Chapter 2

Dimension Reduction Subspaces

2.1 Conditional Independence

In this section we review the definitions and basic properties of conditional independence, which are crucial for Sufficient Dimension Reduction. Our development is in terms of conditional independence of σ-fields — rather than that of random variables — which is more general than can be found in most text books. Besides the generality, it is in fact easier to discuss conditional independence in terms of σ-fields than in terms of random variables. After developing the properties of conditional independence in the general setting, we then discuss the special cases where the relevant σ-fields are generated by random variables. For more information on conditional independence of σ-fields, see Hoffmann-Jorgensen (1994), page 460.

As in Chapter 1, let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $\mathscr{G}_1$ and $\mathscr{G}_2$ be two sub σ-fields of $\mathscr{F}$. We say that $\mathscr{G}_1$ and $\mathscr{G}_2$ are independent, and write $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2$, if, for any $A \in \mathscr{G}_1$ and $B \in \mathscr{G}_2$, A and B are independent; that is, $P(A \cap B) = P(A)P(B)$. Now let $\mathscr{G}_3$ be a third σ-field and, for any $A \in \mathscr{F}$, let $P(A|\mathscr{G}_3)$ be the conditional probability of A given $\mathscr{G}_3$; that is, $P(A|\mathscr{G}_3)$ is a mapping from Ω that (i) is measurable with respect to $\mathscr{G}_3$; (ii) satisfies $\int_G P(A|\mathscr{G}_3)dP = P(G \cap A)$ for all $G \in \mathscr{G}_3$. We say that $\mathscr{G}_1$ and $\mathscr{G}_2$ are conditional independent given $\mathscr{G}_3$, and write $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | \mathscr{G}_3$ if, for every $A \in \mathscr{G}_1$ and $B \in \mathscr{G}_2$, we have

$$P(A \cap B|\mathscr{G}_3) = P(A|\mathscr{G}_3)P(B|\mathscr{G}_3), \quad a.s.\ P.$$

Let $\mathscr{P} = \{A \cap B : A \in \mathscr{G}_1, B \in \mathscr{G}_2\}$. It is easy to show that $\mathscr{P}$ is a π-system. Let $(\mathscr{G}_1, \mathscr{G}_2)$ — or simply $\mathscr{G}_1, \mathscr{G}_2$ — denote the σ-field generated by $\mathscr{P}$. The next proposition gives an equivalent definition of conditional independence.

Proposition 2.1 *Let $\mathscr{G}_1$, $\mathscr{G}_2$, and $\mathscr{G}_3$ be sub-σ-fields of $\mathscr{F}$. Then the following statements are equivalent:*

1. $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | \mathscr{G}_3$;
2. $P(A|\mathscr{G}_2, \mathscr{G}_3) = P(A|\mathscr{G}_3)$ *a.s. P for each* $A \in \mathscr{G}_1$.

Before proving this proposition, we first prove a proposition that will be used many times in later development.

Proposition 2.2 *Let X, Y, and Z be random variables defined on* $(\Omega, \mathscr{F}, P)$*, where the product XY is integrable with respect to P. Then*

$$E(XE(Y|Z)) = E(E(X|Z)Y) = E(E(X|Z)E(Y|Z)). \tag{2.1}$$

PROOF. This is because

$$\begin{aligned} E(XE(Y|Z)) &= E(E(XE(Y|Z)|Z)) \\ &= E(E(X|Z)E(Y|Z)) \\ &= E(E(E(X|Z)Y|Z)) \\ &= E(E(X|Z)Y). \end{aligned}$$

□

We are now ready to prove Proposition 2.1.

PROOF OF PROPOSITION 2.1. *1* $\Rightarrow$ *2*. It suffices to show that, for any $A_1 \in \mathscr{G}_1$, $P(A_1|\mathscr{G}_3)$ is the conditional probability of A_1 given $(\mathscr{G}_2, \mathscr{G}_3)$. Since $P(A_1|\mathscr{G}_3)$ is measurable with respect to $(\mathscr{G}_2, \mathscr{G}_3)$, it suffices to show that, for any $B \in (\mathscr{G}_2, \mathscr{G}_3)$ we have

