13 Inner product spaces

13.1 Dot product

In Calculus III you already considered the dot product of two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^3$, which was defined as

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

This formula implies, using basic geometry, that, alternatively, it can be written as

$$\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}||\boldsymbol{y}| \cos \theta,$$

where $|\boldsymbol{x}|$ and $\boldsymbol{y}$ are the lengthes of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, and θ (any of the two) angle between these two vectors. The length of a vector was calculated using the usual Pythagoras theorem as

$$|\boldsymbol{x}| = \left(x_1^2 + x_2^2 + x_3^2\right)^2 = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}.$$

That is, if we are given the dot product then we also have length of vectors and angles between them.

It is quite natural to generalize introduced formula for the real n-dimensional vector space $\mathbf{R}^n$:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{j=1}^{n} x_j y_j,$$

and also introduce length and angles.

Recall that the second important vector space in which we do the actual calculations is $\mathbf{C}^n$. The introduced formulas will not work for it, because the length can be an imaginary number — something, which I would like to avoid. To mend the situation, I introduce the following dot product in $\mathbf{C}^n$:

$$\boldsymbol{w} \cdot \boldsymbol{z} = w_1 \bar{z}_1 + \ldots + w_n \bar{z}_n, \quad \boldsymbol{w}, \boldsymbol{z} \in \mathbf{C}^n.$$

By this definition I have

$$|\boldsymbol{z}| = \sqrt{\boldsymbol{z} \cdot \boldsymbol{z}} = |z_1|^2 + \ldots + |z_n|^2 \geq 0,$$

which is a natural property for the length.

I also note that introduced in this way dot product has the following natural properties:

1. $\boldsymbol{w} \cdot \boldsymbol{z} = \overline{\boldsymbol{z} \cdot \boldsymbol{w}}$, $\boldsymbol{w}, \boldsymbol{z} \in \mathbf{C}^n$;
2. $(\alpha \boldsymbol{v} + \beta \boldsymbol{w}) \cdot \boldsymbol{z} = \alpha(\boldsymbol{v} \cdot \boldsymbol{z}) + \beta(\boldsymbol{w} \cdot \boldsymbol{z})$;
3. $\boldsymbol{z} \cdot \boldsymbol{z} \geq 0$ for any $\boldsymbol{z} \in \mathbf{C}^n$;
4. $\boldsymbol{z} \cdot \boldsymbol{z} = 0$ if an only if $\boldsymbol{z} = 0$.

Now my goal is to introduce something more abstract and general for an arbitrary vector space over $\mathbf{C}$ or $\mathbf{R}$.

Math 329: *Intermediate Linear Algebra* by Artem Novozhilov©
e-mail: artem.novozhilov@ndsu.edu. Fall 2016

13.2 Inner product space

Definition 13.1. *Let V be a vector space over $\mathbf{C}$ (or over $\mathbf{R}$). Assume that the function*

$$\langle \cdot\,, \cdot \rangle : V \times V \longrightarrow \mathbf{C} \text{ or } \langle \cdot\,, \cdot \rangle : V \times V \longrightarrow \mathbf{R}$$

is defined for any $u, v \in V$ and satisfies the following axioms

1. $\langle u\,, v \rangle = \overline{\langle v\,, u \rangle}$ *or, for the field* $\mathbf{R}$, $\langle u\,, v \rangle = \langle v\,, u \rangle$;
2. $\langle \alpha u + \beta v\,, w \rangle = \alpha \langle u\,, w \rangle + \beta \langle v\,, w \rangle$;
3. $\langle u\,, u \rangle \geq 0$;
4. $\langle u\,, u \rangle = 0$ *if and only if* $u = 0$.

Then the pair $(V, \langle \cdot\,, \cdot \rangle)$ is called an inner product space. The function $\langle \cdot\,, \cdot \rangle$ is called an inner product.

Example 13.2. First example we already saw, the usual dot product in $\mathbf{R}^n$ is an example of an inner product in $\mathbf{R}^n$, I just want to emphasize that there are many *other* possible inner products in $\mathbf{R}^n$. As an exercise, I invite you to check that the function

$$\langle \boldsymbol{x}\,, \boldsymbol{y} \rangle_{\boldsymbol{D}} = \boldsymbol{D}\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{j=1}^{n} d_j x_j y_j$$

is a legitimate inner product that satisfies the four axioms above, for any diagonal $\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_n)$, where all $d_j > 0$.

