Linear Algebra
Week 10

G-07

28 XI 2024

This week is about the Pseudoinverse a.k.a. Moore-Penrose inverse, which tries to achieve an analogous effect to the inverse for not invertible matrices. After Pseudoinverses we will change the focus and look at projections of sets and the Farkas Lemma.

1 The Pseudoinverse

Consider the vector b such that $Ax = b$. We want to find x as always. If A is invertible, we can write $x = A^{-1}b$. Let's consider the case when A is not invertible. We want to find a matrix $A^\dagger$ that has a similar job as the actual inverse such that $A^\dagger b$ gives us x. This is achieved via the Least Squares Method as we have seen, but only if A has linearly independent columns. What if this is not the case? Before we delve into the definitions, here is some intuition about what the pseudoinverse achieves for us.

1.1 Another Point of View

Let $A \in \mathbb{R}^{m\times n}$. It helps to think of A as a linear transformation, in other words a function that maps vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$.

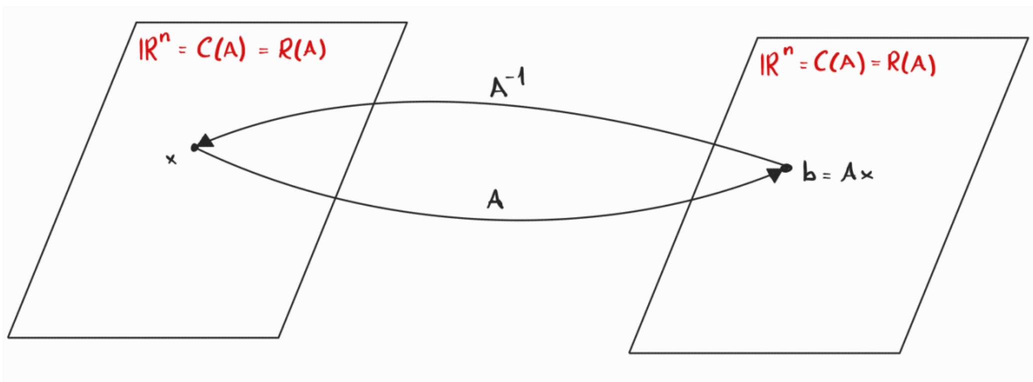
$$A : \mathbb{R}^n \mapsto \mathbb{R}^m$$
$$x \mapsto Ax$$

Now our task is to find the inverse function $A^\dagger$ that maps vectors from $\mathbb{R}^m$ back to $\mathbb{R}^n$ in a way that reverses the effect of A, if such a function exists.

$$A^\dagger : \mathbb{R}^m \mapsto \mathbb{R}^n$$
$$Ax \mapsto x$$

A is invertible

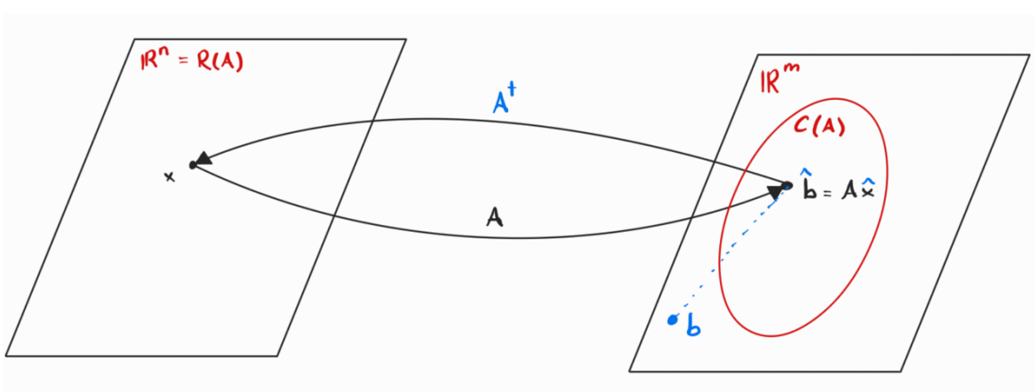
Let's visualize this for the case when A is invertible. If A is invertible then A must be square, so let $A \in \mathbb{R}^{n \times n}$.



The vector $x \in \mathbb{R}^n$ is taken to the vector b by A. Since A is invertible $rank(A) = r = n = \#columns = \#rows$. Therefore the column space and the row space of A are both equal to the vector space $\mathbb{R}^n$. Seen as a function, A is bijective. For every vector b in $\mathbb{R}^n$ you can find a vector x in $\mathbb{R}^n$ such that $Ax = b$ (surjective) and for $x_1 \neq x_2$ we have $Ax_1 \neq Ax_2$ (injective).[1]

A has independent columns

Let $A \in \mathbb{R}^{m \times n}$ again. We can visualize the mapping by the *function* A as follows:



[1]This is not a formal proof!

