Polynomial Spectral Decomposition of Conditional Expectation Operators

Anuran Makur and Lizhong Zheng

EECS Department, Massachusetts Institute of Technology

Allerton Conference 2016
Outline

1. Introduction
   - Motivation: Regression and Maximal Correlation
   - Preliminaries
   - Spectral Characterization of Maximal Correlation

2. Polynomial Decompositions of Compact Operators

3. Illustrations of Polynomial SVDs
Motivation: Regression and Maximal Correlation

Fix a joint distribution $P_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$.

Regression: [Breiman and Friedman, 1985]

$$\inf_{f \in F, g \in G} \mathbb{E} \left[ (f(X) - g(Y))^2 \right]$$

where we minimize over:

$F \equiv \{ f : \mathcal{X} \rightarrow \mathbb{R} | \mathbb{E}[f(X)] = 0, \mathbb{E}[f^2(X)] = 1 \}$

$G \equiv \{ g : \mathcal{Y} \rightarrow \mathbb{R} | \mathbb{E}[g(Y)] = 0, \mathbb{E}[g^2(Y)] = 1 \}$

Maximal Correlation: [Rényi, 1959]

$$\rho(X; Y) \equiv \sup_{f \in F, g \in G} \mathbb{E}[f(X)g(Y)]$$

Equivalence:

$$\mathbb{E}[(f(X) - g(Y))^2] = 2 - 2 \mathbb{E}[f(X)g(Y)]$$

Maximal correlation is a singular value of an operator!
Motivation: Regression and Maximal Correlation

Fix a joint distribution \( P_{X,Y} \) on \( \mathcal{X} \times \mathcal{Y} \).

**Regression:** [Breiman and Friedman, 1985]
Find \( f^* \in \mathcal{F} \) and \( g^* \in \mathcal{G} \) that minimize the mean squared error:

\[
\inf_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E} \left[ (f(X) - g(Y))^2 \right]
\]

where we minimize over:

\[
\mathcal{F} \triangleq \{ f : \mathcal{X} \rightarrow \mathbb{R} | \mathbb{E}[f(X)] = 0, \mathbb{E}[f^2(X)] = 1 \}
\]

\[
\mathcal{G} \triangleq \{ g : \mathcal{Y} \rightarrow \mathbb{R} | \mathbb{E}[g(Y)] = 0, \mathbb{E}[g^2(Y)] = 1 \}
\]

Maximal Correlation:
[Rényi, 1959]
Find \( f^* \in \mathcal{F} \) and \( g^* \in \mathcal{G} \) that maximize the correlation:

\[
\rho(X; Y) \equiv \sup_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E}[f(X)g(Y)]
\]

Equivalence:
\[
\mathbb{E}[ (f(X) - g(Y))^2 ] = 2 - 2 \mathbb{E}[f(X)g(Y)]
\]

Maximal correlation is a singular value of an operator!
Motivation: Regression and Maximal Correlation

Fix a joint distribution $P_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$.

**Regression:** [Breiman and Friedman, 1985]
Find $f^* \in \mathcal{F}$ and $g^* \in \mathcal{G}$ that minimize the mean squared error:

$$\inf_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E} \left[ (f(X) - g(Y))^2 \right]$$

where we minimize over:

$$\mathcal{F} \triangleq \{ f : \mathcal{X} \to \mathbb{R} | \mathbb{E}[f(X)] = 0, \mathbb{E}[f^2(X)] = 1 \}$$
$$\mathcal{G} \triangleq \{ g : \mathcal{Y} \to \mathbb{R} | \mathbb{E}[g(Y)] = 0, \mathbb{E}[g^2(Y)] = 1 \}$$

**Maximal Correlation:** [Rényi, 1959]
Find $f^* \in \mathcal{F}$ and $g^* \in \mathcal{G}$ that maximize the correlation:

$$\rho(X; Y) \triangleq \sup_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E}[f(X)g(Y)]$$

Equivalence:
$$\mathbb{E}[(f(X) - g(Y))^2] = 2 - 2\mathbb{E}[f(X)g(Y)]$$
Motivation: Regression and Maximal Correlation

Fix a joint distribution $P_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$.