Note also that sometimes it is more convenient to write the dot product using the matrix multiplication:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n,$$

assuming that by default all the vectors $\boldsymbol{x}$ are column vectors.

Example 13.3. Again, the standard dot product in $\mathbf{C}^n$ is (one) example of an inner product in $\mathbf{C}^n$:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{j=1}^{n} x_j \bar{y}_j.$$

Note that I cannot write this product as $\boldsymbol{y}^\top \boldsymbol{x}$, because I also need conjugation. This motivates me to introduce the operation of transposition and taking complex conjugate at the same time:

$$\boldsymbol{y}^* = \overline{(\boldsymbol{y}^\top)} = (\bar{\boldsymbol{y}})^\top.$$

Using this notation I have

$$\boldsymbol{y}^* \boldsymbol{x} = \boldsymbol{x} \cdot \boldsymbol{y}.$$

In general, for the future I introduce *Hermitian adjoint* $\boldsymbol{A}^*$ of a matrix $\boldsymbol{A}$ as follows

$$\boldsymbol{A}^* = \bar{\boldsymbol{A}}^\top.$$

Example 13.4. Consider a real vector space $\mathbf{P}$ of polynomials defined on the interval $[a, b]$, and let $p_1, p_2 \in \mathbf{P}$. I define

$$\langle p_1 \,, p_2 \rangle = \int_a^b p_1(x) p_2(x) \mathrm{d}x.$$

It is easy to check that this function satisfies the axioms 1–4 and hence defines an inner product making $\mathbf{P}$ an inner product space. For a complex vector space of polynomials I can take

$$\langle p_1 \,, p_2 \rangle = \int_a^b p_1(x) \overline{p_2(x)} \mathrm{d}x.$$

Example 13.5. Let $\mathbf{M}_{m \times n}(\mathbf{C})$ be the complex vector space of $m \times n$ complex matrices. I define

$$\langle \boldsymbol{A} \,, \boldsymbol{B} \rangle = \operatorname{tr}(\boldsymbol{B}^* \boldsymbol{A}).$$

Check the axioms.

In the following I will use the notation $\langle \cdot \,, \cdot \rangle$ to emphasize the abstract inner product, but for all the computations I will use the *standard* inner or dot product for $\mathbf{R}^n$ or $\mathbf{C}^n$ if not specified otherwise.

The four axioms of inner product space allow me to get some simple corollaries almost immediately.

Corollary 13.6. *Let $(V, \langle \cdot \,, \cdot \rangle)$ be an inner product space. Then*

1. $\langle u \,, \alpha v + \beta w \rangle = \bar{\alpha} \langle u \,, v \rangle + \bar{\beta} \langle u \,, w \rangle$.
2. $\langle 0 \,, u \rangle = \langle u \,, 0 \rangle = 0$.
3. *Let $v \in V$. Then $v = 0$ if and only if $\langle v \,, u \rangle = 0$ for any $u \in V$.*
4. *Let $u, v \in V$. Then $u = v$ if and only if $\langle u \,, w \rangle = \langle v \,, w \rangle$ for all $w \in V$.*
5. *Let $\mathscr{A}, \mathscr{B} \colon V \longrightarrow U$ be two linear transformations. Then $\mathscr{A} = \mathscr{B}$ if and only if*

$$\langle \mathscr{A}(v) \,, u \rangle = \langle \mathscr{B}(v) \,, u \rangle$$

for all $v \in V$, $u \in U$. Note that the inner product is taken in U.

Exercise 1. Prove this corollary.

Since $\langle u \,, u \rangle \geq 0$ and equal to zero only if $u = 0$ then I can define

$$|u| = \sqrt{\langle u \,, u \rangle}$$

the *length* or *norm* of the vector u. (Very often it is denoted as $\|u\|$ but I will stick to the simple notation that should remind you of the absolute value.) Before proceeding, however, I would like to make sure that my norm has the properties which I so got used to in our usual 3 dimensional Euclidian space. Clearly, my norm is nonnegative and equal to zero only if $u = 0$, which is good. I also would like that if I multiply my vector by a constant, the length of this vector would be multiplied by the absolute value of this constant. And fortunately I have, from the properties of the inner product, that

$$|\alpha u| = \sqrt{\langle \alpha u \,, \alpha u \rangle} = \sqrt{|\alpha|^2 \langle u \,, u \rangle} = |\alpha| |u|,$$

note that the meaning of the absolute values is different in the rightmost formula.

