*The Annals of Statistics*
1993, Vol. 21, No. 2, 867–889

# ON ALMOST LINEARITY OF LOW DIMENSIONAL PROJECTIONS FROM HIGH DIMENSIONAL DATA

BY PETER HALL AND KER-CHAU LI[1]

*Australian National University and CSIRO, and UCLA*

This paper studies the shapes of low dimensional projections from high dimensional data. After standardization, let $\mathbf{x}$ be a $p$-dimensional random variable with mean zero and identity covariance. For a projection $\beta'\mathbf{x}$, $\|\beta\| = 1$, find another direction $b$ so that the regression curve of $b'\mathbf{x}$ against $\beta'\mathbf{x}$ is as nonlinear as possible. We show that when the dimension of $\mathbf{x}$ is large, for most directions $\beta$ even the most nonlinear regression is still nearly linear.

Our method depends on the construction of a pair of $p$-dimensional random variables, $\mathbf{w}_1, \mathbf{w}_2$, called the *rotational twin*, and its density function with respect to the standard normal density. With this, we are able to obtain closed form expressions for measuring deviation from normality and deviation from linearity in a suitable sense of average. As an interesting by-product, from a given set of data we can find simple unbiased estimates of $E(f_{\beta'\mathbf{x}}(t)/\phi_1(t) - 1)^2$ and $E[(\|E(\mathbf{x}|\beta,\ \beta'\mathbf{x} = t)\|^2 - t^2) f^2_{\beta'\mathbf{x}}(t)/\phi_1^2(t)]$, where $\phi_1$ is the standard normal density, $f_{\beta'\mathbf{x}}$ is the density for $\beta'\mathbf{x}$ and the "$E$" is taken with respect to the uniformly distributed $\beta$. This is achieved without any smoothing and without resorting to any laborious projection procedures such as grand tours. Our result is related to the work of Diaconis and Freedman.

The impact of our result on several fronts of data analysis is discussed. For example, it helps establish the validity of regression analysis when the link function of the regression model may be grossly wrong. A further generalization, which replaces $\beta'\mathbf{x}$ by $B'\mathbf{x}$ with $B = (\beta_1, \dots, \beta_k)$ for $k$ randomly selected orthonormal vectors $(\beta_i,\ i = 1, \dots, k)$, helps broaden the scope of application of sliced inverse regression (SIR).

**1. Introduction and summary of main results.** Recent advances in computer technology have greatly enhanced our ability to extract useful information from high dimensional data. There are several procedures for optimally seeking interesting features in the data [e.g., Friedman and Stuetzle (1981), Huber (1985), Friedman (1987), Donoho and Johnstone (1989), Haerdle and Stoker (1989), Hall (1989a, b) and Chen (1991)]. Equipped with modern graphics facilities, statisticians can now easily interact with the high dimensional data by 3-D rotation plots, scatterplot matrices, contour plots, colors, brushing, slicing and many animation techniques [e.g., Wegman and Depriest (1986), Cleveland (1988), Cleveland and MacGill (1988) and Tierney (1990)]. Facing this new trend of statistical activities, theoretical investigation

Received September 1991; revised June 1992.
[1]Partially supported by NSF Grant DMS-89-02494.
*AMS* 1991 *subject classifications*. Primary 60F99; secondary 62H99.
*Key words and phrases*. Projections, projection pursuit, data visualization, dimension reduction, sliced inverse regression, regression analysis, link violation.

of properties and structures of high dimensional data is certainly no less demanding than its empirical counterpart.

In this paper, some properties about the distributions of low dimensional projections from high dimensional data will be studied. Specifically, let $\mathbf{x}$ be a $p$-dimensional random variable with mean zero and the identity covariance:

$$(1.1) \qquad E\mathbf{x} = 0, \quad \operatorname{cov}\mathbf{x} = I.$$

For a direction $\beta$, $\|\beta\| = 1$, and consider the projected variable $\beta'\mathbf{x}$. Find another direction $b$ so that the regression curve of $b'\mathbf{x}$ against $\beta'\mathbf{x}$, $E(b'\mathbf{x}|\beta'\mathbf{x})$, is as nonlinear as possible. When the dimensionality $p$ of $\mathbf{x}$ is large, we show that for most directions $\beta$ even the most nonlinear regression is nearly linear.

To state our result more precisely, consider $E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)$. If the distribution of $\mathbf{x}$ satisfies the condition

$$(1.2) \qquad E(b'\mathbf{x}|\beta, \beta'\mathbf{x} = t) \text{ is linear in } t, \quad \text{for any } b,$$

then the scatterplot of $b'\mathbf{x}$ against $\beta'\mathbf{x}$ would show linear patterns. Equations (1.1) and (1.2) imply

$$(1.3) \qquad E(\mathbf{x}|\beta, \beta'\mathbf{x} = t) = t\beta,$$

or equivalently,

$$(1.4) \qquad \|E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)\|^2 - t^2 = 0.$$

Of course, we do not expect to have (1.4) all the time. The left-hand side, which is always nonnegative, will be used as the measure of deviation from linearity. Given any small number $\varepsilon > 0$, we are interested in the size of the set $\{\beta$: $\|E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)\|^2 - t^2 > \varepsilon$, $\|\beta\| = 1\}$. We shall show that the measure of this set, relative to the entire sphere $\{\beta$: $\|\beta\| = 1\}$, converges to zero as the dimension $p$ tends to infinity. Equivalently, if $\beta$ is treated as a random vector from the uniform distribution on the unit sphere in $R^p$, then in probability

$$(1.5) \qquad \|E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)\|^2 - t^2 \to 0.$$

The conditions on $\mathbf{x}$ (see Theorem 3.2 in subsection 3.2) are discussed in Section 4. We argue that it is unusual to violate any of these conditions without violating the normalization condition (1.1), which can be achieved by an affine transformation.

A generalization of this result, which replaces $\beta'\mathbf{x}$ by $B'\mathbf{x}$ with $B = (\beta_1, \dots, \beta_k)$ for $k$ randomly selected orthonormal vectors, is also valid:

$$(1.6) \qquad \|E(\mathbf{x}|B, B'\mathbf{x} = \mathbf{t})\|^2 - \|\mathbf{t}\|^2 \to 0,$$

where $\mathbf{t}$ is a $k$ dimensional vector.

