SYSTEMS OF LINEAR EQUATIONS CAN BE REGARDED AS PROJECTIONS

RAMON MASSONI AND FERRAN MUIÑOS

ABSTRACT. Solving a system of linear equations is equivalent to conducting an orthogonal projection in a sense that will be made precise below.

1. Problem

1.1. Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, find a vector $v \in \mathbb{R}^n$ such that $Av = b$.

1.2. This problem has a solution if and only if $\mathrm{rank}(A) = \mathrm{rank}(A \,|\, b)$, or equivalently, if $b \in \mathrm{span}\{A_1, \ldots, A_n\}$.

1.3. If A is a full-rank, square matrix, then A^{-1} exists and the solution can be given as follows $\hat{v} = A^{-1}b$.

1.4. The entries of a solution vector $\hat{v} = (v_1, \ldots, v_n)$ can be thought of as a concrete way of carrying out a linear combination of the column vectors of A.

1.5. What if $\mathrm{rank}(A) < \mathrm{rank}(A \,|\, b)$? In this case $b \notin \mathrm{span}\{A_1, \ldots, A_n\}$, i.e., b cannot be given as a linear combination of the column vectors of A.

1.6. One possible way of relaxing the problem in the later situation would go as follows: we no longer ask for a solution vector $\hat{v}$ such that $A\hat{v} = b$, but such that $A\hat{v} - \pi(b) = 0$, with $\pi(b)$ being some convenient transform of b such that:

    i $\pi(b)$ is not too far from b;
    ii $\mathrm{rank}(A) = \mathrm{rank}(A \,|\, \pi(b))$.

1.7. Whenever they exist, the solutions of the strict version of the problem should be solutions for the more relaxed version, although we cannot expect the converse to be true in general.

2. Orthogonal Projection

2.1. A linear map $\pi_v : V \to V$ is deemed an orthogonal projection onto a vector v if it maps v onto itself and any vector orthogonal to v to zero.

2.2. What is the matrix of π_v? It depends on the basis of V we employ to represent vector coordinates. Fixing an orthonormal basis of V, the matrix of π_v is simply the rank one square matrix given by uu^t, where $u = v/\|v\|$.

2.3. Let S be a vector subspace of V. We can define the orthogonal projection onto S as the linear map $\pi_S : V \to V$ that maps any vector of S to itself and any vector orthogonal to S to zero.

2.4. Suppose that $S = \mathrm{span}\{u_1, \ldots, u_n\}$ and let's further assume that the vectors $u_1, \ldots, u_n$ are an orthogonal set. Then the orthogonal projection onto S is given by the following linear map:

$$(2.1) \qquad \pi_S = \sum_{i=1}^n \pi_{u_i} = \sum_{i=1}^n \frac{1}{\|u_i\|^2} u_i u_i^t.$$

2.5. First, we must check that any vector v of S is mapped to itself by π_S thus defined. Since v belongs to S we can write $v = \lambda_1 u_1 + \ldots + \lambda_n u_n$. Then

$$
\begin{aligned}
\pi_S(v) &= \sum_{i=1}^n \frac{1}{\|u_i\|^2} u_i u_i^t v \\
&= \sum_{i=1}^n \frac{\lambda_i}{\|u_i\|^2} u_i u_i^t u_i = \sum_{i=1}^n \frac{\lambda_i \|u_i\|^2}{\|u_i\|^2} u_i \\
&= \sum_{i=1}^n \lambda_i u_i = v.
\end{aligned}
$$

2.6. Second, any vector v orthogonal to S is mapped to zero by π_S. To justify that note that if v is orthogonal to S, it is orthogonal to all its generators.

2.7. The former projection π_S as written in equation 2.1 can be written in matrix form as $A(A^t A)^{-1} A^t$, where A denotes the matrix with $u_1, \ldots, u_n$ as column vectors.

2.8. Note that if the vectors $u_1, \ldots, u_n$ were an orthonormal set, then the matrix expression for π_S would be just AA^t, since the middle factor $A^t A$ would be the identity matrix.