Finally, since I have the geometric picture that the vectors $u, v, u+v$ form a triangle, I would like to have that the length of $u+v$ would not exceed the sum of the lengthes of u and v:

$$|u+v| \leq |u| + |v|.$$

To show that my norm actually satisfies this *triangle inequality*, first I will prove an auxiliary lemma.

Lemma 13.7 (Cauchy–Schwarz inequality)**.** *Let* $(V, \langle \cdot\,, \cdot \rangle)$ *be an inner product space. Then for any* $u, v \in V$

$$|\langle u\,, v\rangle| \leq |u||v|,$$

and the equality holds if and only if the collection (u, v) *is linearly dependent.*

Proof. I will prove this lemma for the case $\mathbf{F} = \mathbf{R}$ and leave it as an exercise for $\mathbf{F} = \mathbf{C}$ (this will require a couple more steps due to the difference in the first axiom).

Let $t \in \mathbf{R}$ be a scalar and consider $u - tv \in V$. By the properties of the inner product

$$0 \leq |u - tv|^2 = \langle u - tv\,, u - tv\rangle = \langle u\,, u\rangle - 2t\langle u\,, v\rangle + t^2\langle v\,, v\rangle\,.$$

On the right I have a quadratic polynomial with respect to variable t. Since this polynomial is nonnegative it cannot have more then one root and hence its discriminant is nonpositive:

$$(\langle u\,, v\rangle)^2 - \langle u\,, u\rangle\langle v\,, v\rangle \leq 0,$$

which implies the required inequality. ■

Now it is a simple matter to show that the norm of a vector satisfies the triangle inequality:

$$\begin{aligned} |u+v|^2 = \langle u+v\,, u+v\rangle = |u|^2 + \langle u\,, v\rangle + \langle v\,, u\rangle + |v|^2 \leq \\ \leq |u|^2 + 2|\langle v\,, u\rangle| + |v|^2 \leq \\ \leq |u|^2 + 2|u||v| + |v|^2 = \\ = (|u| + |v|)^2. \end{aligned}$$

To finish this subsection I will state one more result without proof (which is left as an exercise).

Lemma 13.8 (Parallelogram identity)**.**

$$|u+v|^2 + |u-v|^2 = 2(|u^2| + |v|^2).$$

13.3 Orthogonality

Note that the familiar $\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}||\boldsymbol{y}|\cos\theta$ can be used to *define* the angle between two vectors $u, v \in V$, where V is an inner product space as

$$\cos\theta = \frac{\langle u\,, v\rangle}{|u||v|},$$

and the Cauchy–Schwarz inequality guaranties that at least for real vector spaces $-1 \leq \cos\theta \leq 1$ and hence the angle is well defined. But the most important angle for us will be $\pi/2$, for which $\cos\pi/2 = 0$. Hence I am motivated to introduce the key definition.

Definition 13.9. *Two vectors $u, v \in V$ are called orthogonal if*

$$\langle u \, , v \rangle = 0.$$

Quite often the orthogonal vectors are denoted $u \perp v$.

To show immediate power of this definition, consider

Proposition 13.10 (Pythagoras' theorem)**.** *Let V be an inner product space. Then if $u \perp v$ for $u, v \in V$ then*

$$|u + v|^2 = |u|^2 + |v|^2.$$

Proof.

$$|u + v|^2 = \langle u + v \, , u + v \rangle = \langle u \, , u \rangle + \langle v \, , v \rangle = |u|^2 + |v|^2.$$

∎

Definition 13.11. *A collection of vectors $(v_1, v_2, \ldots, v_k)$ is called orthogonal if*

$$v_j \perp v_l, \quad j \neq l.$$

In addition, if for each v_j $|v_j| = 1$ then this system is called orthonormal.

