = b^{(1)}$$

Si quisiéramos usar esto para obtener la solución deberíamos conocer a $b^{(1)}$. Para evitar esto, el próximo resultado caracteriza a la solución en función de los datos originales: A y b.

Proposición: *Sean $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\bar{x} \in \mathbb{R}^n$ solución de cuadrados mínimos lineales y $r = b - A\bar{x}$ entonces*

$$A^t r = 0$$
$$A^t A\bar{x} = A^t b$$

Demostración:
Sabemos que $A\bar{x} = b^{(1)} = b - b^{(2)}$, entonces $b - A\bar{x} = b^{(2)}$. Como $b^{(2)} \in Nu(A^t)$, entonces

$$0 = A^t b^{(2)} = A^t(b - A\bar{x}) \Rightarrow A^t r = 0$$

y además que

$$0 = A^t b^{(2)} = A^t(b - A\bar{x}) \Rightarrow A^t b = A^t A\bar{x}$$

Este último sistema es conocido bajo el nombre de *ecuaciones normales*. ■

De esta manera nos quedó caracterizada la solución del problema de cuadrados mínimos lineales como la solución de un sistema de ecuaciones. ¿Qué propiedades podemos decir del sistema de ecuaciones normales?

- $A^t A \in \mathbb{R}^{n\times n}$
- $A^t A$ es simétrica ya que $(A^t A)^t = A^t (A^t)^t = A^t A$
- $A^t A$ es semidefinida positiva ya que $x^t A^t Ax = ||Ax||_2^2$ que por definción de norma es ≥ 0 para todo x.

- A^tA es definida positiva si A tiene columnas linealmente independientes. De la propiedad anterior deducimos que es ≥ 0 y será nula en el caso que $||Ax||_2^2 = 0$. Si las columnas de A son li, entonces no existe x no nulo en estas condiciones.

Esta caracterización de la solución y las propiedades de la matriz asociada al sistema nos dan una excelente herramienta para resolver el problema. Se pueden usar algoritmos apropiados como la factorizacion de Choleski. Sin embargo, debido a la aritmética finita se nos pueden presentar algunos problemas.
Por ejemplo:

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \text{ entonces } A^tA = \begin{bmatrix} 1 & \epsilon & 0 \\ 1 & 0 & \epsilon \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix}$$

Si $\epsilon > 0$ y representable en aritmética finita, las columnas de A son li y por lo tanto la solución del problema debería ser única. Sin embargo, tomándolo suficientemente chico, la aritmética finita obtendría que $1+\epsilon^2 = 1$ y tendríamos un sistema de ecuaciones normales donde la matriz A^tA sería singular.
Además podemos tener problemas de condicionamiento. Si calculamos el número de condición relativo a la norma 1, tenemos que $\kappa_1(A^tA) = ||A^tA||_1||(A^tA)^-1||_1$.

La matrix $(A^tA)^{-1} = \begin{bmatrix} \frac{1+\epsilon^2}{\epsilon^2(2+\epsilon^2)} & \frac{-1}{\epsilon^2(2+\epsilon^2)} \\ \frac{-1}{\epsilon^2(2+\epsilon^2)} & \frac{1+\epsilon^2}{\epsilon^2(2+\epsilon^2)} \end{bmatrix}$

Sabemos que la norma 1 de una matriz B resulta ser $\max_{j=1,\ldots,n} \sum_{i=1}^n |b_{ij}|$

Entonces $||(A^tA)^-1||_1 = 2+\epsilon^2$ y $||(A^tA)^-1||_1 = \frac{1}{\epsilon^2}$, de donde surge que $\kappa_1(A) = \frac{(2+\epsilon^2)}{\epsilon^2}$. Para valores chicos de ϵ, $\kappa_1(A)$ resulta grande y el sistema estará mal condicionado.

Métodos Numéricos modo virtual
(pandemia COVID-19)
Material Complementario

CML - QR - versión 1.0

Este es material complementario de las diapos de la clase de cuadrados mínimos usadas durante el dictado virtual (pandemia COVID-19). En este documento deducimos la metodología de solución basada en la factorización QR.

