Orthogonality

Example

Remember the example

$$A = [a_1, a_2, a_3, a_4] = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ -10 & -5 & 0 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

The Gauss elimination method on $[A|b\ c]$ (in Matlab: `rref([A b c])`) gives

$$[\hat{A}|\hat{b}\ \hat{c}] = \left[\begin{array}{cccc|cc} 1 & 0 & -1 & -2 & 0 & -1/5 \\ 0 & 1 & 2 & 3 & 0 & 3/5 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

We conclude that $\{a_1, a_2\}$ are linearly independent and that they are a basis for the linear space

$$V = R(A) = S(a_1, a_2, a_3, a_4) = S(a_1, a_2), \quad \dim(V) = 2.$$

We also see that $b \notin V$ because $Ax = b$ has no solution, and that $c \in V$ because $Ax = c$ has the multiple solutions

$$x = \underbrace{s \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}}_{=x_h \in N(A)} + \underbrace{\begin{bmatrix} -1/5 \\ 3/5 \\ 0 \\ 0 \end{bmatrix}}_{=x_p} = x_h + x_p.$$

We write the basis $\{a_1, a_2\}$ as a matrix

$$B = [a_1, a_2] = \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ -10 & -5 \end{bmatrix}.$$

and note that $V = R(A) = R(B) = S(a_1, a_2)$ is a two-dimensional subspace of $\mathbb{R}^3$ (a plane through the origin).

Figure is missing!!

In order to express c as a linear combination of the basis, i.e.,

$$c = Bx = x_1 a_1 + x_2 a_2,$$

we solve the equation $Bx = c$ and get the unique solution $x = \begin{bmatrix} -1/5 \\ 3/5 \end{bmatrix}$, so that

$$c = -\frac{1}{5} \begin{bmatrix} 1 \\ 4 \\ -10 \end{bmatrix} + \frac{3}{5} \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Ortogonal projektion på V

(AMBS 42.35–38) I allmänhet betraktar vi en bas $B = [a_1, \dots, a_n]$, $m \times n$, för ett underrum $V = R(B) = S(a_1, \dots, a_n)$ till $\mathbb{R}^m$, $n \le m$, och en vektor $b \in \mathbb{R}^m$, $b \notin V$. I exemplet: $m = 3$, $n = 2$.

Vi vill hitta en vektor $Pb \in V$ sådan att $b - Pb \perp V$, dvs

$$(b - Pb, v) = 0 \quad \forall v \in V. \tag{1}$$

Det räcker att testa ortogonalitet mot alla basvektorer, dvs

$$(b - Pb, a_j) = 0, \quad j = 1, \dots, n. \tag{2}$$

För om (2) gäller så skriver vi $v = \sum_{j=1}^n \alpha_j a_j$ och

$$(b - Pb, \sum_{j=1}^n \alpha_j a_j) = \sum_{j=1}^n \alpha_j \underbrace{(b - Pb, a_j)}_{=0} = 0,$$

så att (1) gäller. Det är klart att (1) medför (2).

Nu beräknar vi Pb. Vi gör en ansats: $Pb = \sum_{i=1}^n x_i a_i = Bx$. Vi måste bestämma koefficienterna x_i. Ekvation (2) ger

$$\begin{aligned} &(Pb, a_j) = (b, a_j), \quad j = 1, \dots, n \\ &\sum_{i=1}^n x_i (a_i, a_j) = (b, a_j), \quad j = 1, \dots, n. \end{aligned}$$

Detta är ett ekvationsystem för x_i. Alternativt kan man skriva detta på matrisform

$$\begin{aligned} &(Bx, a_j) = (b, a_j), \quad j = 1, \dots, n, \\ &a_j^T Bx = a_j^T b, \quad j = 1, \dots, n, \\ &B^T Bx = B^T b. \end{aligned}$$

Detta är ett system av n ekvationer i n obekanta med symmetrisk matris: $B^T B$.