In this case the row space of A is equal to the whole $\mathbb{R}^n$ but the column space of A is just a subspace of $\mathbb{R}^m$. This might be counterintuitive at first but let's see why.[2] Let $rank(A) = r$. Since A has linearly independent columns we have $r = n$. Now if A has $rank(A) = r$ then we know there are at least r rows because column rank = row rank, so $r \leq m$. If $r = m$ then A would be invertible, but we handled that case[3], so observe $r < m$. The column space of A is a subspace of $\mathbb{R}^m$ and the row space of A is a subspace of $\mathbb{R}^n$ as we already know.[4] We have n linearly independent columns and n linearly independent rows because column rank is equal to the row rank. This is not enough to build a basis of $\mathbb{R}^m$ and therefore $\mathbf{C}(A) \neq \mathbb{R}^m$ but it is enough to build a basis of $\mathbb{R}^n$ and hence $\mathbf{R}(A) = \mathbb{R}^n$. In the example below where $m = 7$ and $n = 4$, A definitely has linearly independent columns and linearly dependent rows. But see its column space is not the whole $\mathbb{R}^m$, whereas the row space is the whole $\mathbb{R}^n$.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our function takes x from $\mathbb{R}^n$ and maps it only to a small subspace of $\mathbb{R}^m$. This means there will be vectors $b \in \mathbb{R}^m$ for which there exists no x such that $Ax = b$ - like the blue b in the illustration. In this case we project b onto the column space of A to get $proj_{C(A)}(b) = \hat{b}$. Then we map $\hat{b}$ back to $\hat{x}$. We find $\hat{x}$ that brings us closest to b. This is nothing else but *Least Squares*. If you observe A as a function and only take the vectors in $\mathbf{C}(A) \subset \mathbb{R}^m$ into consideration for the **Im**(A) then you have a bijective function from $\mathbf{R}(A) = \mathbb{R}^n$ to $\mathbf{C}(A)$.

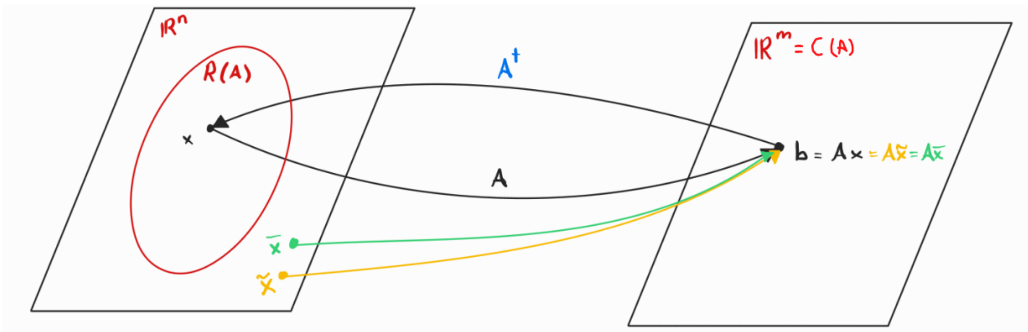
[2] The blue part is good to know and you should be able to understand this with your knowledge so far. But it is blue since it argues about an intermediate result to explain pseudoinverse and also blue looks cool.

[3] What we will do is also valid if A is invertible, but the focus here is to see what $A^\dagger$ does for us when we don't have any actual inverse.

[4] See sections 4.3.1 and 4.3.2 in the notes from first half.

A has independent rows

Let $A \in \mathbb{R}^{m \times n}$. We can visualize the mapping done by the *function* A as follows:



With a similar argumentation as above, we can see that the column space is equal to the whole $\mathbb{R}^m$ and the row space of A is just a subspace of $\mathbb{R}^n$ in the case where A has linearly independent rows.[5] Here the issue is that there exists multiple x such that $Ax = b$. We have to pick one $\tilde{x}$ so that $A^\dagger b = \tilde{x}$. So we pick the one with the smallest norm by multiplying b with $A^\dagger$. **Lemma 5.5.5.** clarifies for us that the vector x in $Ax = b$ has the smallest norm when $x \in \mathbf{R}(A) = \mathbf{C}(A^\top)$. If A has linearly independent rows but linearly dependent columns, A has a non trivial (= not just $\mathbf{0}$) nullspace. Then the solution space for $Ax = b$ is a particular solution + a vector from $\mathbf{N}(A)$ as we have seen in **Theorem 4.39** in the first half. From the solution space we take the vector with the smallest norm, *i.e.* the one closest to the origin which is the particular solution.