Dada $A \in \mathbb{R}^{m\times n}$, sabemos que existen P matriz de permutació, $Q \in \mathbb{R}^{m\times m}$ matriz ortogonal y $R \in \mathbb{R}^{m\times n}$ matriz tal que $r_{ij} = 0$ si $i > j$, tal que $AP = QR$. Además, si $p = rango(A)$, entonces $p = rango(R)$ y la estructura de R es la siguiente:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} & r_{1p+1} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2p} & r_{2p+1} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & r_{pp} & r_{pp+1} & \dots & r_{pn} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}$$

¿Cómo obtenemos esta factorización? Si recordamos los algoritmos de factorización QR, la idea es ir multiplicando por matrices ortogonales de tal manera de anular elementos de la matriz que se encuentran por debajo del término a_{ii}. Este proceso siempre está bien definido. Lo que hemos agregado en este punto es mover todas aquellas columnas que nos hayan quedado con términos nulos en a_{ii} para el final. Eso es lo que hace la matriz P.
Notación: llamaremos R_1 a la submatriz de R con las primeras p columnas y p filas y R_2 a la submatriz de R con las últimas $n-p$ columnas y p filas. Notar que R_1 es triangular superior con elementos no nulos en la diagonal.

Veamos como podemos usar esta factorización para resolver el problema de cuadrados mínimos lineales.
Sabemos que multiplicar por una matriz ortogonal no cambia el valor de la norma 2. Entonces

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ||Q^t(Ax-b)||_2^2 = \min_{x\in\mathbb{R}^n} ||Q^tAx-Q^tb)||_2^2 = \min_{x\in\mathbb{R}^n} ||Q^tAPP^{-1}x-Q^tb)||_2^2 = \min_{x\in\mathbb{R}^n} ||RP^{-1}x-Q^tb)||_2^2$$

Notemos $P^{-1}x = y$ y $Q^tb = (b^{(p)}, b^{(m-p)})$. Analizamos dos casos:

- A tiene columnas li, por lo tanto $p = n$ y $P = I$.

$$\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2 = \min_{x\in\mathbb{R}^n} ||Rx - Q^t b)||_2^2 = \min_{x\in\mathbb{R}^n} ||R_1 x - b^{(n)}||_2^2 + ||b^{(m-n)}||_2^2$$

El segundo término es una constante y el primer término lo mínimo que puede valer es 0. Como R_1 es triangular superior con elementos no nulos en la diagonal, entonces podemos afirmar que existe x tal que $R_1 x - b^{(n)} = 0$ y esa es la solución del problema.

- A tiene columnas ld, por lo tanto $p < n$.

$$\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2 = \min_{x\in\mathbb{R}^n} ||RP^{-1}x - Q^t b)||_2^2 \underset{\substack{\downarrow \\ y = P^{-1}x}}{=} \min_{y\in\mathbb{R}^n} ||Ry - Q^t b||_2^2$$

$$\min_{x\in\mathbb{R}^n} ||Ax - b||_2^2 \underset{\substack{\downarrow \\ y = P^{-1}x}}{=} \min_{y\in\mathbb{R}^n} ||(R_1, R_2)y - Q^t b||_2^2 = \min_{y\in\mathbb{R}^n} ||(R_1, R_2)y - b^{(p)}||_2^2 + ||b^{(m-p)}||_2^2$$

El segundo término es una constante y el primer término lo mínimo que puede valer es 0. Veamos si es posible encontrar valores para y que hagan nula la norma y de esa forma obtener el mínimo. Entonces, buscamos y tal que $(R_1, R_2)y - b^{(p)} = 0$. Si notamos $y^{(p)}$ a las primeras p componentes de y y $y^{(n-p)}$ al resto, tenemos que

$$(R_1, R_2)y - b^{(p)} = R_1 y^{(p)} + R_2 y^{(n-p)} - b^{(p)} = 0$$
$$R_1 y^{(p)} = b^{(p)} - R_2 y^{(n-p)}$$

Es suficiente tomar valores arbitrarios para las coordenadas de $y^{(n-p)}$ y se determinan en forma unívoca las p componentes de $y^{(p)}$. De esta manera se pueden obtener las soluciones x del problema recordando que $y = P^{-1}x$.

DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Métodos Numéricos modo virtual
(pandemia COVID-19)
Material Complementario

CML - SVD - versión 1.0

Este es material complementario de las diapos de la clase de cuadrados mínimos usadas durante el dictado virtual (pandemia COVID-19). En este documento deducimos la metodología de solución basada en la factorización SVD.

Dada $A \in \mathbb{R}^{m\times n}$ con $p = rango(A)$, por la descomposición SVD sabemos que existen $U \in \mathbb{R}^{m\times m}$ matriz ortogonal, $V \in \mathbb{R}^{n\times n}$ matriz ortogonal y $\Sigma \in \mathbb{R}^{m\times n}$ tal que $A = U\Sigma V^t$ con

$$\Sigma = \begin{bmatrix} \sigma^1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^p & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

con $\sigma^1 \geq \sigma^2 \geq, \ldots, \geq \sigma^p > 0$.
Veamos como podemos usar esta factorización para resolver el problema de cuadrados mínimos lineales.
Sabemos que multiplicar por una matriz ortogonal no cambia el valor de la norma 2. Entonces

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ||U^t(Ax-b)||_2^2 = \min_{x\in\mathbb{R}^n} ||U^tAx - U^tb)||_2^2 = \min_{x\in\mathbb{R}^n} ||\Sigma V^t x - U^t b)||_2^2 =$$

Notemos $V^t x = y$ y $U^t b = (b^{(p)}, b^{(m-p)})$. Analizamos dos casos:

- A tiene columnas li, por lo tanto $p = n$.

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ||\Sigma V^t x - U^t b)||_2^2 \underset{\substack{\downarrow \\ V^t x = y}}{=} \min_{y\in\mathbb{R}^n} ||\Sigma y - b^{(n)}||_2^2 + ||b^{(m-n)}||_2^2$$

  El segundo término es una constante y el primer término lo mínimo que puede valer es 0. Como Σ tiene n elementos no nulos, entonces podemos afirmar que existe un único y tal que $\Sigma y - b^{(n)} = 0$. Basta tomar $y_i = \frac{b_i^{(n)}}{\sigma^i}$ para $i = 1, \ldots, n$ y recordar que $V^t x = y$ para obtener la solución del problema.

- A tiene columnas ld, por lo tanto $p < n$. Sea $\Sigma^{(p)}$ la submatriz de Σ con las primeras p filas y p columnas, $y^{(p)}$ el vector de las primeras p componentes de y y $y^{(n-p)}$ al resto.

$$\min_{x\in\mathbb{R}^n} ||Ax-b||_2^2 = \min_{x\in\mathbb{R}^n} ||\Sigma V^t x - U^t b)||_2^2 \underset{\substack{\downarrow \\ V^t x = y}}{=} \min_{y\in\mathbb{R}^n} ||\Sigma^{(p)} y^{(p)} - b^{(p)}||_2^2 + ||b^{(m-p)}||_2^2$$

El segundo término es una constante y el primer término lo mínimo que puede valer es 0. Veamos si es posible encontrar valores para $y^{(p)}$ que hagan nula la norma y de esa forma obtener el mínimo. Entonces, buscamos $y^{(p)}$ tal que $\Sigma^{(p)} y^{(p)} - b^{(p)} = 0$. Como $\sigma^i \neq 0$ para $i = 1, \ldots, p$, podemos obtener

$y_i = \frac{b_i^{(n)}}{\sigma^i}$ para $i = 1, \ldots, p$

El resto de las componente de y no influyen en el mínimo y pueden tomar valores arbitrarios. De esta manera se pueden obtener las soluciones x del problema recordando que $V^t x = y$.

