A General Theory for Nonlinear Sufficient Dimension Reduction: Formulation and Estimation

K.Y. Lee
B. Li
F. Chiaromonte

May 2, 2014
Outline

1. $\sigma$-Field as the condition for DR;
2. Sufficiency and minimal sufficiency;
3. Unbiasedness and exhaustiveness;
4. Population criterion;
Set-up

$$(\Omega, \mathcal{F}, P)$$ be a probability space,

$$\Rightarrow (\Omega_X, \mathcal{F}_X, P_X)$$

$$\Rightarrow (\Omega_Y, \mathcal{F}_Y, P_Y)$$

$$(\Omega_{XY}, \mathcal{F}_{XY}, P_{XY})$$ where $$\Omega_{XY} = \Omega_X \times \Omega_Y, \mathcal{F}_{XY} = \mathcal{F}_X \times \mathcal{F}_Y$$

$$\Rightarrow \sigma(X) = X^{-1}(\mathcal{F}_X)$$

$$\sigma(Y) = Y^{-1}(\mathcal{F}_Y)$$

$$\sigma(X, Y) = (X, Y)^{-1}(\mathcal{F}_{XY})$$
Let $G \subseteq \sigma(X)$ be a sub $\sigma$-field, if

$$Y \perp X|G,$$

then $G$ is called a **SDR $\sigma$-field** for $Y$ vs $X$.

**Remark:** $G$ can be induced by some rv, say $U$, with the measurable space $(U, \mathcal{F}_U)$, i.e.

$$G = U^{-1}(\mathcal{F}_U).$$

Then, SDR is achieved by using $U$, which is a transformation of $X$. Since the transformation is not necessarily to be linear, non-linear SDR can be achieved.
Example 1: Let

\[ \Omega_X = \mathbb{R}^P, \Omega_Y = \mathbb{R}^q \]

\[ \mathcal{F}_X, \mathcal{F}_Y, \mathcal{F}_{XY} \] are Borel \( \sigma \)-fields

If \( U = B^T X \) and \( \mathcal{G} = \sigma(U) \), then we have the usual linear DR

\[ Y \independent X | B^T X. \]
Example 2: Let $\lambda$ be the Lebesgue on $[a, b]$ and

$$\Omega_X = L^2_\lambda, \Omega_Y = \mathbb{R}.$$ 

If $\{h_1, \ldots, h_d\} \subset L^2_\lambda$ and $U = \left( \langle X, h_1 \rangle_{L^2_\lambda}, \ldots, \langle X, h_d \rangle_{L^2_\lambda} \right)$, then we have the functional DR problem considered by [Ferre and Yao 2003]

$$Y \perp X|\langle X, h_1 \rangle_{L^2_\lambda}, \ldots, \langle X, h_d \rangle_{L^2_\lambda}.$$ 

Remarks: Generalize SDR to the infinite-dimensional case, but still linear in $X$. 


Minimal Sufficiency

\( \mathcal{G} \) is not unique, for example, \( \mathcal{G} = \sigma(X) \) is valid but not reduction. We want to find the smallest \( \mathcal{G} \).

The following Theorem shows existence and uniqueness of the minimal sufficient \( \mathcal{G} \).

**Theorem 1**

Suppose that the family of probability measures \( \{ P_{X|Y}(\cdot|y) : y \in \Omega_Y \} \) is dominated by a \( \sigma \)-finite measure. Then there is a unique sub \( \sigma \)-field \( \mathcal{G}^* \) of \( \sigma(X) \) such that:

1. \( Y \perp X | \mathcal{G}^* \);
2. if \( \mathcal{G} \) is a sub \( \sigma \)-field of \( \sigma(X) \) such that \( Y \perp X | \mathcal{G} \), then \( \mathcal{G}^* \subseteq \mathcal{G} \).

\( \mathcal{G}^* \) (\( = \mathcal{G}_Y|_X \)) is called central \( \sigma \)-field.
Adding More Structures

Let $L^2_{P_X}, L^2_{P_Y},$ and $L^2_{P_{XY}}$ be the function spaces on $\Omega_X, \Omega_Y$ and $\Omega_{XY}$. They are all 0 mean functions.

$$\mathcal{M}_G = \{f \in L^2_{P_{XY}} : f \text{ is } G\text{-measurable}\}.$$  

Remark: $G$ is a linear sub-space of $L^2_{P_{XY}}$.