If I have an orthogonal (or orthonormal) collection then I know that this collection is linearly independent, since

Proposition 13.12. *Let $S = (v_1, \ldots, v_k)$ be an orthogonal collection, where $v_j \neq 0$. Then S is linearly independent.*

Proof. Consider

$$\alpha_1 v_1 + \ldots \alpha_k v_k = 0.$$

Take the inner product of the both sides of this equality with v_1:

$$\alpha_1 \langle v_1 \, , v_1 \rangle = 0,$$

and since $v_1 \neq 0$ then $\alpha_1 = 0$. Similarly I show that all the rest of scalars are also zero. ∎

Remark 13.13. Note that zero vector is orthogonal to any vector and hence can be added to any orthogonal collection without ruining orthogonality. However, in most cases when someone talks about "orthogonal collection or system" they mean "orthogonal collection of non-zero vectors." I will use the same abuse of language.

So, from the last proposition I have that if I have k vectors in an orthogonal collection in k-dimensional vector space V then they form *orthogonal* or *orthonormal* basis. The advantage to work with an orthogonal basis is the simplicity to calculate the coordinates of an arbitrary vector with respect to this basis. Indeed, for any $v \in V$, the coordinates are the scalars in the following linear combination

$$v = \alpha_1 v_1 + \ldots + \alpha_k v_k.$$

To find α_j I take the inner product of both sides with v_j and get

$$\langle v \, , v_j \rangle = \alpha_j \langle v_j \, , v_j \rangle \implies \alpha_j = \frac{\langle v \, , v_j \rangle}{\langle v_j \, , v_j \rangle} \, .$$

This formula becomes especially simple if I deal with an *orthonormal* basis:

$$\alpha_j = \langle v , v_j \rangle .$$

For the following I will need the notion of a subspace orthogonal to a given vector, and orthogonal subspaces.

Definition 13.14. *Let $E \subseteq V$. I will say that $v \perp E$ if v is orthogonal to any vector in E. I say that subspaces F and E are orthogonal if any vector from F is orthogonal to any vector in E.*

Lemma 13.15. *Let $E \subseteq V$ be spanned by $v_1, \dots, v_k$. Then $v \in V$ is orthogonal to E if and only if $\langle v , v_j \rangle = 0$ for all j.*

Exercise 2. Prove the lemma above.

Now I can introduce an orthogonal projection.

Definition 13.16. *Let $v \in V$ and $E \subseteq V$. The orthogonal projection $\mathscr{P}_E(v)$ onto E is the vector $w \in V$ such that $w \in E$ and $v - w \perp E$.*

Remark 13.17. The meaning of the orthogonal projection is quite simple: It is the best approximation of an arbitrary vector $v \in V$ by a vector w from some subspace E of V. That is we are looking for a vector $w \in E$ such that the length $|v - w|$ is minimally possible. Let me show that indeed what I defined provides the best approximation for a vector v. I first prove the implied geometric meaning and uniqueness of the orthogonal projection, and after this I will show, by an explicit formula, that it always exists.

Theorem 13.18. *The orthogonal projection $w = \mathscr{P}_E(v)$ minimizes the distance from v to E, i.e., for all $u \in E$*

$$|w - v| \leq |u - v|.$$

Moreover, if for some $u \in E$

$$|w - v| = |u - v|$$

then $w = u$, that is, the orthogonal projection is unique.

Proof. Consider

$$v - u = v - w + w - u.$$

Since $v - w \perp E$ and $w - u \in E$ then, by the Pythagoras theorem,

$$|v - u|^2 = |v - w|^2 + |w - u|^2 \geq |v - w|^2,$$

which finishes the proof. ∎

The next proposition shows that if E has an orthogonal basis than I can explicitly calculate my projection.

