Linear least squares problem: Example

We want to determine n unknown parameters $c_1,\dots,c_n$ using m measurements where $m \geq n$.
Here $\|\cdot\|$ always denotes the 2-norm $\|v\|_2 = (v_1^2+\cdots+v_n^2)^{1/2}$.

Example problem

Fit the experimental data

t	0	1	2	3
y	0	1	4	7

with a curve of the form $g(t) = c_1\cdot 1 + c_2\cdot t + c_3\cdot t^2$.

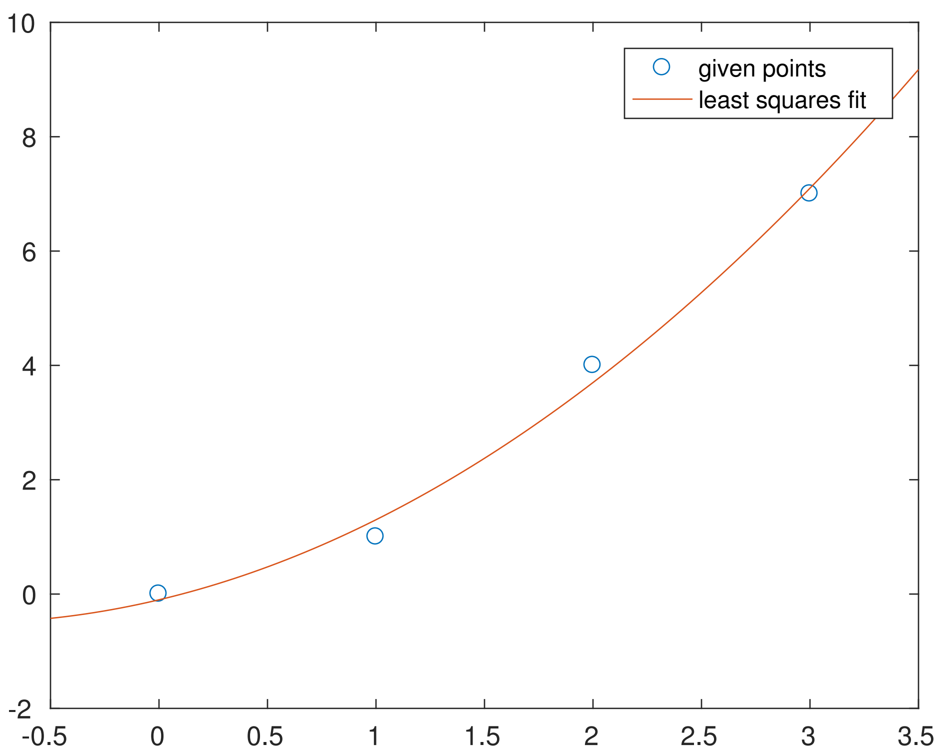
Here $n=3$, $m=4$ and $g_1(t)=1$, $g_2(t)=t$, $g_3(t)=t^2$. We define the matrix $A := \begin{bmatrix} g_1(t_1) & \cdots & g_n(t_1) \\ \vdots & & \vdots \\ g_1(t_m) & \cdots & g_n(t_m) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$.

We want to find $c \in \mathbb{R}^3$ such that $\|Ac-y\|_2 = \min$.

Find solution using normal equations: Find $M := A^\top A = \begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix}$ and $b := A^\top y = \begin{bmatrix} 12 \\ 30 \\ 80 \end{bmatrix}$. Then solve the 3×3 linear system $Mc = b$ using Gaussian elimination, yielding the solution vector $c = \begin{bmatrix} -0.1 \\ 0.9 \\ 0.5 \end{bmatrix}$.

```
t = [0;1;2;3]; y = [0;1;4;7];
A = [t.^0,t,t.^2];
c = (A'*A)\(A'*y)                  % use normal equations
c = A\y                            % Matlab shortcut (actually uses QR decomposition)

te = linspace(-.5,3.5,1e2)';       % points for plotting
plot(t,y,'o',te,[te.^0,te,te.^2]*c) % plot given points and fitted curve
legend('given points','least squares fit')
```