$$\underbrace{\underbrace{B^T}_{n\times m}\underbrace{B}_{m\times n}}_{n\times n}\underbrace{x}_{n\times 1} = \underbrace{\underbrace{B^T}_{n\times m}\underbrace{b}_{m\times 1}}_{n\times 1}$$

I exemplet:

$$B^T B = \begin{bmatrix} 1 & 4 & -10 \\ 2 & 3 & -5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ -10 & -5 \end{bmatrix} = \begin{bmatrix} 117 & 64 \\ 64 & 38 \end{bmatrix}$$

$$B^T b = \begin{bmatrix} 1 & 4 & -10 \\ 2 & 3 & -5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ 0 \end{bmatrix}$$

Ekvationssystemet $B^T Bx = B^T b$ blir

$$\begin{bmatrix} 117 & 64 \\ 64 & 38 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -5 \\ 0 \end{bmatrix}$$

och leder till trappstegsmatrisen $\left[\begin{array}{cc|c} 117 & 64 & -5 \\ 0 & \frac{350}{117} & \frac{320}{117} \end{array}\right]$. Vi får unik lösning

$$\begin{cases} \hat{x}_2 = \frac{320}{350} \approx 0.9143 \\ \hat{x}_1 \approx -0.5429 \end{cases}$$

dvs $\hat{x} = \begin{bmatrix} -0.543 \\ 0.914 \end{bmatrix}$ och $Pb = B\hat{x} \approx \begin{bmatrix} 1.286 \\ 0.571 \\ 0.857 \end{bmatrix}$

Vi går tillbaka till det allmänna fallet. Vi vill lösa $B^T Bx = B^T b$. Vi ska visa att matrisen $B^T B$ är icke-singulär. Eftersom den är en kvadratisk matris, $n \times n$, räcker det att visa att $N(B^T B) = \{0\}$ enligt Sats 42.6.

Vi kollar detta genom att lösa det homogena systemet $B^T Bx = 0$. Multiplicera med x^T:

$$\begin{aligned}
&x^T B^T Bx = 0 \\
&(Bx)^T (Byx) = 0 \\
&\|Bx\|^2 = 0 \\
&Bx = 0 \\
&x = 0 \quad \text{ty } a_1, \ldots, a_n \text{ är linjärt oberoende.}
\end{aligned}$$

Homogena ekvationen har alltså endast den triviala lösningen $x = 0$. Vi drar slutsatsen att $B^T B$ är icke-singulär och vår inhomogena ekvation $B^T Bx = B^T b$ har unik lösning för varje högerled. Vi kallar lösningen $\hat{x}$. Den ges av formeln $\hat{x} = (B^T B)^{-1} B^T b$. Vi har därmed bestämt en unik vektor $Pb \in V$ sådan att $b - Pb \perp V$. Den kallas för den ortogonala projektionen av b på V och ges av formeln

$$Pb = B\hat{x} = B(B^T B)^{-1} B^T b.$$

Projektionen är en linjär operator och ges av matrisen

$$P = B(B^T B)^{-1} B^T.$$

Sats 42.10 ger en annan karaktärisering av den ortogonala projektionen: Pb är det element i V som ligger närmast b.

Sats 42.10

$$(b - Pb, v) = 0 \quad \forall v \in V \qquad \Leftrightarrow \qquad \|b - Pb\| \leq \|b - v\| \quad \forall v \in V.$$

Läs beviset!

Exemplet: $c = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \in V$

$$\|b - Pb\| = \left\| \begin{bmatrix} 1 - 1.286 \\ 1 - 0.571 \\ 1 - 0.857 \end{bmatrix} \right\| \approx 0.5345 \leq \|b - c\| = \left\| \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\| = 1.$$

ON-bas

Det är lättare att räkna om vi använder en ON-bas för V. Vektorerna $\{g_1, \dots, g_n\}$ är en ON-bas om de är ortogonala och normerade dvs

$$(g_i, g_j) = \begin{cases} 0, & i \neq j \\ \|g_i\|^2 = 1, & i = j \end{cases}$$

Med Kroneckers delta: $\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$ kan vi skriva detta $(g_i, g_j) = \delta_{ij}$.