Proposition 13.19. *Let $(v_1, \dots, v_k)$ be an orthonormal basis of E then*

$$w = \mathscr{P}_E(v) = \sum_{j=1}^{k} \frac{\langle v , v_j \rangle}{|v_j|^2} v_j.$$

Proof. Since w is a linear combination of v_k hence $w \in E$. I need to show that $v - w \perp E$, for this it is enough to show that $\langle v - w\, , v_j \rangle = 0$ for all j. Direct calculations yield

$$\langle v - w\, , v_j \rangle = \langle v\, , v_j \rangle - \langle w\, , v_j \rangle = \langle v\, , v_j \rangle - \langle v\, , v_j \rangle = 0.$$

■

Remark 13.20. The formula above shows that $\mathscr{P}_E$ is a linear operator (can you prove this fact directly?). Hence it is a good question to ask what is the matrix of $\mathscr{P}_E$ in the standard basis for the standard inner products. It turns our that the matrix is given by

$$\boldsymbol{P}_E = \sum_{j=1}^{k} \frac{1}{|\boldsymbol{v}_j|^2} \boldsymbol{v}_j \boldsymbol{v}_j^*.$$

Finally, what if I only have a basis for E and not an orthogonal basis? Actually, there is a procedure that allows to *orthogonalize* any basis. This procedure is called *Gram–Schmidt orthogonalization* algorithm. Here is this algorithm. I start with a basis $(u_1, u_2, \ldots, u_k)$.

1. Take $v_1 = u_1$ and let $E_1 = \operatorname{span}(v_1) = \operatorname{span}(u_1)$.
2. Now I want to get my second vector in the form $v_2 = u_2 - \alpha v_1$ such that $\langle v_2\, , v_1 \rangle = 0$. This can be accomplished only if $\alpha v_1 = \mathscr{P}_{E_1}(u_2)$. Let $E_2 = \operatorname{span}(v_1, v_2)$ and note that $\operatorname{span}(u_1, u_2) = E_2$ by construction.

...

r+1 Now, assume that I already have r vectors in my orthogonal system: $(v_1, \ldots, v_r)$. Let $E_r = \operatorname{span}(v_1, \ldots, v_r)$ and also I have that $\operatorname{span}(u_1, \ldots, u_r) = E_r$. The $r + 1$-th vector is found as

$$v_{r+1} = u_{r+1} - \mathscr{P}_{E_r}(u_{r+1}).$$

The only thing to note that $v_{r+1} \neq 0$ since $u_{r+1} \notin E_r$.

Example 13.21. Let $V = \mathbf{R}^3$, let $\langle \cdot\, , \cdot \rangle$ be the standard dot product, and consider

$$\boldsymbol{u}_1 = (1, 1, 1)^\top, \quad \boldsymbol{u}_2 = (0, 1, 2)^\top, \quad \boldsymbol{u}_3 = (1, 0, 2)^\top.$$

It is easy to check that these are three linearly independent vectors and hence a basis of $\mathbf{R}^3$. I want to use the Gram–Schmidt algorithm to find an orthogonal basis.

Let $\boldsymbol{v}_1 = \boldsymbol{u}_1$. Now,

$$\boldsymbol{v}_2 = \boldsymbol{u}_2 - \mathscr{P}_{E_1}(\boldsymbol{u}_2) = \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 = (-1, 0, 1)^\top.$$

Similarly,

$$\boldsymbol{v}_3 = \boldsymbol{u}_3 - \mathscr{P}_{E_2}(\boldsymbol{u}_3) = \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2} \boldsymbol{v}_2 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 = (1/2, -1, 1/2)^\top.$$

Remark 13.22. Since if I am given a nonzero vector v I can always normalize it $u = v/|v|$ such that $|u| = 1$, from a theoretical point of view I can always speak of orthonormal bases, because in this case many formulas look much simpler.

13.4 Least square method

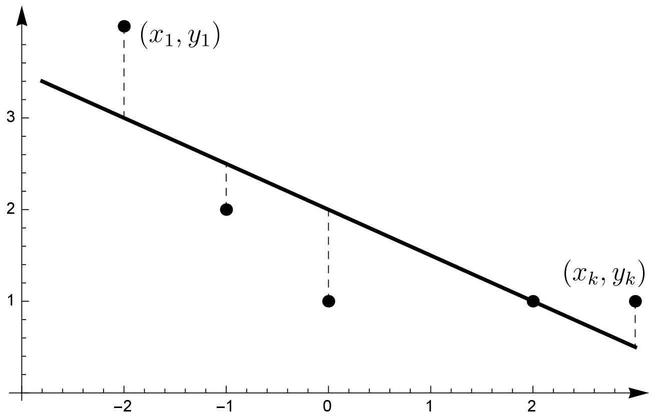
Now consider one application of the discussed abstract construction. Assume we are given a set of points with coordinates

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \quad \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \quad \ldots, \quad \begin{bmatrix} x_k \\ y_k \end{bmatrix},$$

and we need to find a straight line with the equation $y = ax + b$ that approximates the data in some "best" sense. Of course, definition of "best" can easily change the line, but one very reasonable assumption is to ask so that the expression

$$\sum_{j=1}^{k} (a + bx_j - y_j)^2$$

be minimal (see the figure where the expression being minimized is the sum of the squares of the lengthes of dashed intervals).