A does not have full column or row rank

Then we have both problems as above. If you choose a random b that can be outside of $\mathbf{C}(A)$ since $\mathbf{C}(A)$ is not the whole $\mathbb{R}^m$ but even if you choose some $b \in \mathbf{C}(A)$ then there will be more than one x such that $Ax = b$. This is why we decompose A into C and R in this case where C has full column rank and R has full row rank. After we write the *function* A into two functions we can apply what we have done above.

[5] You can play the $\mathbf{R}(A) = \mathbf{C}(A^\top)$ card here and argue based on that.

1.2 Definitions

First 4 pages can be seen as an extended look into **Proposition 5.5.12.** that says if you view A as a function from the row space of A to the column space of A, then it is a bijection. Let's see other findings from the chapter briefly.

We have seen that the pseudoinverse has the same effect as applying the least squares method, if the columns of A are linearly independent. So in this case we define $A^\dagger$ such that $A^\dagger b$ is the least squares solution.

Definition 5.5.1. ($A^\dagger$ for matrices with full column rank). For $A \in \mathbb{R}^{m\times n}$ with $rank(A) = n$ we define the pseudoinverse $A^\dagger \in \mathbb{R}^{n\times m}$ of A as

$$A^\dagger = (A^\top A)^{-1} A^\top$$

Proposition 5.5.2. tells us that in this case where A has linearly independent columns, $A^\dagger$ as we defined it above is a *left inverse of A*. We have $A^\dagger A = I$. (Multiply & see!)

If A has full row rank, then $A^\top$ has full column rank. Exploiting this fact, we can find the pseudoinverse of $A^\top$ if A itself has linearly independent rows.

$$(A^\top)^\dagger = ((A^\top)^\top (A^\top))^{-1} (A^\top)^\top = (AA^\top)^{-1} A$$

This is the pseudoinverse of $A^\top$. Since $A^\top$ has linearly independent columns $(A^\top)^\dagger$ is the left inverse so we have

$$(A^\top)^\dagger A^\top = I$$

If we transpose both sides

$$((A^\top)^\dagger A^\top)^\top = A((A^\top)^\dagger)^\top = I^\top = I$$

As we can see here the pseudoinverse of A is then

$$((A^\top)^\dagger)^\top = ((AA^\top)^{-1} A)^\top = A^\top (AA^\top)^{-1}$$

Definition 5.5.3 ($A^\dagger$ for matrices with full row rank). For $A \in \mathbb{R}^{m\times n}$ with $rank(A) = m$ we define the pseudoinverse $A^\dagger \in \mathbb{R}^{n\times m}$ of A as

$$A^\dagger = A^\top (AA^\top)^{-1}$$

Lemma 5.5.4. tells us that in this case where A has linearly independent rows, $A^\dagger$ as we defined it above is a *right inverse of* A. We have $AA^\dagger = I$. (Multiply & see!)

The lecture notes have **Lemma 5.5.5.** and **Lemma 5.5.6.** at this point to explain what the pseudoinverse achieves for us in the case where A has full row rank. We made a similar argumentation above when we handled this case but it is important that you read and understand that part of the lecture notes to see what $A^\dagger$ does if A has full row rank. Summarizing again, if A has full row rank and is not invertible, then it has more columns than its rank which makes its nullspace nontrivial. Therefore there exists infinitely many solutions to $Ax = b$. They are all in form *(the particular solution + something in* $\boldsymbol{N}(A)$*)*. If you cut out the part in $\mathbf{N}(A)$ you are left with the solution with the smallest norm which is the particular solution and that is always in $C(A^\top)$. See **Theorem 4.39** and further explanations there.

Definition 5.5.7 ($A^\dagger$ for all matrices). For $A \in \mathbb{R}^{m\times n}$ with $rank(A) = r$ and CR decomposition $A = CR$ we define the pseudoinverse $A^\dagger$ as

$$A^\dagger = R^\dagger C^\dagger$$

which can be rewritten as

$$A^\dagger = R^\top (RR^\top)^{-1}(C^\top C)^{-1}C^\top = R^\top (C^\top CRR^\top)^{-1}C^\top = R^\top (C^\top AR^\top)^{-1}C^\top$$

Lemma 5.5.8 Given $A \in \mathbb{R}^{m\times n}$ and vector $b \in \mathbb{R}^n$, the (unique) solution to

$$\min_{x\in\mathbb{R}^n} \|x\|^2 \quad \text{s.t.} \quad A^\top Ax = A^\top b$$

is given by $\hat{x} = A^\dagger b$.

You can prove the equality by direct calculation. To formally prove that x also has minimum norm you should prove that $x \in \mathbf{C}(A^\top A)$ using **Lemma 5.5.5.**

We used CR decomposition of A to define the pseudoinverse in the general case. We can actually use any other full rank factorization as well. There is nothing special about CR.

