Conversation 41: Least Squares Solutions

Winfried Just
Department of Mathematics, Ohio University

MATH3200: Applied Linear Algebra

Existence of orthonormal bases

Denny: So if $B = \{\vec{\mathbf{b}}_1, \dots, \vec{\mathbf{b}}_k\}$ is an orthonormal basis of a vector space V, we can calculate the alternative coordiantes $\vec{\mathbf{c}} = [c_1, \dots, c_k]$ of a vector $\vec{\mathbf{x}}$ in V with respect to B by simply letting $c_i = \langle \vec{\mathbf{b}}_i, \vec{\mathbf{x}} \rangle$.

I like that. It makes the calculations a lot easier!

Frank: If the vector space V does have an orthonormal basis. I doubt whether many vector spaces have such bases.

Theo: Every vector space has an orthonormal basis. There is a procedure called *Gram-Schmidt orthonormalization* that works by starting with any basis $A = \{\vec{\mathbf{a}}_1, \vec{\mathbf{a}}_2, \dots, \vec{\mathbf{a}}_k\}$ of V and building step-by-step an orthonormal basis $B = \{\vec{\mathbf{b}}_1, \vec{\mathbf{b}}_2, \dots, \vec{\mathbf{b}}_k\}$ of V.

This wasn't covered in class, but I'll be happy to show you the algorithm.

Alice: Could just explain to us the idea behind it and skip the details, Theo?

Gram-Schmidt orthonormalization: The idea

Theo: We start with any basis $A = \{\vec{\mathbf{a}}_1, \vec{\mathbf{a}}_2, \dots, \vec{\mathbf{a}}_k\}$ of V and build step-by-step an orthonormal basis $B = \{\vec{\mathbf{b}}_1, \vec{\mathbf{b}}_2, \dots, \vec{\mathbf{b}}_k\}$ of V as follows:

(1) Let $\vec{\mathbf{b}}_1$ be the normalization of $\vec{\mathbf{a}}_1$.

(2) Let $\vec{\mathbf{b}}_2$ be the normalization of the *orthogonal complement of $\vec{\mathbf{a}}_2$ with respect to $\vec{\mathbf{b}}_1$.*

$\dots$

$(\ell + 1)$ Suppose we have already constructed $B_\ell = \{\vec{\mathbf{b}}_1, \dots, \vec{\mathbf{b}}_\ell\}$. Then we let $\vec{\mathbf{b}}_{\ell+1}$ be the normalization of the *orthogonal complement of $\vec{\mathbf{a}}_{\ell+1}$ with respect to span(B_ℓ).*

Cindy: What is that "orthogonal complement of $\vec{\mathbf{a}}_{\ell+1}$ with respect to *span*(B_ℓ)"?

What, exactly, does this mean?

Bob: This wasn't formally defined in class yet.

Alice: But perhaps we can figure out what it means when we use an analogy with something that was defined in class.

Cindy: Would

"orthogonal complement of $\vec{\mathbf{a}}_{\ell+1}$ with respect to *span*(B_ℓ)"

be similar to

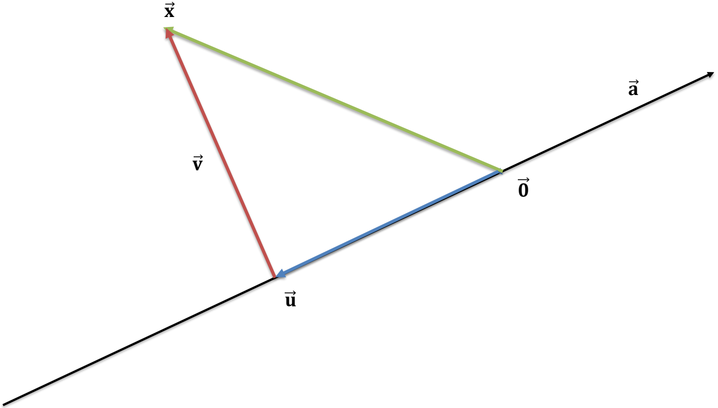
"orthogonal complement of $\vec{\mathbf{a}}_2$ with respect to $\vec{\mathbf{b}}_1$"?

Bob: This one was defined in terms of orthogonal projections.

Cindy: And there was a picture of it in Lecture 43.

Let's look again at this picture.