Definitions:

1) If $G$ is sufficient, then $\mathcal{M}_G$ is called a **SDR class**. $G^*$ is minimal sufficient (central $\sigma$-algebra), then $\mathcal{M}_{G^*}$ is called the **central class**.

2) If $G = \sigma(U)$, then we also use $\mathcal{M}_U = \mathcal{M}_G$.

Remark: $\mathcal{M}_{G^*}$ is the generalization of the usual central space $S_{Y|X}$ in linear SDR.
Unbiasedness and Exhaustiveness

If $\mathcal{M} \subset L^2_{P_X}$ is a collection of $G^*$ measurable function, then $\mathcal{M}$ is unbiased for $M_{G^*}$.

If the members of $\mathcal{M}$ generate $G^*$, then it is exhaustive.

Example:

In linear DR, if $B$ is a DR matrix and $\text{span}(B) \subset S_{Y|X}$, then it is unbiased.

If $\text{span}(B) = S_{Y|X}$, it is exhaustive.
Unbiasedness and Exhaustiveness

If $\mathcal{M} \subset L^2_{P_X}$ is a collection of $\mathcal{G}^*$ measurable function, then $\mathcal{M}$ is **unbiased** for $\mathcal{M}_{\mathcal{G}^*}$.

If the members of $\mathcal{M}$ generate $\mathcal{G}^*$, then it is **exhaustive**.

**Example:**

In linear DR, if $B$ is a DR matrix and $\text{span}(B) \subset \mathcal{S}_{Y|X}$, then it is unbiased.

If $\text{span}(B) = \mathcal{S}_{Y|X}$, it is exhaustive.
For two spaces \( S_1 \) and \( S_2 \), denote \( S_1 \ominus S_2 = S_1 \cap S_2 \).

**Theorem 2**

If the family \( \{ \Pi_y : y \in \Omega_Y \} \) is dominated by a \( \sigma \)-finite measure, then

\[
L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}] \subseteq \mathcal{M}_{G^*},
\]

i.e. unbiased for \( \mathcal{M}_{G^*} \).

**Proof:**

\[
\Leftrightarrow L^2_{P_X} \ominus \mathcal{M}_{G^*} \subseteq L^2_{P_X} \ominus L^2_{P_Y}
\]

\[
f \in L^2_{P_X} \ominus \mathcal{M}_{G^*} \Rightarrow f \perp \mathcal{M}_{G^*} \Rightarrow \mathbb{E} [f(X)|\mathcal{G}^*] = 0
\]

\[
\Rightarrow \mathbb{E} [f(X)|Y] = 0 \Rightarrow f \perp \mathcal{M}_Y \Rightarrow f \in L^2_{P_X} \ominus L^2_{P_Y}
\]

**Remarks:**

1) \( L^2_{P_X} \ominus L^2_{P_Y} \) resembles the residual in a regression.

2) \( L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}] \) is the orthogonal complement of the residual class, called regression class.
For two spaces $S_1$ and $S_2$, denote $S_1 \ominus S_2 = S_1 \cap S_2$.

**Theorem 2**

If the family $\{\Pi_y : y \in \Omega_Y\}$ is dominated by a $\sigma$-finite measure, then

$$L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}] \subseteq \mathcal{M}_{G^*},$$

i.e. unbiased for $\mathcal{M}_{G^*}$.

**Proof:** \iff $L^2_{P_X} \ominus \mathcal{M}_{G^*} \subseteq L^2_{P_X} \ominus L^2_{P_Y}$

$$f \in L^2_{P_X} \ominus \mathcal{M}_{G^*} \Rightarrow f \perp \mathcal{M}_{G^*} \Rightarrow \mathbb{E} [f(X)|\mathcal{G}^*] = 0$$

$$\Rightarrow \mathbb{E} [f(X)|Y] = 0 \Rightarrow f \perp \mathcal{M}_Y \Rightarrow f \in L^2_{P_X} \ominus L^2_{P_Y}$$

Remarks:

1) $L^2_{P_X} \ominus L^2_{P_Y}$ resembles the residual in a regression.

2) $L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}]$ is the orthogonal complement of the residual class, called regression class.
For two spaces $S_1$ and $S_2$, denote $S_1 \ominus S_2 = S_1 \cap S_2$.

**Theorem 2**

If the family $\{\Pi_y : y \in \Omega_Y\}$ is dominated by a $\sigma$-finite measure, then

$$L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}] \subseteq \mathcal{M}g^*,$$

i.e. unbiased for $\mathcal{M}g^*$.