Proposition 5.5.9. For $A \in \mathbb{R}^{m\times n}$, with $\text{rank}(A) = r$, and let $S \in \mathbb{R}^{m\times r}$ and $T \in \mathbb{R}^{r\times n}$ such that $A = ST$. Then,

$$A^\dagger = T^\dagger S^\dagger.$$

Remark 5.5.10. tells us: if $A = ST$ and $rank(A) = r$, then $rank(S) \geq r$ and $rank(T) \geq r$. So the matrices S/T in the factorization $A = ST$ above are in fact full column/row rank. The reason is that the matrix multiplication ST can be seen either as the linear combinations of the columns of S or the linear combinations of the rows of T. If there were less than r linearly independent columns of S then you could not generate r linearly independent combinations using those columns, so A could not have r linearly independent columns. And similarly, if there were less than r linearly independent rows of T, you could not have r linearly independent rows in A. Now read this once more and keep in mind that column rank = row rank.

There are useful properties of pseudoinverses that you should keep in mind (or on your cheatsheet). You can find all proofs in lecture notes under **Theorem 5.5.11** or in the solution of assignment 10 exercise 1.
Theorem 5.5.11 Let $A \in \mathbb{R}^{m\times n}$

1. $AA^\dagger A = A$
2. $A^\dagger A A^\dagger = A^\dagger$
3. $AA^\dagger$ is symmetric. It is the projection matrix for projection on $\mathbf{C}(A)$
4. $A^\dagger A$ is symmetric. It is the projection matrix for projection on $\mathbf{C}(A^\top)$
5. $(A^\top)^\dagger = (A^\dagger)^\top$
6. (Asgn 10.1.(1)) If $rank(A) = rank(B) = n$, we have $(AB)^\dagger = B^\dagger A^\dagger$

2 Projection of Sets

So far we have projected points but what about projecting sets? Before we go on and project sets, it is important to know how we represent sets of points in our vector space. In this chapter we deal with sets of points given by inequalities in $\mathbb{R}^n$.

(Not in lecture material) If you have two vectors $x, y \in \mathbb{R}^n$ then

$$x \leq y \iff \begin{array}{c} x_1 \leq y_1, \\ x_2 \leq y_2, \\ \vdots \\ x_n \leq y_n \end{array}$$

The comparison must hold element wise. In particular, for a vector $x \in \mathbb{R}^n$ it holds that $\mathbf{0} \leq x$ if and only if all elements of x are non-negative.

We will look at more than one linear inequality by considering systems of linear inequalities in form of $Ax \leq b$. The set of vectors -or points- satisfying this type of systems of linear inequalities has a special name: polyhedron.

Definition 5.6.1. Let $A \in Q^{m \times n}, b \in Q^m$ and $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. P is called a *polyhedron*. Let $S = \{1, \ldots, s\}$. The **projection of P** on the subspace $\mathbb{R}^s$ associated with the variables in the subset S is

$$proj_S(P) := \{x \in \mathbb{R}^s \mid \exists y \in \mathbb{R}^{n-s} \text{ such that} (x, y) \in P\}$$

(x, y) means we have a tuple of two vectors. This makes more sense when you see the polyhedron in this point of view (*not* from lecture notes):

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \{(x, y) \in \mathbb{R}^s \times \mathbb{R}^{n-s} \mid A_x x + A_y y \leq b\}$$

and if you want to project P then you want to find the polyhedron P_x that is defined by A_x and b_x such that

$$P_x = \{x \in \mathbb{R}^n \mid A_x x \leq b_x\}$$

An example of this P_x, the projected polyhedron, is the polyhedron Q in the proof of Theorem 5.6.3.

2.1 One Dimensional Example

Let's look at the 1 dimensional to 0 dimensional projection example in the lecture notes. 1 dimensional means the A in our system of inequalities has only one column, i.e. it is a vector. And 0 dimensional means 0. So concretely we project the polyhedron of form $ax \leq b$ where $a, b \in \mathbb{R}^n$ are vectors and $x \in \mathbb{R}$ is just a scalar, to $0 \leq c$ for some scalar $c \in \mathbb{R}$.

Let $a \in \mathbb{Q}^m$, $a_i \neq 0$ for all i, and $b \in \mathbb{Q}^m$. Consider $P = \{x \in \mathbb{R} \mid ax \leq b\} \subseteq \mathbb{R}$. This means

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} x \leq \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \text{ that is } \begin{matrix} a_1 x \leq b_1, \\ a_2 x \leq b_2, \\ \vdots \\ a_n x \leq b_n \end{matrix} \text{ we can write } \begin{matrix} x \leq or \geq \frac{b_1}{a_1}, \\ x \leq or \geq \frac{b_2}{a_2}, \\ \vdots \\ x \leq or \geq \frac{b_n}{a_n} \end{matrix}$$