**Regression:** [Breiman and Friedman, 1985]
Find $f^* \in \mathcal{F}$ and $g^* \in \mathcal{G}$ that minimize the mean squared error:

$$\inf_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E} \left[ (f(X) - g(Y))^2 \right]$$

where we minimize over:

- $\mathcal{F} \triangleq \{ f : \mathcal{X} \rightarrow \mathbb{R} | \mathbb{E}[f(X)] = 0, \mathbb{E}[f^2(X)] = 1 \}$
- $\mathcal{G} \triangleq \{ g : \mathcal{Y} \rightarrow \mathbb{R} | \mathbb{E}[g(Y)] = 0, \mathbb{E}[g^2(Y)] = 1 \}$

**Maximal Correlation:** [Rényi, 1959]
Find $f^* \in \mathcal{F}$ and $g^* \in \mathcal{G}$ that maximize the correlation:

$$\rho(X; Y) \triangleq \sup_{f \in \mathcal{F}, g \in \mathcal{G}} \mathbb{E}[f(X)g(Y)]$$

Equivalence: $\mathbb{E}[\left( f(X) - g(Y) \right)^2] = 2 - 2\mathbb{E}[f(X)g(Y)]$

Maximal correlation is a singular value of an operator!
Source random variable $X \in \mathcal{X} \subseteq \mathbb{R}$ with probability density $P_X$ on the measure space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \lambda)$
Preliminaries

- **Source** random variable $X \in \mathcal{X} \subseteq \mathbb{R}$
  with probability density $P_X$
  on the measure space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \lambda)$

- **Output** random variable $Y \in \mathcal{Y} \subseteq \mathbb{R}$
Preliminaries

- **Source** random variable $X \in \mathcal{X} \subseteq \mathbb{R}$
  with probability density $P_X$
  on the measure space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \lambda)$

- **Output** random variable $Y \in \mathcal{Y} \subseteq \mathbb{R}$

- **Channel** conditional probability densities $\{P_{Y|X=x} : x \in \mathcal{X}\}$
  on the measure space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \mu)$. 
**Preliminaries**

- **Source** random variable \( X \in \mathcal{X} \subseteq \mathbb{R} \)
  with probability density \( P_X \)
  on the measure space \((\mathcal{X}, \mathcal{B}(\mathcal{X}), \lambda)\)

- **Output** random variable \( Y \in \mathcal{Y} \subseteq \mathbb{R} \)

- **Channel** conditional probability densities \( \{ P_{Y|X=x} : x \in \mathcal{X} \} \)
  on the measure space \((\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \mu)\).

- **Marginal probability laws**: \( P_X \) and \( P_Y \)
Preliminaries

- **Hilbert spaces:**

\[ L^2(\mathcal{X}, \mathbb{P}_X) \triangleq \{ f : \mathcal{X} \to \mathbb{R} | \mathbb{E}[f^2(X)] < +\infty \} \]

\[ L^2(\mathcal{Y}, \mathbb{P}_Y) \triangleq \{ g : \mathcal{Y} \to \mathbb{R} | \mathbb{E}[g^2(Y)] < +\infty \} \]
## Preliminaries

- **Hilbert spaces:**

  \[
  \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) \triangleq \{ f : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}[f^2(X)] < +\infty \}
  \]

  \[
  \mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y) \triangleq \{ g : \mathcal{Y} \to \mathbb{R} \mid \mathbb{E}[g^2(Y)] < +\infty \}
  \]

- **Correlation as inner products**

  \[
  \langle f_1, f_2 \rangle_{\mathbb{P}_X} \triangleq \mathbb{E}[f_1(X)f_2(X)]
  \]

  \[
  \langle g_1, g_2 \rangle_{\mathbb{P}_Y} \triangleq \mathbb{E}[g_1(Y)g_2(Y)]
  \]
Preliminaries

- **Hilbert spaces:**

\[
\mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) \triangleq \{ f : \mathcal{X} \to \mathbb{R} \mid \mathbb{E}[f^2(X)] < +\infty \}
\]

\[
\mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y) \triangleq \{ g : \mathcal{Y} \to \mathbb{R} \mid \mathbb{E}[g^2(Y)] < +\infty \}
\]

- **Conditional Expectation Operators:**

\[
C : \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) \to \mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y) : (C(f))(y) \triangleq \mathbb{E}[f(X) \mid Y = y]
\]

\[
C^* : \mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y) \to \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) : (C^*(g))(x) \triangleq \mathbb{E}[g(Y) \mid X = x]
\]
Preliminaries

Proposition (Conditional Expectation Operators)

$C$ and $C^*$ are bounded linear operators with operator norms $\|C\|_{op} = \|C^*\|_{op} = 1$. Moreover, $C^*$ is the adjoint operator of $C$. 

Operator Norm: $\|C\|_{op} \triangleq \sup_{f \in L^2(X, \mathbb{P}_X)} \|C(f)\|_{L^2(Y, \mathbb{P}_Y)} \frac{\|f\|_{L^2(X, \mathbb{P}_X)}}{\|f\|_{L^2(Y, \mathbb{P}_Y)}} \leq 1$ by Jensen's inequality:

$\|C(f)\|_{L^2(Y, \mathbb{P}_Y)} = \mathbb{E}\left[\mathbb{E}[f(X) | Y]\right] \leq \mathbb{E}\left[\mathbb{E}[f^2(X) | Y]\right] = \|f\|_{L^2(X, \mathbb{P}_X)}$. 