The results (1.5) and (1.6) are closely related to the work of Diaconis and Freedman (1984) where they show that for most high dimensional data sets, almost all low dimensional projections are nearly normal. In our context, this suggests that with probability approaching 1, the conditional distribution of $B'\mathbf{x}$ given $B$, is asymptotically normal as $p$ tends to infinity. Normal densities have linear conditional expectations. Taking $k = 2$, it is natural to conject that if we choose a pair of orthonormal directions $\beta, b$ at random, then with probability approaching 1,

$$E(b'\mathbf{x}|\beta, \beta'\mathbf{x} = t) \to 0. \tag{1.7}$$

By comparison, result (1.5) is equivalent to

$$\sup_{b \perp \beta, \|b\| = 1} |E(b'\mathbf{x}|\beta, \beta'\mathbf{x} = t)| \to 0. \tag{1.8}$$

Condition (1.8) is much stronger than (1.7). Instead of picking the second direction $b$ at random, what (1.8) considers is the most nonlinear direction. Thus Diaconis and Freedman's result does not provide clues to whether (1.8) [equivalently (1.5)] might be true or false.

Our strategy of establishing (1.5) and (1.6) is rather different from the characteristic function approach employed in Diaconis and Freedman (1984). We seek direct ways of computing the expected value for the aforementioned nonlinearity measure [the left-hand side of (1.4)]. A pair of random vectors, called the *rotational twin*, is constructed in subsection 2.2. The marginal and the joint density functions of the rotational twin help us to find closed form expressions for the crucial terms (2.1), (2.2) and (2.3), which assess average deviation from normality and from linearity for the projected variables in a suitable sense. As a by-product, we can easily estimate these terms from an i.i.d. sample by the method of moments. This is done without going through laborious random projection procedures like grand tours, and without applying any smoothing technique at all.

The impact of our results on several fronts of data analysis is discussed next.

1.1. *Impact on data analysis.* The first impact is on regression analysis. Suppose that in addition to $\mathbf{x}$, we have a variable $y$ and want to regress it against $\mathbf{x}$, using, say, multiple linear regression. Since there is no guarantee that the linear model

$$y = \alpha + \beta'\mathbf{x} + \varepsilon$$

is valid, the result might not always be useful. In fact, one often tries other models, such as Box–Cox transformations, or generalized linear models, to increase the goodness of fit. However, if (1.2) holds then it has been shown that the slope estimate obtained from the multiple linear regression still estimates $\beta$ consistently up to a constant of proportionality, even if the true model is nonlinear:

$$y = g(\beta'\mathbf{x}, \varepsilon),$$

see, for example, Brillinger (1977, 1983), Li and Duan (1989) and the references given there. The same result holds for estimates obtained by most commonly used regression methods based on the minimization of some convex criterion functions [Li and Duan (1989)]. The associated inference procedures for confidence intervals and hypothesis testing also remain valid after some simple modification to accommodate the possibility that the specified link function may be grossly wrong.

When $\mathbf{x}$ is spherically symmetric, (1.2) holds for any $\beta$. But (1.2) is much weaker than spherical symmetry because *only one fixed* $\beta$, although unknown to the statistician, is required there. A practical question is how often we can ensure (1.2) (at least approximately) for many data sets that may have obviously violated the spherical symmetry condition. We can answer this question now via a Bayesian argument.

Suppose that Nature is playing an average game, instead of a minimax game, with the statistician. So the unknown $\beta$, normalized to have length 1, may be viewed as being chosen at random from the unit sphere $S^{p-1}$ in $R^p$. Now, using the result of this paper, we can claim that as long as the dimension of $\mathbf{x}$ is large, the chance is good that (1.2) may approximately hold. Therefore, for many data sets, a blind application of standard regression procedure, without checking (1.2), and without knowing exactly what the link function is, may still yield an approximately correct answer for estimating $\beta$ up to a constant of proportionality.

We should not take this favorable result as a certificate for developing a cavalier attitude toward model checking. To the contrary, it has been well recognized that possible violation of the link function can often be detected by residual plots, the plots of the residuals against the predicted values, $\hat{\beta}'\mathbf{x}$. But if $\hat{\beta}$ deviates substantially from the direction of the true $\beta$, then the residual plot may not be informative enough in suggesting the correct form of the link function. By showing that $\hat{\beta}$ is often proportional to $\beta$, we have in part explained why data analysts can often recover the right structure of the data by a careful study of residual plots.

The role of graphics is certainly not restricted to the final or the intermediate stage of data analysis. It may be even more important at the beginning stage. Quite often, data browsing can help us to rectify the focus of our study and to avoid attacking wrong problems. For example, if the scatterplot of $y$ against $\mathbf{x}$ shows a significant heteroscedastic pattern but no significant trend, then fitting a mean curve to the data might not be as essential as studying the pattern and the size of the conditional variance, $\mathrm{var}(y|\mathbf{x})$.

An important issue of data visualization quickly emerges when the dimension of $\mathbf{x}$ is large. There are so many plots to inspect that without proper statistical guidance about which ones to concentrate on first, one will soon lose patience and may fail to synthesize what has been found.

To address this issue, Li (1991) formulates it as a dimension reduction problem:

$$y = g(\beta_1'\mathbf{x}, \dots, \beta_k'\mathbf{x}, \varepsilon).$$

Here $g$ is an unknown function and $\varepsilon$ is an unknown random error independent of $\mathbf{x}$. The goal is to estimate the *effective dimension reduction* (e.d.r.) space, the space spanned by the $\beta$'s. When $k$ is small, we can effectively reduce the data by projecting $\mathbf{x}$ along the e.d.r. directions for further studying their relationship with $y$. A method, sliced inverse regression (SIR), is proposed for this estimation. While there are several possible variations for implementation, the basic principle is to reverse the roles of $y$ and $\mathbf{x}$. Instead of regressing the univariate $y$ against the multivariate $\mathbf{x}$, the multivariate $\mathbf{x}$ is regressed against the univariate $y$. Estimates based on the first moment $E(\mathbf{x}|y)$ have been studied more extensively [e.g., Carroll and Li (1992), Duan and Li (1991), Hsing and Carroll (1992) and Li (1990, 1991, 1992b)] and estimates based on the second moments were also suggested [Cook and Weisberg (1991) and Li (1990, 1991, 1992a)]. One crucial condition for the success of SIR is the $k$ components version of (1.2):

$$E(b'\mathbf{x}|\beta_1'\mathbf{x}, \dots, \beta_k'\mathbf{x}) \text{ is linear for any } b. \tag{1.9}$$

Using the Bayesian argument again as in the beginning of this section, we can infer that (1.9) is expected to hold approximately for many high dimensional data sets. Thus, without checking (1.9), a blind application of SIR can still be helpful in finding the most informative directions for viewing the data. This is a desirable situation because in many cases where the distribution of $\mathbf{x}$ is not elliptically symmetric, it does not seem possible for us to verify a complicated condition like (1.9) which even involves the unknown directions of main interest. However, a diagnostic check is recommended after using SIR.

