- New concepts:
  - Use $\boldsymbol{QR}$ factorization to solve least squares problems
  - Comparison to normal system

Revisit polynomial least squares problem and its geometric interpretation

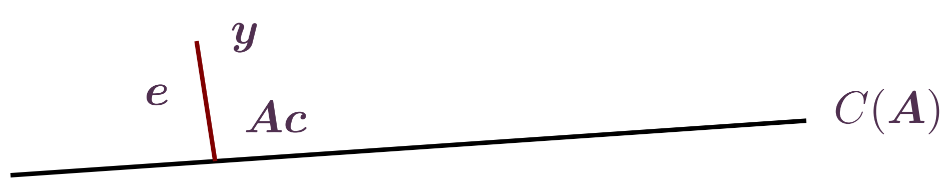
- Given data $\mathcal{D} = \{(x_i, y_i), i = 1, \dots, \}$ the problem of finding a polynomial

$$p(x) = c_0 + c_1 x + c_2 x^2 + \dots + c_n x^n$$

that passes as close as possible to the data points led to the minimization of the norm of the error vector

$$\boldsymbol{e} = \boldsymbol{A}\boldsymbol{c} - \boldsymbol{y} = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{x} & \boldsymbol{x}^2 & \dots & \boldsymbol{x}^n \end{pmatrix} - \boldsymbol{y}, \text{with}\, \boldsymbol{c} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \boldsymbol{x}^k = \begin{pmatrix} x_0^m \\ x_1^m \\ x_2^m \\ \vdots \\ x_m^m \end{pmatrix}$$

- Note dimensions: $\boldsymbol{e}, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, but $\boldsymbol{c} \in \mathbb{R}^n$. Typically $m \gg n$, "lots of data, few basis column vectors". The minimum error norm is obtained when $\boldsymbol{e} = \boldsymbol{A}\boldsymbol{c} - \boldsymbol{y} \perp C(\boldsymbol{A})$



- Recall that in the $\boldsymbol{Q}\boldsymbol{R}$ factorization, the orthonormal column vectors within $\boldsymbol{Q}$ are obtained as linear combinations of the columns of $\boldsymbol{A}$, and hence the column spaces are the same $C(\boldsymbol{Q}) = C(\boldsymbol{A})$
- Also recall that $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^T$ is the projector onto the column space of $\boldsymbol{Q}$.

- From above, recognize that $\boldsymbol{A}\boldsymbol{c}$, the linear combination of columns of $\boldsymbol{A}$ that gets as close as possible to $\boldsymbol{y}$ is simply the projection of $\boldsymbol{y}$ onto $C(\boldsymbol{A})$

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{y}$$

- Since $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, obtain

$$\boldsymbol{Q}\boldsymbol{R}\boldsymbol{c} = \boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{y} \Rightarrow (\boldsymbol{Q}^T\boldsymbol{Q})\,\boldsymbol{R}\boldsymbol{c} = (\boldsymbol{Q}^T\boldsymbol{Q})\,\boldsymbol{Q}^T\boldsymbol{y} \Rightarrow \boldsymbol{I}_n\boldsymbol{R}\boldsymbol{c} = \boldsymbol{I}_n\boldsymbol{Q}^T\boldsymbol{y} \Rightarrow \boldsymbol{R}\boldsymbol{c} = \boldsymbol{Q}^T\boldsymbol{y}$$

Example

1. Generate some data on a line and perturb it by some random quantities

```
octave> m=1000; x=(0:m-1)/m; c0=2; c1=3; yex=a0+a1*x; y=(yex+rand(1,m)-0.5)';
```

2. Form the matrices $\boldsymbol{A}$, $\boldsymbol{N} = \boldsymbol{A}^T \boldsymbol{A}$, vector $\boldsymbol{b} = \boldsymbol{A}^T \boldsymbol{y}$

```
octave> A=ones(m,2); A(:,2)=x(:); N=A'*A; b=A'*y;
```

3. Solve the system $\boldsymbol{N}\boldsymbol{a} = \boldsymbol{b}$, and form the linear combination $\tilde{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{a}$ closest to $\boldsymbol{y}$

```
octave> c=N\b; disp(c'); ytilde=A*c;
```

1.9921 3.0292

4. Factorize $\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{A}$, and solve system $\boldsymbol{R}\boldsymbol{c} = \boldsymbol{Q}^T \boldsymbol{y}$. Obtain the same $\boldsymbol{c}$ vector

```
octave> [Q R]=qr(A); d=Q(:,1:2)'*y;
```

```
octave> c=R(1:2,1:2)\d; disp(c');
```

1.9921 3.0292

When n is large the $\boldsymbol{Q}\boldsymbol{R}$ solution of a least squares problem is less affected by floating point representation errors than the normal equation solution.