$$\int_B P(A_1|\mathscr{G}_3) dP = P(A_1 \cap B).$$

Because $\mathscr{P} = \{A_2 A_3 : A_2 \in \mathscr{G}_2, A_3 \in \mathscr{G}_3\}$ is a π system generating $(\mathscr{G}_2, \mathscr{G}_3)$, by the π-λ Theorem (see, for example, Billingsley (1995), page 42), it suffices to show the above inequality for all $B \in \mathscr{P}$. Let $B = A_2 \cap A_3$ where $A_2 \in \mathscr{G}_2$, $A_3 \in \mathscr{G}_3$. Then

$$\int_B P(A_1|\mathscr{G}_3) dP = E[I_{A_2} I_{A_3} E(I_{A_1}|\mathscr{G}_3)]$$

By Proposition 2.1, the right-hand side can be written as

$$E[E(I_{A_2} I_{A_3}|\mathscr{G}_3) E(I_{A_1}|\mathscr{G}_3)] = E[I_{A_3} E(I_{A_2}|\mathscr{G}_3) E(I_{A_1}|\mathscr{G}_3)].$$

By *1*, we have $E(I_{A_2}|\mathscr{G}_3) E(I_{A_1}|\mathscr{G}_3) = E(I_{A_2} I_{A_1}|\mathscr{G}_3)$. Hence the right-hand side above reduces to $E[I_{A_3} E(I_{A_2} I_{A_1}|\mathscr{G}_3)]$, which, by Proposition 2.1, is the same as

$$E[E(I_{A_3}|\mathscr{G}_3) I_{A_2} I_{A_1}] = E[I_{A_3} I_{A_2} I_{A_1}] = E(I_{A_1} I_B) = P(A_1 \cap B),$$

as desired.
2 $\Rightarrow$ *1*. The relation in statement *1* is equivalent to

$$E(I_{A_1} I_{A_2}|\mathscr{G}_3) = E(I_{A_1}|\mathscr{G}_3) E(I_{A_2}|\mathscr{G}_3)$$

for all $A_1 \in \mathscr{G}_1$ and $A_2 \in \mathscr{G}_2$. Because $\mathscr{G}_3 \subseteq (\mathscr{G}_2, \mathscr{G}_3)$, the left-hand side is

$$\begin{aligned} E[E(I_{A_1} I_{A_2} | \mathscr{G}_2, \mathscr{G}_3)|\mathscr{G}_3] &= E[E(I_{A_1} I_{A_2} | \mathscr{G}_2, \mathscr{G}_3)|\mathscr{G}_3] \\ &= E[I_{A_2} E(I_{A_1} | \mathscr{G}_2, \mathscr{G}_3)|\mathscr{G}_3]. \end{aligned}$$

By statement *2*, $E(I_{A_1}|\mathscr{G}_2, \mathscr{G}_3) = E(I_{A_1}|\mathscr{G}_3)$. Hence the right-hand side reduces to $E(I_{A_2}|\mathscr{G}_3)E(I_{A_1}|\mathscr{G}_3)$. □

Conditional independence satisfies a set of axioms — called semigraphoid axioms — which will be extremely useful for our discussions. These axioms were first proved in Dawid (1979) as properties for conditional independence. They were proposed as axioms for constructing graphical models and for making causal inference by Pearl and Verma (1987) and Pearl et al. (1989). While in its original form the semigraphoid axioms were stated in terms of random variables, here we state them in terms of sub σ-fields of $\mathscr{F}$, which makes the statements more general and the proof more transparent.

Theorem 2.1 *Let $\mathscr{G}_1, \ldots, \mathscr{G}_4$ be sub-σ-fields of $\mathscr{F}$. Then the following statements hold true:*

1. $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | \mathscr{G}_3 \Rightarrow \mathscr{G}_2 \perp\!\!\!\perp \mathscr{G}_1 | \mathscr{G}_3$;
2. $\mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3) | \mathscr{G}_4 \Rightarrow \mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | \mathscr{G}_4$;
3. $\mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3) | \mathscr{G}_4 \Rightarrow \mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | (\mathscr{G}_3, \mathscr{G}_4)$;
4. $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | (\mathscr{G}_3, \mathscr{G}_4)$, $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_3 | \mathscr{G}_4 \Rightarrow \mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3) | \mathscr{G}_4$.