Following the spirit of Diaconis and Freedman (1984), we consider severe violation of (1.9) as an important feature of the data, due to its unusualness. We should never ignore this possibility. In the rejoinder of Li (1991), discussion of how to detect this using SIR methodology, and how to resolve the confounding issue in modelling, is commenced.

**2. General strategy.** Let $f_p(\cdot)$ be the density function of $\mathbf{x}$, and $\phi_p(\cdot)$ be the $p$-dimensional standard normal density function. For any $\beta$ with $\|\beta\| = 1$, let $f_{\beta'\mathbf{x}}(t)$ be the density of $\beta'\mathbf{x}$ at $t$. In this section we shall find closed form expressions for the following quantities:

$$\mathbf{A}_1(t) = E\left(\frac{f_{\beta'\mathbf{x}}(t)}{\phi_1(t)} - 1\right)^2, \tag{2.1}$$

$$\mathbf{A}_2(t) = E\big(\|E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)\|^2 - t^2\big)\frac{f_{\beta'\mathbf{x}}^2(t)}{\phi_1^2(t)}, \tag{2.2}$$

$$\mathbf{A}_3(t) = \frac{\mathbf{A}_2(t)}{Ef_{\beta'\mathbf{x}}^2(t)/\phi_1^2(t)}, \tag{2.3}$$

where the expectation in each quantity is taken with respect to the uniform distribution of $\beta$.

Formula (2.1) quantifies the average departure from normality for the distribution of a random one-dimensional projection of $\mathbf{x}$. The result of Diaconis and Freedom (1984) suggests, but does not imply, that $\mathbf{A}_1(t)$ is close to zero for large $p$ under suitable conditions. The closed form expression we shall obtain later makes this suggestion more transparent. It also allows us to construct a simple estimate from a given data set for $\mathbf{A}_1(t)$. We can assess whether the asymptotics have become effective or not.

For describing the departure from linearity for the regression curve $E(\mathbf{x}|\beta, \beta'\mathbf{x})$, a natural quantity would be the simple average

$$E\big(\|E(\mathbf{x}|\beta, \beta'\mathbf{x} = t)\|^2 - t^2\big).$$

We are unable to obtain a simple expression for this quantity. As a substitute, weighted versions given by (2.2) or (2.3) are suggested. Even though $\mathbf{A}_2(t)$ appears to be rather complicated, a closed form expression can be obtained. As explained in Section 1, (2.2) is always nonnegative. We can derive (1.5) by establishing that $\mathbf{A}_2(t)$ converges to zero.

2.1. *Normal density as a base.* It is more convenient to work with a normal density. The following lemma serves as a bridge for relating the conditional expectation taken with respect to $\mathbf{x}$ and that with respect to a normal random vector $\mathbf{z}$. First define the density ratio

$$h_p(\mathbf{x}) = \frac{f_p(\mathbf{x})}{\phi_p(\mathbf{x})}. \tag{2.4}$$

LEMMA 2.1. *With definition* (2.4), *we have*

$$f_{\beta'\mathbf{x}}(t) = \phi_1(t) E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big), \tag{2.5}$$

$$E(\mathbf{x}|\beta, \beta'\mathbf{x} = t) = \frac{E\big(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)}{E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)}, \tag{2.6}$$

*where* $\mathbf{z}$ *follows the p-dimensional standard normal distribution and is independent of* $\beta$.

PROOF. To get (2.5), observe that

$$\begin{aligned} f_{\beta'\mathbf{x}}(t) &= \int_{\beta'\mathbf{x}=t} f_p(\mathbf{x})\, d\mathbf{x} = \int_{\beta'\mathbf{z}=t} h_p(\mathbf{z})\phi_p(\mathbf{z})\, d\mathbf{z} \\ &= \phi_1(t) E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big), \end{aligned}$$

where the last identity is due to the rotational invariance of the normal density. Relation (2.6) can be obtained similarly. This completes the proof. □

Now we can rewrite (2.1)–(2.3) as

$$\mathbf{A}_1(t) = E\big[E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\big]^2 - 2E\big[E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\big] + 1,$$

$$\mathbf{A}_2(t) = E\|E\big(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\|^2 - t^2 E\big[E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\big]^2,$$

$$\mathbf{A}_3(t) = \frac{E\|E\big(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\|^2}{E\big[E\big(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\big)\big]^2} - t^2.$$

To proceed, we need to assess the three terms involving the conditional expectations. One approach we have attempted is to compute them directly by making an Edgeworth type of expansion for the density ratio $h_p(\mathbf{z})$. In fact, this is our original motivation for changing $\mathbf{x}$ to $\mathbf{z}$. However, the formulae obtained appear rather complicated. Eventually we take an alternative approach which attempts to convert the conditional expectations to the unconditional ones. This motivates our construction of the rotational twin as to be discussed next. One of the referees asks if this idea has been applied in other contexts. To the best of our knowledge, we are not aware of any precedent.

2.2. *The rotational twin.* Let $t$ be a fixed real number and $\beta$ be uniformly distributed on $S^{p-1}$. Given $\beta$, let $v_1$ and $v_2$ be independent standard normal random vectors on the orthogonal complement of $\beta$:

$$v_1, v_2|\beta \sim N(0, I - \beta\beta') \times N(0, I - \beta\beta').$$

The rotational twin is defined as

$$\mathbf{w}_1 = t\beta + v_1,$$

$$\mathbf{w}_2 = t\beta + v_2.$$

The distribution of $\mathbf{w}_1, \mathbf{w}_2$ is given below.

LEMMA 2.2. *The law of* $\mathbf{w}_1$ *is spherically symmetric, with* $\|\mathbf{w}_1\|^2$ *distributed as* $t^2 + \chi^2_{p-1}$, *where* $\chi^2_{p-1}$ *denotes a chi-squared distribution with* $p - 1$ *degrees of freedom.*

PROOF. The spherical symmetry of $\mathbf{w}_1$ is obvious because of the rotational invariance of $\beta$ and $v_1$. The orthogonality between $\beta$ and $v_1$ implies $\|\mathbf{w}_1\|^2 = t^2 + \|v_1\|^2$. The proof is completed upon observing that $\|v_1\|^2$ follows a chi-squared distribution with $p - 1$ degrees of freedom. □

LEMMA 2.3. *Conditional on* $\mathbf{w}_1$, *the distribution of* $\mathbf{w}_2$ *can be described as follows*:

(i) *Decompose* $\mathbf{w}_2$ *into two parts*:

$$\mathbf{w}_2 = k\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} + \mathbf{w}_3,$$

*where* $\mathbf{w}_3'\mathbf{w}_1 = 0$.