A. Makur & L. Zheng (MIT)

Polynomial Spectral Decomposition  Allerton Conference 2016 7 / 25
Proposition (Conditional Expectation Operators)

\( C \) and \( C^* \) are \textit{bounded linear operators} with operator norms \( \| C \|_{op} = \| C^* \|_{op} = 1 \). Moreover, \( C^* \) is the adjoint operator of \( C \).

\textbf{Operator Norm:} \( \| C \|_{op} \triangleq \sup_{f \in \mathcal{L}^2(\mathcal{X}, \mathcal{P}_X)} \frac{\| C(f) \|_{\mathcal{P}_Y}}{\| f \|_{\mathcal{P}_X}} \)
Proposition (Conditional Expectation Operators)

C and C* are bounded linear operators with operator norms \( \| C \|_{op} = \| C^* \|_{op} = 1 \). Moreover, C* is the adjoint operator of C.

- **Operator Norm:** \( \| C \|_{op} \triangleq \sup_{f \in \mathcal{L}^2(\mathcal{X}, \mathcal{P}_X)} \frac{\| C(f) \|_{\mathcal{P}_Y}}{\| f \|_{\mathcal{P}_X}} \)

- \( \| C \|_{op} \leq 1 \) by Jensen’s inequality:

\[
\| C(f) \|_{\mathcal{P}_Y}^2 = \mathbb{E} \left[ \mathbb{E} [f(X) | Y]^2 \right] \leq \mathbb{E} \left[ \mathbb{E} [f^2(X) | Y] \right] = \| f \|_{\mathcal{P}_X}^2.
\]
Proposition (Conditional Expectation Operators)

$C$ and $C^*$ are bounded linear operators with operator norms $\|C\|_{op} = \|C^*\|_{op} = 1$. Moreover, $C^*$ is the adjoint operator of $C$.

- **Operator Norm:**
  \[
  \|C\|_{op} \triangleq \sup_{f \in \mathcal{L}^2(\mathcal{X}, \mathcal{P}_X)} \frac{\|C(f)\|_{\mathcal{P}_Y}}{\|f\|_{\mathcal{P}_X}}
  \]

- $\|C\|_{op} \leq 1$ by Jensen’s inequality.

- Let $1_S : S \to \mathbb{R}$ denote the everywhere unity function: $1_S(x) = 1$. $C(1_X) = 1_Y$ and $\|1_X\|_{\mathcal{P}_X}^2 = \|1_Y\|_{\mathcal{P}_Y}^2 = 1 \Rightarrow \|C\|_{op} = 1$.
Prop (Spectral Characterization of Maximal Correlation) [Rényi, 1959]

For random variables $X$ and $Y$ as defined earlier:

$$
\rho(X; Y) = \sup_{f \in L^2(X, \mathbb{P}_X)} \frac{\| C(f) \|_{\mathbb{P}_Y}}{\| f \|_{\mathbb{P}_X}}
$$

where the supremum is achieved by some $f^* \in L^2(X, \mathbb{P}_X)$ if $C$ is compact.

$C$ has largest singular value $\| C \|_{\text{op}} = 1$:

$$
C(1_X) = 1_Y, \quad C^*(1_Y) = 1_X.
$$

$\rho(X; Y)$ is the second largest singular value of $C$ with singular vectors $f^* \perp 1_X$ and $g^* = C(f^*) / \rho(X; Y) \perp 1_Y$ that maximize correlation.
Prop (Spectral Characterization of Maximal Correlation) [Rényi, 1959]

For random variables $X$ and $Y$ as defined earlier:

$$
\rho(X; Y) = \sup_{f \in L^2(X, \mathbb{P}_X): \mathbb{E}[f(X)] = 0} \frac{\|C(f)\|_{\mathbb{P}_Y}}{\|f\|_{\mathbb{P}_X}}
$$

where the supremum is achieved by some $f^* \in L^2(X, \mathbb{P}_X)$ if $C$ is compact.

- $C$ has largest singular value $\|C\|_{op} = 1$: $C(1_X) = 1_Y$, $C^*(1_Y) = 1_X$.
Prop (Spectral Characterization of Maximal Correlation) [Rényi, 1959]

For random variables $X$ and $Y$ as defined earlier:

$$
\rho(X; Y) = \sup_{f \in L^2(X, \mathbb{P}_X): \mathbb{E}[f(X)] = 0} \frac{\|C(f)\|_{\mathbb{P}_Y}}{\|f\|_{\mathbb{P}_X}}
$$

where the supremum is achieved by some $f^* \in L^2(X, \mathbb{P}_X)$ if $C$ is compact.

