eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

ZIENTZIA ETA TEKNOLOGIA FAKULTATEA
FACULTAD DE CIENCIA Y TECNOLOGÍA

6. The Least Squares Problem.

Jon Asier Bárcena-Petisco (University of the Basque Country UPV/EHU) - Numerical Methods I (2023/24)

Contents on this chapter

- Introduction to the Least Square Problem.
- Obtention of the normal equations.
- Resolution of the normal equations.
- Resolution of the Least Square Problem using the QU decomposition.
- Extra: Householder reflections (not part of this year's syllabus).

6.1. Resolution of the Least Square Problem.

Least Square Problem: introduction

Let us consider $n, m \in \mathbb{N}$ with $m > n$. We consider $A \in \mathbb{R}^{m\times n}$ a matrix of rank n; that is, of linearly independent columns. Let $b \in \mathbb{R}^m$. We want to solve: $Ax = b$. However, $x \mapsto Ax$ is not be surjective; thus, it may not have a solution. However, we can still define $J(x) = \|Ax - b\|_2^2$ and pose the following questions:

- Does J has a unique minimum?
- If so, can we find a characterization of the minimum?
- If so, how can we compute the minimum?

Finding the minimum of J is called the *Least Square Problem*.

Remark

The euclidean norm is used because it is induced by a scalar product, which will help us in many steps. With other norms, we may not even ensure the uniqueness of the minimum for the analogue problem (this is clear with the supremum norm).

Some technical results (i)

Lemma

Let $A \in \mathbb{R}^{m\times n}$, $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$. Then,

$$(Au)\cdot v = u\cdot (A'v).$$

Proof.

Both sides of the equality are: $\sum_{i=1}^{m}\sum_{j=1}^{n} u(j)A(i,j)v(i)$ □

Lemma

Let $n, m \in \mathbb{N}$ with $m > n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n. Then, $Ax \neq 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

This result was shown in the Linear Algebra course.

Some technical results (ii)

Lemma

Let $A \in \mathbb{R}^{n\times n}$. Then, the following statements are equivalent:

- *A is non-singular,*
- *the columns of A are linearly independent,*
- *$Ax \neq 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.*

This result was shown in the Linear Algebra course.

Lemma

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then, A is positive definite if and only if all its eigenvalues are strictly positive.

This result was shown in the Linear Algebra course.

Characterization of the minimum (i)

Proposition

Let $n, m \in \mathbb{N}$ with $m > n$, $A \in \mathbb{R}^{m\times n}$ a matrix of rank n and $b \in \mathbb{R}^m$. Let us consider $\overline{x} \in \mathbb{R}^n$ such that:

$$A'A\overline{x} = A'b.$$

Then, if $J(x) = \|Ax - b\|_2^2$:

$$J(\overline{x}) < J(x), \quad \forall x \in \mathbb{R}^n \setminus \{\overline{x}\}.$$

Definition

Let $n, m \in \mathbb{N}$ with $m > n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n, and $b \in \mathbb{R}^m$. The system:

$$A'Ax = A'b.$$

is called the *normal equations of the Least Square Problem*.

Characterization of the minimum (ii)

Proof.

The proposition is equivalent to showing that for all $h \in \mathbb{R}^n$:

$$J(\overline{x}) < J(\overline{x} + h), \quad \forall h \in \mathbb{R}^n \setminus \{0\}.$$

We have that:

$$\begin{aligned} J(\overline{x} + h) - J(\overline{x}) &= \|A\overline{x} + Ah - b\|_2^2 - \|A\overline{x} - b\|_2^2 \\ &= \|A\overline{x} - b\|_2^2 + 2(A\overline{x} - b) \cdot Ah + \|Ah\|_2^2 - \|A\overline{x} - b\|_2^2 \\ &= 2(A\overline{x} - b) \cdot Ah + \|Ah\|_2^2 \\ &= 2[A'(A\overline{x} - b)] \cdot h + \|Ah\|_2^2 \\ &= 2(A'A\overline{x} - A'b) \cdot h + \|Ah\|_2^2 \\ &= \|Ah\|_2^2 > 0. \end{aligned}$$

The last inequality follows from the fact that the rank of A is n. $\square$

Well-posedness of the normal equations.

Proposition

Let $n, m \in \mathbb{N}$ with $m > n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n. Then, $A'A$ is non-singular.

Proof.