(ii) *The distribution of* $k$ *is normal with mean* $t^2/\|\mathbf{w}_1\|$ *and variance* $(\|\mathbf{w}_1\|^2 - t^2)/\|\mathbf{w}_1\|^2$.

(iii) *Conditional on* $k$, $\mathbf{w}_3$ *is spherically symmetric on the orthogonal complement of* $\mathbf{w}_1$, *with* $\|\mathbf{w}_3\|^2$ *being distributed as*

$$\chi^2_{p-2} + \frac{t^2(\|\mathbf{w}_1\| - k)^2}{\|\mathbf{w}_1\|^2 - t^2}.$$

PROOF. First decompose $t\beta$ into two orthogonal parts:

$$t\beta = \frac{t^2}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + |t|\sqrt{1 - \frac{t^2}{\|\mathbf{w}_1\|^2}}\,\mathbf{e}, \tag{2.7}$$

where $\mathbf{e}$ is a unit vector orthogonal to $\mathbf{w}_1$. By an invariance argument we see that conditional on $\mathbf{w}_1$, $\mathbf{e}$ is uniformly distributed. Then decompose $v_2$ into two orthogonal parts:

$$v_2 = v_3 + z\frac{v_1}{\|v_1\|},$$

where $z$, the projection on $v_1$, follows the standard normal distribution, and $v_3$ is normal on the orthogonal complement of the space spanned by $v_1$ and $t\beta$. Now we can write $\mathbf{w}_2$ as

$$\begin{aligned}
\mathbf{w}_2 &= t\beta + z\frac{v_1}{\|v_1\|} + v_3\\
&= t\beta + z\frac{\mathbf{w}_1 - t\beta}{\sqrt{\|\mathbf{w}_1\|^2 - t^2}} + v_3\\
&= \left(\frac{t^2}{\|\mathbf{w}_1\|} + z\sqrt{1 - \frac{t^2}{\|\mathbf{w}_1\|^2}}\right)\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} + \left(t\sqrt{1 - \frac{t^2}{\|\mathbf{w}_1\|^2}} - z\frac{t}{\|\mathbf{w}_1\|}\right)\mathbf{e} + v_3.
\end{aligned}$$

It follows that

$$k = \frac{t^2}{\|\mathbf{w}_1\|} + z\sqrt{1 - \frac{t^2}{\|\mathbf{w}_1\|^2}}\,. \tag{2.8}$$

This verifies (ii).

Conditioning on $k$ does not affect the distribution of $\mathbf{e}$, which is still uniform on a $p - 1$ dimensional sphere, perpendicular to $\mathbf{w}_1$. Neither does it affect the distribution of $v_3$, which is still normal on the orthogonal complement of $\mathbf{e}$ and $\mathbf{w}_1$. Furthermore, using (2.8) we can derive

$$\mathbf{w}_3 = \left(t\frac{\|\mathbf{w}_1\| - k}{\sqrt{\|\mathbf{w}_1\|^2 - t^2}}\right)\mathbf{e} + v_3.$$

Following the same argument as in the proof of Lemma 2.2 we can obtain (iii). This completes the proof of Lemma 2.3. □

After a straightforward manipulation as outlined in Appendix A, we can obtain the density function of $\mathbf{w}_1, \mathbf{w}_2$.

COROLLARY 2.1. *Let $g_p(\mathbf{w}_1|t)$ be the density of $\mathbf{w}_1$ and $g_p(\mathbf{w}_1, \mathbf{w}_2|t)$ be the joint density of $\mathbf{w}_1, \mathbf{w}_2$. Then, the ratios of these density functions to the* *corresponding normal density functions,*

$$Q_p(\mathbf{w}_1|t) = g_p(\mathbf{w}_1|t)/\phi_p(\mathbf{w}_1),$$

$$Q_p(\mathbf{w}_1, \mathbf{w}_2|t) = g_p(\mathbf{w}_1, \mathbf{w}_2|t)/\phi_p(\mathbf{w}_1)\phi_p(\mathbf{w}_2),$$

*are given by*

$$Q_p(\mathbf{w}_1|t) = \frac{\sqrt{2}\,\Gamma(p/2)}{\Gamma((p-1)/2)} \cdot \frac{1}{\|\mathbf{w}_1\|}\left(1 - \frac{t^2}{\|\mathbf{w}_1\|^2}\right)^{(p-3)/2} e^{t^2/2}, \tag{2.9}$$

$$Q_p(\mathbf{w}_1, \mathbf{w}_2|t) = \frac{p-2}{\|\mathbf{w}_1\|\,\|\mathbf{w}_2\|\,|\sin\theta|}\left(1 - \frac{t^2\|\mathbf{w}_1 - \mathbf{w}_2\|^2}{\|\mathbf{w}_1\|^2\|\mathbf{w}_2\|^2\sin^2\theta}\right)^{(p-4)/2} e^{t^2}, \tag{2.10}$$

*where*

$$\theta = \textit{angle between } \mathbf{w}_1 \textit{ and } \mathbf{w}_2,$$

*for the region satisfying the constraints*

$$\|\mathbf{w}_1\|^2, \|\mathbf{w}_2\|^2 > t^2,$$

$$t^2\|\mathbf{w}_1 - \mathbf{w}_2\|^2 \le \|\mathbf{w}_1\|^2\|\mathbf{w}_2\|^2\sin^2\theta;$$

*elsewhere, the ratios are zero.*

Now we have the key theorem for our approach.