Be careful, a_i can be negative for some i. This is why the equality sign at a given line is $\leq$ if a_i is positive and $\geq$ if a_i is negative. As you can see we have many upper and lower bounds for x. The most restrictive ones are the minimum of the upper bounds and the maximum of the lower bounds. We name them u and l respectively. Set

$$u := \min \left\{ \frac{b_i}{a_i} \mid a_i > 0 \right\}, \quad l := \max \left\{ \frac{b_i}{a_i} \mid a_i < 0 \right\}. \text{ Then}$$

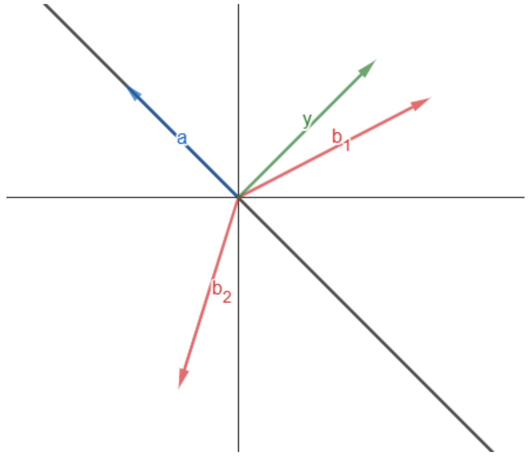
$$P = \{x \in \mathbb{R} \mid x \leq \frac{b_i}{a_i} \text{ if } a_i > 0,\ x \geq \frac{b_i}{a_i} \text{ if } a_i < 0\} = \{x \in \mathbb{R} \mid x \leq u,\ x \geq l\}.$$

Now we want to determine if this condition is feasible, or in other words can we find x that can satisfy the conditions $l \leq x \leq u$? This is the case if the upper bound is actually greater than the lower bound: $l \leq u$. We can now use this as a certificate to see if our initial polyhedron is empty or not. This is **Proposition 5.6.2.**:

$$P \neq \emptyset \iff l \leq u \iff 0 \leq u - l \iff 0 \leq y^\top b \text{ for all } y \geq 0 \text{ s.t. } y^\top a = 0$$

The last statement needs some explanation. It is actually a special case of the famous Farkas Lemma. So we explain what the Farkas Lemma is trying to achieve for us. $0 \leq y^\top b$ for all $y \geq 0$ s.t. $y^\top a = 0$ means we can find a vector y that is orthogonal to the line defined by the vector a such that

$0 \leq y^\top b$. We can write $\{ax \mid x \in \mathbb{R}\} = \{v \mid v^\top \cdot y = 0\}$. So the line defined by our vector a is a hyperplane, in which all vectors are orthogonal to y. Now if $0 \leq y^\top b$ this would mean b is on the *correct* side of the line defined by a.



In this visualization for $a, b_1, b_2, y \in \mathbb{R}^2$ you can see the line that represents all possible ax for $x \in \mathbb{R}$. We have "for all $0 \leq y$" in the above expression so y is in the positive quarter of the 2 dimensional space. Also see that y is orthogonal to the line ax. You can choose x such that $ax \leq b_1$. In our case this means that both coordinates of ax are smaller or equal to b_1 and yes there exists such a point on the gray line. But there is no x such that $ax \leq b_2$. No matter how you choose x, at least one of the coordinates of ax will be greater than the corresponding coordinate of b_2. In one sentence, we define a **separating hyperplane** using y that separates the b's for which there exist a solution to $ax \leq b$, from the b's for which there doesn't exist a solution for $ax \leq b$.

2.2 Elimination of One Variable

Now we will see the general case where we project from n dimensions to $n-1$ dimensions by eliminating one variable. Throughout this section let $A \in \mathbb{Q}^{m \times n}, b \in \mathbb{Q}^m$ and $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$.

$$\text{Our A is} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \text{ we define } \bar{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1,n-1} \\ a_{21} & a_{22} & \dots & a_{2,n-1} \\ & \vdots & & \\ a_{m1} & a_{m2} & \dots & a_{m,n-1} \end{bmatrix}$$

as the first n-1 columns of A. Similarly $\bar{x} = \begin{bmatrix} x_1 & , \dots , & x_{n-1} \end{bmatrix}^\top$ denotes the first $n-1$ elements of a vector x. Our inequality $Ax \leq b$ looks like the following:

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m \end{gathered}$$

Step 1. Partition the indices $M = \{1, \dots, m\}$ of the rows of the matrix A into three subsets:

$$M_0 = \{i \in M \mid a_{i,n} = 0\}, \quad M_+ = \{i \in M \mid a_{i,n} > 0\}, \quad M_- = \{i \in M \mid a_{i,n} < 0\}.$$

After you partitioned the rows based on the sign of the elements in the last column of A you continue.