PROOF. *1*. Let $A_1 \in \mathscr{G}_1$, $A_2 \in \mathscr{G}_2$. Then

$$P(A_1 \cap A_2 | \mathscr{G}_3) = P(A_1 | \mathscr{G}_3) P(A_2 | \mathscr{G}_3)$$

if and only if

$$P(A_2 \cap A_1 | \mathscr{G}_3) = P(A_2 | \mathscr{G}_3) P(A_1 | \mathscr{G}_3).$$

2. Since $\mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3) | \mathscr{G}_4$, we have, for any $A_1 \in \mathscr{G}_1$ and $B = A_2 \cap A_3$ where $A_2 \in \mathscr{G}_2$ and $A_3 \in \mathscr{G}_3$, we have $P(A_1 \cap B | \mathscr{G}_4) = P(A_1 | \mathscr{G}_4) P(B | \mathscr{G}_4)$. Take $A_3 = \Omega$. Then $B = A_2$ and we have $P(A_1 \cap A_2 | \mathscr{G}_4) = P(A_1 | \mathscr{G}_4) P(A_2 | \mathscr{G}_4)$. Since this is true for all $A_1 \in \mathscr{G}_1$ and $A_2 \in \mathscr{G}_2$, we have $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2 | \mathscr{G}_4$.
3. We need to show that, for any $A_1 \in \mathscr{G}_1$, $A_2 \in \mathscr{G}_2$, we have

$$E(I_{A_1} I_{A_2} | \mathscr{G}_3, \mathscr{G}_4) = E(I_{A_1} | \mathscr{G}_3, \mathscr{G}_4) E(I_{A_2} | \mathscr{G}_3, \mathscr{G}_4).$$

We will show that the right-hand side is the conditional expectation on the left-hand side. Since the right-hand side is measurable with respect to $(\mathscr{G}_3, \mathscr{G}_4)$, we only need to show that, for any $B \in (\mathscr{G}_3, \mathscr{G}_4)$,

$$E[I_B E(I_{A_1} | \mathscr{G}_3, \mathscr{G}_4) E(I_{A_2} | \mathscr{G}_3, \mathscr{G}_4)] = P(A_1 \cap A_2 \cap B).$$

Since $\mathscr{P}$ is a π-system generating $(\mathscr{G}_3, \mathscr{G}_4)$, by the π-λ Theorem, it suffices to show that the above equality holds for all $B = A_3 \cap A_4$ where $A_3 \in \mathscr{G}_3$ and $A_4 \in \mathscr{G}_4$. For such B, the left-hand side becomes

$$E[I_{A_3} I_{A_4} E(I_{A_1}|\mathscr{G}_3, \mathscr{G}_4) E(I_{A_2}|\mathscr{G}_3, \mathscr{G}_4)].$$

Because $\mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3)|\mathscr{G}_4$, by statement *2* we have $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_3|\mathscr{G}_4$. Hence, by Proposition 2.1, $E(I_{A_1}|\mathscr{G}_3, \mathscr{G}_4) = E(I_{A_1}|\mathscr{G}_4)$ and the above becomes

$$\begin{aligned} E[I_{A_3} I_{A_4} E(I_{A_1}|\mathscr{G}_4) E(I_{A_2}|\mathscr{G}_3, \mathscr{G}_4)] &= E[I_{A_4} E(I_{A_1}|\mathscr{G}_4) E(I_{A_2} I_{A_3}|\mathscr{G}_3, \mathscr{G}_4)] \\ &= E[E(I_{A_4} E(I_{A_1}|\mathscr{G}_4) I_{A_2} I_{A_3}|\mathscr{G}_3, \mathscr{G}_4)] \\ &= E[E(I_{A_1}|\mathscr{G}_4) I_{A_4} I_{A_2} I_{A_3}]. \end{aligned}$$

By Proposition 2.1, the right-hand side can be rewritten as

$$E[I_{A_1} E(I_{A_4} I_{A_2} I_{A_3}|\mathscr{G}_4)] = E[I_{A_1} I_{A_4} E(I_{A_2} I_{A_3}|\mathscr{G}_4)] = E[E(I_{A_1} I_{A_4}|\mathscr{G}_4) E(I_{A_2} I_{A_3}|\mathscr{G}_4)].$$

By $\mathscr{G}_1 \perp\!\!\!\perp (\mathscr{G}_2, \mathscr{G}_3)|\mathscr{G}_4$ again, the right-hand side can be rewritten as

$$E[E(I_{A_1} I_{A_4} I_{A_2} I_{A_3}|\mathscr{G}_4)] = E(I_{A_1} I_{A_4} I_{A_2} I_{A_3}) = P(A_1 \cap A_2 \cap B).$$

This proves statement *3*.
4. We need to show that, for any $A_1 \in \mathscr{G}_1$, $B \in (\mathscr{G}_2, \mathscr{G}_3)$, we have

$$E(I_{A_1} I_B|\mathscr{G}_4) = E(I_{A_1}|\mathscr{G}_4) E(I_B|\mathscr{G}_4).$$