Least squares problem with orthogonal basis

For a least squares problem we are given n linearly independent vectors $a^{(1)}, \ldots, a^{(n)} \in \mathbb{R}^m$ which form a basis for the subspace $V = \operatorname{span}\{a^{(1)}, \ldots, a^{(n)}\}$. For a given right hand side vector $y \in \mathbb{R}^m$ we want to find $u \in V$ such that $\|u - y\|$ is minimal. We can write $u = c_1 a^{(1)} + \cdots + c_n a^{(n)} = Ac$ with the matrix $A = \left[a^{(1)}, \ldots, a^{(n)}\right] \in \mathbb{R}^{m \times n}$. Hence we want to find $c \in \mathbb{R}^n$ such that $\|Ac - y\|$ is minimal.

Solving this problem is much simpler if we have an **orthogonal basis for the subspace** V: Assume we have vectors $p^{(1)}, \ldots, p^{(n)}$ such that

- $\operatorname{span}\left\{p^{(1)}, \ldots, p^{(n)}\right\} = V$
- the vectors are orthogonal on each other: $p^{(i)} \cdot p^{(j)} = 0$ for $i \neq j$

We can then write $u = d_1 p^{(1)} + \cdots + d_n p^{(n)} = Pd$ with the matrix $P = [p^{(1)}, \ldots, p^{(n)}] \in \mathbb{R}^{m \times n}$. Hence we want to find $d \in \mathbb{R}^n$ such that $\|Pd - b\|$ is minimal. The normal equations for this problem give

$$(P^\top P)d = P^\top b \tag{1}$$

where the matrix

$$P^\top P = \begin{bmatrix} p^{(1)\top} \\ \vdots \\ p^{(n)\top} \end{bmatrix} \left[p^{(1)}, \ldots, p^{(n)}\right] = \begin{bmatrix} p^{(1)} \cdot p^{(1)} & \cdots & p^{(1)} \cdot p^{(n)} \\ \vdots & & \vdots \\ p^{(n)} \cdot p^{(1)} & \cdots & p^{(n)} \cdot p^{(n)} \end{bmatrix} = \begin{bmatrix} p^{(1)} \cdot p^{(1)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p^{(n)} \cdot p^{(n)} \end{bmatrix}$$

is now diagonal since $p^{(i)} \cdot p^{(j)} = 0$ for $i \neq j$. Therefore the normal equations (1) are actually decoupled

$$\left(p^{(1)} \cdot p^{(1)}\right) d_1 = p^{(1)} \cdot y$$

$$\vdots$$

$$\left(p^{(n)} \cdot p^{(n)}\right) d_n = p^{(n)} \cdot y$$

and have the solution

$$d_i = \frac{p^{(i)} \cdot y}{p^{(i)} \cdot p^{(i)}} \qquad \text{for } i = 1, \ldots, n$$

Gram-Schmidt orthogonalization

We still need a method to construct from a given basis $a^{(1)}, \ldots, a^{(n)}$ an orthogonal basis $p^{(1)}, \ldots, p^{(n)}$.

Given n linearly independent vectors $a^{(1)}, \ldots, a^{(n)} \in \mathbb{R}^m$ we want to find vectors $p^{(1)}, \ldots, p^{(n)}$ such that

- $\operatorname{span}\left\{p^{(1)}, \ldots, p^{(n)}\right\} = \operatorname{span}\left\{a^{(1)}, \ldots, a^{(n)}\right\}$
- the vectors are orthogonal on each other: $p^{(i)} \cdot p^{(j)} = 0$ for $i \neq j$

Step 1: $\boxed{p^{(1)} := a^{(1)}}$

Step 2: $\boxed{p^{(2)} := a^{(2)} - s_{12} p^{(1)}}$ where we choose s_{12} such that $p^{(1)} \cdot p^{(2)} = 0$:

$$p^{(1)} \cdot a^{(2)} - s_{12} p^{(1)} \cdot p^{(1)} = 0 \qquad \Longleftrightarrow \qquad \boxed{s_{12} = \frac{p^{(1)} \cdot a^{(2)}}{p^{(1)} \cdot p^{(1)}}}$$