- $C$ has largest singular value $\|C\|_{op} = 1$: $C(1_X) = 1_Y$, $C^*(1_Y) = 1_X$.
- $\rho(X; Y) =$ second largest singular value of $C$ with singular vectors $f^* \perp 1_X$ and $g^* = C(f^*)/\rho(X; Y) \perp 1_Y$ that maximize correlation.
Outline

1 Introduction

2 Polynomial Decompositions of Compact Operators
   - The Hermite SVD
   - Assumptions and Definitions
   - Polynomial EVD of Compact Self-Adjoint Operators
   - Polynomial SVD of Conditional Expectation Operators

3 Illustrations of Polynomial SVDs
The Hermite SVD

Gaussian Channel: $P_{Y|X=x} = \mathcal{N}(x, \nu)$ with expectation parameter $x \in \mathbb{R}$ and fixed variance $\nu \in (0, \infty)$

$$\forall x, y \in \mathbb{R}, \quad P_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\nu}} \exp \left( -\frac{(y - x)^2}{2\nu} \right)$$

Gaussian Source: $P_X = \mathcal{N}(0, \nu)$ with fixed variance $\nu \in (0, \infty)$

$$\forall x \in \mathbb{R}, \quad P_X(x) = \frac{1}{\sqrt{2\pi\nu}} \exp \left( -\frac{x^2}{2\nu} \right)$$
The Hermite SVD

**Gaussian Channel:** \( P_{Y|X=x} = \mathcal{N}(x, \nu) \) with expectation parameter \( x \in \mathbb{R} \) and fixed variance \( \nu \in (0, \infty) \)

\[
\forall x, y \in \mathbb{R}, \quad P_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{(y-x)^2}{2\nu}\right)
\]

**Gaussian Source:** \( P_X = \mathcal{N}(0, p) \) with fixed variance \( p \in (0, \infty) \)

\[
\forall x \in \mathbb{R}, \quad P_X(x) = \frac{1}{\sqrt{2\pi p}} \exp\left(-\frac{x^2}{2p}\right)
\]

**Remark:** (AWGN channel) \( Y = X + W \) with \( X \perp\!\!\!\!\perp W \sim \mathcal{N}(0, \nu) \)

**Gaussian Output Marginal:** \( P_Y = \mathcal{N}(0, p + \nu) \)

\[
\forall y \in \mathbb{R}, \quad P_Y(y) = \frac{1}{\sqrt{2\pi (p + \nu)}} \exp\left(-\frac{y^2}{2(p + \nu)}\right)
\]
The Hermite SVD

Prop (Hermite SVD) [Abbe & Zheng, 2012], [Makur & Zheng, 2016]

For the Gaussian channel $P_{Y|X}$ and Gaussian source $P_X$, the conditional expectation operator $C : L^2(\mathbb{R}, P_X) \rightarrow L^2(\mathbb{R}, P_Y)$ has SVD:

$$\forall k \in \mathbb{N}, \quad C \left( H^{(p)}_k \right) = \sigma_k H^{(p+\nu)}_k$$

with singular values: $\{\sigma_k \in (0, 1] : k \in \mathbb{N}\}$ where $\sigma_0 = 1$ and $\lim_{k \to \infty} \sigma_k = 0$,

and singular vectors:

- $\{H^{(p)}_k\}$ with degree $k$ : $k \in \mathbb{N}$ - Hermite polynomials that are orthonormal with respect to $P_X$,

- $\{H^{(p+\nu)}_k\}$ with degree $k$ : $k \in \mathbb{N}$ - Hermite polynomials that are orthonormal with respect to $P_Y$. 

For which joint distributions $P_X, P_Y$ are the singular vectors of $C$ orthonormal polynomials?
The Hermite SVD

Prop (Hermite SVD) [Abbe & Zheng, 2012], [Makur & Zheng, 2016]

For the Gaussian channel $P_{Y|X}$ and Gaussian source $P_X$, the conditional expectation operator $C : \mathcal{L}^2(\mathbb{R}, P_X) \to \mathcal{L}^2(\mathbb{R}, P_Y)$ has SVD:

$$\forall k \in \mathbb{N}, C \left( H^{(p)}_k \right) = \sigma_k H^{(p+\nu)}_k$$

with singular values: $\{ \sigma_k \in (0, 1] : k \in \mathbb{N} \}$ where $\sigma_0 = 1$ and $\lim_{k \to \infty} \sigma_k = 0$,

and singular vectors:

- $\{ H^{(p)}_k \}$ with degree $k : k \in \mathbb{N}$ - Hermite polynomials that are orthonormal with respect to $P_X$,

- $\{ H^{(p+\nu)}_k \}$ with degree $k : k \in \mathbb{N}$ - Hermite polynomials that are orthonormal with respect to $P_Y$.

For which joint distributions $P_{X,Y}$ are the singular vectors of $C$ orthonormal polynomials?
Assumptions and Definitions

- $L^2(\mathcal{X}, \mathbb{P}_X)$ and $L^2(\mathcal{Y}, \mathbb{P}_Y)$ are infinite-dimensional.
Assumptions and Definitions

- \( L^2(\mathcal{X}, \mathbb{P}_X) \) and \( L^2(\mathcal{Y}, \mathbb{P}_Y) \) are infinite-dimensional.