We will show that the right-hand side is the conditional expectation on the left-hand side, for which it suffices to show that, for any $A_4 \in \mathscr{G}_4$,

$$E(I_{A_4} E(I_{A_1}|\mathscr{G}_4) E(I_B|\mathscr{G}_4)] = E(I_{A_4} I_{A_1} I_B). \tag{2.2}$$

By the π-λ Theorem, it suffices to verify this equality for all $B = A_2 A_3$ where $A_2 \in \mathscr{G}_2$ and $A_3 \in \mathscr{G}_3$. For such a set B, the left-hand side is

$$E(I_{A_4} E(I_{A_1}|\mathscr{G}_4) E(I_{A_2} I_{A_3}|\mathscr{G}_4)] = E(E(I_{A_1}|\mathscr{G}_4) E(I_{A_4} I_{A_2} I_{A_3}|\mathscr{G}_4)]. \tag{2.3}$$

By Proposition 2.1, the condition $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_3|\mathscr{G}_4$, and Proposition 2.1, we can rewrite the right-hand side of (2.3) as

$$\begin{aligned} E(E(I_{A_1}|\mathscr{G}_4) I_{A_4} I_{A_2} I_{A_3}] &= E(E(I_{A_1}|\mathscr{G}_3, \mathscr{G}_4) I_{A_4} I_{A_2} I_{A_3}] \\ &= E(E(I_{A_1}|\mathscr{G}_3, \mathscr{G}_4) E(I_{A_4} I_{A_2} I_{A_3}|\mathscr{G}_3, \mathscr{G}_4)] \\ &= E(E(I_{A_1}|\mathscr{G}_3, \mathscr{G}_4) E(I_{A_2}|\mathscr{G}_3, \mathscr{G}_4) I_{A_4} I_{A_3}]. \end{aligned}$$

By the condition $\mathscr{G}_1 \perp\!\!\!\perp \mathscr{G}_2|\mathscr{G}_3, \mathscr{G}_4$, the right-hand side can be rewritten as

$$\begin{aligned} E(E(I_{A_1} I_{A_2}|\mathscr{G}_3, \mathscr{G}_4) I_{A_4} I_{A_3}] &= E(E(I_{A_1} I_{A_2} I_{A_4} I_{A_3}|\mathscr{G}_3, \mathscr{G}_4)] \\ &= E(I_{A_1} I_{A_2} I_{A_4} I_{A_3}) \\ &= E(I_{A_1} I_{A_4} I_B), \end{aligned}$$

which is the right-hand side of (2.2). □

Now let us consider the conditional independence of random elements. Suppose X_1, …, X_4 are random elements defined on $(\Omega, \mathscr{F}, P)$ taking values in $(\Omega_{X_1}, \mathscr{F}_{X_1}), \ldots, (\Omega_{X_4}, \mathscr{F}_{X_4})$, respectively. Let $\sigma(X_i)$ be the sub-σ-field of $\mathscr{F}$ generated by X_i; that is, $\sigma(X_i) = X_i^{-1}(\mathscr{F}_{X_i})$. We say that X_1 and X_2 are independent if $\sigma(X_1) \perp\!\!\!\perp \sigma(X_2)$. In this case we write $X_1 \perp\!\!\!\perp X_2$. We say that X_1 and X_2 are independent given X_3 if $\sigma(X_1) \perp\!\!\!\perp \sigma(X_2) | \sigma(X_3)$. In this case we write $X_1 \perp\!\!\!\perp X_2 | X_3$. We say that X_1 and X_2 are conditionally independent given (X_3, X_4) if

$$\sigma(X_1) \perp\!\!\!\perp \sigma(X_2) | (\sigma(X_3), \sigma(X_4)).$$

In this case we write $X_1 \perp\!\!\!\perp X_2 | (X_3, X_4)$.

The following corollaries state the special cases of Proposition 2.1 and Theorem 2.1 for random elements.

Corollary 2.1 *Let X_1, X_2, and X_3 be random elements defined on $(\Omega, \mathscr{F}, P)$. Then the following statements are equivalent:*

1. $X_1 \perp\!\!\!\perp X_2 | X_3$;
2. $P(A|X_2, X_3) = P(A|X_3)$ *a.s. P for each* $A \in \sigma(X_1)$.