Step 3: $\boxed{p^{(3)} := a^{(3)} - s_{13} p^{(1)} - s_{23} p^{(2)}}$ where we choose s_{13}, s_{23} such that

- $p^{(1)} \cdot p^{(3)} = 0$, i.e., $p^{(1)} \cdot a^{(3)} - s_{13} p^{(1)} \cdot p^{(1)} - s_{23} \underbrace{p^{(1)} \cdot p^{(2)}}_{0} = 0$, hence $\boxed{s_{13} = \frac{p^{(1)} \cdot a^{(3)}}{p^{(1)} \cdot p^{(1)}}}$

- $p^{(2)} \cdot p^{(3)} = 0$, i.e., $p^{(2)} \cdot a^{(3)} - s_{13} \underbrace{p^{(2)} \cdot p^{(1)}}_{0} - s_{23} p^{(2)} \cdot p^{(2)} = 0$, hence $\boxed{s_{23} = \dfrac{p^{(2)} \cdot a^{(3)}}{p^{(2)} \cdot p^{(1)}}}$

$\vdots$

Step n**:** $\boxed{p^{(n)} := a^{(n)} - s_{1n} p^{(1)} - \cdots - s_{n-1,n} p^{(n-1)}}$ where we choose $s_{1n}, \ldots, s_{n-1,n}$ such that $p^{(j)} \cdot p^{(n)} = 0$ for $j = 1, \ldots, n-1$ which yields

$$\boxed{s_{jn} = \frac{p^{(j)} \cdot p^{(n)}}{p^{(j)} \cdot p^{(j)}} \quad \text{for } j = 1, \ldots, n-1}$$

Example: We are given the vectors $a^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $a^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$, $a^{(3)} = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 9 \end{bmatrix}$. Use Gram-Schmidt orthogonalization to find an orthogonal basis $p^{(1)}, p^{(2)}, p^{(3)}$ for the subspace $V = \text{span}\left\{a^{(1)}, a^{(2)}, a^{(3)}\right\}$.

Step 1: $\boxed{p^{(1)} := a^{(1)}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Step 2: $\boxed{p^{(2)} := a^{(2)} - \dfrac{p^{(1)} \cdot a^{(2)}}{p^{(1)} \cdot p^{(1)}} p^{(1)}} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} - \dfrac{6}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{3}{2} \end{bmatrix}$

Step 3: $\boxed{p^{(3)} := a^{(3)} - \dfrac{p^{(1)} \cdot a^{(3)}}{p^{(1)} \cdot p^{(1)}} p^{(1)} - \dfrac{p^{(2)} \cdot a^{(3)}}{p^{(2)} \cdot p^{(2)}} p^{(2)}} = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 9 \end{bmatrix} - \dfrac{14}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \dfrac{15}{5} \begin{bmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$

Note that we have

$$\begin{aligned} a^{(1)} &= p^{(1)} \\ a^{(2)} &= p^{(2)} + \frac{6}{4} p^{(1)} \\ a^{(3)} &= p^{(3)} + \frac{14}{4} p^{(1)} + \frac{15}{5} p^{(2)} \end{aligned}$$

which we can write as

$$\left[a^{(1)}, a^{(2)}, a^{(3)}\right] = \left[p^{(1)}, p^{(2)}, p^{(3)}\right] \begin{bmatrix} 1 & \frac{6}{4} & \frac{14}{4} \\ 0 & 1 & \frac{15}{5} \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & -1.5 & 1 \\ 1 & -0.5 & -1 \\ 1 & 0.5 & -1 \\ 1 & 1.5 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 1.5 & 3.5 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}}_{S}$$