- \( L^2(\mathcal{X}, \mathbb{P}_X) \) admits a unique countable orthonormal basis of polynomials, \( \{p_k : k \in \mathbb{N}\} \subseteq L^2(\mathcal{X}, \mathbb{P}_X) \), where \( p_k : \mathcal{X} \to \mathbb{R} \) is an orthonormal polynomial with degree \( k \).
Assumptions and Definitions

- $\mathcal{L}^2(\mathcal{X}, \mathbb{P}_X)$ and $\mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y)$ are infinite-dimensional.

- $\mathcal{L}^2(\mathcal{X}, \mathbb{P}_X)$ admits a unique countable orthonormal basis of polynomials, $\{p_k : k \in \mathbb{N}\} \subseteq \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X)$, where $p_k : \mathcal{X} \to \mathbb{R}$ is an orthonormal polynomial with degree $k$.

- $\mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y)$ admits a unique countable orthonormal basis of polynomials, $\{q_k : k \in \mathbb{N}\} \subseteq \mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y)$, where $q_k : \mathcal{Y} \to \mathbb{R}$ is an orthonormal polynomial with degree $k$. 
Assumptions and Definitions

Definition (Closure over Polynomials and Degree Preservation)

An operator $T : L^2(\mathcal{X}, \mathbb{P}_X) \rightarrow L^2(\mathcal{Y}, \mathbb{P}_Y)$ is closed over polynomials if for any polynomial $p \in L^2(\mathcal{X}, \mathbb{P}_X)$, $T(p)$ is also a polynomial. Furthermore, $T$ is degree preserving if:

$$\deg (T(p)) \leq \deg (p),$$

and $T$ is strictly degree preserving if:

$$\deg (T(p)) = \deg (p).$$
Assumptions and Definitions

**Definition (Closure over Polynomials and Degree Preservation)**

An operator $T : \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) \rightarrow \mathcal{L}^2(\mathcal{Y}, \mathbb{P}_Y)$ is closed over polynomials if for any polynomial $p \in \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X)$, $T(p)$ is also a polynomial. Furthermore, $T$ is **degree preserving** if:

$$\deg (T(p)) \leq \deg (p),$$

and $T$ is **strictly degree preserving** if:

$$\deg (T(p)) = \deg (p).$$

**Gaussian Channel Example:** $Y = X + W$ with $X \perp \perp W \sim \mathcal{N}(0, \nu)$

$$\mathbb{E} [g(Y)|X = x] = \frac{1}{\sqrt{2\pi\nu}} \int_{\mathbb{R}} g(y) \exp \left( -\frac{(y - x)^2}{2\nu} \right) \, d\mu(y)$$

Convolution preserves polynomials!
Definition (Closure over Polynomials and Degree Preservation)

An operator $T : \mathcal{L}^2 (\mathcal{X}, \mathbb{P}_X) \to \mathcal{L}^2 (\mathcal{Y}, \mathbb{P}_Y)$ is closed over polynomials if for any polynomial $p \in \mathcal{L}^2 (\mathcal{X}, \mathbb{P}_X)$, $T(p)$ is also a polynomial. Furthermore, $T$ is degree preserving if:

$$\text{deg} (T(p)) \leq \text{deg} (p),$$

and $T$ is strictly degree preserving if:

$$\text{deg} (T(p)) = \text{deg} (p).$$

Gaussian Channel Example: $Y = X + W$ with $X \perp W \sim \mathcal{N}(0, \nu)$

$$\mathbb{E} [g(Y) | X = x] = \frac{1}{\sqrt{2\pi\nu}} \int_{\mathbb{R}} g(y) \exp \left( -\frac{(y - x)^2}{2\nu} \right) d\mu(y)$$

Convolution preserves polynomials!
Theorem (Condition for Orthonormal Polynomial Eigenbasis) [Makur and Zheng, 2016]

Let $T : \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X) \rightarrow \mathcal{L}^2(\mathcal{X}, \mathbb{P}_X)$ be a compact self-adjoint operator. $T$ is closed over polynomials and degree preserving if and only if:

$$\forall k \in \mathbb{N}, \quad T(p_k) = \alpha_k p_k$$

where $\{\alpha_k \in \mathbb{R} : k \in \mathbb{N}\}$ are eigenvalues satisfying $\lim_{k \to \infty} \alpha_k = 0$. 
Suppose $C : \mathcal{L}^2 (\mathcal{X}, \mathbb{P}_X) \to \mathcal{L}^2 (\mathcal{Y}, \mathbb{P}_Y)$ is compact and $C^* : \mathcal{L}^2 (\mathcal{Y}, \mathbb{P}_Y) \to \mathcal{L}^2 (\mathcal{X}, \mathbb{P}_X)$ is its adjoint operator. $C$ and $C^*$ are closed over polynomials and strictly degree preserving if and only if:

$$\forall k \in \mathbb{N}, \quad C(p_k) = \beta_k q_k$$

where $\{\beta_k \in (0, \infty) : k \in \mathbb{N}\}$ are the singular values such that $\lim_{k \to \infty} \beta_k = 0$. 