Corollary 2.2 *Let $X_1, \ldots, X_4$ be random variables defined on $(\Omega, \mathscr{F}, P)$. Then the following statements hold true:*

1. $X_1 \perp\!\!\!\perp X_2 | X_3 \Rightarrow X_2 \perp\!\!\!\perp X_1 | X_3$;
2. $X_1 \perp\!\!\!\perp (X_2, X_3) | X_4 \Rightarrow X_1 \perp\!\!\!\perp X_2 | X_4$;
3. $X_1 \perp\!\!\!\perp (X_2, X_3) | X_4 \Rightarrow X_1 \perp\!\!\!\perp X_2 | (X_3, X_4)$;
4. $X_1 \perp\!\!\!\perp X_2 | (X_3, X_4),\ X_1 \perp\!\!\!\perp X_3 | X_4 \Rightarrow X_1 \perp\!\!\!\perp (X_2, X_3) | X_4$.

2.2 Sufficient Dimension Reduction Subspace

Now let $X : \Omega \to \mathbb{R}^p$ be a random vector measurable with respect to $\mathscr{R}^p$, the Borel σ-field in $\mathbb{R}^p$, and let $Y : \Omega \to \mathbb{R}$ be a random variable measurable with respect to $\mathscr{R}$, the Borel σ-field in $\mathbb{R}$. Sufficient Dimension Reduction is concerned with the situations where the distribution of Y given X depends on X only through a set of linear combinations of X. That is, there is a matrix $\beta \in \mathbb{R}^{p \times r}$, where $r \le p$, such that

$$Y \perp\!\!\!\perp X | \beta^\mathsf{T} X. \tag{2.4}$$

This relation is unchanged if we replace β by any βA, where $A \in \mathbb{R}^{r \times r}$ is any nonsingular matrix; that is,

$$Y \perp\!\!\!\perp X | \beta^\mathsf{T} X \Leftrightarrow Y \perp\!\!\!\perp X | (\beta A)^\mathsf{T} X.$$

This is because there is a one-to-one relation between $\beta^\mathsf{T} X$ and $A^\mathsf{T} \beta^\mathsf{T} X = (\beta A)^\mathsf{T} X$. Thus the identifiable parameter in (2.4) is the space spanned by the columns of β, rather than β itself. We denote the column space of β by $\mathrm{span}(\beta)$.

The conditional independence relation (2.4) is extracted from useful regression models such as the Generalized Linear Model, where the conditional density of $f_{Y|X}$ depends on X through $\beta^{\mathsf{T}}X$, where $\beta \in \mathbb{R}^p$. That is,

$$f_{Y|X}(y|x) = h(y, \beta^{\mathsf{T}}x). \tag{2.5}$$

for some function h. See Li (1991). It is easy to see that (2.5) implies (2.4). However, (2.4) is much more general than (2.5) as specified by the Generalized Linear Model in Chapter 1, because (i) β does not have to be vector, and (ii) no assumption is imposed on the form of h. Some other popular regression models can also be stated, at least in form, as special cases of the Sufficient Dimension Reduction problem (2.4). For example, the Single Index Model is of the form

$$Y = f(\beta^{\mathsf{T}}X) + \varepsilon,$$

where $\beta \in \mathbb{R}^p$ ad $X \perp\!\!\!\perp \varepsilon$. See Ichimura (1993). Similarly, the Multiple Index Model

$$Y = f(\beta_1^{\mathsf{T}}X, \ldots, \beta_d^{\mathsf{T}}X) + \varepsilon,$$

where $\beta_1, \ldots, \beta_d \in \mathbb{R}^p$ and $\varepsilon \perp\!\!\!\perp X$, is also a special case of (2.4). See, for example, Yin et al. (2008).

The original form of Sufficient Dimension Reduction proposed by Li (1991) is

$$Y = h(\beta^{\mathsf{T}}X, \varepsilon), \tag{2.6}$$

where $\varepsilon \perp\!\!\!\perp X$. It is again easy to see that (2.6) implies (2.4). The form (2.4) was proposed in Cook (1994). See also Cook (1998).

If condition (2.4) is satisfied for β, then we call $\text{span}(\beta)$ a Sufficient Dimension Reduction subspace, or SDR subspace. Conversely, a subspace $\mathscr{S}$ of $\mathbb{R}^p$ is an SDR subspace if (2.4) is satisfied for any β such that $\mathscr{S} = \text{span}(\beta)$. Obviously, SDR subspace always exists, because $Y \perp\!\!\!\perp X | X$ always holds, which means $\mathbb{R}^p$ is an SDR subspace. Furthermore, the SDR subspace is not unique, because of the following fact.