In the general case we have

$$\begin{aligned} a^{(1)} &= p^{(1)} \\ a^{(2)} &= p^{(2)} + s_{12} p^{(1)} \\ a^{(3)} &= p^{(3)} + s_{13} p^{(1)} + s_{13} p^{(2)} \\ &\vdots \\ a^{(n)} &= p^{(n)} + s_{1n} p^{(1)} + \cdots + s_{n-1,n} p^{(n-1)} \end{aligned}$$

which we can write as

$$\left[a^{(1)},a^{(2)},\ldots,a^{(n)}\right]=\left[p^{(1)},p^{(2)},\ldots,p^{(n)}\right]\begin{bmatrix}1 & s_{12} & \cdots & s_{1n}\\ 0 & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & s_{n-1,n}\\ 0 & \cdots & 0 & 1\end{bmatrix}$$

Therefore we obtain a decomposition $A=PS$ where

- $P\in\mathbb{R}^{m\times n}$ has orthogonal columns
- $S\in\mathbb{R}^{n\times n}$ is upper triangular, with 1 on the diagonal.

Note that the vectors $p^{(1)},\ldots,p^{(n)}$ are different from $\vec{0}$:

Assume, e.g., that $p^{(3)}=a^{(3)}-s_{13}p^{(1)}-s_{23}p^{(2)}=\vec{0}$, then $a^{(3)}=s_{13}p^{(1)}+s_{23}p^{(2)}$ is in span $\{p^{(1)},p^{(2)}\}=$ span $\{a^{(1)},a^{(2)}\}$. This is a contradiction to the assumption that $a^{(1)},a^{(2)},a^{(3)}$ are linearly independent.

Solving the least squares problem $\|Ac-y\|=\min$ using orthogonalization

We are given $A\in\mathbb{R}^{m\times n}$ with linearly independent columns, $b\in\mathbb{R}^n$. We want to find $c\in\mathbb{R}^n$ such that $\|Ac-y\|=\min$.

From the Gram-Schmidt method we get $A=PS$, hence we want to find c such that

$$\|P\underbrace{Sc}_{d}-y\|=\min$$

This gives the following method:

Algorithm: solve least squares problem $\|Ac-y\|_2=\min$ using orthogonalization

- use Gram-Schmidt to find decomposition $A=PS$
- solve $\|Pd-y\|=\min$: $d_i:=\dfrac{p^{(i)}\cdot y}{p^{(i)}\cdot p^{(i)}}$ for $i=1,\ldots,n$
- solve $Sc=d$ by back substitution

Example: Solve the least squares problem $\|Ac-b\|=\min$ for $A=\begin{bmatrix}1&0&0\\1&1&1\\1&2&4\\1&3&9\end{bmatrix}$, $b=\begin{bmatrix}0\\1\\4\\7\end{bmatrix}$.

- Gram-Schmidt gives $A=\underbrace{\begin{bmatrix}1&-1.5&1\\1&-0.5&-1\\1&0.5&-1\\1&1.5&1\end{bmatrix}}_{P}\underbrace{\begin{bmatrix}1&1.5&3.5\\0&1&3\\0&0&1\end{bmatrix}}_{S}$ (see above)
- $d_1=\dfrac{p^{(1)}\cdot b}{p^{(1)}\cdot p^{(1)}}=\dfrac{12}{4}=3$, $d_2=\dfrac{p^{(2)}\cdot b}{p^{(2)}\cdot p^{(2)}}=\dfrac{12}{5}=2.4$, $d_3=\dfrac{p^{(3)}\cdot b}{p^{(3)}\cdot p^{(3)}}=\dfrac{2}{4}=0.5$
- solving $\begin{bmatrix}1&1.5&3.5\\0&1&3\\0&0&1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}3\\2.4\\0.5\end{bmatrix}$ by back substitution gives $c_3=0.5$, $c_2=0.9$, $c_1=-1.1$

Hence the solution of our least squares problem is the vector $c=\begin{bmatrix}-1.1\\0.9\\0.5\end{bmatrix}$.

Note: If you want to solve a least squares problem by hand with pencil and paper, it is usually easier to use the normal equations. But for numerical computation on a computer using orthogonalization is usually more efficient and more accurate.