Let us consider the normal equations of the Least Square Problem with $b = 0$; that is, when $J(x) = \|Ax\|_2^2$. Then, $A'A0 = A'0$. Thus,

$$0 = J(0) < J(x) = \|Ax\|_2^2 = (Ax)\cdot(Ax) = x\cdot(A'Ax), \quad \forall x \neq 0.$$

Consequently, $A'Ax \neq 0$ for all $x \neq 0$, and thus $A'A$ is non-singular. □

Well-posedness of the normal equations: a direct proof (i)

Let us now show a more direct proof for the previous result:

Proof. *Step 1: a sufficient condition.* To see that $A'A$ is non-singular, it suffices to prove that the columns of $A'A$ are free; that is, that if:

$$\lambda_1(A'A)(:,1) + \cdots + \lambda_n(A'A)(:,n) = 0,$$

then $\lambda_1, \ldots, \lambda_n = 0$.

Step 2: elements of $A'A$. We have that:

$$(A'A)(i,j) = \sum_{k=1}^{m} A'(i,k)A(k,j) = \sum_{k=1}^{m} A(k,i)A(k,j) = A(:,i) \cdot A(:,j).$$

Well-posedness of the normal equations: a direct proof (ii)

Step 3: rows of the linear combination. Let

$$B = \lambda_1 (A'A)(:,1) + \cdots + \lambda_n (A'A)(:,n).$$

Then,

$$B(i) = \lambda_1 (A'A)(i,1) + \cdots + \lambda_n (A'A)(i,n) = A(:,i) \cdot \left(\sum_{j=1}^{n} \lambda_j A(:,j) \right).$$

Step 4: simplifying the equation. Since $B(i) = 0$, we have that $\sum_{i=1}^{n} \lambda_i B(i) = 0$, which implies that:

$$\sum_{i=1}^{n} \lambda_i A(:,i) = 0.$$

Step 5: conclusion. Since the rank of A is n, its columns are linearly independent, which together with the previous equation implies:

$$\lambda_i = 0 \quad \forall i = 1, \ldots, n.$$

Thus, the columns of $A'A$ are linearly independent, so $A'A$ is non-singular. $\square$

Resolution of the least square problem

- Compute $A'A$ and $A'b$.
- Solve the problem $A'Ax = b$. This can be done, for instance, with the LUP factorization.

$A'A$ is symmetric, so a natural question arise:

Can we use Cholesky?

As we are going to see, the answer is positive because $A'A$ is positive definite.

Positive definition of $A'A$

Proposition

Let $n, m \in \mathbb{N}$ with $m > n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n. Then, $A'A$ is positive definite.

Proof.

Since $A'A$ is symmetric it can be diagonalized and all the eigenvalues are real. Thus, if λ is an eigenvalue and v its corresponding eigenvector with $\|v\|_2 = 1$, then:

$$\lambda = \lambda v \cdot v = (A'Av) \cdot v = (Av) \cdot (Av) = \|Av\|_2^2 \geq 0.$$

Moreover, since $A'A$ is non-singular $\lambda \neq 0$, so $\lambda > 0$. Consequently, $A'A$ is positive definite. $\square$

Application of the least square problem

- Function approximation
- Interpolation
- Statistics

An example is shown in Exercise 6.3.

6.2. A direct way of solving the Least Square Problem.

Some technical results

Lemma

Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal matrix and $x \in \mathbb{R}^n$. Then:

$$\|Qx\|_2 = \|x\|_2.$$

Lemma

Let $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ be orthogonal matrices. Then $Q_1 Q_2$ is an orthogonal matrix.

Both results were proved in the Linear Algebra course.

QU decomposition and least square problem

Let $n, m \in \mathbb{N}$ with $m > n$, $Q \in \mathbb{R}^{m\times m}$ be an orthogonal matrix, and $U \in \mathbb{R}^{m\times n}$ be an upper triangular matrix (a matrix with $U(i,j) = 0$ for all $i > j$) such that $U(i,i) \neq 0$ for all $i = 1, \ldots, n$. Then, if $A = QU$ and $b \in \mathbb{R}^m$:

$$\|Ax - b\|_2 = \|QUx - b\|_2 = \|Q'(QUx - b)\|_2 = \|Ux - Q'b\|_2,$$

so the minimum is obtained with the solution of the system:

$$U(1{:}n, 1{:}n)x = (Q'b)(1{:}n).$$

Since $U(1{:}n, 1{:}n)$ is non-singular, this system has a unique solution.

