Math 672 Lecture 34

The best answers to the question, "What is the structure of a linear operator $T: V \to V$?" are provided by the spectral theorems:

- If V is complex and $T \in \mathcal{L}(V)$ is normal, then there exist $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ and an orthonormal basis $\{v_1, \dots, v_n\}$ of V such that
$$T(v) = \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle v_j \quad \forall v \in V.$$

- If V is real and $T \in \mathcal{L}(V)$ is self-adjoint, then there exist $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ and an orthonormal basis $\{v_1, \dots, v_n\}$ of V such that
$$T(v) = \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle v_j \quad \forall v \in V.$$

Application: V is real, $T \in \mathcal{L}(V)$ is self-adjoint, and the eigenvalues of T are all positive, say $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$. Suppose we wish to solve $T(v) = y$ for v, where $y \in V$ is given but may be noisy.

Suppose v^*, y^* are the exact values ($T(v^*) = y^*$), y is a (noisy) measurement of y^*, and v is the solution of $T(v) = y$. We would like to compare
$$\|v - v^*\|_V \quad \text{and} \quad \|y - y^*\|_V$$

or, even better,

$$\frac{\|v-v^*\|}{\|v^*\|} \quad \text{and} \quad \frac{\|y-y^*\|}{\|y^*\|}.$$

We have

$$T(v) = y$$

$$\iff \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle v_j = \sum_{j=1}^{n} \langle y, v_j \rangle v_j$$

$$\iff \lambda_j \langle v, v_j \rangle = \langle y, v_j \rangle, \quad j = 1, 2, \ldots, n \quad \text{(why?)}$$

$$\iff \langle v, v_j \rangle = \frac{\langle y, v_j \rangle}{\lambda_j}, \quad j = 1, 2, \ldots, n$$

$$\iff v = \sum_{j=1}^{n} \frac{\langle y, v_j \rangle}{\lambda_j} v_j$$

Thus

$$T^{-1}(y) = \sum_{j=1}^{n} \frac{\langle y, v_j \rangle}{\lambda_j} v_j.$$

It follows that

$$v - v^* = T^{-1}(y) - T^{-1}(y^*) = T^{-1}(y - y^*)$$

$$= \sum_{j=1}^{n} \frac{\langle y - y^*, v_j \rangle}{\lambda_j} v_j$$

$$\Rightarrow \|v-v^*\|^2 = \sum_{j=1}^{n} \frac{|\langle y-y^*, v_j\rangle|^2}{\lambda_j^2}$$

$$\leq \frac{1}{\lambda_n^2} \sum_{j=1}^{n} |\langle y-y^*, v_j\rangle|^2$$

$$= \frac{\|y-y^*\|^2}{\lambda_n^2}$$

$$\Rightarrow \|v-v^*\| \leq \frac{\|y-y^*\|}{\lambda_n}.$$

Note what this inequality says: The error in the data (y) can be magnified by as much as λ_n^{-1} when the equation is solved for v.

- If λ_n is not too small ($\lambda_n \gtrsim 1$), then the error in the solution is, at worst, not much bigger than the error in the data.
- But if $\lambda_n \ll 1$, then the error in the solution can be a lot bigger than the error in the data.

Actually, though, comparing the absolute errors is not so informative. Note that

$$\|y^*\|^2 = \|T(v^*)\|^2 = \left\| \sum_{j=1}^{n} \lambda_j \langle v^*, v_j\rangle v_j \right\|^2$$

$$= \sum_{j=1}^{n} \lambda_j^2 |\langle v^*, v_j\rangle|^2$$

$$\leq \lambda_1^2 \sum_{j=1}^{n} |\langle v^*, v_j \rangle|^2$$

$$= \lambda_1^2 \|v^*\|^2$$

$$\Rightarrow \|y^*\| \leq \lambda_1 \|v^*\|$$

$$\Rightarrow \frac{1}{\|v^*\|} \leq \frac{\lambda_1}{\|y^*\|}.$$

But then

$$\|v-v^*\| \leq \frac{\|y-y^*\|}{\lambda_n}, \quad \frac{1}{\|v^*\|} \leq \frac{\lambda_1}{\|y^*\|}$$

$$\Rightarrow \frac{\|v-v^*\|}{\|v^*\|} \leq \frac{\lambda_1}{\lambda_n} \frac{\|y-y^*\|}{\|y^*\|}.$$

This is quite meaningful — the relative error in the data can be magnified by as much as $\frac{\lambda_1}{\lambda_n}$ when the equation is solved. We call $\frac{\lambda_1}{\lambda_n}$ the <u>condition number</u> of T (or of the equation $T(v)=y$). It is a measure of how sensitive the problem (of solving $T(v)=y$) is to noise in the data.

The condition number is a very powerful concept. It would be advantageous to have it available for $T \in \mathcal{L}(V,W)$ with no special property.