THEOREM 2.1. *Let* $\mathbf{x}_1, \mathbf{x}_2$ *be independent random variables with the same density function,* $f_p(\mathbf{x})$, *as* $\mathbf{x}$. *Then we have the following identities:*

$$\mathbf{A}_1(t) = EQ_p(\mathbf{x}_1, \mathbf{x}_2|t) - 2EQ_p(\mathbf{x}_1|t) + 1, \tag{2.11}$$

$$\mathbf{A}_2(t) = E\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t) - t^2 EQ_p(\mathbf{x}_1, \mathbf{x}_2|t), \tag{2.12}$$

$$\mathbf{A}_3(t) = \frac{E\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t)}{EQ_p(\mathbf{x}_1, \mathbf{x}_2|t)} - t^2. \tag{2.13}$$

PROOF. From the expressions given in the end of subsection 2.1, we need to show that

$$E\left[E\left(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\right)\right] = EQ_p(\mathbf{x}|t), \tag{2.14}$$

$$E\left[E\left(h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\right)\right]^2 = EQ_p(\mathbf{x}_1, \mathbf{x}_2|t), \tag{2.15}$$

$$E\|E\left(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t\right)\|^2 = E\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t). \tag{2.16}$$

To obtain (2.14), write the left-hand side as

$$E\big(E(h_p(t\beta + v_1)|\beta)\big) = Eh_p(\mathbf{w}_1) = \int \frac{f_p(\mathbf{w}_1)}{\phi_p(\mathbf{w}_1)} g_p(\mathbf{w}_1|t)\, d\mathbf{w}_1$$

$$= \int f_p(\mathbf{x}) \frac{g_p(\mathbf{x}|t)}{\phi_p(\mathbf{x})}\, d\mathbf{x} = EQ_p(\mathbf{x}|t).$$

To obtain (2.15), write the left-hand side as

$$\begin{aligned} &E\big[E(h_p(t\beta + v_1)|\beta) \cdot E(h_p(t\beta + v_2)|\beta)\big] \\ &\quad = E\big[E(h_p(\mathbf{w}_1)|\beta)E(h_p(\mathbf{w}_2)|\beta)\big] \\ &\quad = E\big[E(h_p(\mathbf{w}_1)h_p(\mathbf{w}_2)|\beta)\big] \\ &\quad = Eh_p(\mathbf{w}_1)h_p(\mathbf{w}_2) = \int \frac{f_p(\mathbf{w}_1)}{\phi_p(\mathbf{w}_1)} \frac{f_p(\mathbf{w}_2)}{\phi_p(\mathbf{w}_2)} g_p(\mathbf{w}_1|t) g_p(\mathbf{w}_2|t)\, d\mathbf{w}_1\, d\mathbf{w}_2 \\ &\quad = EQ_p(\mathbf{x}_1, \mathbf{x}_2). \end{aligned}$$

To derive (2.16), we write the left-side in terms of the trace operator:

$$\begin{aligned} &E\|E(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t)\|^2 \\ &\qquad = \operatorname{trace}\Big(E\big[E(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t)E(\mathbf{z}h_p(\mathbf{z})|\beta, \beta'\mathbf{z} = t)'\big]\Big) \end{aligned}$$

and then carry out the same argument as before to complete the proof. □

We can easily estimate the quantities (2.14)–(2.16) from the data. Specifically, suppose that $\mathbf{x}_1, \dots, \mathbf{x}_n$ is an i.i.d sample of size $n$ from the unknown density function $f_p(\mathbf{x})$. Then we can construct estimates by

$$\hat{E}Q_p(\mathbf{x}|t) = n^{-1} \sum_{i=1}^{n} Q_p(\mathbf{x}_i|t),$$

$$\hat{E}Q_p(\mathbf{x}_1, \mathbf{x}_2|t) = (n^2 - n)^{-1} \sum_{i \neq j} Q_p(\mathbf{x}_i, \mathbf{x}_j|t),$$

$$\hat{E}\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t) = (n^2 - n)^{-1} \sum_{i \neq j} \mathbf{x}_i'\mathbf{x}_j Q_p(\mathbf{x}_i, \mathbf{x}_j|t).$$

This leads to useful estimates of $\mathbf{A}_1(t), \mathbf{A}_2(t), \mathbf{A}_3(t)$, and does not involve going through laborious projections using tools like grand tours. Neither does it require smoothing.

**3. Conditions on the density of x.** In this section we shall find conditions on $f_p(\mathbf{x})$, the density function of $\mathbf{x}$, so that

$$E\mathbf{A}_1(t) \to 0, \tag{3.1}$$

$$E\mathbf{A}_2(t) \to 0. \tag{3.2}$$

By Markov's theorem, (3.1) and (3.2) imply (1.5). According to Theorem 2.1, it suffices to show that

$$EQ_p(\mathbf{x}|t) \to 1, \tag{3.3}$$

$$EQ_p(\mathbf{x}_1, \mathbf{x}_2|t) \to 1, \tag{3.4}$$

$$E\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t) \to t^2. \tag{3.5}$$

Before going into the detailed discussion (see subsection 3.2), we take a quick look at the rotational-twin functions, $Q_p(\cdot|t)$ and $Q_p(\cdot,\cdot|t)$. First, by Stirling's formula, $\Gamma(p/2)/\Gamma((p-1)/2)$ is $\sqrt{p/2} + o(1)$. Thus if

$$\frac{\|\mathbf{x}\|^2}{p} \to 1 \quad \text{in probability,} \tag{3.6}$$

then

$$Q_p(\mathbf{x}|t) \to 1 \quad \text{in probability.} \tag{3.7}$$

In addition to (3.6), if

$$\cos\theta \to 0 \quad \text{in probability,} \tag{3.8}$$

then

$$Q_p(\mathbf{x}_1, \mathbf{x}_2|t) \to 1 \quad \text{in probability.} \tag{3.9}$$

With additional regularity conditions, convergence in probability can imply convergence in expectation. This resolves (3.3) and (3.4). In subsection 3.1, we shall further argue that in general (3.8) is a natural consequence of (3.6).

Formula (3.5) is harder to derive. We need to find the dominating term for

$$\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t).$$

By Taylor expansion as outlined in Appendix B, we can derive the following.

LEMMA 3.1. *Under* (3.6) *and* (3.8),

$$\mathbf{x}_1'\mathbf{x}_2 Q(\mathbf{x}_1, \mathbf{x}_2|t) = p\cos\theta\left(1 + t^2\left(\cos\theta + \frac{1}{2}\left(2 - \frac{p}{\|\mathbf{x}_1\|^2} - \frac{p}{\|\mathbf{x}_2\|^2}\right)\right)\right.$$
$$\left.\times\big(1 + o_p(1)\big) + o_p(|\cos\theta|)\right) + O_p(p^{-1}). \tag{3.10}$$

In subsection 3.1 below we shall examine the angle $\theta$ more closely. As it turns out, in general the leading term in (3.10) has its expected value converging to $t^2$.

All conditions on $f_p(\mathbf{x})$ needed for establishing (3.1) and (3.2) will be put together in subsection 3.2.