Proposition 2.3 *If $\mathscr{S}_1$ is an SDR subspace and $\mathscr{S}_2$ is any subspace of $\mathbb{R}^p$ that contains $\mathscr{S}_1$, then $\mathscr{S}_2$ is also an SDR subspace.*

PROOF. Let β_1 and β_2 be matrices such that $\text{span}(\beta_1) = \mathscr{S}_1$ and $\text{span}(\beta_2) = \mathscr{S}_2$. Because the σ-fields $(X, \beta_2^{\mathsf{T}}X)$ and $\sigma(X)$ are identical, we have $Y \perp\!\!\!\perp (X, \beta_2^{\mathsf{T}}X) | \beta_1^{\mathsf{T}}X$. By Theorem 2.1, statement (iii), we have

$$Y \perp\!\!\!\perp X | (\beta_1^{\mathsf{T}}X, \beta_2^{\mathsf{T}}X).$$

However, since the σ-fields $(\beta_1^{\mathsf{T}}X, \beta_2^{\mathsf{T}}X)$ and $\beta_2^{\mathsf{T}}X$ are the same, we have $Y \perp\!\!\!\perp X | \beta_2^{\mathsf{T}}X$, which means $\mathscr{S}_2$ is an SDR subspace. $\square$

Since the SDR subspace is not unique, we naturally prefer the smallest SDR subspace, which achieves maximum dimension reduction. Under some mild assumptions, if $\mathscr{S}_1$ and $\mathscr{S}_2$ are SDR subspaces, then so is $\mathscr{S}_1 \cap \mathscr{S}_2$. The assumptions for this to hold is very weak: it was shown in Yin et al. (2008) that, if X is supported by an M-set (or matching set), then this condition is satisfied. Since the result is rather technical we omit its proof and take it for granted in this book. That is, we make the following assumption.

Assumption 2.1 *Let $\mathfrak{A}$ be the class of all SDR subspaces. We assume that*

$$\cap\{\mathscr{S} : \mathscr{S} \in \mathfrak{A}\} \in \mathfrak{A}.$$

That is, the intersection of all SDR subspaces is itself an SDR subspace.

Under this assumption we define the intersection of all SDR subspaces as the target of estimation in Sufficient Dimension Reduction.

Definition 2.1 *Suppose Assumption 2.1. The subspace*

$$\cap\{\mathscr{S} : \mathscr{S} \in \mathfrak{A}\}$$

is called central SDR subspace or the central subspace, and is written as $\mathscr{S}_{Y|X}$. The dimension d of $\mathscr{S}_{Y|X}$ is called the structural dimension.

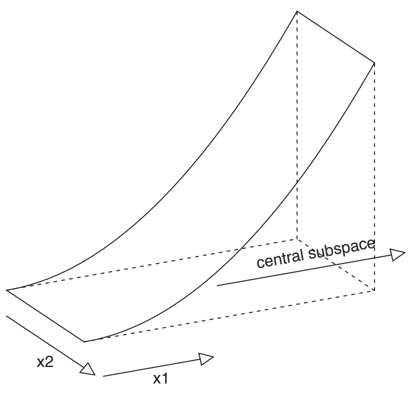

Figure 2.1 *Sufficient Dimension Reduction subspace.*

Figure 2.1 illustrates the idea of the central subspace. The surface in the figure represents the dependence of $E(Y|X)$ on X, where $X = (X_1, X_2)$ is a random vector in $\mathbb{R}^2$. In this case, $E(Y|X_1, X_2)$ is a function of X_1 alone; that is, $E(Y|X_1, X_2) = E(Y|X_1)$.

Under the additional assumption that the distribution of Y given X is determined by the conditional mean $E(Y|X)$, such as the regression model

$$Y = f(X) + \varepsilon,$$

where $X \perp\!\!\!\perp \varepsilon$, the central subspace is spanned by $(1,0)$, as indicated by the long arrow in the plot. The random variable $\beta^\top X$, where $\beta = (1,0)^\top$, carries all the information about the conditional distribution of $Y|X$.

2.3 Transformation Laws of Central Subspace

A useful property of the central subspace is that it transforms equivariantly under affine transformation of X (see, for example, Cook (1998)). Using this property we can work with standardized X in estimation of $\mathscr{S}_{Y|X}$, as we usually do in Sufficient Dimension Reduction.