[Makur and Zheng, 2016]
Theorem (Condition for Orthonormal Polynomial Singular Vectors)
[Makur and Zheng, 2016]

Suppose $C \triangleq \mathbb{E} [\cdot | Y] : L^2 (\mathcal{X}, \mathbb{P}_X) \to L^2 (\mathcal{Y}, \mathbb{P}_Y)$ is compact and $C^* = \mathbb{E} [\cdot | X] : L^2 (\mathcal{Y}, \mathbb{P}_Y) \to L^2 (\mathcal{X}, \mathbb{P}_X)$ is its adjoint operator.

For every $n \in \mathbb{N}$, $\mathbb{E} [X^n | Y]$ is a polynomial in $Y$ with degree $n$ and $\mathbb{E} [Y^n | X]$ is polynomial in $X$ with degree $n$ if and only if:

$$\forall k \in \mathbb{N}, \quad C (p_k) = \beta_k q_k$$

where $\{\beta_k \in (0, 1) : k \in \mathbb{N}\}$ are the singular values such that $\beta_0 = 1$ and $\lim_{k \to \infty} \beta_k = 0$. 

Gaussian Example Proof Sketch:
$Y = X + W$ with $X \sim \mathcal{N}(0, \mathbb{P}) \perp \perp W \sim \mathcal{N}(0, \nu)$.

$C$, $C^*$ are defined by convolution kernels which preserve polynomials. By above theorem, $C$ has Hermite polynomial singular vectors.
Theorem (Condition for Orthonormal Polynomial Singular Vectors) [Makur and Zheng, 2016]

Suppose $C \triangleq \mathbb{E} \left[ \cdot | Y \right] : L^2 (\mathcal{X}, \mathbb{P}_X) \to L^2 (\mathcal{Y}, \mathbb{P}_Y)$ is compact and $C^* = \mathbb{E} \left[ \cdot | X \right] : L^2 (\mathcal{Y}, \mathbb{P}_Y) \to L^2 (\mathcal{X}, \mathbb{P}_X)$ is its adjoint operator.
For every $n \in \mathbb{N}$, $\mathbb{E} \left[ X^n \mid Y \right]$ is a polynomial in $Y$ with degree $n$ and $\mathbb{E} \left[ Y^n \mid X \right]$ is polynomial in $X$ with degree $n$ if and only if:

$$\forall k \in \mathbb{N}, \quad C \left( p_k \right) = \beta_k q_k$$

where $\{\beta_k \in (0, 1] : k \in \mathbb{N}\}$ are the singular values such that $\beta_0 = 1$ and $\lim_{k \to \infty} \beta_k = 0$.

Gaussian Example Proof Sketch:

- $Y = X + W$ with $X \sim \mathcal{N} (0, P) \perp \perp W \sim \mathcal{N} (0, \nu)$. 

$C$, $C^*$ are defined by convolution kernels which preserve polynomials. By above theorem, $C$ has Hermite polynomial singular vectors.
Theorem (Condition for Orthonormal Polynomial Singular Vectors) [Makur and Zheng, 2016]

Suppose \( C \triangleq \mathbb{E} [\cdot | Y] : L^2 (\mathcal{X}, \mathbb{P}_X) \to L^2 (\mathcal{Y}, \mathbb{P}_Y) \) is compact and \( C^* = \mathbb{E} [\cdot | X] : L^2 (\mathcal{Y}, \mathbb{P}_Y) \to L^2 (\mathcal{X}, \mathbb{P}_X) \) is its adjoint operator. For every \( n \in \mathbb{N} \), \( \mathbb{E} [X^n | Y] \) is a polynomial in \( Y \) with degree \( n \) and \( \mathbb{E} [Y^n | X] \) is polynomial in \( X \) with degree \( n \) if and only if:

\[
\forall k \in \mathbb{N}, \quad C (p_k) = \beta_k q_k
\]

where \( \{ \beta_k \in (0, 1] : k \in \mathbb{N} \} \) are the singular values such that \( \beta_0 = 1 \) and \( \lim_{k \to \infty} \beta_k = 0 \).