Recall that the standard inner product in $\mathbf{R}^n$ allows to compute the distance between two vectors as follows $|\boldsymbol{x} - \boldsymbol{y}|^2 = \sum_{j=1}^{k} (x_j - y_j)^2$, and hence the above condition means that I am minimizing the distance between the vector with coordinates $a + bx_j$ and the vector with coordinates y_j. Finally, the vector with coordinates $a + bx_j$ can be written as a product of matrix $\boldsymbol{A}$ and vector $[a \;\; b]^\top$:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_k \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix},$$

therefore I am minimizing the distance

$$|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{y}|^2,$$

where $\boldsymbol{u} = [a \;\; b]^\top$. Of course, if this distance would be 0, then I would be talking about simply finding the solution to $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{y}$, but in general my system has no solution (no straight line can pass through all the given points).

Now, geometrically, $\boldsymbol{A}\boldsymbol{u}$ for all possible $\boldsymbol{u}$ is exactly the image of $\boldsymbol{A}$, and my minimum is hence the distance from $\boldsymbol{y}$ to $\operatorname{im} \boldsymbol{A}$. This distance, as we already know, is minimal, when $\boldsymbol{A}\boldsymbol{u} = \mathscr{P}_{\operatorname{im} \boldsymbol{A}}(\boldsymbol{y})$, and hence I am led to the following solution of my problem: To minimize the distance from $\boldsymbol{Au}$ to $\boldsymbol{y}$ I need to solve the system of linear algebraic equation

$$\boldsymbol{Au} = \mathscr{P}_{\mathrm{im}\,\boldsymbol{A}}(\boldsymbol{y}),$$

which by construction always has a solution (it can happen however that we will have infinitely many solutions), and this is the solution to my original problem.

Naively, to achieve my goal, I can 1) find a basis for the image of $\boldsymbol{A}$, 2) orthogonalize it, 3) calculate the corresponding orthogonal projection, 4) solve the system. It turns out, however, that these steps are not necessary. Indeed, $\boldsymbol{Au}$ is the orthogonal projection of $\boldsymbol{y}$ if and only if $\boldsymbol{y} - \boldsymbol{Au} \perp \mathrm{im}\,\boldsymbol{A}$, which implies that

$$\boldsymbol{y} - \boldsymbol{Au} \perp \boldsymbol{a}_j,$$

where $\boldsymbol{a}_j$ is the j-th column of $\boldsymbol{A}$. I hence must have

$$0 = \langle \boldsymbol{y} - \boldsymbol{Au}\,, \boldsymbol{a}_j \rangle = \boldsymbol{a}_j^*(\boldsymbol{y} - \boldsymbol{Au})$$

for any j. Putting these equalities together in the matrix form, I get

$$\boldsymbol{A}^*(\boldsymbol{y} - \boldsymbol{Au}) = \boldsymbol{0},$$

which implies the so-called *normal equation*

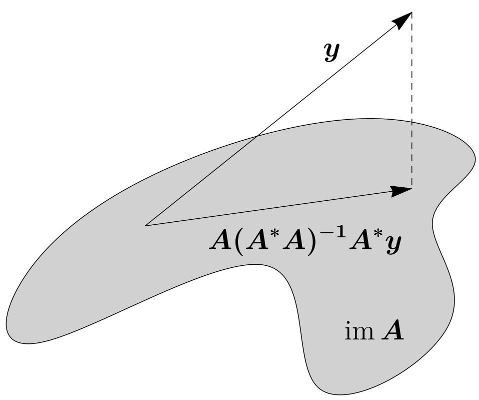
$$\boldsymbol{A}^*\boldsymbol{Au} = \boldsymbol{A}^*\boldsymbol{y}.$$

A solution to this problem will give me the least square solution to the original approximation problem.

