Projections and Least Squares

Solving the unsolvable!

by Javier Rodríguez Laguna, Univ. Carlos III de Madrid

Last Update: Feb 29, 2008

Introduction: Solving the Unsolvable

You have a problem and you prove it is unsolvable. For example, an incompatible linear system. What do you do next? It depends. If it is a class assignment, you're done. But what if it is a *real* problem? Imagine you like two boys/girls (or one of each!), and they're jealous. Since you can't have both, what do you do? You go home alone? No, you try to get the "minimum disturbance/pain/cost solution". In real life, to have a proof that a problem has no solution is no excuse to drop it. You find the best pseudo-solution. And this way you get a new problem, which certainly *has* a solution.

Orthogonal Projection on a Line

Consider the following problem in $\mathbf{R}^3$: to find the point X in a 1D subspace r (a line through the origin) which is closest to a certain point B outside that line. This problem might be reformulated this way: find the X in a certain "easy subspace" (i.e.: the subspace where we can work nicely) which "simulates best" a very very complex element B.

Line r will be given by a certain vector v, ok? So, an arbitrary element P of that line will be given by $P = \lambda v$, for any value whatsoever of λ. In components: $P = (\lambda v_x, \lambda v_y, \lambda v_z)$. The (squared) distance of this point to point $B = (B_x, B_y, B_z)$ will be

$$D^2(\lambda) = (\lambda v_x - B_x)^2 + (\lambda v_y - B_y)^2 + (\lambda v_z - B_z)^2$$

Now we want to minimize that (squared) distance, so we borrow the idea from calculus to derivate and set to zero:

$$D^2(\lambda)' = 2(\lambda v_x - B_x)v_x + 2(\lambda v_y - B_y)v_y + 2(\lambda v_z - B_z)v_z = 0$$

Solving for λ we get:

$$\lambda(v \cdot v) - B \cdot v = 0 \qquad \rightarrow \qquad \lambda = \frac{B \cdot v}{v \cdot v}$$

Now the desired point is

$$X = \lambda v = \left(\frac{B \cdot v}{v \cdot v}\right) v$$

This X is so important that it deserves a name of its own. We call it the **orthogonal projection** of B on Span(v): $\text{proj}_v B$.

E1. Find the point in the line denoted by vector $v = (0, 0, 1)$ which is closest to $(2, 3, 4)$. Give a geometrical interpretation.

We have used calculus, but can do without! What we've done is to **project** point B on line r. Abusing of notation, we'll use vector v and line r as interchangeable. See figure 1 and realize that vector $B - X$ is orthogonal to vector v!

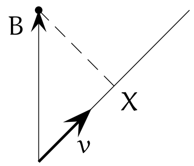

Figure 1. Projecting B on v.

So, how to obtain X? It must be parallel to v, right? Let us define $\hat{v} \equiv v/\|v\|$ to be the normalized version of v. Now, let's find the length of X. See that $\|B\|$ is the hypotenuse of a right triangle, and $\|X\|$ is one of the sides. Let α be the angle between $\hat{v}$ and B. Then,

$$\|X\| = \|B\| \cos(\alpha)$$

But if you remember the formula $\langle u, v\rangle = \|u\| \cdot \|v\| \cdot \cos(\alpha)$, then we get

$$\|X\| = \langle B, \hat{v}\rangle \qquad \text{(remember that } v \text{ is normalized!)}$$

So, in other words,

$$X = \langle B, \hat{v}\rangle\, \hat{v}$$

If v is not unitary, you have to substitute $\hat{v} = v/\|v\|$, so you get

$$X = \frac{\langle B, v \rangle}{\|v\|^2} \, v$$

which is the formula we got by minimization! That's fine.

E2. Prove and explain the formula $(B - \mathrm{proj}_v B) \perp v$. Give an example.

E3. Consider the vector space $C[0,1]$. Let $f(x) = x^3$ and $g(x) = x$. Find $\mathrm{proj}_g f$ and give an interpretation of the result.

Orthogonal Projection on a General Subspace

Now the problem is to find the point X in a more general subspace which is closest to a given point B. In other words: to project a point B on any subspace. We'll start with a plane for simplicity, but generalization is very easy.