Vi uttrycker en allmän vektor $v \in V$ med den nya basen, $v = \sum_{i=1}^{n} y_i g_i$. Koefficienterna ges av

$$(v, g_j) = (\sum_i y_i g_i, g_j) = \sum_i y_i \underbrace{(g_i, g_j)}_{=\delta_{ij}} = y_j.$$

Vi får $y_j = (v, g_j)$, enkelt eller hur!

Normen av vektorn uttrycks också enkelt i ON-basen. Vi har Pythagoras sats:

$$\|v\|^2 = (v, v) = (\sum_i y_i g_i, \sum_j y_j g_j) = \sum_i \sum_j y_i y_j (g_i, g_j) = \sum_i y_i^2.$$

Dessa räkningar kan också skrivas på matrisform. Vi bildar matrisen $Q = [g_1, \dots, g_n]$, $m \times n$. Att detta är en ON-bas betyder $Q^T Q = I$, $n \times n$. Vektorn $v \in V$ uttrycks med ON-basen, $v = Qy$. Koefficienterna y bestäms genom att multiplicera med Q^T från vänster: $Q^T v = Q^T Q y = I y = y$. Pythagoras sats blir $\|v\|^2 = v^T v = (Qy)^T (Qy) = y^T Q^T Q y = y^T I y = y^T y = \|y\|^2$.

Hur skapar man en ON-bas? En ”vanlig” bas kan ortonormaliseras med hjälp av Gram-Schmidts metod. Vi visar hur det går till i vårt exempel.

Exemplet:

1. $g_1 = \dfrac{a_1}{\|a_1\|} = \dfrac{1}{\sqrt{117}} \begin{bmatrix} 1 \\ 4 \\ -10 \end{bmatrix} \approx \begin{bmatrix} 0.0925 \\ 0.3698 \\ -0.9245 \end{bmatrix}$ (normering av första basvektorn)
   $V_1 = S(a_1) = S(g_1)$
2. Ansats: $\tilde{g}_2 = a_2 - \alpha_1 g_1$ (linjär kombination av gamla basvektorn a_2 och den nya g_1).

Bestäm α_1 så att $\tilde{g}_2 \perp V_1$ dvs

$$(a_2 - \alpha_1 g_1, g_1) = 0$$
$$(a_2, g_1) - \alpha_1 \underbrace{(g_1, g_1)}_{=1} = 0$$

$$\alpha_1 = (a_2, g_1) = g_1^T a_2 = \frac{1}{\sqrt{117}}[1 \quad 4 \quad -10] \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix} = \frac{64}{\sqrt{117}}$$

$$\tilde{g}_2 = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix} - \frac{64}{117} \begin{bmatrix} 1 \\ 4 \\ -10 \end{bmatrix} \approx \begin{bmatrix} 1.45 \\ 0.81 \\ 0.47 \end{bmatrix}$$

$$g_2 = \frac{\tilde{g}_2}{\|\tilde{g}_2\|} = \begin{bmatrix} 0.8401 \\ 0.4695 \\ 0.2718 \end{bmatrix}$$

Vi har nu en ON-bas $Q = [g_1, g_2] = \begin{bmatrix} 0.0925 & 0.8401 \\ 0.3698 & 0.4695 \\ -0.9245 & 0.2718 \end{bmatrix}$, som spänner upp samma rum som den gamla basen $B = [a_1, a_2] = \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ -10 & -5 \end{bmatrix}$.

Ekvationssystemet för ortogonal projektion blir nu: $Q^T Q \hat{y} = Q^T b$, dvs $\hat{y} = Q^T b$, och projektionen blir $Pb = Q\hat{y} = QQ^T b$, dvs $P = QQ^T$.