If I multiply this solution by $\boldsymbol{A}$ I will find the orthogonal projection of $\boldsymbol{y}$ on the image of $\boldsymbol{A}$. If I assume that the square matrix $\boldsymbol{A}^*\boldsymbol{A}$ is invertible, then this projection takes especially compact and simple form

$$\mathscr{P}_{\mathrm{im}\,\boldsymbol{A}}(\boldsymbol{y}) = \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{y}.$$

and since it is true for any $\boldsymbol{y}$, what I found is actually the matrix of the orthogonal projection onto the image of $\boldsymbol{A}$ (see the graphical illustration).



Example 13.23. Now consider actual example: Let my data consist of 5 data points:

$$(-2,4),\quad (-1,2),\quad (0,1),\quad (2,1),\quad (3,1).$$

In my notations I have

$$\boldsymbol{A}=\begin{bmatrix}1 & -2\\ 1 & -1\\ 1 & 0\\ 1 & 2\\ 1 & 3\end{bmatrix},\quad \boldsymbol{y}=\begin{bmatrix}4\\2\\1\\1\\1\end{bmatrix}.$$

I have

$$\boldsymbol{A}^*\boldsymbol{A}=\begin{bmatrix}5 & 2\\ 2 & 18\end{bmatrix},\quad \boldsymbol{A}^*\boldsymbol{y}=\begin{bmatrix}9\\-5\end{bmatrix}$$

Now, solving

$$\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{u}=\boldsymbol{A}^*\boldsymbol{y},$$

I get $\boldsymbol{u}=(2,-1/2)$, and hence my best least square approximation is given by

$$y=2-1/2x,$$

which is shown, together with the data, on the figure above.

In the same way it is possible to consider approximation by polynomials of any degree. For example, let me find the best possible approximation of the data from the example above by quadratic polynomial

$$y=a+bx+cx^2.$$

Now the difference is that my matrix $\boldsymbol{A}$

$$\begin{bmatrix}1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\\ \vdots & & \\ 1 & x_k & x_k^2\end{bmatrix},$$

but the rest of the calculations are the same.

I have for this case

$$\boldsymbol{A}=\begin{bmatrix}1 & -2 & 4\\ 1 & -1 & 1\\ 1 & 0 & 0\\ 1 & 2 & 4\\ 1 & 3 & 9\end{bmatrix},$$

and

$$\boldsymbol{A}^*\boldsymbol{A}=\begin{bmatrix}5 & 2 & 18\\ 2 & 18 & 26\\ 18 & 26 & 114\end{bmatrix},\quad \boldsymbol{A}^*\boldsymbol{y}=\begin{bmatrix}9\\-5\\31\end{bmatrix},$$

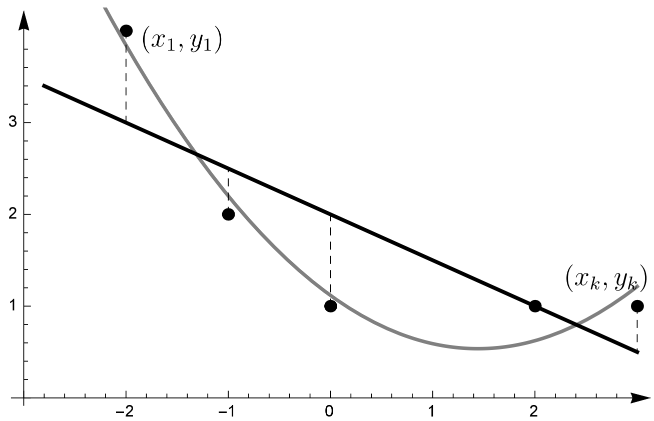
hence my solution to

$$\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{u}=\boldsymbol{A}^*\boldsymbol{y}$$

is

$$a = \frac{86}{77}, \quad b = -\frac{62}{77}, \quad c = \frac{43}{154},$$

which can be graphically illustrated as follows (the quadratic approximation is now in grey).



You can see that at least for this example if there is a good reason to believe that quadratic approximation meaningful then it definitely should be preferred.