This whole exercise of understanding the "structure" of a linear operator $T \in \mathcal{L}(V)$ essentially reduces to the question of choosing a basis $\mathcal{B}$ for V such that $\mathcal{M}_{\mathcal{B},\mathcal{B}}(T)$ is as simple as possible.

If we allow different bases for the domain and codomain of T, it turns out that we can handle any $T \in \mathcal{L}(V)$, or even any $T \in \mathcal{L}(V,W)$.

<u>Theorem</u>: Let V, W be finite-dimensional inner product spaces over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) and let $T \in \mathcal{L}(V,W)$. Then there exist orthonormal bases $\mathcal{B} = \{v_1, \dots, v_n\}$ of V and $\mathcal{C} = \{w_1, \dots, w_m\}$ of W and real numbers $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_t \geq 0$, $t = \min\{m,n\}$, such that

$$T(v) = \sum_{j=1}^{t} \sigma_j \langle v, v_j \rangle_V \, w_j$$

<u>Proof</u>: Note that $T^*T \in \mathcal{L}(V)$ is self-adjoint:

$$\begin{aligned} \langle (T^*T)(v), u \rangle_V = \langle T^*(T(v)), u \rangle_V &= \langle T(v), T(u) \rangle_W \\ &= \langle v, T^*(T(u)) \rangle_V \\ &= \langle v, (T^*T)(u) \rangle_V . \end{aligned}$$

Thus there exists an orthonormal basis $\mathcal{B} = \{v_1, \dots, v_n\}$ of V and corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ such that

$$(T^*T)(v) = \sum_{j=1}^{n} \lambda_j \langle v, v_j \rangle v_j \quad \forall v \in V.$$

Wlog we can assume that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Note that

$$\lambda_j = \lambda_j \langle v_j, v_j \rangle_V = \langle \lambda_j v_j, v_j \rangle_V = \langle (T^*T)(v_j), v_j \rangle_V$$

$$= \langle T(v_j), T(v_j) \rangle_V \geq 0$$

$$\Rightarrow \lambda_j \geq 0$$

Define $r \geq 0$ such that

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0$$

($r=0$ in the trivial case that $T=0$; $r=n$ is possible).

Define $\sigma_j = \sqrt{\lambda_j}$ for $j=1,\dots,r$ and

$$w_j = \sigma_j^{-1} T(v_j), \quad j=1,\dots,r.$$

We then have

$$\langle w_j, w_j \rangle_W = \langle \sigma_j^{-1} T(v_j), \sigma_j^{-1} T(v_j) \rangle_W$$

$$= \sigma_j^{-2} \langle T(v_j), T(v_j) \rangle_W$$

$$= \lambda_j^{-1} \langle T^*T(v_j), v_j \rangle_V$$

$$= \lambda_j^{-1} \langle \lambda_j v_j, v_j \rangle_V = 1, \quad j=1,\dots,r,$$

and

$$\begin{aligned}
\langle w_i, w_j \rangle_W &= \langle \sigma_i^{-1} T(v_i), \sigma_j^{-1} T(v_j) \rangle_W \\
&= \sigma_i^{-1} \sigma_j^{-1} \langle T(v_i), T(v_j) \rangle_W \\
&= \sigma_i^{-1} \sigma_j^{-1} \langle T^*T(v_i), v_j \rangle_V \\
&= \sigma_i^{-1} \sigma_j^{-1} \langle \lambda_i v_i, v_j \rangle_V \\
&= \sigma_i \sigma_j^{-1} \langle v_i, v_j \rangle_V = 0, \quad 1 \le i, j \le r,\ i \ne r.
\end{aligned}$$

Thus $\{w_1, \dots, w_r\}$ is an orthonormal subset of W, which implies that $r \le m$ (also $r \le n$ by definition). Let us extend $\{w_1, \dots, w_r\}$ to an orthonormal basis $\{w_1, \dots, w_m\}$ of W and define $\sigma_j = 0$ for $j = r+1, \dots, \min\{m, n\}$. If $r < n$ and $r < j \le n$, then $\lambda_j = 0$, which implies that

$$\begin{aligned}
& (T^*T)(v_j) = 0 \\
\Rightarrow\ & \langle v_j, (T^*T)(v_j) \rangle_V = 0 \\
\Rightarrow\ & \langle T(v_j), T(v_j) \rangle_W = 0 \\
\Rightarrow\ & T(v_j) = 0.
\end{aligned}$$

We then have

$$\begin{aligned}
T(v) &= T\Big(\sum_{j=1}^{n} \langle v, v_j \rangle_V v_j \Big) \\
&= \sum_{j=1}^{n} \langle v, v_j \rangle_V T(v_j)
\end{aligned}$$

$$= \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_V w_j \quad \text{(since } T(v_j) = \sigma_j w_j \text{ for } j = 1, \dots, r \text{ and } T(v_j) = 0 \text{ for } j = r+1, \dots, n)$$

$$= \sum_{j=1}^{t} \sigma_j \langle v, v_j \rangle_V w_j \quad \text{(since } \sigma_j = 0 \text{ for } r < j \le t). \;/\!/$$