Orthogonal projection $\vec{u}$ onto $\vec{a}$ with orthogonal complement $\vec{v}$



Orthogonal projections and distances

Denny: Among all vectors in *span*$(\vec{\mathbf{a}})$, the vector $\vec{\mathbf{u}}$ is the one with the shortest distance to $\vec{\mathbf{x}}$. This distance is the norm $\|\vec{\mathbf{v}}\|$ of the orthogonal complement $\vec{\mathbf{v}}$ of $\vec{\mathbf{x}}$ with respect to $\vec{\mathbf{a}}$.

Question C41.1: Did Denny get this right?

Theo: Exactly!

Cindy: But what if *span*$(\vec{\mathbf{a}}_1, \dots, \vec{\mathbf{a}}_k)$ has dimension higher than 1, if it is, for example, a plane? Would then the orthogonal projection of a vector $\vec{\mathbf{x}}$ onto *span*$(\vec{\mathbf{a}}_1, \dots, \vec{\mathbf{a}}_k)$ also be the vector $\vec{\mathbf{u}}$ in this span that has the smallest distance from $\vec{\mathbf{x}}$?
And would the orthogonal complement still be $\vec{\mathbf{v}} = \vec{\mathbf{x}} - \vec{\mathbf{u}}$?

Alice: Right, Cindy! Let's look at an interesting special case.

Question C41.2: What would that orthogonal projection $\vec{\mathbf{u}}$ be if $\vec{\mathbf{x}}$ happens to be in *span*$(\vec{\mathbf{a}}_1, \dots, \vec{\mathbf{a}}_k)$?

In this case, $\vec{\mathbf{u}} = \vec{\mathbf{x}}$ and $\vec{\mathbf{v}} = \vec{\mathbf{0}}$.

Cindy: Great! Now I know how to think geometrically about these orthogonal projections.

Frank: OK, but is this good for anything?
I mean, does it have any engineering applications?

Alice: It does. Think about a system of linear equations $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ that is overconstrained.

Frank: Means you cannot solve it. So just forget it.

Alice: Not exactly. But you could still try to find the vector $\vec{\mathbf{x}}$ for which the distance $\|\mathbf{A}\vec{\mathbf{x}} - \vec{\mathbf{b}}\|$ is as small as possible.

Denny: Do you mean, the vector $\vec{\mathbf{x}}$ that best fudges it so that the constraints are violated as little as possible?

Theo: I wouldn't say it this way, but basically this is what Alice meant. Since the Euclidean distance is the square root of the sum of squares of the coordinates, this is called a *least squares solution.*

What are least squares solutions good for?

Frank: But if the system $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is overconstrained, then Theo's "least squares solution" isn't really a solution of this system!

Question C41.3: Did Frank get this right?

Theo: You are absolutely right about this, Frank. But so-called least squares "solutions" have many applications in science and engineering. For example, if you have a large set of data one the dependence of one variable y on another variable x and want to fit a line that best *approximates* this dependence, a so-called *regression line*, then we can find this line as a least squares solution. I will be happy to show you how—

Bob: Thank you Theo, but maybe not now.

Denny: Yeah. Let's talk about something else. I am still curious about this guy Marvin. Alice had promised us that we would learn how he could best fudge it and compose his meal from his favorite ingredients Losit-Quick and Losit-Easy without violating the recommendations of that Dr. Losit too much.

Review: Marvin's problem

Cindy: Can you remind us, Denny, about the problem?

Denny: Sure. To put it in a mathematical nutshell:
He needs coefficients d_Q, d_E such that $d_Q\vec{\mathbf{v}}_Q + d_E\vec{\mathbf{v}}_E = \vec{\mathbf{v}}_M$, or, if we write it out coordinatewise,

$$d_Q \begin{bmatrix} 20 \\ 200 \\ 20 \end{bmatrix} + d_E \begin{bmatrix} 25 \\ 150 \\ 60 \end{bmatrix} = \begin{bmatrix} d_Q 20 + d_E 25 \\ d_Q 200 + d_E 150 \\ d_Q 20 + d_E 60 \end{bmatrix} = \begin{bmatrix} 50 \\ 300 \\ 100 \end{bmatrix} = \vec{\mathbf{v}}_M$$

Here d_Q and d_E are the numbers of servings of Losit-Quick and Losit-Easy, respectively, and $\vec{\mathbf{v}}_M$ is the vector of nutrients that this guy Dr. Losit had recommended.

But when we translated this into a system of linear equations, the system didn't have a solution.

So we were wondering how Marvin could best fudge it and find values for d_Q and d_E so that the resulting meal would be as close to the recommendations as possible.

How about a least squares solution for Marvin's problem?

Frank: Are you saying, Denny, that we wanted to find the least squares solution for Marvin's problem?