There is a different view of the projection on a line which can be useful. Notice that $(B - X) \perp v$. So, $\langle B - X, v \rangle = 0$. This can be thought as an equation to find X. The problem is that, e.g. in $\mathbf{R}^3$, X has three components and we have a single equation! What to do? We know that X *must* belong to the line spanned by v, so $X = \lambda v$. Thus, $\langle B - \lambda X, v \rangle = 0$. This is an equation for λ, a single number, that we can easily solve. This way of thinking can be generalized in order to project on planes or higher subspaces.

For example, imagine that we want to project B on a plane spanned by a_1 and a_2. Figure 2 can help you. Then, $B - X$ must be orthogonal to the plane. To be concrete, it has to be orthogonal both to a_1 and a_2, so:

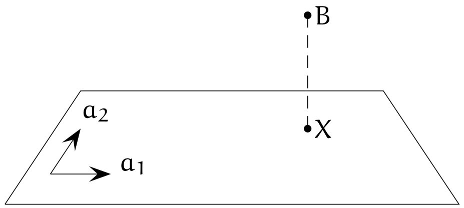

Figure 2. A plane and a point B outside it. Point X is the closest point to B within the plane.

$$\langle a_1, B - X \rangle = 0 \qquad \langle a_2, B - X \rangle = 0$$

But if X belongs to the plane, there must be coefficients λ_1 and λ_2 such that $X = \lambda_1 a_1 + \lambda_2 a_2$. Therefore,

$$\begin{cases} \langle a_1, B - (\lambda_1 a_1 + \lambda_2 a_2) \rangle = 0 \\ \langle a_2, B - (\lambda_1 a_1 + \lambda_2 a_2) \rangle = 0 \end{cases}$$

Two equations and two unknowns, so we can solve! Let's try an example...

E4. Project vector $(1,2,3)$ on the subspace spanned by $(1,0,0)$ and $(0,1,0)$. Is the result reasonable?

But let's try to simplify a little bit the equations. The expression $X = \lambda_1 a_1 + \lambda_2 a_2$ can be written in a nicer way. Let us create a matrix A by putting a_1 and a_2 together as columns, $A = [a_1, a_2]$. Then,

$$X = A \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

Now we can say the target subspace is $\mathrm{Col}(A)$, the subspace spanned by the columns of A. Now let us define vector $\lambda \equiv (\lambda_1, \lambda_2)^T$. This way we have $X = A\lambda$, and the previous equations become

$$\begin{cases} \langle a_1, B - A\lambda \rangle = 0 \\ \langle a_2, B - A\lambda \rangle = 0 \end{cases}$$

Wait, wait, we have seen this thing before, when discussing orthogonal complements! Then we said that the equations $a_1 \cdot w = 0$, $a_2 \cdot w = 0$ equivalent to $A^T w = 0$. So, *if the inner product is the usual dot product* in $\mathbf{R}^n$, then this is equivalent to

$$A^T (B - A\lambda) = 0$$

Or, in other terms,

$$(A^T A) \lambda = A^T B$$

These are called the **normal equations**. Once you have λ, you find X easily: $X = A\lambda$.

E5. Write down the normal equations for the previous example and solve them. Check that you get the same solution to the projection problem.

E6. In $\mathbf{R}^4$, consider the subspace given by $W = \mathrm{Span}(u_1, u_2, u_3)$, where $u_1 = (1,0,0,0)$, $u_2 = (1,1,0,0)$ and $u_3 = (0,0,1,1)$. Find the orthogonal projection of $B = (1,2,3,4)$ on W.

E7. The previous normal equations are given for $\mathbf{R}^n$ with the usual dot product. What changes should be made for a general vector space with a general inner product?

Least Squares Solution to a Problem

Imagine that you are given an incompatible linear system, $Ax = b$. As we discussed in the introduction, what do you do once you've proved it is incompatible? You go home? No, if it is a real life problem. You may be interested in an *approximate* solution, i.e.: to find out the value of x which makes $\|Ax - b\|$ minimum. This can be achieved with calculus, but a little bit of geometry may be more interesting!

The main insight is the following. The fact that $Ax = b$ is an incompatible system means that there is no x such that $Ax = b$. Let $a_1, \cdots, a_m$ be the columns of A. Then, the expression Ax can be rewritten as $a_1x_1 + \cdots + a_mx_m$, OK? This means that Ax *always belongs to the subspace spanned by the columns of* A. We call this subspace, as usual, $\mathrm{Col}(A)$.

