The Singular Value Decomposition

Recap

The linear eqn. $Tu=f$ for b'dd lin. op. $T: H \to H$ (Hilbert Space) is well posed (soln. exists, unique, cont. in data) if

$$\|Tu\| \geq \delta \|u\| \quad \text{for all } u \in H$$

(fixed $\delta > 0$)

We can often convert diff. eqn's (unb'dd ops) into integral eqn's (b'dd ops) of the form

$$u(x) + \int_{\Omega_1} K(x,y)\,u(y)\,dy = f(x), \qquad x \in \Omega_2 \subset \mathbb{R}^d$$

This is a Fredholm integral eqn. (2^{nd} Kind), which has the special operator $T = I + K$.

The integral op $[Ku](x) = \int_{\Omega_1} K(x,y)\,u(y)\,dy$ is called Hilbert-Schmidt if $K \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$ or $(\Omega_1 \times \Omega_2)$:

$$\int_{\Omega_2}\int_{\Omega_1} |K(x,y)|^2\,dy\,dx < \infty.$$

Properties of HS Ops ($\Omega_1 = \Omega_2 = \mathbb{R}^d$)

Fact 1: If $f \in L^2(\mathbb{R}^d)$, then $y \to K(x,y) f(y)$ is integrable for a.e. $x \in \mathbb{R}^d$.

$$\Rightarrow \quad g(x) = \int_{\mathbb{R}^d} K(x,y) f(y) \, dy \quad \text{is well-defined a.e.}$$

Fact 2: The operator $K : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is b'dd,

$$\|K\|_{L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)} \leq \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |K(x,y)|^2 \, dx \, dy \right]^{1/2} = \|K\|_{L^2(\mathbb{R}^d \times \mathbb{R}^d)} < \infty.$$

$\Rightarrow$ Operator norm is controlled by kernel's L^2 norm.

Fact 3: The adjoint of K is

$$[K^* f](x) = \int_{\mathbb{R}^d} \overline{K(y,x)} f(y) \, dy$$

Pf $\langle Kf, g \rangle = \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} K(x,y) f(y) \, dy \right] \overline{g(x)} \, dx$

switch int. limits (Fubini) $= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} K(x,y) \overline{g(x)} \, dx \right] f(y) \, dy$

$$= \int_{\mathbb{R}^d} \Big[\underbrace{\overline{\int_{\mathbb{R}^d} \overline{k(x,y)}\, g(x)\, dx}}_{K^* g} \Big] f(y)\, dy$$

$$= \langle f, K^* g \rangle$$

<u>Note:</u> This may not be true when $\Omega_1 \neq \Omega_2 \neq \mathbb{R}^d$!

$$\int_{\Omega_2} \Big[\underbrace{\int_{\Omega_1} k(x,y) f(y)\, dy}_{Kf} \Big] \overline{g(x)}\, dx = \int_{\mathbb{R}^d} \Big[\int_{\mathbb{R}^d} \chi_{\Omega_1}(y) k(x,y) f(y)\, dy \Big] \chi_{\Omega_2}(x) \overline{g(x)}\, dx$$

$$= \int_{\mathbb{R}^d} \Big[\int_{\mathbb{R}^d} \chi_{\Omega_2}(x) k(x,y) \overline{g(x)}\, dx \Big] \chi_{\Omega_1}(y) f(y)\, dy$$

$$= \int_{\Omega_1} \Big[\underbrace{\overline{\int_{\Omega_2} \overline{k(x,y)}\, g(x)\, dx}}_{K^* g} \Big] f(y)\, dy$$

So domain/int. limits of K^* are <u>different</u> than those of K!

The Singular Value Decomposition (SVD)

Just like a matrix, every Hilbert-Schmidt operator is diagonal if we choose the right orthonormal bases for its input and output spaces. This decomposition allows us to "slice up" an inf-dim operator into countably many decoupled 1D operators!