Step 2. You want to have x_n on one side of the inequality and everything else on the other side. For the equations where there is no x_n term because $a_n = 0$ you want to have 0 on one side and everything else on the other side. You do this separately for M_+, M_- and M_0 because dividing by a negative number flips the inequality and you can't divide by zero.

- M_+**:** Let's assume the i-th row is in M_+. Then $0 \leq a_{in}$ and we have

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i \tag{1}$$

  You can leave x_n on one side of the inequality alone by subtracting $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{i,n-1}x_{n-1}$ from both sides and then dividing by a_{in}.

$$x_n \leq \frac{b_i}{a_{in}} - \frac{a_{i1}}{a_{in}}x_1 - \frac{a_{i2}}{a_{in}}x_2 - \cdots - \frac{a_{i,n-1}}{a_{in}}x_{n-1} \tag{2}$$

  We can say the two conditions (1) and (2) are equivalent.

- M_-**:** Now we do the same thing for another row let's call it the l-th row. But this row is in M_- so the coefficient a_{ln} is negative. Therefore we have to flip the inequality sign. The inequality

$$a_{l1}x_1 + a_{l2}x_2 + \cdots + a_{ln}x_n \leq b_l \tag{3}$$

becomes the following:

$$\frac{b_l}{a_{ln}} - \frac{a_{l1}}{a_{ln}} x_1 - \frac{a_{l2}}{a_{ln}} x_2 - \ldots - \frac{a_{l,n-1}}{a_{ln}} x_{n-1} \leq x_n \tag{4}$$

- M_0**:** Again the same process. Let the k-th row be in M_0 This means you have $a_{kn} = 0$ now. So we have

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{k,n-1}x_{n-1} + 0x_n \leq b_k \tag{5}$$

  We write it slightly different so that we have 0 on one side and everything on the other:

$$0 \leq b_k - a_{k1}x_1 - a_{k2}x_2 - \ldots - a_{k,n-1}x_{n-1} \tag{6}$$

Step 3. After the 2. step we have a lot of upper bounds for x_n as in inequalities of type (2) which were initially in M_+. Also we have lots of lower bounds for x_n as in the inequalities of type (4) which were initially in M_-. This is the critical step: ***If our system of inequalities is feasible then all the lower bounds for x_n are smaller than all the upper bounds.***

Since we have variables we can't just take the maximum of the lower bounds and minimum of the upper bounds as in the one dimensional example. Therefore we compare them pairwise. For this there is a notation in the lecture notes. It basically packs up the inequalities we have above into vectors. The constants d are for the constant terms which are shown in red above. The vectors f are for the coefficients of $x_1, \ldots, x_{n-1}$ which are shown in blue above (the - signs are included). So

- equations in M_+ like (2) become $x_n \leq d_i + f_i^\top \bar{x}$
- equations in M_- like (4) become $d_l + f_l^\top \bar{x} \leq x_n$
- equations in M_0 like (6) become $0 \leq d_k + f_k^\top \bar{x}$

We can eliminate x_n and use the fact that all lower bounds should be less than all upper bounds. We say following must hold: $d_l + f_l^\top \bar{x} \leq d_i + f_i^\top \bar{x}$ for all $l \in M_-$ and for all $i \in M_+$. So we return

$$\begin{aligned} Q = \{\bar{x} \in \mathbb{R}^{n-1} \mid 0 \leq d_k + f_k^T \bar{x} \quad & \text{for all } k \in M_0, \\ d_l + f_l^T \bar{x} \leq d_i + f_i^T \bar{x} \quad & \text{for all } l \in M_-,\ i \in M_+\} \end{aligned}$$

Congratulations! You have successfully projected the polyhedron from n dimensions to n-1 dimensions. Theorem 5.6.3. tells that the returned set Q is indeed a polyhedron because we can write it as $Q = \{\bar{x} \in \mathbb{R}^{n-1} \mid F\bar{x} \leq \delta\}$, which is how we defined a polyhedron.

An important note is that *you can choose any variable to eliminate*, so it doesn't necessarily need to be the last one. And if you have to go through this process manually then you might want to choose the variable with less work. In the proof of Theorem 5.6.3. you see that the number of inequalities $k = |M_0| + |M_-||M_+|$. This is because we compare the lower bounds and upper bounds pairwise and carry on the inequalities in M_0 as they are.

Let $S_1 = \{1, \ldots, n-1\}$ and $S_2 = \{1, \ldots, n-2\}$. Lemma 5.6.4. tells us that projecting a polyhedron using S_1 first and S_2 second in separate steps is actually equivalent to projecting directly using S_2.

3 The Farkas Lemma

Before we see the Farkas Lemma let's develop a compact algebraic notation for the projection process above.