**Proof:**

$\Leftrightarrow L^2_{P_X} \ominus \mathcal{M}g^* \subseteq L^2_{P_X} \ominus L^2_{P_Y}$

$$f \in L^2_{P_X} \ominus \mathcal{M}g^* \Rightarrow f \perp \mathcal{M}g^* \Rightarrow \mathbb{E}[f(X)|\mathcal{G}^*] = 0$$

$$\Rightarrow \mathbb{E}[f(X)|Y] = 0 \Rightarrow f \perp \mathcal{M}_Y \Rightarrow f \in L^2_{P_X} \ominus L^2_{P_Y}$$

**Remarks:**

1) $L^2_{P_X} \ominus L^2_{P_Y}$ resembles the **residual** in a regression.

2) $L^2_{P_X} \ominus [L^2_{P_X} \ominus L^2_{P_Y}]$ is the orthogonal complement of the residual class, called **regression class**.
Completeness and Exhaustiveness

**Definition 5**

Let $\mathcal{G} \subseteq \sigma(X)$ be a sub $\sigma$-field. The class $\mathcal{M}_g$ is said to be **complete** if, for any $g \in \mathcal{M}_g$,

$$
\mathbb{E} [g(X) | Y] = 0 \quad a.s.\, P \quad \Rightarrow \quad g(X) = 0 \quad a.s.\, P.
$$

**Examples:**

1) **[Forward regression]** Suppose there exists a function $h \in \left[ L^2_{P_X} \right]^q$ such that

$$
Y = h(X) + \epsilon,
$$

where $\epsilon \perp X$ and $\mathbb{E} [\epsilon] = 0$. Then $\mathcal{M}_{h(X)}$ is a complete and sufficient dimension reduction class for $Y$ versus $X$. 

---

K.Y. Lee B. Li F. Chiaromonte

May 2, 2014 11 / 27
Completeness and Exhaustiveness

**Definition 5**

Let \( \mathcal{G} \subseteq \sigma(X) \) be a sub \( \sigma \)-field. The class \( \mathcal{M}_G \) is said to be **complete** if, for any \( g \in \mathcal{M}_G \),

\[
E [g(X)|Y] = 0 \quad a.s.P \quad \Rightarrow \quad g(X) = 0 \quad a.s.P.
\]

**Examples:**

1) **[Forward regression]** Suppose there exists a function \( h \in \left[ L^2_{P_X} \right]^q \)
such that

\[
Y = h(X) + \epsilon,
\]

where \( \epsilon \perp X \) and \( E [\epsilon] = 0 \). Then \( \mathcal{M}_{h(X)} \) is a complete and sufficient
dimension reduction class for \( Y \) versus \( X \).
2) **[Inverse regression]** Suppose \( q < p \), \( \Omega_Y \) has a nonempty interior, and \( P_Y \) is dominated by the Lebesgue measure on \( \mathbb{R}^q \). Suppose there exists functions \( g \in \left[ L_{P_X}^2 \right]^q \) and \( h \in \left[ L_{P_X}^2 \right]^{p-q} \) such that:

1. \( g(X) = Y + \epsilon \), where \( Y \perp \epsilon \), and \( \epsilon \sim N(0, \Sigma) \);
2. \( \sigma(g(X), h(X)) = \sigma(X) \);
3. \( h(X) \perp (Y, g(X)) \);
4. the induced measure \( P_X \circ g^{-1} \) is dominated by the Lebesgue measure on \( \mathbb{R}^q \).

Then \( \mathcal{M}_{g(X)} \) is a complete sufficient dimension reduction class for \( Y \) versus \( X \).
When a complete and sufficient dimension reduction class exists, it is unique and coincides with the central class.

**Theorem 3**

Suppose $\{\Pi_y : y \in \Omega_Y\}$ is dominated by a $\sigma$-finite measure, and $\mathcal{G}$ is a sub $\sigma$-field of $\sigma(X)$. If $\mathcal{M}_G$ is a complete and sufficient dimension reduction class, then

$$\mathcal{M}_G = \mathcal{G}_Y|_X = \mathcal{M}_G^*,$$

i.e. it is exhaustive.
Summary of Sufficiency, Completeness, Unbiasedness and Exhaustiveness

With the fairly general assumption of the function spaces, we see that

$$L^2_{P_X} \ominus L^2_{P_Y} \quad \text{(residual class)}$$

plays a crucial role in nonlinear DR. If its **orthogonal complement** in
$$L^2_{P_X}$$ is a complete and sufficient DR class for $$Y$$ versus $$X$$, then it is the **central class**, i.e.

$$L^2_{P_X} \cap \{ L^2_{P_X} \ominus L^2_{P_Y} \}^\perp = \mathcal{M}g^*.$$

Without completeness, it is still **unbiased**, i.e.

$$L^2_{P_X} \cap \{ L^2_{P_X} \ominus L^2_{P_Y} \}^\perp \subseteq \mathcal{M}g^*.$$
Characterization of the Regression Class

Definition

For two sets $A$ and $B$, we say $A \subseteq B$ modulo constants if for each $f \in A$ there is $c \in \mathbb{R}$ such that $f + c \in B$.