Gaussian Example Proof Sketch:
- \( Y = X + W \) with \( X \sim \mathcal{N} (0, \sigma) \perp \perp W \sim \mathcal{N} (0, \nu) \).
- \( C, C^* \) are defined by convolution kernels which preserve polynomials.
Theorem (Condition for Orthonormal Polynomial Singular Vectors) [Makur and Zheng, 2016]

Suppose $C \triangleq \mathbb{E}[\cdot|Y] : L^2(\mathcal{X}, \mathbb{P}_X) \to L^2(\mathcal{Y}, \mathbb{P}_Y)$ is compact and $C^* = \mathbb{E}[\cdot|X] : L^2(\mathcal{Y}, \mathbb{P}_Y) \to L^2(\mathcal{X}, \mathbb{P}_X)$ is its adjoint operator. For every $n \in \mathbb{N}$, $\mathbb{E}[X^n|Y]$ is a polynomial in $Y$ with degree $n$ and $\mathbb{E}[Y^n|X]$ is polynomial in $X$ with degree $n$ if and only if:

$$\forall k \in \mathbb{N}, \quad C(p_k) = \beta_k q_k$$

where $\{\beta_k \in (0, 1] : k \in \mathbb{N}\}$ are the singular values such that $\beta_0 = 1$ and $\lim_{k \to \infty} \beta_k = 0$.

Gaussian Example Proof Sketch:
- $Y = X + W$ with $X \sim \mathcal{N}(0, \sigma) \perp \perp W \sim \mathcal{N}(0, \nu)$.
- $C$, $C^*$ are defined by convolution kernels which preserve polynomials.
- By above theorem, $C$ has Hermite polynomial singular vectors.
1. Introduction

2. Polynomial Decompositions of Compact Operators

3. Illustrations of Polynomial SVDs
   - The Laguerre SVD
   - The Jacobi SVD
   - Natural Exponential Families and Conjugate Priors
The Laguerre SVD

Poisson Channel: \( P_{Y|X=x} = \text{Poisson}(x) \) with rate parameter \( x \in (0, \infty) \)

\[
\forall x \in (0, \infty), \forall y \in \mathbb{N}, \quad P_{Y|X}(y|x) = \frac{x^y e^{-x}}{y!}
\]

Gamma Source: \( P_X = \text{gamma}(\alpha, \beta) \) with shape parameter \( \alpha \in (0, \infty) \) and rate parameter \( \beta \in (0, \infty) \)

\[
\forall x \in (0, \infty), \quad P_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}
\]
The Laguerre SVD

**Poisson Channel:** \( P_{Y|X=x} = \text{Poisson}(x) \) with rate parameter \( x \in (0, \infty) \)

\[
\forall x \in (0, \infty), \forall y \in \mathbb{N}, \quad P_{Y|X}(y|x) = \frac{x^y e^{-x}}{y!}
\]

**Gamma Source:** \( P_X = \text{gamma}(\alpha, \beta) \) with shape parameter \( \alpha \in (0, \infty) \) and rate parameter \( \beta \in (0, \infty) \)

\[
\forall x \in (0, \infty), \quad P_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}
\]

**Negative Binomial Output Marginal:**
\( P_Y = \text{negative-binomial} \left( p = \frac{1}{\beta+1}, \alpha \right) \) with success probability parameter \( p \in (0, 1) \) and number of failures parameter \( \alpha \in (0, \infty) \)

\[
\forall y \in \mathbb{N}, \quad P_Y(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!} \left( \frac{1}{\beta + 1} \right)^y \left( \frac{\beta}{\beta + 1} \right)^\alpha
\]
The Laguerre SVD

**Proposition (Laguerre SVD) [Makur and Zheng, 2016]**

For the Poisson channel $P_{Y|X}$ and gamma source $P_X$, the conditional expectation operator $C : \mathcal{L}^2((0, \infty), \mathbb{P}_X) \rightarrow \mathcal{L}^2(\mathbb{N}, \mathbb{P}_Y)$ has SVD:

$$\forall k \in \mathbb{N}, \quad C \left( L_k^{(\alpha, \beta)} \right) = \sigma_k M_k^{\left(\alpha, \frac{1}{\beta+1}\right)}$$

with singular values: $\{ \sigma_k \in (0, 1] : k \in \mathbb{N} \}$ where $\sigma_0 = 1$ and $\lim_{k \to \infty} \sigma_k = 0$, and singular vectors:

- $\{ L_k^{(\alpha, \beta)} \text{ with degree } k : k \in \mathbb{N} \}$ - generalized Laguerre polynomials that are orthonormal with respect to $\mathbb{P}_X$,
- $\{ M_k^{\left(\alpha, \frac{1}{\beta+1}\right)} \text{ with degree } k : k \in \mathbb{N} \}$ - Meixner polynomials that are orthonormal with respect to $\mathbb{P}_Y$. 
The Jacobi SVD

**Binomial Channel:** \( P_{Y|X=x} = \text{binomial}(n, x) \) with number of trials parameter \( n \in \mathbb{N}\setminus\{0\} \) and success probability parameter \( x \in (0, 1) \)

\[
\forall x \in (0, 1), \forall y \in [n] \triangleq \{0, \ldots, n\}, \quad P_{Y|X}(y|x) = \binom{n}{y} x^y (1 - x)^{n-y}
\]