Theorem 2.2 *If $A \in \mathbb{R}^{p\times p}$ be a nonsingular matrix and $b \in \mathbb{R}^p$, then*

$$\mathscr{S}_{Y|X} = A^\top \mathscr{S}_{Y|AX+b}.$$

PROOF. Let

$$\begin{aligned}
\mathfrak{A}_{Y|X} &= \{\mathscr{S} : \mathscr{S} \text{ is an SDR space of } Y \text{ versus } X\},\\
A^{-\top}\mathfrak{A}_{Y|X} &= \{A^{-\top}\mathscr{S} : \mathscr{S} \in \mathfrak{A}_{Y|X}\},\\
\mathfrak{A}_{Y|AX+b} &= \{\mathscr{S} : \text{ is an SDR subspace of } Y \text{ versus } AX+b\},
\end{aligned}$$

where, for a matrix A, $A^{-\top}$ denote $(A^{-1})^\top$ or $(A^\top)^{-1}$. We first show that

$$A^{-\top}\mathfrak{A}_{Y|X} = \mathfrak{A}_{Y|AX+b}. \tag{2.7}$$

Let $\mathscr{S}' \in A^{-\top}\mathfrak{A}_{Y|X}$. Then $\mathscr{S}' = A^{-\top}\mathscr{S}$, where $\mathscr{S} \in \mathfrak{A}_{Y|X}$. Let β be a matrix such that $\text{span}(\beta) = \mathscr{S}$. Then $Y \perp\!\!\!\perp X|\beta^\top X$, which implies

$$\begin{aligned}
Y \perp\!\!\!\perp |\beta^\top A^{-1}[(AX+b)-b] &\Rightarrow Y \perp\!\!\!\perp X|\beta^\top A^{-1}(AX+b)\\
&\Rightarrow Y \perp\!\!\!\perp X|(A^{-\top}\beta)^\top(AX+b)\\
&\Rightarrow Y \perp\!\!\!\perp (AX+b)|(A^{-\top}\beta)^\top(AX+b).
\end{aligned}$$

The last line means $\text{span}(A^{-\top}\beta) = A^{-\top}\mathscr{S} = \mathscr{S}' \in \mathfrak{A}_{AX+b}$. Hence $A^{-\top}\mathfrak{A}_{Y|X} \subseteq \mathfrak{A}_{Y|AX+b}$. Next, let $\mathscr{S}' \in \mathfrak{A}_{Y|AX+b}$ and let β' be a matrix such that $\text{span}(\beta') = \mathscr{S}'$. Then

$$Y \perp\!\!\!\perp AX+b|(\beta')^\top(AX+b) \Rightarrow Y \perp\!\!\!\perp X|(A^\top\beta')^\top X.$$

So $\text{span}(A^\top\beta') \in \mathfrak{A}_{Y|X}$, which implies $\mathfrak{A}_{Y|AX+b} \subseteq A^{-\top}\mathscr{S}_{Y|X}$. Thus we have proved (8.3).

The relation (8.3) can be equivalently written as

$$\{\mathscr{S} : \mathscr{S} \in \mathfrak{A}_{Y|X}\} = A^\top\{\mathscr{S}' : \mathscr{S}' \in \mathfrak{A}_{Y|AX+b}\} = \{A^\top\mathscr{S}' : \mathscr{S}' \in \mathfrak{A}_{Y|AX+b}\}.$$

Take intersection of the classes of sets on both sides to obtain

$$\begin{aligned}\mathscr{S}_{Y|X} &= \cap\{\mathscr{S} : \mathscr{S} \in \mathfrak{A}_{Y|X}\} \\ &= \cap\{A^{\mathsf{T}}\mathscr{S}' : \mathscr{S}' \in \mathfrak{A}_{Y|AX+b}\} \\ &= A^{\mathsf{T}} \cap \{\mathscr{S}' : \mathscr{S}' \in \mathfrak{A}_{Y|AX+b}\} = A^{\mathsf{T}}\mathscr{S}_{Y|AX+b},\end{aligned}$$

which is the desired equality. □

Typically, to estimate the central subspace $\mathscr{S}_{Y|X}$, we start with an estimator of a positive semidefinite matrix M such that $\text{span}(M)$ is a subspace of $\mathscr{S}_{Y|Z}$, where $Z = \Sigma^{-1/2}(X-\mu)$, $\Sigma = \text{var}(X)$, and $\mu = E(X)$. Let $s \leq d$ be the rank of M, and let $v_1, \ldots, v_s$ be the vectors of M corresponding the nonzero eigenvalues. Then we use the sample-level counterpart of $\mathscr{S} = \text{span}(v_1, \ldots, v_s)$ as estimator of $\mathscr{S}_{Y|Z}$. The space $\Sigma^{-1/2}\mathscr{S}$ is a subspace of $\Sigma^{-1/2}\mathscr{S}_{Y|Z}$ which, by Theorem 2.2, is the central subspace $\mathscr{S}_{Y|X}$.