Vi uttrycker vektorn $c \in V$ i ON-basen: $c = \alpha_1 g_1 + \alpha_2 g_2$, koefficienterna blir

$$\alpha_1 = (c, g_1) = g_1^T c \approx \begin{bmatrix} 0.0925 & 0.3698 & -0.9245 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 1.3868,$$

$$\alpha_2 = (c, g_2) = g_2^T c \approx \begin{bmatrix} 0.8401 & 0.4695 & 0.2718 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 1.0377.$$

Pythagoras sats: $\|c\|^2 = \alpha_1^2 + \alpha_2^2 \approx 3$.

Eller i matrisform:

$$c = Q\alpha, \quad \alpha = Q^T c = \begin{bmatrix} 0.0925 & 0.3698 & -0.9245 \\ 0.8401 & 0.4695 & 0.2718 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \approx \begin{bmatrix} 1.3868 \\ 1.0377 \end{bmatrix}.$$

The least squares method

We now return to the original example

$$A = [a_1, a_2, a_3, a_4] = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ -10 & -5 & 0 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The equation $Ax = b$ has no solution. It is an *underdetermined system* because it has 4 unknowns but only 3 equations. We might also consider *overdetermined systems*, with more equations than unknowns.

For a system of equations $Ax = b$, with A of type $m \times n$, which has no solution we might instead try to find x such that the norm of the residual $b - Ax$ is minimal, i.e., such that

$$\|b - Ax\|$$

is minimal. From Theorem 42.10 we know that there is a unique vector $Pb \in V = R(A)$ which minimizes the distance between b and V. Also, since Pb belongs to $R(A)$ we can express it as at linear combination of the columns of A, i.e., $Pb = Ax$ for some $x \in \mathbb{R}^n$. How do we compute x? The condition that $b - Ax \perp V$ leads to

$$\begin{aligned}
&(Ax, a_j) = (b, a_j), \quad j = 1, \ldots, n,\\
&a_j^T Ax = a_j^T b, \quad j = 1, \ldots, n,
\end{aligned}$$

which is the same as

$$A^T Ax = A^T b.$$

This is a symmetric system of equations with n equations and n unknowns. We have already seen that it has a solution $x \in \mathbb{R}^n$, but it may be non-unique. It is unique if the columns of A are linearly independent (i.e., a basis), remember that we proved this for $B^T B$ before. If there are more than one solution it is customary to choose the one which is closest to the origin, i.e., the one with minimal norm $\|x\|$. This is called the least-squares solution, $\hat{x}$.

In Matlab: `x=b\A` computes the unique solution by the Gauss elimination method when A is non-singular, otherwise it automatically uses the least-squares method. If the least-squares method has multiple solutions, then it selects one with as many zero components as possible. This is ususally not the same as the one with minimal norm. To obtain this we use the so-called pseudo-inverse: `x=pinv(A)*b`.

In our example: Matlab calculations give

$$A^T A = \begin{bmatrix} 117 & 64 & 11 & -42 \\ 64 & 38 & 12 & -14 \\ 11 & 12 & 13 & 14 \\ -42 & -14 & 14 & 42 \end{bmatrix}, \quad A^T b = \begin{bmatrix} -5 \\ 0 \\ 5 \\ 10 \end{bmatrix}.$$

The command `rref( [A’*A, A’*b] )` gives the echelon matrix

$$\left[\begin{array}{rrrr|r} 1 & 0 & -1 & -2 & -0.5429 \\ 0 & 1 & 2 & 3 & 0.9143 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

and we see that the system $A^T Ax = A^T b$ has multiple solutions

$$x = s \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -0.5429 \\ 0.9143 \\ 0 \\ 0 \end{bmatrix}$$

The command `x=b\A` gives

$$x = \begin{bmatrix} 0.0667 \\ 0 \\ 0 \\ 0.3048 \end{bmatrix}.$$

Here we have $s = 0$, $t = 0.3048$. The command `x=pinv(A)*b` gives

$$\hat{x} = \begin{bmatrix} -0.0143 \\ 0.0571 \\ 0.1286 \\ 0.2000 \end{bmatrix}.$$

Here we have $s = 0.1286$, $t = 0.2000$.

`2004-02-07 /stig`