**Beta Source:** \( P_X = \text{beta}(\alpha, \beta) \) with shape parameters \( \alpha \in (0, \infty) \) and \( \beta \in (0, \infty) \)

\[
\forall x \in (0, 1), \quad P_X(x) = \frac{x^{\alpha-1}(1 - x)^{\beta-1}}{B(\alpha, \beta)}
\]
The Jacobi SVD

**Binomial Channel:** \( P_{Y|X=x} = \text{binomial}(n, x) \) with number of trials parameter \( n \in \mathbb{N}\backslash\{0\} \) and success probability parameter \( x \in (0, 1) \)

\[
\forall x \in (0, 1), \forall y \in [n] \triangleq \{0, \ldots, n\}, \quad P_{Y|X}(y|x) = \binom{n}{y} x^y (1 - x)^{n-y}
\]

**Beta Source:** \( P_X = \text{beta}(\alpha, \beta) \) with shape parameters \( \alpha \in (0, \infty) \) and \( \beta \in (0, \infty) \)

\[
\forall x \in (0, 1), \quad P_X(x) = \frac{x^{\alpha-1}(1 - x)^{\beta-1}}{B(\alpha, \beta)}
\]

**Beta-Binomial Output Marginal:** \( P_Y = \text{beta-binomial}(n, \alpha, \beta) \)

\[
\forall y \in [n], \quad P_Y(y) = \binom{n}{y} \frac{B(\alpha + y, \beta + n - y)}{B(\alpha, \beta)}
\]
For the binomial channel $P_{Y|X}$ and beta source $P_X$, the conditional expectation operator $C : L^2((0, 1), \mathbb{P}_X) \rightarrow L^2([n], \mathbb{P}_Y)$ has SVD:

$$\forall k \in [n], \quad C \left( J_{k}^{(\alpha, \beta)} \right) = \sigma_k Q_{k}^{(\alpha, \beta)}$$

$$\forall k \in \mathbb{N}\setminus[n], \quad C \left( J_{k}^{(\alpha, \beta)} \right) = 0$$

with singular values: $\{\sigma_k \in (0, 1] : k \in [n]\}$ where $\sigma_0 = 1$, and singular vectors:

- $\{J_{k}^{(\alpha, \beta)} \text{ with degree } k : k \in \mathbb{N}\}$ - Jacobi polynomials that are orthonormal with respect to $\mathbb{P}_X$,
- $\{Q_{k}^{(\alpha, \beta)} \text{ with degree } k : k \in [n]\}$ - Hahn polynomials that are orthonormal with respect to $\mathbb{P}_Y$. 
Why are these joint distributions special?

- $P_{Y|X}$ is a natural exponential family with quadratic variance function (introduced in [Morris, 1982]):

$$
\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \quad P_{Y|X}(y|x) = \exp(xy - \alpha(x) + \beta(y))
$$

where $P_{Y|X}(y|0) = \exp(\beta(y))$ is the base distribution, $\alpha(x)$ is the log-partition function with $\alpha(0) = 0$, and $\text{VAR}(Y|X = x)$ is a quadratic function of $\mathbb{E}[Y|X = x]$. 

A. Makur & L. Zheng (MIT)  
Polynomial Spectral Decomposition  
Allerton Conference 2016
Why are these joint distributions special?

- $P_{Y|X}$ is a natural exponential family with quadratic variance function (introduced in [Morris, 1982]):
  \[
  \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \quad P_{Y|X}(y|x) = \exp(xy - \alpha(x) + \beta(y))
  \]

- $P_X$ belongs to the corresponding conjugate prior family:
  \[
  \forall x \in \mathcal{X}, \quad P_X(x; y', n) = \exp(y'x - n\alpha(x) - \tau(y', n))
  \]

  where $\tau(y', n)$ is the log-partition function.
Why are these joint distributions special?

- $P_{Y|X}$ is a **natural exponential family** with quadratic variance function (introduced in [Morris, 1982]):

  \[
  \forall x \in X, \forall y \in Y, \quad P_{Y|X}(y|x) = \exp(xy - \alpha(x) + \beta(y))
  \]

- $P_X$ belongs to the corresponding **conjugate prior** family:

  \[
  \forall x \in X, \quad P_X(x; y', n) = \exp(y'x - n\alpha(x) - \tau(y', n))
  \]

- All moments exist and are finite:
  - Gaussian likelihood with Gaussian prior,
  - Poisson likelihood with gamma prior,
  - binomial likelihood with beta prior.
Conclusion

Summary:

1. Regression and maximal correlation
   \[\Rightarrow\] conditional expectation operators

2. Closure over polynomials and degree preservation
   \[\Leftrightarrow\] orthogonal polynomial eigenvectors or singular vectors

3. Check conditional moments are polynomials
   \[\Rightarrow\] Gaussian-Gaussian, Gamma-Poisson, Beta-Binomial examples

4. Examples have natural exponential family/conjugate prior structure
That's all Folks!