2.9. Remark that if a vector u is in the null space of π_S, then it must be orthogonal to S. Can you tell why?

2.10. Now assume that the column vectors of A are linearly independent, but not necessarily orthonormal. Then we know there is a full-rank, square matrix E such that the matrix $Q = AE$ has orthonormal vectors as columns. Then the projection π_S has the following matrix expression:

$$
\begin{aligned}
QQ^t &= Q(Q^t Q)^{-1} Q^t = (AE)((AE)^t(AE))^{-1}(AE)^t \\
&= AE(E^t A^t AE)^{-1} E^t A^t = AEE^{-1}(A^t A)^{-1}(E^t)^{-1} E^t A^t \\
&= A(A^t A)^{-1} A^t
\end{aligned}
$$

2.11. In other words, the matrix of the orthogonal projection onto a vector subspace S does not depend on the choice of a linearly independent set spanning S. Take any matrix A with a linearly independent generating set of S as columns, then the matrix of the orthogonal projection onto S is just $A(A^t A)^{-1} A^t$.

3. BACK TO THE PROBLEM

3.1. Can the reader suggest one possible transform that serves the purpose of solving the linear system problem with the requirements referred to in section 1.6? The orthogonal projection onto the column space of A will do. Let's denote this linear map as π_A.

3.2. The orthogonal projection of b onto the column space of A can be given as $\pi_A(b) = A(A^t A)^{-1} A^t b$. In view of this expression, we can already conclude that the vector $\hat{v} = (A^t A)^{-1} A^t b$ is a solution to the problem, i.e., $A\hat{v} = \pi_A(b)$. If A has linearly independent columns, then $\hat{v}$ is the unique solution.

3.3. Last but not least, it remains to be discussed what the "not too far" part from the requirements in section 1.6 is achieved with the π_A projection.

3.4. Given a vector subspace $S \subset V$ and any vector $v \in V$, then the orthogonal projection $\hat{v} = \pi_S(v)$ is the closest possible vector in S to v, in other words, $\|w - v\| \geq \|\pi_S(v) - v\|$ for any $w \in S$. Let's prove it:

Suppose the non-trivial case where $v \notin S$. Then $v = \pi_S(v) + u$ and we can conclude that $u \in \ker(\pi_S)$, which in turn implies that u is orthogonal to S. On the other hand, for any $w \in S$ we have $w = \pi_S(v) + r$, for some other $r \in S$. Then $w - v = \pi_S(v) + r - \pi_S(v) + u = r + u$ with u orthogonal to r, whereas $\pi_S(v) - v = u$. Consequently,

$$\|w - v\|^2 = (r + u)^t (r + u) = \|r\|^2 + \|u\|^2 \geq \|\pi_S(v) - v\|^2 = \|u\|^2.$$

4. A MORE GENERAL PERSPECTIVE ON PROJECTIONS

4.1. Although not explicitly stated, we have seen that $\pi_S : V \to V$ is a linear map with a remarkable property: whenever composed with itself, it gives the same linear map, i.e., $\pi_S \circ \pi_S = \pi_S$.

4.2. A linear map fulfilling the condition in section 4.1 is known as a "projection".

4.3. Orthogonal projections with respect to some vector subspace are projections of a particular kind, but of course not the only ones we can conceive.

5. APPLICATION 1: LEAST SQUARES REGRESSION

6. APPLICATION 2: PROBABILITY

6.1. The joint probability distribution on the finite sets X and Y sets a probability $p(x, y)$ for each $x \in X$ and $y \in Y$ such that $\sum_{x,y} p(x, y) = 1$.

6.2. Define the vector $v \in \mathbb{R}^{|X| \times |Y|}$ with entries $\sqrt{p(x, y)}$. Think of it as reshaping the joint probability table of p into a vector in some convenient order– this can be formalized using the tensor semantics.

6.3. The orthogonal projection onto v is a rank one linear map that encodes quite a bit of information about the joint probability distribution p. Check out the passage from classical probability to quantum probability.