3.1 Compact Algebraic Notation for Projection of Sets

Definition 5.6.5 Let $A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^m$ and $P = \{x \in \mathbb{R}^n \mid\}$. For $k \in \{1, \ldots, j\}$ let $A^{(j)}$ to be the submatrix of A with column vectors $A_{.k}$ Let $P^{(0)} = P$ and $C^{(0)} = \mathbb{R}^m_+$. Define for $i \in \{1, \ldots, n\}$

$$C^{(i)} = \left\{ y \in \mathbb{R}^m_+ \;\middle|\; y^\top A_{.k} = 0 \text{ for all } k = n-i+1, \ldots, n \right\}.$$

$$P^{(i)} = \left\{ \bar{x} \in \mathbb{R}^{n-i} \;\middle|\; y^\top A^{(n-i)}\bar{x} \leq y^\top b \text{ for all } y \in C^{(i)} \right\}.$$

First of all $A_{.k}$ means the k-th column not the submatrix. So if we let $a_1, \ldots, a_n$ be the columns of A, we can write $C^{(i)}$ as:

$$C^{(i)} = \{y \in \mathbb{R}^m_+ \mid y^\top a_k = 0 \text{ for all } k = n-i+1, \ldots, n\}$$

This means for an elimination step we have all the vectors $y \in \mathbb{R}^m_+$ that zero out the last i columns of our matrix A in the set $C^{(i)}$.

Since y is not negative, it does not flip the sign of the inequality for arbitrary rows. Therefore if we have $Ax \leq b$, we can multiply with $y^\top$ from left on both sides of the inequality. We get $y^\top Ax \leq y^\top b$. But since we choose our y from the set $C^{(i)}$, we know that $y^\top a_j = 0$ for the last i columns of A will be 0. We can just ignore the last i columns and observe the submatrix $A^{(n-i)}$, the matrix A with the first $n-i$ columns. We also have to adjust x accordingly. We crop the last i entries because they would be multiplied with 0 anyways. Which leaves us with $y^\top A^{(n-i)}\bar{x} \leq y^\top b$.

How does this correspond to what we have done above? For this we have a theorem and a proof.

Theorem 5.6.6. $proj_{S_{n-i}}(P) = P^{(i)}$

As always we show two directions. First look at $proj_{S_{n-i}}(P) \subseteq P^{(i)}$. So let $\bar{x} \in proj_{S_{n-i}}(P)$ which means that there exists a real vector z of dimension i, which complements the vector $\bar{x}$ such that $(\bar{x}, z) \in P$. This is the definition 5.6.1. This means for $(\bar{x}, z)$ we have

$$P = \{(\bar{x}, z) \in \mathbb{R}^n \times \mathbb{R}^{n-i} \mid A_k x + A_z z \leq b\}$$

or in a more visual way:

$$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_{n-i} \\ | & | & & | \end{bmatrix} \bar{x} + \begin{bmatrix} | & | & & | \\ a_{n-i+1} & a_{n-i+1} & \dots & a_n \\ | & | & & | \end{bmatrix} z \leq b$$

If you choose $y \in C^{(i)}$ then by definition $y^\top a_j = 0$ for the last i columns of A. So we have:

$$y^\top \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_{n-i} \\ | & | & & | \end{bmatrix} \bar{x} + y^\top \begin{bmatrix} | & | & & | \\ a_{n-i+1} & a_{n-i+1} & \dots & a_n \\ | & | & & | \end{bmatrix} z \leq y^\top b$$

$$\iff y^\top \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_{n-i} \\ | & | & & | \end{bmatrix} \bar{x} + \mathbf{0} z \leq y^\top b$$

But this is just the set $P^{(i)}$ written in a slightly different way. If you write $A^{(n-i)}$ for the submatrix of A with the first $n-i$ columns, then you have $y^\top A^{(n-i)}\bar{x} \leq y^\top b$. This is how we defined $P^{(i)}$. We can conclude $\bar{x} \in P^{(i)}$.

We proved $proj_{S_{n-i}}(P) \subseteq P^{(i)}$, now we want to prove $P^{(i)} \subseteq proj_{S_{n-i}}(P)$. This is an induction proof, but here we will go over the base case just as in the lecture notes. You can find the rest of the proof in the solutions of assignment 10. Let a_n be the last column of A. We observe

$$C^{(1)} = \{y \in \mathbb{R}^m_+ \mid y^\top a_n = 0\},\ P^{(1)} = \{\bar{x} \in \mathbb{R}^{n-1} \mid y^\top A^{(n-1)}\bar{x} \leq y^\top b \text{ for all } y \in C^{(1)}\}$$

We want to show $P^{(1)} \subseteq proj_{S_{n-1}}(P)$. The proof in the lecture notes argues that any vector y that zeroes out the last column of A, which is a_n, when multiplied $y^\top a_n = 0$, should either pick rows from M_0 or add some rows from M_- and some rows from M_+ to get 0. This is done by choosing y as a standard unit vector. Because if you multiply a matrix A with e_i from left as in $e_i^\top A$, then you get the i-th row. This is what we describe in Q from Theorem 5.6.3. Try writing the matrix as rows of inequalities as in page 11 above and multiply it with the proposed vector from the lecture notes $-\frac{1}{a_{ln}}e_l + \frac{1}{a_{in}}e_i \in C^{(1)}$ for $l \in M_-$ and $i \in M+$ to see why this vector actually yields zero when multiplied with the last column of A.