So, b does not belong to $\mathrm{Col}(A)$. There should be a certain $\hat{b}$, within $\mathrm{Col}(A)$, for which the distance to b is minimum... Hey! We have solved *that* problem before! It's just the projection problem! So, we have to find the $\hat{b}$ such that $\hat{b} - b$ is orthogonal to $\mathrm{Col}A$.

Go again to figure 2. The subspace is $\mathrm{Col}(A)$, b is B, and $\hat{b}$ is X. What is our x then? We said that $\hat{b}$ must be in the subspace $\mathrm{Col}(A)$, right? Then, it must be expandable as a linear combination of the columns of A: $\hat{b} = Ax$, for some x. In the previous section, we called it λ.

So, now the solution of the previous problem can be mapped to ours. The **normal equations** are just

$$(A^T A)\, x = A^T b$$

That's all. The solution to this equation is the so-called **least-squares** solution to the original $Ax = b$ problem. Let us go through a worked example...

So, let us consider the following problem: to find the line going through the points $(0,1)$, $(1,2)$ and $(2,4)$. Two points make up a line, and here we have three, and they're not aligned. Therefore, the problem, as it is, has no

solution. Let $y = mx + n$ be the desired line. Then, we would have to solve the following system of equations:

$$\begin{cases} 1 = m \cdot 0 + n \\ 2 = m \cdot 1 + n \\ 4 = m \cdot 2 + n \end{cases}$$

Written in matrix form we get

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$$

The system is incompatible, as you can readily check. This should be obvious from the fact that the three points are *not* in a line. But we might try to find the line that makes the "best fit" to all three points. This means to find the solution to $Ax = b$ in the least-squares sense.

So, we apply the formula:

$$A^T A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix} \qquad A^T b = \begin{pmatrix} 10 \\ 7 \end{pmatrix}$$

And get the **normal equations**:

$$(A^T A)x = A^T b \qquad \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} 10 \\ 7 \end{pmatrix}$$

That has solution $m = 1'5$, $b = 5/6$. The solution is reasonable, if you plot the line along with the points. Of course, what we have obtained is the **linear regression** for the given set of points.

E8. Get a set of four points in the plane, approximately on a line, and find the best-fit line.

Q1. Will the dimension of the normal equations for the fitting to a line problem depend on the number of points? And if you do it in 3D?

Q2. Do the normal equations always have a solution?

E9. Find the best-fitting parabola to the following four points: $(0,1)$, $(1,0)$, $(2,1)$ and $(3,2)$. Where is the vertex?

Projection Operator

Consider a certain subspace W. If we're going to project many vectors on it, it would be nice to have an operator, P_W, which does the trick. How to obtain it?

Let's start with a (more or less) trivial example. Suppose that we're working on $\mathbf{R}^3$ and we want to find the projector on the xy plane. So,

$$P_W \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

With this equation we mean that, for any point in 3D space, the x and y components are kept equal, while the z component is "projected out", to be zero. It's not hard to find out that the corresponding matrix is

$$P_W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

You can check easily that it gives the correct result. But, how to find it in a more general case? In the naming convention of the section on projection on a general subspace,

$$P_W B = X$$

Let's consider the normal equations:

$$(A^T A)\lambda = A^T B$$

But we know that $A\lambda = X$. And this equation *has* a solution. Therefore, we can invert it, at least formally!

$$\lambda = A^{-1} X$$

(please, be extremely careful with that equation!) Then,

$$(A^T A) A^{-1} X = A^T B$$

We want to "find" X from B. Therefore, we take all the operators to the right side, in the correct order:

$$A^{-1} X = (A^T A)^{-1} A^T B$$

(again, we should be careful, but that thing can be proved rigorously). Last step,

$$X = A(A^T A)^{-1} A^T B$$

Therefore, we can read the projector from that:

$$P_W = A(A^T A)^{-1} A^T$$

E10. Compute the projector on the xy plane for vectors in $\mathbf{R}^3$ using the previous formula.

E11. Compute the projector matrix on the subspace of $\mathbf{R}^3$ spanned by $u_1 = (1, 2, 0)$ and $u_2 = (0, 1, -1)$.

E12. For the previous projector, check that $P_W^2 = P_W$.

The equation in the last exercise, $P_W^2 = P_W$, is true for *any projector*. The reason is that once you have projected a vector on a subspace W, the result is on the subspace. Therefore, projecting again (i.e.: applying again the projector) will not do anything new! So, doing it twice is the same as doing it once. Operators which fulfill this condition ($A^n = A$ for all n), are called *idempotent*.