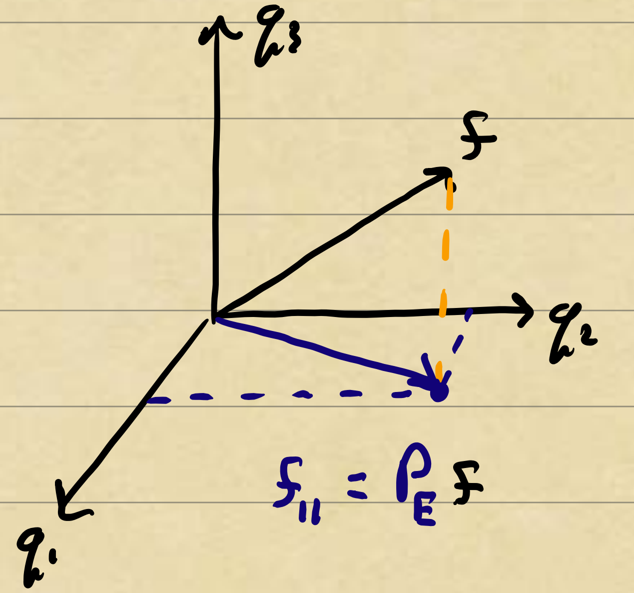
What Makes a "Good" Basis?

Goal: Minimize $\|f - E\underline{c}\|$ in Hilbert norm.

$\uparrow$ $E : \mathbb{R}^N \to H$ (quasimatrix)



Two step procedure:

$\Rightarrow$ Project $\quad f_\parallel = P_E f$

$\Rightarrow$ Solve $\quad E\underline{c} = f_\parallel$

Here, $P_E : H \to \mathrm{col}(E)$ is the orthogonal projection of H onto $\mathrm{col}(E)$.

Algorithm: Compute best approx. via QR.

Step 1. QR decomposition of dictionary:

$$
\begin{array}{ccccc}
E & & Q & & R \\
\begin{bmatrix} | & & | \\ e_1 & \cdots & e_N \\ | & & | \end{bmatrix} & = & \begin{bmatrix} | & & | \\ q_1 & \cdots & q_N \\ | & & | \end{bmatrix} & & \begin{bmatrix} \times & \cdots & \times \\ 0 & \times \cdots & \times \\ \vdots & \ddots & \vdots \\ 0 & \cdots\; 0 & \times \end{bmatrix} \\
\text{dictionary} & & \text{ONB} & & \text{change of basis}
\end{array}
$$

Step 2. Solve upper triangular system

$$R\underline{c} = Q^* f$$

Change-of-Basis and the Gram matrix

R is the change-of-basis matrix $E \to Q$

$$E\underline{c} = QR\underline{c} = Q\underline{b}, \qquad \underline{b} = R\underline{c}$$

$\underline{b}$ = coords in Q basis, $\underline{c}$ = coords in E basis.

Question: How do we change basis between two non-orthogonal bases for col(E)?

Suppose we have two basis sets for finite-dimensional subspace V of Hilbert space $\mathcal{H}$.

$$E = \begin{bmatrix} | & & | \\ e_1 & \cdots & e_N \\ | & & | \end{bmatrix} \qquad F = \begin{bmatrix} | & & | \\ f_1 & \cdots & f_N \\ | & & | \end{bmatrix}$$

Given $g \in V$, we can write

$$g = \begin{bmatrix} | & & | \\ e_1 & \cdots & e_N \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} | & & | \\ f_1 & \cdots & f_N \\ | & & | \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}$$

To express the F-coords of $\underline{g}$ in terms of the E coords of $\underline{g}$, we calculate

(4) $$\underline{c} = \underbrace{(E^T E)^{-1}(E^T F)}_{\text{change-of-basis matrix}}\underline{d}$$

The matrix $E^T E$ is the Gram matrix.

Question: Why/When is the Gram matrix invertible, i.e., $(E^T E)^{-1}$ exist?

$\Rightarrow$ Homework 3, Question 1 (b).

"Good" Basis vs. "Bad" Basis

In demo01.m, we saw that the monomial basis performed poorly for best approximation while the Legendre basis performed well.

Question: What makes a good basis?

Intuitively, a good basis should not be "too far" from orthonormal/orthogonal.

More precisely, a good basis should provide a stable representation of $f \in V$ under perturbations to coordinates of f.

Suppose $f = E\underline{c}$ and $\hat{f} = E\hat{\underline{c}}$, where

$$\|\underline{c} - \hat{\underline{c}}\| \leq \varepsilon \|\underline{c}\|.$$

Can $\|f - \hat{f}\|$ be much bigger than $\varepsilon\|f\|$?

$$\|f - \hat{f}\| = \|E\underline{c} - E\hat{\underline{c}}\| \leq \|E\| \|\underline{c} - \hat{\underline{c}}\|$$
$$\leq \varepsilon \|E\| \|\underline{c}\|$$

The operator norm $\|E\| = \sup_{v \in \mathbb{R}^N} \frac{\|E\underline{v}\|}{\|\underline{v}\|}$ measures the ability of $\underline{v} \mapsto E\underline{v}$ to amplify vectors in its domain. For an ONB Q, the operator norm $\|Q\| = 1$.

To conclude our stability analysis, we would like to bound $\|\underline{c}\| \leq (\text{constant}) \|\underline{f}\|$.

From change-of-basis in (*) with $F\underline{d} = \underline{f}$

$$\underline{c} = (E^T E)^{-1} E^T \underline{f}$$

$$\Rightarrow \quad \|\underline{c}\| \leq \|(E^T E)^{-1}\| \, \|E^T\| \, \|\underline{f}\|$$

The Gram matrix inverse $(E^T E)^{-1}$ tells us how much amplification happens in the map $\underline{f} \to \underline{c}$! Or put another way, how much $E: \underline{c} \to \underline{f}$ can "shrink" the coordinates of $\underline{f}$ to produce $\underline{f}$ itself.

For an ONB Q, $(Q^T Q)^{-1} = I$ has norm 1.

Using $\|E\| \|E^T\| = \|E^T E\|$, we find

$$\boxed{\|\underline{f} - \hat{\underline{f}}\| \leq \varepsilon \, \|E^T E\| \, \|(E^T E)^{-1}\|.}$$

The constant $\kappa_E = \|E^T E\| \|(E^T E)^{-1}\|$ is called the condition number of $E^T E$.

$$\hat{f} = E(c + \underbrace{\varepsilon}_{= \hat{c} - c}) = \overbrace{Ec}^{\text{"signal"}} + \overbrace{E\varepsilon}^{\text{"noise"}}$$

$\|(E^TE)^{-1}\|$ = "how much signal can shrink"

$\|E^TE\|$ = "how much noise can amplify"

In summary, the condition number

$$\kappa_E = \|E^TE\|\,\|(E^TE)^{-1}\|$$

of the Gram matrix tells us how much perturbations to the coordinates of f in the basis E can be amplified to produce larger perturbations in f itself.

Good bases typically have Gram matrices with relatively low condition numbers.

ONB's are the ideal case, because

$$\kappa_Q = \|Q^TQ\|\,\|(Q^TQ)^{-1}\| = \|I\|\,\|I\| = 1.$$

When E is "close" to an ONB, E^TE is "close" to the identity matrix and $\kappa_E \approx 1$.