Another type of transformation is that on the response Y. Because

$$Y \perp\!\!\!\perp X | \beta^{\mathsf{T}} X \Rightarrow g(Y) \perp\!\!\!\perp X | \beta^{\mathsf{T}} X,$$

any SDR subspace of Y versus X is also a subspace of $g(Y)$ versus X. Let $\mathfrak{A}$ be the collection of all SDR subspaces for Y versus X, and $\mathfrak{B}$ be the collection of all SDR subspaces of $g(Y)$ versus X. Then $\mathfrak{A} \subseteq \mathfrak{B}$. It follows that $\cap\{\mathscr{S} : \mathscr{S} \in \mathfrak{A}\} \supseteq \{\mathscr{S} : \mathscr{S} \in \mathfrak{B}\}$. Thus we have proved the following theorem.

Theorem 2.3 *For any measurable function $g(Y)$, we have $\mathscr{S}_{g(Y)|X} \subseteq \mathscr{S}_{Y|X}$.*

2.4 Fisher Consistency, Unbiasedness, and Exhaustiveness

In estimation theory, a statistic can often be represented as a mapping defined on the family of all distributions of a random element W to the parameter space. Specifically, let $\mathfrak{F}$ be the family of all distributions of W, let Θ be the parameter space, and let F_n be the empirical distribution based on an i.i.d. sample $W_1, \ldots, W_n$ of observations on W. For example, in our context, W is (X,Y) and W_i is (X_i,Y_i). A statistical functional T is a mapping from $\mathfrak{F}$ to some space $\mathbb{S}$. Many statistics can be represented as $T(F_n)$. Let $\theta_0 \in \Theta$ be the parameter to be estimated.

Definition 2.2 *We say that a statistic $T(F_n)$ is Fisher consistent for estimating θ_0 if $T(F_0) = \theta_0$.*

For example, suppose $T(F) = E_F(X)$, and we are interested in estimating $\mu_0 = E_{F_0}(X)$. Then $T(F_n) = \bar{X}$ and $T(F_0) = \mu_0$. So $\bar{X}$ is Fisher consistent estimator of μ_0. Similarly, if $\Sigma(F) = \text{var}_F(X)$, then $\Sigma(F_n)$, the sample covariance matrix of $X_1, \ldots, X_n$, is a Fisher consistent estimate of true covariance matrix $\text{var}_{F_0}(X)$.

In the context of Sufficient Dimension Reduction, let $\mathfrak{F}$ represent the collection of all distributions of (X,Y), let F_n represent the empirical distribution based on an i.i.d. sample $(X_1,Y_1), \ldots, (X_n,Y_n)$ of (X,Y), and let F_0 be the true distribution of (X,Y). The parameter space is the collection of all p dimensional subspace of $\mathbb{R}^p$. This is called the Grassman manifold $\mathbb{G}_{p,d}$. The central subspace $\mathscr{S}_{Y|X}$ is a member of $\mathbb{G}_{p,d}$. An estimator of the central subspace typically takes the form $M(F_n)$, where M is a mapping from $\mathfrak{F}$ to $\mathbb{R}^{p\times r}$. Following Ye and Weiss (2003), we call the estimator $M(F_n)$ a *candidate matrix*.

Definition 2.3 *We say that an estimator $M(F_n)$ is an unbiased estimator of $\mathscr{S}_{Y|X}$ if* $\mathrm{span}(M(F_0)) \subseteq \mathscr{S}_{Y|X}$. *We say that $M(F_n)$ is an exhaustive estimator of $\mathscr{F}_{Y|X}$ if* $\mathrm{span}(M(F_0)) \supseteq \mathscr{S}_{Y|X}$. *We say that $M(F_n)$ is a Fisher consistent estimate of $\mathscr{S}_{Y|X}$ if* $\mathrm{span}(M(F_0)) = \mathscr{S}_{Y|X}$.