Finding an ortho*normal* basis $q^{(1)}, \ldots, q^{(n)}$: the QR decomposition

The Gram-Schmidt method gives an orthogonal basis $p^{(1)}, \ldots, p^{(n)}$ for $V = \text{span}\left\{a^{(1)}, \ldots, a^{(n)}\right\}$

Often it is convenient to have a so-called ortho*normal* basis $q^{(1)}, \ldots, q^{(n)}$ where the basis vectors have length 1: Define

$$q^{(j)} = \frac{1}{\left\|p^{(j)}\right\|} p^{(j)} \qquad \text{for } j = 1, \ldots, n$$

then we have

- $\text{span}\left\{q^{(1)}, \ldots, q^{(n)}\right\} = V$
- $q^{(j)} \cdot q^{(k)} = \begin{cases} 1 & \text{for } j = k \\ 0 & \text{otherwise} \end{cases}$

This means that the matrix $Q = [q^{(1)}, \ldots, q^{(n)}]$ satisfies $Q^\top Q = I$ where I is the $n \times n$ identity matrix.

Since $p^{(j)} = \left\|p^{(j)}\right\| q^{(j)}$ we have

$$\begin{aligned}
a^{(1)} &= \underbrace{\left\|p^{(1)}\right\|}_{r_{11}} q^{(1)} \\
a^{(2)} &= \underbrace{\left\|p^{(2)}\right\|}_{r_{22}} q^{(2)} + \underbrace{s_{12}\left\|p^{(1)}\right\|}_{r_{12}} p^{(1)} \\
&\vdots \\
a^{(n)} &= \underbrace{\left\|p^{(n)}\right\|}_{r_{nn}} q^{(n)} + \underbrace{s_{1n}\left\|p^{(1)}\right\|}_{r_{1n}} p^{(1)} + \cdots + \underbrace{s_{n-1,n}\left\|p^{(n-1)}\right\|}_{r_{n-1,n}} q^{(n-1)}
\end{aligned}$$

which we can write as

$$\left[a^{(1)}, a^{(2)}, \ldots, a^{(n)}\right] = \left[q^{(1)}, q^{(2)}, \ldots, q^{(n)}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{n-1,n} \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

$$A = QR$$

where the $n \times n$ matrix R is given by

$$\begin{bmatrix} \text{row } 1 \text{ of } R \\ \vdots \\ \text{row } n \text{ of } R \end{bmatrix} = \begin{bmatrix} \left\|p^{(1)}\right\| \cdot (\text{row 1 of } S) \\ \vdots \\ \left\|p^{(n)}\right\| \cdot (\text{row } n \text{ of } S) \end{bmatrix}$$

We obtain the so-called **QR decomposition** $A = QR$ where

- the matrix $Q \in \mathbb{R}^{m \times n}$ has orthonormal columns, $\text{range}\, Q = \text{range}\, A$
- the matrix $R \in \mathbb{R}^{n \times n}$ is upper triangular, with nonzero diagonal elements

Example: In our example we have $p^{(1)} \cdot p^{(1)} = 4$, $p^{(2)} \cdot p^{(2)} = 5$, $p^{(3)} \cdot p^{(3)} = 4$, hence

$$q^{(1)} = \frac{1}{2} p^{(1)} = \begin{bmatrix} .5 \\ .5 \\ .5 \\ .5 \end{bmatrix}, \qquad q^{(2)} = \frac{1}{\sqrt{5}} p^{(2)} = \frac{1}{\sqrt{5}} \begin{bmatrix} -1.5 \\ -.5 \\ .5 \\ 1.5 \end{bmatrix}, \qquad q^{(3)} = \frac{1}{2} p^{(3)} = \begin{bmatrix} .5 \\ -.5 \\ -.5 \\ .5 \end{bmatrix}$$

and we obtain the QR decomposition

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} .5 & -1.5/\sqrt{5} & .5 \\ .5 & -.5/\sqrt{5} & -.5 \\ .5 & .5/\sqrt{5} & -.5 \\ .5 & 1.5/\sqrt{5} & .5 \end{bmatrix} \begin{bmatrix} 2 & 3 & 7 \\ 0 & \sqrt{5} & 3\sqrt{5} \\ 0 & 0 & 2 \end{bmatrix}$$