3.1. *The angle $\theta$.* In this section, (3.10) will be further analyzed in order to find the expectation of the leading term.

The following lemma is instructive for this purpose.

LEMMA 3.2. *Under* (1.1) *and the condition that*

$$\|\mathbf{x}\|^2 = p, \tag{3.11}$$

*we have*

$$E\cos\theta = 0, \tag{3.12a}$$

$$E\cos^2\theta = p^{-1}. \tag{3.12b}$$

PROOF. Observe that $\cos\theta = \mathbf{x}_1'\mathbf{x}_2/(\|\mathbf{x}_1\|\,\|\mathbf{x}_2\|) = \mathbf{x}_1'\mathbf{x}_2/p$. Then (3.12a) follows from $E\mathbf{x} = 0$. Furthermore,

$$\begin{aligned} E\cos^2\theta &= p^{-2}E\mathbf{x}_1'\mathbf{x}_2\mathbf{x}_2'\mathbf{x}_1 \\ &= p^{-2}E\mathbf{x}_1'I\mathbf{x}_1 \\ &= p^{-2}\cdot p = p^{-1}, \end{aligned}$$

where we have used $\operatorname{cov}\mathbf{x} = I$ to obtain the second identity. This completes the proof. □

This lemma shows that as $p$ tends to infinity, $\cos\theta$ converges to zero at the rate of root $p$. Hence if two independent replicates are from a density satisfying (1.1) and (3.11), they should be almost perpendicular to each other. By this lemma, we can see that the leading term in (3.10) has expectation equal to $t^2$, as desired. But (3.11) is of course much stronger than (3.6).

To loosen condition (3.11), we may consider the Taylor expansion:

$$\begin{aligned} \cos\theta &= p^{-1}\mathbf{x}_1'\mathbf{x}_2\left(1 - \left(\frac{\|\mathbf{x}_1\|}{\sqrt{p}} - 1\right)(1 + o_p(1))\right) \\ &\quad\times\left(1 - \left(\frac{\|\mathbf{x}_2\|}{\sqrt{p}} - 1\right)(1 + o_p(1))\right). \end{aligned} \tag{3.13}$$

In order to ignore the $o_p(1)$ terms, we further assume that

$$\frac{\|\mathbf{x}\|^2}{p} = 1 + o_p(p^{-1/4}). \tag{3.14}$$

Continuing the expansion in (3.10), we get

$$\begin{aligned} &\mathbf{x}_1'\mathbf{x}_2 Q(\mathbf{x}_1, \mathbf{x}_2 | t) \\ &\quad = \mathbf{x}_1'\mathbf{x}_2\left(1 + t^2\left(\frac{\mathbf{x}_1'\mathbf{x}_2}{p} - \frac{1}{2}\left(\frac{\|\mathbf{x}_1\|}{\sqrt{p}} + \frac{\|\mathbf{x}_2\|}{\sqrt{p}} - 2\right)\right) + o_p(p^{-1/2})\right). \end{aligned} \tag{3.15}$$

The leading term is seen to have mean $t^2$, as desired.

LEMMA 3.3. *Under* (1.1) *and* (3.14), *the leading term in* $\mathbf{x}_1'\mathbf{x}_2 Q(\mathbf{x}_1, \mathbf{x}_2 | t)$ [*namely the right-hand side in* (3.15) *without* $o_p(p^{-1/2})$] *has expected value equal to* $t^2$.

Another way to analyze (3.10) will be given below, based on a more detailed study of $\cos\theta$. This will lead to a better result in the next section. All proofs will be given in Appendix B.

First, we provide a bound on $\cos\theta$.

LEMMA 3.4. *Under* (1.1), *the angle* $\theta$ *between two i.i.d. random vectors* $\mathbf{x}_1, \mathbf{x}_2$ *generated from the distribution of* $\mathbf{x}$, *satisfies the following inequalities*:

$$pE\cos\theta = \left\| E\frac{\mathbf{x}}{\|\mathbf{x}\|} \right\|^2 \le \left[ \operatorname{Var}\left( \frac{\sqrt{p}}{\|\mathbf{x}\|} \right) \right]^{1/2}, \tag{3.16}$$

$$\frac{1}{p} \le E\cos^2\theta = \operatorname{trace}\left[ \left( E\frac{\mathbf{x}\mathbf{x}'}{\|\mathbf{x}\|^2} \right)^2 \right] \le \Lambda_p E\|\mathbf{x}\|^{-4}, \tag{3.17}$$

$$\Lambda_p = \sup_{b_i} \sum_{i=1}^{p} E(b_i'\mathbf{x})^4,$$

*where the supremum is taken over all possible choices of the orthonormal basis* $\{b_i,\ i = 1, \dots, p\}$.

Typically the magnitude of $E\|\mathbf{x}\|^{-4}$ is $p^{-2}$ and the magnitude of $\Lambda_p$ is in the order $p$ since $Eb_i'\mathbf{x}^2 = 1$. This helps explain why $\cos\theta$ tends to zero.

Now define

$$B_1(c) = \left\{ \mathbf{x}: \frac{\|\mathbf{x}\|^2}{p} \le 1 - c \right\} \tag{3.18}$$

and consider the following conditions:

$$E\frac{p}{\|\mathbf{x}\|^2} 1(B_1(c)) = o(1), \text{ for some } c,\ 1 > c > 0, \tag{3.19}$$

$$E\frac{p}{\|\mathbf{x}\|^2} 1(B_1(c)) = O(p^{-1/2}), \quad \text{for some } c,\ 1 > c > 0, \tag{3.20}$$

$$P\left\{ \left| \frac{\|\mathbf{x}\|^2}{p} - 1 \right| \ge c \right\} = o\left( \frac{1}{p} \right), \text{ for each } c,\ 1 > c > 0, \tag{3.21}$$

where $1(\cdot)$ is the indicator function.