Remark

This decomposition is also known as the QR decomposition. U stands for upper, and R for right.

Algorithms requiring QU decomposition

The algorithms that require the QU decomposition are hard to implement by hand. In fact, you can implement them using a computer. **This year Householder reflections will not be part of the syllabus. I have left the slides in case you want to learn about Householder reflections.**

Householder reflections (i)

Proposition

Let $u \in \mathbb{R}^n$. Then, the folowing matrix:

$$H = I_n - \frac{2}{\|u\|_2^2} uu'$$

satisfies the following properties:

1. *H is symmetric.*
2. *H is orthogonal.*
3. *$Hx = x$ for all $x \in u^{\perp}$ (space of vectors orthogonal to u).*
4. *$Hu = -u$.*

Definition

Let $u \in \mathbb{R}^n$. Then, the folowing matrix:

$$H = I_n - \frac{2}{\|u\|_2^2} uu'$$

is called the *Householder matrix* induced by u.

Householder reflections (ii)

Proof.

1. The symmetry follows form $(uu')' = (u')'u' = uu'$.
2. The orthogonality of H follows from:

$$\begin{aligned} HH' = HH &= I_n I_n - \frac{4}{\|u\|_2^2} I_n uu' + \frac{4}{\|u\|_2^4} uu'uu' \\ &= I_n - \frac{4}{\|u\|_2^2} I_n uu' + \frac{4}{\|u\|_2^4} u(u'u)u' \\ &= I_n - \frac{4}{\|u\|_2^2} I_n uu' + \frac{4}{\|u\|_2^2} uu' \\ &= I_n. \end{aligned}$$

3. If $x \in u^{\perp}$, $uu'x = u(u'x) = 0$, so $Hx = x$.
4. $Hu = u - \frac{2}{\|u\|_2^2} u(u'u) = u - 2u = -u$.

□

Householder reflections for simplifying vectors

Proposition

Let $x \in \mathbb{R}^n$, *let* $u = x + \operatorname{sign}(x_1)\|x\|_2 e_1$ *and let* H *the Householder matrix induced by* u. *Then:*

$$Hx = -\operatorname{sign}(x_1)\|x\|_2 e_1$$

Proof.

Let us first compute the square of the norm of u:

$$\begin{aligned}\|u\|_2^2 = u'u &= \|x\|_2^2 + 2\operatorname{sign}(x_1)\|x\|_2 x(1) + \|x\|_2^2\|e_1\|_2^2 \\ &= 2\|x\|_2(\|x\|_2 + |x(1)|).\end{aligned}$$

Thus:

$$\begin{aligned}Hx = x - 2\frac{u'x}{\|u\|_2}u &= x - \frac{\|x\|_2^2 + \|x\|_2\operatorname{sign}(x_1)e_1'x}{\|x\|_2(\|x\|_2 + |x(1)|)}u \\ &= x - \frac{\|x\|_2(\|x\|_2 + |x(1)|)}{\|x\|_2(\|x\|_2 + |x(1)|)}u \\ &= x - u = -\operatorname{sign}(x_1)\|x\|_2 e_1.\end{aligned}$$

□

Householder reflections for computing a QU decomposition

Theorem (QU decomposition with Householder reflections)

Let $n, m \in \mathbb{N}$ with $m \geq n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n. Then, there are $Q \in \mathbb{R}^{m\times m}$ an orthogonal matrix, and $U \in \mathbb{R}^{m\times n}$ an upper triangular matrix (a matrix with $U(i,j) = 0$ for all $i > j$) with $U(i,i) \neq 0$ for all $i = 1, \dots, n$ such that $A = QU$.

As in the LUP factorization, the proof is by an induction with an algorithm that can be implemented in a computer.

Step 1: stating the induction

Proposition

Let $n, m \in \mathbb{N}$ with $m \geq n$ and $A \in \mathbb{R}^{m\times n}$ a matrix of rank n. Then, for all $k = 1, \ldots, n$ there is an orthogonal matrix Q_k and a matrix U_k such that $U_k = Q_k A$ and $U_k(i,j) = 0$ if $j \leq k$ and $i > j$.

Proof of the Theorem asuming the proposition.