Let's start with an $m \times n$ matrix A ($m \geq n$):

$$\begin{array}{cc} & n\text{ columns} \\ \begin{array}{c} m \\ \text{rows} \end{array} & \begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix} \end{array} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} — & v_1^* & — \\ & \vdots & \\ — & v_r^* & — \end{bmatrix}$$

$$A \quad = \quad U \quad \times \quad \Sigma \quad \times \quad V^*$$

Here, r = rank of A ($\leq n$) and

U = ONB for column space of A ($C(A)$)

V = ONB for row space of A ($C(A^T)$)

Σ = "size" of A coupling inputs to outputs along directions $v_j \to u_j$

Given input $x = \alpha v_j$, we get

$$Ax = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} — & v_1^* & — \\ & \vdots & \\ — & v_r^* & — \end{bmatrix} \begin{bmatrix} | \\ x \\ | \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ \alpha \\ \vdots \\ 0 \end{bmatrix} \leftarrow j^{\text{th}} \text{ entry}$$

$$= \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ \alpha \sigma_j \\ \vdots \\ 0 \end{bmatrix} \leftarrow j^{\text{th}} \text{ entry} = \alpha \sigma_j u_j$$

In this way we can unpack $A = USV^*$ as a sum of rank-1 terms

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^* \implies$$

best rank k approx to A

<u>Example:</u> Suppose that $x_1, x_2, x_3, \ldots \in \mathbb{R}^n$ is a sequence of random vectors whose entries are drawn from a multivariate normal distribution with covariance matrix

$$C_{ij} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right],$$

and mean $\mu = \left[\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n)\right]$.

The entries of the matrix capture how the different entries of the vectors $X_1, \ldots, X_n, \ldots$ are correlated with each other.

The SVD of C transforms the random variables $X_1, X_2, X_3, \ldots$ into new random variables with uncorrelated entries and the singular vals reveal "directions" of maximal variance

$$C = U \Sigma U^* \qquad (\text{SPSD})$$

If $y_j = U^* x_j$, then y_j are i.i.d. with diagonal cov matrix Σ.

SVD for H-S ops

For Hilbert-Schmidt ops, we can also find orthonormal bases $\{u_j\}_{j=1}^{\infty} \in L^2(\mathbb{R}^d)$ and $\{v_j\}_{j=1}^{\infty}$ with singular values $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq 0$,

$$Kf = \sum_{j \geq 1}^{\infty} \sigma_j \langle f, v_j \rangle u_j \quad \text{for all } f \in L^2$$

Analogous to the matrix case, the singular values describe the stretching/shrinking of $f \to Kf$.

Therefore, singular vals are related to norms!

Operator norm $$\|K\|_{L^2 \to L^2} = \sigma_1$$

Hilbert-Schmidt norm $$\|K\|_{HS} = \underbrace{\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |k(x,y)|^2 dx dy\right]^{1/2}}_{L^2 \text{ norm of } k(x,y)} = \left[\sum_{k=1}^{\infty} \sigma_k^2\right]^{1/2} < \infty$$

The H-S norm is a continuous analoge of the Frobenius norm for matrices!

$$\|A\|_F = \sqrt{\sum_{k=1}^{m}\sum_{j=1}^{n} |a_{jk}|^2} = \sqrt{\sum_{k=1}^{r} \sigma_k^2}$$

The L^2 kernel and connection to $\{\sigma_k\}_{k=1}$ allow us to form finite-dim. approximations to K that converge in the operator norm.

"finite rank" approx $\Rightarrow$ $$K_n f = \sum_{j=1}^{n} \sigma_j \langle f, v_j \rangle u_j, \text{ then}$$

$$K - K_n = \sum_{j=n+1}^{\infty} \sigma_j \langle f, v_j \rangle u_j, \text{ and}$$

$$\|K-K_n\| = \sigma_{n+1} \to 0 \text{ as } n \to \infty$$

This property, norm-approx by finite rank operators, makes K compact.

Note that K_n above is the best possible rank-n approx of K, in sense that

$$\sigma_{n+1} = \|K-K_n\| \leq \|K-\tilde{K}_n\|$$

for any rank-n $\tilde{K}_n : L^2 \to L^2$.

This is an analogue of Eckart-Young theorem for H-S operators.