LEMMA 3.5. *Assume* (1.1). *Under* (3.6) *and* (3.19), *we have*

$$pE\cos\theta = o(1). \tag{3.22}$$

*Under* (3.21), *we have*

$$pE\cos^2\theta \to 1. \tag{3.23}$$

*Under* (3.20), *we have*

$$E\left(1 - \frac{p}{\|\mathbf{x}_1\|^2}\right)\cos\theta = o(p^{-1}). \tag{3.24}$$

COROLLARY 3.1. *Assume* (3.20) *and* (3.21). *Then the expectation for the leading term in* $\mathbf{x}_1'\mathbf{x}_2 Q(\mathbf{x}_1, \mathbf{x}_2|t)$ *converges to* $t^2$:

$$Ep\cos\theta\left(1 + t^2\left(\cos\theta + \frac{1}{2}\left(2 - \frac{p}{\|\mathbf{x}_1\|^2} - \frac{p}{\|\mathbf{x}_2\|^2}\right)\right)\right) \to t^2.$$

3.2. *Convergence in mean.* In the discussion up to now we have proved the convergence in probability, (3.7) and (3.9). We also have shown that the leading term in $\mathbf{x}_1'\mathbf{x}_2 Q(\mathbf{x}_1, \mathbf{x}_2|t)$ has an expected value converging to $t^2$. Since the convergence in mean is usually implied by the convergence in probability, with some additional conditions so that we can interchange the limit with the expectation, we are able to derive (3.3)–(3.5), which implies (1.5).

For any small positive $c$, define

$$B_2(c) = \{\theta: |\sin\theta| \le c\}. \tag{3.25}$$

THEOREM 3.1. *Under* (1.1), *assume that*

$$E\frac{p}{\|\mathbf{x}\|^2} = O(1), \tag{3.26}$$

$$E\frac{1}{\sin^2\theta} = O(1), \tag{3.27}$$

*and that for any c, the left-hand side of* (3.21) *is of order* $o(p^{-1/2})$. *Then we have*

$$\left|\frac{f_{\beta'\mathbf{x}}(t)}{\phi_1(t)} - 1\right| \to 0 \quad \text{in probability}$$

*and*

$$\int_{-M}^{M}\left(\frac{f_{\beta'\mathbf{x}}(t)}{\phi_1(t)} - 1\right)^2 dt \to 0 \quad \text{in probability for any } M > 0.$$

THEOREM 3.2. *Under* (1.1), *assume that for any c*, (3.21) *holds, and that for some* $c > 0$,

$$E\frac{p}{\|\mathbf{x}\|^2}1(B_1(c)) = o(p^{-1}), \tag{3.28}$$

$$E\frac{1}{|\sin\theta|}1(B_2(c)) = o(p^{-1}). \tag{3.29}$$

*Then* (1.5) *holds. Furthermore, for any positive M,*

$$\int_{-M}^{M}\left(\|E(\mathbf{x}|\beta,\beta'\mathbf{x}=t)\|^2-t^2\right)^{1/2} f_{\beta'\mathbf{x}}(t)\,dt\to 0.$$

**4. Discussion.** In this section we shall argue that the conditions needed in Theorems 3.1 and 3.2 are very mild. Cases that violate any one of these conditions without violating condition (1.1) should be regarded as unusual.

4.1. *Conditions* (3.21) *and* (3.28). Note that (3.21) and (3.28) regulate the tails (left or right) of the norm of the random vector $\mathbf{x}$. The normalization condition (1.1) implies $E\|\mathbf{x}\|^2 = p$. Hence by various forms of the law of large numbers and the central limit theorem, for independent as well as dependent coordinate cases, (3.21) and (3.28) can be expected to hold in general. For data analysis, if the true density violates one of these conditions, we may expect to find either outliers or a cluster of points piling up around the mean. If the data points have been standardized to have zero mean with identity covariance, then plotting the histogram for the radius of the data points can help assess if (3.21) and (3.28) might be violated.

EXAMPLE 4.1. Suppose that $\mathbf{x} = (x_1, \dots, x_p)'$, where the coordinates $x_i$'s are independent with bounded eighth moment. Then by Chebyshev's inequality and bounds of moments [e.g., Whittle (1960)], we see that (3.21) holds.

EXAMPLE 4.2. Condition (3.28) holds for the normal density. In fact, from the density of the chi-squared distribution we can derive the stronger result that for any $1 > c > 0$, there exists a small $c_1 > 0$ such that

$$E\left(\frac{p}{\|\mathbf{z}\|^2}\right)^2 1(B_1(c)) = O(c_1^p). \tag{4.1}$$

Now suppose that the density function $f_p(\mathbf{x})$ of $\mathbf{x}$ has finite information (with respect to the normal density),

$$I_p = \int \frac{f_p(\mathbf{x})^2}{\phi_p(\mathbf{x})}\,d\mathbf{x} = Eh_p^2(\mathbf{z}).$$

By Chebyshev's inequality, we can deduce that if

$$I_p = O(c_2^p) \quad \text{for some } c_2 > 1, \tag{4.2}$$

then (3.28) is satisfied for some $c$ such that the corresponding $c_1$ in (4.1) is smaller than $c_2^{-1}$:

$$\begin{aligned} E\frac{p}{\|\mathbf{x}\|^2}1(B_1(c)) &= E\frac{p}{\|\mathbf{z}\|^2}h_p(\mathbf{z})1(B_1(c)) \\ &\le \left(Eh_p^2(\mathbf{z})\right)^{1/2}\left(E\left(\frac{p}{\|\mathbf{z}\|^2}\right)^2 1(B_1(c))\right)^{1/2} \\ &= \left(I_p O(c_1^p)\right)^{1/2} = \left(O(c_2 c_1^p)\right)^{1/2}. \end{aligned}$$

REMARK 4.1. We can use the same argument as in Example 4.2 to show that if $I_p$ is of polynomial order, then (3.21) holds. The difficulty in handling the exponential information arises because we want (3.21) to hold for "each" $c$, instead of "some" $c$.

4.2. *Condition* (3.29). This condition concerns the distribution of the directional part $\mathbf{r} = \mathbf{x}/\|\mathbf{x}\|$ of $\mathbf{x}$. Let $\gamma(\cdot)$ be the density function of $\mathbf{r}$ with respect to the uniform distribution on $S^{p-1}$. As we have argued in subsection 3.1, the event $B_2(c)$ would occur only with an asymptotically negligible probability. Thus violating (3.29) would mean that the density of $\sin\theta$ in the neighborhood of $\theta = 0$ decayed too slowly. In the following we shall argue that this would not happen unless the density function $\gamma(\mathbf{r})$ is too spiky.

Let $\tilde{\gamma}_c(\mathbf{r})$ be an upper envelope of $\gamma(\mathbf{r})$:

$$\tilde{\gamma}_c(\mathbf{r}) = \sup_{|\sin(\mathbf{r},\mathbf{e})| \le c} \gamma(\mathbf{e}),$$

where $\sin(\mathbf{r},\mathbf{e})$ denotes the sine of the angle between $\mathbf{r}$ and $\mathbf{e}$. This quantity gives the peak value of $\gamma(\cdot)$ in a neighborhood of $\mathbf{r}$. The average $E\tilde{\gamma}_c(\mathbf{r})$ measures how spiky the density function $\gamma(\cdot)$ is on the average.