In Matlab we can find the QR decomposition using **`[Q,R]=qr(A,0)`**

```
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]';
>> [Q,R] = qr(A,0)
Q =
   -0.5000    0.6708    0.5000
   -0.5000    0.2236   -0.5000
   -0.5000   -0.2236   -0.5000
   -0.5000   -0.6708    0.5000
R =
   -2.0000   -3.0000   -7.0000
         0   -2.2361   -6.7082
         0         0    2.0000
```

Note that Matlab returned the basis $-q^{(1)}, -q^{(2)}, q^{(3)}$ (which is also an orthonormal basis) and hence rows 1 and 2 of the matrix R is (-1) times our previous matrix R.

If we want to find an orthonormal basis for $\text{range}A$ and an orthonormal basis for the orthogonal complement $(\text{range}A)^\perp = \text{null}A^\top$ we can use the command **`[Qh,Rh]=qr(A)`** : It returns matrices $\hat{Q} \in \mathbb{R}^{m\times m}$ and $\hat{R} \in \mathbb{R}^{m\times n}$ with

$$\hat{Q} = \begin{bmatrix} \overbrace{q^{(1)},\ldots,q^{(n)}}^{\text{basis for range}A}, \overbrace{q^{(n+1)},\ldots,q^{(m)}}^{\text{basis for }(\text{range}A)^\perp} \end{bmatrix}, \qquad \hat{R} = \begin{bmatrix} R \\ \begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix} \end{bmatrix} \left.\vphantom{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}}\right\} m-n \text{ rows of zeros}$$

```
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]';
>> [Qh,Rh] = qr(A)
Qh =
   -0.5000    0.6708    0.5000    0.2236
   -0.5000    0.2236   -0.5000   -0.6708
   -0.5000   -0.2236   -0.5000    0.6708
   -0.5000   -0.6708    0.5000   -0.2236
Rh =
   -2.0000   -3.0000   -7.0000
         0   -2.2361   -6.7082
         0         0    2.0000
         0         0         0
```

But in most cases we only need an orthonormal basis for $\text{range}A$ and we should use `[Q,R]=qr(A,0)` (which Matlab calls the "economy size" decomposition).

Solving the least squares problem $\|Ac - b\| = \min$ using the QR decomposition

If we use an orthonormal basis $q^{(1)},\ldots,q^{(n)}$ for $\text{span}\{a^{(1)},\ldots,a^{(n)}\}$ we have $Q^\top Q = I$. The solution of $\|Qd - y\| = \min$ is therefore given by the normal equations $(Q^\top Q)d = Q^\top y$, i.e., we obtain $d = Q^\top y$.

This gives the following method:

Algorithm: solve the least squares problem $\|Ac - y\|_2 = \min$ using orthonormalization:

- find the QR decomposition $A = QR$
- let $d = Q^\top y$
- solve $Rc = d$ by back substitution

In Matlab we can do this as follows:

```
[Q,R] = qr(A,0);
d = Q'*y;
c = R\d;
```

In our example we have

```
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]'; y = [0;1;4;7];
>> [Q,R] = qr(A,0);
>> d = Q'*y;
>> c = R\d
c =
   -0.1000
    0.9000
    0.5000
```

We can use the **shortcut `c=A\y`** which actually uses the QR decomposition to find the solution of $\|Ac - y\|_2 = \min$

```
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]'; y = [0;1;4;7];
>> c = A\y
c =
   -0.1000
    0.9000
    0.5000
```

Warning: In **Matlab symbolic mode** the backslash command does **not** find the least squares solution:

```
>> A = sym([1 1 1 1; 0 1 2 3; 0 1 4 9])'; y = sym([0;1;4;7]);
>> c = A\y
Warning: The system is inconsistent. Solution does not exist.
c =
 Inf
 Inf
 Inf
```