A is a dense subset of $B$ modulo constants, if (i) $A \subseteq B$ modulo constants and (ii) any $f \in B$ can be approximated by a sequence $\{f_n + c_n\} \subseteq A$.

Remark: Recall the denseness assumption (AS) in [Fukumizu, Bach and Jordan 2009].

Examples: Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$ with finite variances are dense in $L^2_{P_X}$ and $L^2_{P_Y}$ modulo constants.
Characterization of the Regression Class

**Definition**

For two sets $A$ and $B$, we say $A \subseteq B$ **modulo constants** if for each $f \in A$ there is $c \in \mathbb{R}$ such that $f + c \in B$.

$A$ is a dense subset of $B$ modulo constants, if (i) $A \subseteq B$ modulo constants and (ii) any $f \in B$ can be approximated by a sequence $\{f_n + c_n\} \subseteq A$.

**Remark:** Recall the denseness assumption (AS) in [Fukumizu, Bach and Jordan 2009].

**Examples:** Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$ with finite variances are dense in $L^2_{P_X}$ and $L^2_{P_Y}$ modulo constants.
Characterization of the Regression Class

**Definition**

For two sets $A$ and $B$, we say $A \subseteq B$ **modulo constants** if for each $f \in A$ there is $c \in \mathbb{R}$ such that $f + c \in B$.

$A$ is a dense subset of $B$ modulo constants, if (i) $A \subseteq B$ modulo constants and (ii) any $f \in B$ can be approximated by a sequence $\{f_n + c_n\} \subseteq A$.

**Remark:** Recall the denseness assumption (AS) in [Fukumizu, Bach and Jordan 2009].

**Examples:** Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$ with finite variances are dense in $L^2_{P_X}$ and $L^2_{P_Y}$ modulo constants.
(Recall) Due to Riesz Representation Theorem, we can define cross-covariance operators for RKHSs.

\[ \langle f, \sum_{XX}^{{RKHS}} g \rangle_{\mathcal{H}_X} = \text{cov}(f(X), g(X))_{P_X}, \]
\[ \langle f, \sum_{YY}^{{RKHS}} g \rangle_{\mathcal{H}_Y} = \text{cov}(f(X), g(X))_{P_Y}, \]
\[ \langle f, \sum_{YX}^{{RKHS}} g \rangle_{\mathcal{H}_Y} = \text{cov}(f(X), g(X))_{P_{XY}}. \]

They are bounded and self-adjoint.
But in general, $\mathcal{H}_X$ and $\mathcal{H}_Y$ don’t have to be RKHSs, we can still have $\Sigma_{XX}$ and $\Sigma_{YY}$.

Let $\mathcal{G}_X = \text{Range}(\Sigma_{XX})$, and $\mathcal{G}_Y = \text{Range}(\Sigma_{YY})$. (so $\mathcal{G}_X$ and $\mathcal{H}_Y$ may not be $\subseteq \mathcal{H}_X, \mathcal{H}_Y$, but definitely $\subseteq L^2_{P_X}, L^2_{P_Y}$.)

Under assumptions (A) and (B), we can define (similar to Fukumizu’s RKHSs case):

$$\langle f, \Sigma_{XX}g \rangle_{\mathcal{G}_X} := \langle f, g \rangle_{L^2_{P_X}},$$

$$\langle f, \Sigma_{YY}g \rangle_{\mathcal{G}_Y} := \langle f, g \rangle_{L^2_{P_Y}},$$

$$\langle f, \Sigma_{YX}g \rangle_{\mathcal{G}_Y} := \langle f, g \rangle_{L^2_{P_Y}},$$

why not $L^2_{P_{XY}}$?

and we have

$$\Sigma_{YX} = \Sigma_{YY}^{1/2} R_{YX} \Sigma_{XX}^{1/2}$$

Reminder: our central class is a $L^2_{P_X}$ object.
But in general, $\mathcal{H}_X$ and $\mathcal{H}_Y$ don’t have to be RKHSs, we can still have $\Sigma_{XX}$ and $\Sigma_{YY}$.