Denny: I didn't say that!
But now that you mentioned it, I guess that is what we are after.

Alice: Right! We want to find the least squares solution $[d_Q, d_E]^T$ for the system

$$\begin{aligned} 20d_Q + 25d_E &= 50 \\ 200d_Q + 150d_E &= 300 \\ 20d_Q + 60d_E &= 100 \end{aligned}$$

Denny: Yeah. But how?

Alice: Let's see whether we can figure it out.
Suppose $[d_Q, d_E]^T$ is any vector of servings. Then $\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E$ are the columns of the coefficient matrix of this system, and $d_Q\vec{\mathbf{v}}_Q + d_E\vec{\mathbf{v}}_E$ is in $span(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E)$.

How to find the least squares solution for Marvin?

Denny: But $\vec{\mathbf{v}}_M$ is not in $span(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E)$!! That's Marvin's problem!

Alice: So he only can get a vector $\vec{\mathbf{u}} = d_Q\vec{\mathbf{v}}_Q + d_E\vec{\mathbf{v}}_E$ in $span(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E)$ such that the distance between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}_M$ is as small as possible.

Question C41.4: Which of the vectors in $span(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E)$ is $\vec{\mathbf{u}}$?

This should be the orthogonal projection of $\vec{\mathbf{v}}_M$ onto $span(\vec{\mathbf{v}}_Q, \vec{\mathbf{v}}_E)$.

Theo: There are many methods for calculating $\vec{\mathbf{u}}$.

Alice: We don't have time to go into details of them right now. Can you use one of them in MatLab and quickly give us the result, Theo?

Theo: (Sigh) These methods are really very interesting and I'd be so happy to explain them here. But if you insist—

The orthogonal projection $\vec{\mathbf{u}}$

Theo: MATLAB gives me:

$\vec{\mathbf{u}} = [45.7890, 300.3275, 100.9358]^T$.

Denny: Wow! This is pretty darn close to Dr.Losit's recommendation $\vec{\mathbf{v}}_M = [50, 300, 100]^T$.

Frank: No big deal if Marvin fudges it and uses only his favorite ingredients. I wouldn't treat this $\vec{\mathbf{v}}_M$ too seriously anyway.

Alice: I'm glad that now you recognize why least squares solutions are useful, Frank.

The least squares solution of Marvin's problem

Denny: But wait! How much of his favorite beer-flavored Losit-Easy can Marvin add for the least squares solution?

Alice: For that we need to find the least squares solution itself.

Denny: I thought we already did:

$\vec{\mathbf{u}} = [45.7890,\ 300.3275,\ 100.9358]^T$.

Alice: Not yet. The least squares solution is the vector of coefficients $[d_Q, d_E]^T$ for the linear combination $d_Q\vec{\mathbf{v}}_Q + d_E\vec{\mathbf{v}}_E = \vec{\mathbf{u}}$, not the orthogonal projection $\vec{\mathbf{u}}$ itself.

Question C41.5: How can we find these coefficients?

By solving the system of linear equations $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{u}}$ whose coefficient matrix has $\vec{\mathbf{v}}_Q$ and $\vec{\mathbf{v}}_E$ as its columns.

The least squares solution of Marvin's problem, completed

Bob: We learned this in Lecture 22 and practiced it in Module 42.

Theo: We can see here how various concepts of linear algebra are related and work together.

Cindy: While you guys were talking, I quickly solved the system and found the coefficients:

$d_Q = 0.3199$ and $d_E = 1.5756$.

Denny: Which means Marvin can compose his meal with just adding a little bit of Losit-Quick to his favorite beer-flavored Losit-Easy and practically get the recommended mix of nutrients!!

This linear algebra stuff really is good for something!

Take-home message

This conversation gave a very brief introduction to *least squares solutions* of systems of linear equations $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ that have many important applications.

When the system $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is consistent, then a least squares solution is simply a solution of the system.

When $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ is overconstrained, then a "least squares solution" is not really a solution, but a vector $\vec{\mathbf{x}}$ such that $\mathbf{A}\vec{\mathbf{x}}$ is as close to $\vec{\mathbf{b}}$ as possible; in other words, violates the constraints imposed by the system as little as possible.

For a least squares solution $\vec{\mathbf{x}}$ the vector $\mathbf{A}\vec{\mathbf{x}}$ is the orthogonal projection of the vector $\vec{\mathbf{b}}$ onto the linear span of the the columns of $\mathbf{A}$.

When the columns of $\mathbf{A}$ form a linearly independent set, then $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ has exactly one least squares solution.