DEPARTAMENTO DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Métodos Numéricos modo virtual
(pandemia COVID-19)
Material Complementario

CML - Error - versión 1.0

Este es material complementario de las diapos de la clase de cuadrados mínimos usadas durante el dictado virtual (pandemia COVID-19). En este documento hacemos un análisis del error.

Vamos a analizar la sensibilidad de la solución cuando variamos el término independiente. Queremos determinar la relación entre pequeños cambios en el vector b con los cambios en la solución. La idea es muy similar a la que vimos para sistemas lineales donde el número de condición de la matriz nos permitía establecer esta relación. En este caso vamos a tener una generalización del número de condición.

Proposición: *Sea $A \in \mathbb{R}^{m\times n}$ con $rango(A) = n$. Sean $b, \bar{b} \in \mathbb{R}^m$ y $b = b^1 + b^2$, $\bar{b} = \bar{b}^1 + \bar{b}^2$ con $b^1, \bar{b}^1 \in Im(A)$ y $b^2, \bar{b}^2 \in Nu(A^t)$. Si $b^1 \neq 0$ entonces*

$$\frac{||x^* - \bar{x}^*||_2}{||x^*||_2} = \frac{||(A^tA)^{-1}A^tb - (A^tA)^{-1}A^t\bar{b}||_2}{||(A^tA)^{-1}A^tb||_2} \leq \chi(A)\frac{||b^1 - \bar{b}^1||_2}{||b^1||_2}$$

donde $\chi(A) = ||A||_2||(A^tA)^{-1}A^t||_2$

Demostración:
Como el $rango(A) = n$, la solución del problema de cuadrados mínimos lineales es única y basado en las propiedades que vimos, sabemos que verifica:

$$x^* = (A^tA)^{-1}A^tb \qquad Ax^* = b^1$$
$$\bar{x}^* = (A^tA)^{-1}A^t\bar{b} \qquad A\bar{x}^* = \bar{b}^1$$

$$||(A^tA)^{-1}A^tb - (A^tA)^{-1}A^t\bar{b}||_2 = ||(A^tA)^{-1}A^t(b^1+b^2) - (A^tA)^{-1}A^t(\bar{b}^1+\bar{b}^2)||_2 \underset{\substack{\downarrow \\ b^2, \bar{b}^2 \in Nu(A^t)}}{=} ||(A^tA)^{-1}A^tb^1 - (A^tA)^{-1}A^t\bar{b}^1||_2$$

$$||(A^tA)^{-1}A^tb - (A^tA)^{-1}A^t\bar{b}||_2 = ||(A^tA)^{-1}A^tb^1 - (A^tA)^{-1}A^t\bar{b}^1||_2 = ||(A^tA)^{-1}A^t(b^1 - \bar{b}^1)||_2 \underset{\substack{\downarrow \\ \text{por ser norma inducida}}}{\leq} ||(A^tA)^{-1}A^t||_2||b^1 - \bar{b}^1||_2$$

Por otro lado, $Ax^* = b^1$, entonces $||b^1||_2 = ||Ax^*||_2 \underset{\substack{\downarrow \\ \text{por ser norma inducida}}}{\leq} ||A||_2||x^*||_2$

$$\frac{1}{||x^*||_2} \leq \frac{||A||_2}{||b^1||_2}$$

En conclusión tenemos las dos siguientes desigualdades:

$$||(A^tA)^{-1}A^tb - (A^tA)^{-1}A^t\bar{b}||_2 \leq ||(A^tA)^{-1}A^t||_2||b^1 - \bar{b}^1||_2$$

$$\frac{1}{||x^*||_2} \leq \frac{||A||_2}{||b^1||_2}$$

Multiplicando los términos (son todos positivos) del mismo lado de las desigualdades obtenemos:

$$\frac{||(A^tA)^{-1}A^tb - (A^tA)^{-1}A^t\bar{b}||_2}{||x^*||_2} \leq \frac{||A||_2||(A^tA)^{-1}A^t||_2||b^1 - \bar{b}^1||_2}{||b^1||_2}$$

$\blacksquare$