Let $G_X = \overline{\text{Range}(\Sigma_{XX})}$, and $G_Y = \overline{\text{Range}(\Sigma_{YY})}$. (so $G_X$ and $\mathcal{H}_Y$ may not be $\subseteq \mathcal{H}_X, \mathcal{H}_Y$, but definitely $\subseteq L^2_{P_X}, L^2_{P_Y}$.)

Under assumptions (A) and (B), we can define (similar to Fukumizu’s RKHSs case):

$$\langle f, \Sigma_{XX} g \rangle_{G_X} := \langle f, g \rangle_{L^2_{P_X}},$$

$$\langle f, \Sigma_{YY} g \rangle_{G_Y} := \langle f, g \rangle_{L^2_{P_Y}},$$

$$\langle f, \Sigma_{YX} g \rangle_{G_Y} := \langle f, g \rangle_{L^2_{P_Y}},$$

why not $L^2_{P_{XY}}$?

and we have

$$\Sigma_{YX} = \Sigma_{YY}^{1/2} R_{YX} \Sigma_{XX}^{1/2}$$

Reminder: our central class is a $L^2_{P_X}$ object.
But in general, \( \mathcal{H}_X \) and \( \mathcal{H}_Y \) don’t have to be RKHSs, we can still have \( \Sigma_{XX} \) and \( \Sigma_{YY} \).

Let \( \mathcal{G}_X = \text{Range}(\Sigma_{XX}) \), and \( \mathcal{G}_Y = \text{Range}(\Sigma_{YY}) \). (so \( \mathcal{G}_X \) and \( \mathcal{H}_Y \) may not be \( \subseteq \mathcal{H}_X, \mathcal{H}_Y \), but definitely \( \subseteq L^2_{PX}, L^2_{PY} \).)

Under assumptions (A) and (B), we can define (similar to Fukumizu’s RKHSs case):

\[
\langle f, \Sigma_{XX} g \rangle_{\mathcal{G}_X} := \langle f, g \rangle_{L^2_{PX}},
\]
\[
\langle f, \Sigma_{YY} g \rangle_{\mathcal{G}_Y} := \langle f, g \rangle_{L^2_{PY}},
\]
\[
\langle f, \Sigma_{YX} g \rangle_{\mathcal{G}_Y} := \langle f, g \rangle_{L^2_{PY}},
\]

why not \( L^2_{PXY} \)?

and we have

\[
\Sigma_{YX} = \Sigma_{YY}^{1/2} R_{YX} \Sigma_{XX}^{1/2}
\]

**Reminder:** our central class is a \( L^2_{PX} \) object.
Assumptions:

(A) $\mathcal{H}_X$ and $\mathcal{H}_Y$ are dense in $L^2_{P_X}$ and $L^2_{P_Y}$ modulo constants;  

(B) There are constants $C_1 > 0$ and $C_2 > 0$ such that $\text{Var}(f(X)) \leq C_1 \| f \|_{\mathcal{H}_X}$ and $\text{Var}(g(Y)) \leq C_1 \| g \|_{\mathcal{H}_Y}$.

Theorem 4: Extended Covariance Operators

Under assumptions (A) and (B), there exist unique isomorphisms

$$\tilde{\Sigma}^{1/2}_{XX} : L^2_{P_X} \to \mathcal{G}_X, \quad \tilde{\Sigma}^{1/2}_{YY} : L^2_{P_Y} \to \mathcal{G}_Y$$

that agree with $\Sigma^{1/2}_{XX}$ and $\Sigma^{1/2}_{YY}$ on $\mathcal{G}_X$ and $\mathcal{G}_Y$ in the sense that for all $f \in \mathcal{G}_X$ and $g \in \mathcal{G}_Y$,

$$\tilde{\Sigma}^{1/2}_{XX}(f - \mathbb{E}[f]) = \Sigma^{1/2}_{XX} f, \quad \tilde{\Sigma}^{1/2}_{YY}(f - \mathbb{E}[g]) = \Sigma^{1/2}_{YY} g.$$

Furthermore, for any $f \in L^2_{P_X}$, $g \in L^2_{P_Y}$ we have

$$\langle \tilde{\Sigma}^{1/2}_{YY}(g), R_{YX} \tilde{\Sigma}^{1/2}_{XX}(f) \rangle_{\mathcal{G}_Y} = \text{Cov}(f(X), g(Y)).$$
Examples:

1. For $f' \in \mathcal{G}_X$ and $g' \in \mathcal{G}_Y$, let $f = f' - \mathbb{E} [f']$ and $g = g' - \mathbb{E} [g']$. Then

$$
\langle \hat{\Sigma}_{YY}^{1/2} g, R_{XY} \hat{\Sigma}_{XX}^{1/2} f \rangle_{g_Y} = \text{Cov} (f(X), g(Y)).
$$

2. For all $f, g \in L^2_{PX}$ and $s, t \in L^2_{PY}$, we have

$$
\langle \hat{\Sigma}_{XX}^{1/2} g, \hat{\Sigma}_{XX}^{1/2} f \rangle_{g_X} = \text{Cov} (f(X), g(X))_{PX},
$$

$$
\langle \hat{\Sigma}_{YY}^{1/2} s, \hat{\Sigma}_{YY}^{1/2} t \rangle_{g_Y} = \text{Cov} (s(Y), t(Y))_{PY}.
$$

Remark: The extended covariance operators will be used to characterize the residual class $L^2_{PX} \ominus L^2_{PY}$. 
Examples:

1. For $f' \in \mathcal{G}_X$ and $g' \in \mathcal{G}_Y$, let
   \[ f = f' - \mathbb{E} [f'] \quad \text{and} \quad g = g' - \mathbb{E} [g']. \]
   Then
   \[
   \langle \tilde{\Sigma}_{YY}^{1/2} g, R_{XX} \tilde{\Sigma}_{XX}^{1/2} f \rangle_{g_Y} = \text{Cov} \left( f(X), g(Y) \right). 
   \]

2. For all $f, g \in L^2_{P_X}$ and $s, t \in L^2_{P_Y}$, we have
   \[
   \langle \tilde{\Sigma}_{XX}^{1/2} g, \tilde{\Sigma}_{XX}^{1/2} f \rangle_{g_X} = \text{Cov} \left( f(X), g(X) \right)_{P_X},
   \]
   \[
   \langle \tilde{\Sigma}_{YY}^{1/2} s, \tilde{\Sigma}_{YY}^{1/2} t \rangle_{g_Y} = \text{Cov} \left( s(Y), t(Y) \right)_{P_Y}. 
   \]

Remark: The extended covariance operators will be used to characterize the residual class $L^2_{P_X} \ominus L^2_{P_Y}$. 
Define

$$
\mathbb{E}_{X|Y} := \tilde{\Sigma}_{YY}^{-1/2} R_{YX} \tilde{\Sigma}_{XX}^{1/2}
$$

$$L_{P_X}^2 \rightarrow L_{P_Y}^2$$

**Proposition 3**

Under conditions (A) and (B), we have:

1. $$\forall f \in L_{P_X}^2, \mathbb{E}_{X|Y} f = \mathbb{E} [f(X)|Y];$$
2. $$\forall g \in L_{P_Y}^2, \mathbb{E}_{X|Y}^* f = \mathbb{E} [g(Y)|X].$$
**SIR** finds the DR directions by applying PCA on

\[
[\text{Var} (X)]^{-1} \text{Var} (\mathbb{E} [X|Y]) .
\]

We want to use operators to define the variance of the conditional expectation in functional spaces.

**Corollary 1**

Under conditions (A) and (B), \( \forall f, g \in L^2_{P_X} \),

\[
\langle g, \mathbb{E}^*_X|Y \mathbb{E} X|Y f \rangle_{L^2_{P_X}} = \text{Cov} (\mathbb{E} [g(X)|Y], \mathbb{E} [f(X)|Y]) .
\]

Moreover, \( \mathbb{E}^*_X|Y \mathbb{E} X|Y \) is a bounded linear operator on \( L^2_{P_X} \), and \( \| \mathbb{E}^*_X|Y \mathbb{E} X|Y \| \leq 1 \).

\[
\langle f, \mathbb{E}^*_X|Y \mathbb{E} X|Y f \rangle_{L^2_{P_X}} \quad \text{generalizes Var} (\mathbb{E} [X|Y]) .
\]
- **SIR** finds the DR directions by applying PCA on 
\[ [\text{Var} (X)]^{-1} \text{Var} (\mathbb{E} [X|Y]) . \]

- We want to use operators to define the variance of the conditional expectation in functional spaces.