It suffices to use the result with $k = n$. In that case, $A = Q_n' U_n$. Moreover, $U(i,i) \neq 0$ for all $i = 1, \ldots, n$. Otherwise, the rank of U would be at most $n-1$. Thus, the rank of A would be at most $n-1$, contradicting that it is n. $\square$

Step 2: the induction

For the base case we simply have to consider Q_1 the matrix induced by the Householder reflection of $u_1 = \tilde{u}_1 + \text{sign}(u_1(1))\|\tilde{u}_1\|_2 e_1$, for $u_1 = A(:,1) + \text{sign}(A(1,1))\|A(:,1)\|_2 e_1$.

To continue with, let us suppse that the proposition holds for k ($k \in \{1, \dots, n-1\}$), and prove it for $k+1$. Let us consider $\tilde{H}_{k+1}$ the Householder matrix induced by $u_k = \tilde{u}_k + \text{sign}(u_k(1))\|\tilde{u}_k\|_2 e_1$, for $\tilde{u}_k = Q_k A(k+1{:}m, k+1)$. Let $H_{k+1} \in \mathbb{R}^{m\times m}$ defined by:

$$H_{k+1}(i,j) = \begin{cases} 1 & i = j \leq k, \\ \tilde{H}_{k+1}(i,j) & i,j \geq k+1, \\ 0 & \text{otherwise} \end{cases}$$

Let us fix: $Q_{k+1} = H_{k+1}Q_k$. We have to prove:

- H_{k+1} is orthogonal.
- $Q_{k+1}A(i,j) = 0$ if $j \leq k+1$ and $i > j$.

Step 3: orthogononality of H_{k+1}

First, let us consider $i < j$. Then,

- If $i, j \leq k$, then $H_{k+1}(i,:) \cdot H_{k+1}(j,:) = e_i' \cdot e_j' = 0$.
- If $i \leq k < j$, then $H_{k+1}(i,:) \cdot H_{k+1}(j,:) = e_i' \cdot (0, \tilde{H}_{k+1}(j-k,:)) = 0$.
- If $i, j \geq k+1$, then
  $H_{k+1}(i,:) \cdot H_{k+1}(j,:) = \tilde{H}_{k+1}(i-k,:) \cdot \tilde{H}_{k+1}(j-k,:) = 0$.

In addition, if $i = j$, then:

- If $i \leq k$, then $H_{k+1}(i,:) \cdot H_{k+1}(i,:) = e_i' \cdot e_i' = 1$.
- If $i \geq k+1$, then $H_{k+1}(i,:) \cdot H_{k+1}(i,:) = \tilde{H}_{k+1}(i-k,:) \cdot \tilde{H}_{k+1}(i-k,:) = 0$.

Step 4: the zeros of $Q_{k+1}A$

If $j \leq k$ and $i > j$:

$$(Q_{k+1}A)(i,j) = (H_{k+1}(Q_kA))(i,j) = \sum_{r=1}^{n} H_{k+1}(i,r)(Q_kA)(r,j) = 0.$$

We have used that by the induction hypothesis $(Q_kA)(r,j) = 0$ if $r > j$ and that $H_{k+1}(i,r) = 0$ if $r < \max\{i, k+1\}$, which is the case if $r \leq j$

Finally, let us consider the case $j = k+1$. Then, if $i > k+1$:

$$\begin{aligned}(Q_{k+1}A)(i,k+1) = (H_{k+1}(Q_kA))(i,k+1) &= \sum_{r=1}^{n} H_{k+1}(i,r)(Q_kA)(r,k+1) \\ &= \sum_{r=k+1}^{n} H_{k+1}(i,r)(Q_kA)(r,k+1) \\ &= [\tilde{H}_{k+1} \cdot [(Q_kA)(k+1{:}n, k+1)]](i-k) = 0.\end{aligned}$$

We have used that $H_{k+1}(i,r) = 0$ if $r \leq k$ and the definition of $\tilde{H}_{k+1}$ as the matrix induced by Householder reflection of $(Q_kA)(k+1{:}n, k+1)$. $\square$

The algorithm

- Compute H_1, $H_1 A$, H_2, $H_2(H_1 A)$, etc.
- Compute $[(H_n H_{n-1} \cdots H_1)b](1{:}n)$.
- Solve $U(1{:}n, 1{:}n)x = [H_n H_{n-1} \cdots H_1 b](1{:}n)$.

Indeed, as $Q = (H_n H_{n-1} \cdots H_1)'$, then $Q' = H_n H_{n-1} \cdots H_1$, so we have to use the argument at the beginning of this section.

Normal equations VS QU decomposition

- QU is more stable numerically.
- In both cases the amount of computation is $O(n^3)$, so there is no difference in the time.