3.2 The Farkas Lemma

Theorem 5.6.7. (The Farkas Lemma) Let $A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^n$. Either there exists a vector $x \in \mathbb{R}^n$ such that $Ax \leq b$ or there exists a vector $y \in \mathbb{R}^m$ such that $y \geq 0, y^\top A = 0$ and $y^\top b < 0$.

Proof: Now we can use the algebraic notation from above. We will look at the case where we eliminated all variables and only left with a logical statement. This is the case when we look at $C^{(n)}$ and $P^{(n)}$ for a matrix A with n columns.

$$C^{(n)} = \{y \in \mathbb{R}^m_+ \mid y^\top a_j \text{ for all } j = 1, \ldots, n\} = \{y \geq 0 \mid y^\top A = 0\}$$

which means y is now orthogonal to all columns of A, *i.e.* y is in the Left Nullspace of A. Our set $P^{(n)}$ becomes

$$P^{(n)} = \{0 \leq y^\top b \text{ for all } y \in C^{(n)}\}$$

See that there is no x in the condition since $y^\top A\bar{x}$ is 0 and we only write 0. So this is just a simple logical argument. If all vectors in $C^{(n)}$ fulfill this condition then it is true otherwise it is false. Quting from the lecture notes:

"We conclude with Proposition 5.6.2. that

$$P \neq \emptyset \iff P^{(1)} \neq \emptyset \iff \cdots \iff P^{(n)} \neq \emptyset$$
$$\iff \text{for all } y \geq 0 \text{ with } y^\top A = 0 \text{ we have } y^\top b \geq 0.$$

This leads to the following conclusion. Either $P \neq \emptyset$ or $P = \emptyset$. This is equivalent to saying: either there exists a vector $x \in \mathbb{R}^n$ such that $Ax \leq b$ or there exists a vector $y \in \mathbb{R}^m$ such that $y \geq 0$, $y^\top A = 0$ and $y^\top b < 0$."

□

Geometrically this means that there is a ***separating hyperplane*** defined by y as explained above on page 9-10 with 1 dimensional example. The column space of A is in this hyperplane, since y is in the Left Nullspace of A. The property $0 \leq y^\top b$ or $y^\top b < 0$ determines on which side of this hyperplane the vector b lies. If it lies "above" or inside the hyperplane, then we can satisfy $Ax \leq b$. In this way the Farkas Lemma gives us a way to determine whether or not a polyhedron $P = \emptyset$ or not. It proposes a certificate to see if our system of inequalities is feasible.

4 Hints

1. Solved in class. If you have to prove something about the pseudoinverse proving it with CR decomposition proves it for all matrices. But sometimes you want to prove it for the matrices with linearly independent columns and for the matrices with linearly independent rows separately so that you can then argue for C and R which are also a matrix with linearly independent columns and a matrix with linearly independent rows respectively.

2. See the section *1.1 Another Point of View*, and use the fact that $A^\dagger A$ is the projection matrix that projects vectors onto the Row Space of A which is equal to $C(A^\top)$.

3. You should first prove that $proj_{S_{n-j}}(P)$ is also a polyhedron (by induction). Then you can use the definition of the projection of a polyhedron which consists of all z such that (z, x) is in P. So you take elements $z_1, \dots, z_{n-j}$ and say there exists a vector x that completes this vector z such that $(z, x) \in P$. You then do this piecewise for $n - j$ and $n - k$.

4. We want to do the induction step and show $P^{(i)} = proj_{S_{n-i}}(P)$ for all polyhedra $P \subseteq \mathbb{R}^n$ with some $n \geq i$. Think of a proper induction hypothesis and observe $P' = proj_{S_{n-i+1}}(P) \subseteq R^{n-i+1}$.

5. Think about the polygons defined by the inequalities $A_1\mathbf{x} \leq \mathbf{b}_1$ FOR P_1 and $A_2\mathbf{x} \leq \mathbf{b}_2$ for P_2. Now consider this system: $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \mathbf{x} \leq \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$. Does this one have a solution/is it satisfiable? What does Farkas Lemma imply in that case?

mkilic

♠