**Corollary 1**

Under conditions (A) and (B), \( \forall f, g \in L^2_{P_X} \),

\[ \langle g, \mathbb{E}^*_X|Y \mathbb{E}_X|Y f \rangle_{L^2_{P_X}} = \text{Cov} \left( \mathbb{E} [g(X)|Y], \mathbb{E} [f(X)|Y] \right) . \]

Moreover, \( \mathbb{E}^*_X|Y \mathbb{E}_X|Y \) is a bounded linear operator on \( L^2_{P_X} \), and \( \|\mathbb{E}^*_X|Y \mathbb{E}_X|Y\| \leq 1 \).

- \( \langle f, \mathbb{E}^*_X|Y \mathbb{E}_X|Y f \rangle_{L^2_{P_X}} \) generalizes \( \text{Var} (\mathbb{E} [X|Y]) \).
**SIR** finds the DR directions by applying PCA on
\[ [\text{Var}(X)]^{-1} \text{Var}(\mathbb{E}[X|Y]). \]

We want to use operators to define the variance of the conditional expectation in functional spaces.

**Corollary 1**

Under conditions (A) and (B), \( \forall f, g \in L^2_{P_X} \),
\[ \langle g, \mathbb{E}^*_{X|Y} \mathbb{E}_{X|Y} f \rangle_{L^2_{P_X}} = \text{Cov}(\mathbb{E}[g(X)|Y], \mathbb{E}[f(X)|Y]). \]

Moreover, \( \mathbb{E}^*_{X|Y} \mathbb{E}_{X|Y} \) is a bounded linear operator on \( L^2_{P_X} \), and
\[ \| \mathbb{E}^*_{X|Y} \mathbb{E}_{X|Y} \| \leq 1. \]

\[ \langle f, \mathbb{E}^*_{X|Y} \mathbb{E}_{X|Y} f \rangle_{L^2_{P_X}} \] generalizes \( \text{Var}(\mathbb{E}[X|Y]). \)
Relate operators with central class:

**Theorem 5**

If conditions (A) and (B) are satisfied and $\mathcal{M}_{G^*}$ is complete, then

$$\text{Range} \left( \mathbb{E}^*_{X|Y} \mathbb{E}_{X|Y} \right) = \mathcal{M}_{G^*}.$$ 

**Remark:**

1) $L^2_{P_X}$ inner product absorbs the marginal variance in the predictor vector.

2) Sample estimator of the directions from $\mathcal{M}_{G^*}$ is called GSIR.
$T : \mathcal{H}_X \to L^2_{P_X}, \quad f \to f - \mathbb{E}[f],$

$T_j : \text{Range}(T) \to \mathbb{R}, \quad g \to \mathbb{E}[g(X)|Y \in J_i],$

where $\{J_i\}_{i=1,...,h}$ is a partition of $\Omega_Y$ (i.e. slicing)

$\mu_1, \ldots, \mu_h \in \text{Range}(T)$ are Riesz representations of $T_j'$s.

Then use

$$\text{span}\left\{\Sigma_{XX}^{-1} \mu_1, \ldots, \Sigma_{XX}^{-1} \mu_h\right\} \subseteq \mathcal{C}_Y|X \subseteq \mathcal{M}g^*.$$
Sample Estimator for GSIR

\[ \widehat{E}_{X|Y} \widehat{E}_{X|Y} \]

\[ = (G_X + \epsilon_X I_n)^{-3/2} G_X^{3/2} (G_Y + \epsilon_Y I_n)^{-1} G_Y^2 (G_Y + \epsilon_Y I_n)^{-1} G_X^{3/2} (G_X + \epsilon_X I_n)^{-3/2} \]

where

\[ G_X : \text{centered Gram matrix induced by pd function } k_X(\cdot, \cdot). \]

Then

\[ \hat{f}_i = \hat{\phi}_i^T (G_Y + \epsilon_Y I_n)^{-1}, \]

where \( \hat{\phi}_i \) is the \( i^{th} \) leading eigen-vector of \( \widehat{E}_{X|Y} \widehat{E}_{X|Y}. \)
Define:

\[
\mathbb{E}_{Y|X}^{(nc)} : L^2_{P_X}(nc) \rightarrow L^2_{P_Y}(nc)
\]
such that

\[
\langle g, \mathbb{E}_{Y|X}^{(nc)} f \rangle_{L^2_{P_X}(nc)} = \mathbb{E} [g(Y)f(X)]
\]

Then there exists an operator

\[
V_{X|Y} : \Omega_Y \rightarrow \mathcal{B}(L^2_{P_X})
\]

to represent \[
\left( \mathbb{E}_{Y|X}^{(nc)} [fg] - \mathbb{E}_{Y|X}^{(nc)} [f] \mathbb{E}_{Y|X}^{(nc)} [g] \right) (y).
\]

So we have:

\[
\langle f, V_{X|Y} f \rangle_{L^2_{P_X}} = \text{Var} (f(X)|Y)
\]
Define:

\[
\mathbb{E}^{(nc)}_{Y|X} : L^2_{P_X} \to L^2_{P_Y} \quad \text{such that}
\]

\[
\langle g, \mathbb{E}^{(nc)}_{Y|X} f \rangle_{L^2_{P_X}} = \mathbb{E} [g(Y)f(X)]
\]

Then there exists an operator

\[
V_{X|Y} : \Omega_Y \to \mathcal{B}(L^2_{P_X})
\]

to represent \((\mathbb{E}^{(nc)}_{Y|X}[fg] - \mathbb{E}^{(nc)}_{Y|X}[f]\mathbb{E}^{(nc)}_{Y|X}[g])(y)\).

So we have:

\[
\langle f, V_{X|Y} f \rangle_{L^2_{P_X}} = \text{Var} (f(X)|Y)
\]
Define:

\[ E_{Y|X}^{(nc)} : L^2_{P_X}^{(nc)} \rightarrow L^2_{P_Y}^{(nc)} \] such that

\[ \langle g, E_{Y|X}^{(nc)} f \rangle_{L^2_{P_X}^{(nc)}} = E[g(Y)f(X)] \]

Then there exists an operator

\[ V_{X|Y} : \Omega_Y \rightarrow \mathcal{B}(L^2_{P_X}) \]

to represent \( \left( E_{Y|X}^{(nc)} [fg] - E_{Y|X}^{(nc)} [f] E_{Y|X}^{(nc)} [g] \right) (y) \).

So we have:

\[ \langle f, V_{X|Y} f \rangle_{L^2_{P_X}} = \text{Var} \left( f(X) | Y \right) \]
Define

\[ V : \langle f, Vg \rangle = \text{Cov} \left( f(X), g(X) \right), \]
\[ S = \mathbb{E} \left[ (V - V_{X|Y})^2 \right]. \]

\( S \) generalizes \( \Sigma^{-1} \mathbb{E} [ \text{Var} (X) - \text{Var} (X|Y)]^2 \Sigma^{-1} \).

Then,

\[ \mathcal{C}_{X|Y} \subseteq \overline{\text{Range}(S)} \subseteq M_{\mathcal{G}^*}. \]

Remark: GSAVE is expected to discover functions outside \( \mathcal{C}_{X|Y} \).
Define

\[ V : \langle f, Vg \rangle = \text{Cov} (f(X), g(X)), \]

\[ S = \mathbb{E} \left[ (V - V_{X|Y})^2 \right]. \]

\( S \) generalizes \( \Sigma^{-1} \mathbb{E} \left[ \text{Var} (X) - \text{Var} (X|Y) \right]^2 \Sigma^{-1}. \)

Then,

\[ \mathcal{C}_{X|Y} \subseteq \text{Range}(S) \subseteq \mathcal{M}_{G^*}. \]

**Remark:** GSAVE is expected to discover functions outside \( \mathcal{C}_{X|Y}. \)
Sample Estimator for GS SAVE

\[ \hat{E}^{(nc)}_{X|Y} = (L_Y L_Y^T)^+ (L_Y L_X^T) \]

where

\[ L_X = (1_n, K_X)^T \], i.e. non-centered Gram matrix plus a intercept column

\[ L_Y (y) = (1, k_Y(y, Y_1), \ldots, k_Y(y, Y_n))^T \]

\[ C_Y(y) = L_Y^T (L_Y L_Y^T)^+ L_Y (y) \]

\[ \Lambda(y) = \text{diag}(C_Y(y)) - C_Y(y) C_Y^T(y) \]

Then

\[ \hat{S} = \frac{1}{n} \sum_{i=1}^{n} \left( L_X Q L_X^T + \epsilon_X I_{n+1} \right)^{-1/2} L_X Q \Gamma_i Q \Gamma_i Q L_X^T \left( L_X Q L_X^T + \epsilon_X I_{n+1} \right)^{-1/2} \]

where

\[ \Gamma_i = \left( Q/n - \Lambda(Y_i) \right), \quad Q = I_n - 1_n 1_n^T / n \]