First we shall find an upper bound for the cumulative distribution, $F_p(c) = P\{|\sin\theta| \le c\}$, in the neighborhood of zero.

Now let $\mathbf{e}_1, \mathbf{e}_2$ be two independent random vectors generated from the uniform distribution on $S^{p-1}$. Define the indicator function $\delta_c(\mathbf{e}_1, \mathbf{e}_2) = 1$ or 0, depending on whether the absolute value of the sine of the angle between $\mathbf{e}_1$ and $\mathbf{e}_2$ is less than $c$ or not. An upper bound for $F_p(c)$ is given by

$$\begin{aligned} F_p(c) &= E\gamma(\mathbf{e}_1)\gamma(\mathbf{e}_2)\delta_c(\mathbf{e}_1,\mathbf{e}_2) \\ &= E\big[\gamma(\mathbf{e}_1)E(\gamma(\mathbf{e}_2)\delta_c(\mathbf{e}_1,\mathbf{e}_2)|\mathbf{e}_1)\big] \\ &\le E\big[\gamma(\mathbf{e}_1)\cdot\tilde{\gamma}_c(\mathbf{e}_1)E\delta_t(\mathbf{e}_1,\mathbf{e}_2)\big] \\ &= E\tilde{\gamma}_c(\mathbf{r})\cdot E\delta_c(\mathbf{e}_1,\mathbf{e}_2). \end{aligned} \tag{4.3}$$

Furthermore, we can show (see Appendix D) that

$$E\delta_c(\mathbf{e}_1,\mathbf{e}_2) \le \sqrt{\frac{2}{\pi}}\,p^{-1/2}c^{p-1}(1+o(1)). \tag{4.4}$$

From this we obtain an approximate upper bound for the left-hand side of (3.29):

$$\begin{aligned} E\frac{1}{|\sin\theta|}B_2(c) &= \int_0^c t^{-1}\,dF_p(t) \\ &= c^{-1}F_p(c) + \int_0^c t^{-2}F_p(t)\,dt \\ &\le c^{-1}F_p(c) + E\tilde{\gamma}_c(\mathbf{r})\int_0^c t^{-2}E\delta_t(\mathbf{e}_1,\mathbf{e}_2)\,dt \\ &\le E\tilde{\gamma}_c(\mathbf{r})\sqrt{\frac{2}{\pi}}\,p^{-1/2}c^{p-2}(1+o(1)). \end{aligned}$$

Putting these together, we have:

LEMMA 4.1. *If*

$$E\tilde{\gamma}_c(\mathbf{r}) = o(p^{-1/2}c^{-p})$$

*for some small value* $c \le 1$, *then* (3.29) *holds.*

Thus, unless the density for the directional part in $\mathbf{x}$ is too spiky, we typically would expect (3.29) to hold.

REMARK 4.2. The normal density clearly satisfies (3.29). Furthermore, we can show that

$$E\frac{1}{\sin^2\theta}1(B_2(c)) = O(c_1^p)$$

for some $c_1$, $1 > c_1 > 0$. Now we can use the same argument as in Example 4.2 to prove that if (4.2) holds then (3.29) also holds.

**5. Extension.** We can extend our results to the more general situation (1.6).

First the result in subsection 2.1 needs only very minor changes. Lemma 2.1 is still valid if $\beta$ is replaced by $B$ and $\phi_1(t)$ by $\phi_k(\mathbf{t})$. Next, in subsection 2.2, we can define the rotational twin by

$$\mathbf{w}_1 = B\mathbf{t} + v_1,$$
$$\mathbf{w}_2 = B\mathbf{t} + v_2,$$

where

$$v_1, v_2|B \sim N(0, I - BB') \times N(0, I - BB').$$

Lemma 2.2 is still valid if we change $\chi^2_{p-1}$ to $\chi^2_{p-k}$. Similarly, in Lemma 2.3, we need to change $t^2$ to $\|\mathbf{t}\|^2$, and $\chi^2_{p-2}$ to $\chi^2_{p-k-1}$. In the proof of Lemma 2.3, the key decomposition (2.7) still holds, and $\mathbf{e}$ is still uniform on the unit sphere orthogonal to $\mathbf{w}_1$. For Corollary 2.1, we can replace (2.9) and (2.10) by

$$(2.9') \qquad Q_p(\mathbf{w}_1|\mathbf{t}) = \frac{2^{k/2}\Gamma(p/2)}{\Gamma((p-k)/2)}\cdot\frac{1}{\|\mathbf{w}_1\|^k}\left(1-\frac{\|\mathbf{t}\|^2}{\|\mathbf{w}_1\|^2}\right)^{(p-k-2)/2} e^{\|\mathbf{t}\|^2/2},$$

$$(2.10') \qquad \begin{aligned} Q_p(\mathbf{w}_1, \mathbf{w}_2|\mathbf{t}) = {}& \frac{2^k\Gamma(p/2)}{\Gamma((p-k)/2)}\cdot\frac{\Gamma((p-1)/2)}{\Gamma((p-k-1)/2)} \\ &\times\frac{1}{\|\mathbf{w}_1\|^k\|\mathbf{w}_2\|^k|\sin\theta|^k} \\ &\times\left(1-\frac{\|\mathbf{t}\|^2\|\mathbf{w}_1-\mathbf{w}_2\|^2}{\|\mathbf{w}_1\|^2\|\mathbf{w}_2\|^2\sin^2\theta}\right)^{(p-k-3)/2} e^{\|\mathbf{t}\|^2}. \end{aligned}$$

These changes do not alter the main argument in Section 3. Theorems 3.1 and 3.2 can be generalized by substituting $B$ for $\beta$. Therefore, (1.6) is seen to hold in general.

## APPENDIX A

**Proof of Corollary 2.1.** Result (2.9) follows from Lemma 2.2. To obtain (2.10), we first use Lemma 2.3 to write $Q_p(\mathbf{w}_1, \mathbf{w}_2|t)$ as the product of $Q_p(\mathbf{w}_1|t)$, the density of $k$ given $\mathbf{w}_1$ over $\phi_1(k)$, and the density of $\|\mathbf{w}_3\|^2$ given $\mathbf{w}_1$ and $k$ over a chi-squared density with $p - 2$ degrees of freedom:

$$\frac{\sqrt{2}\,\Gamma(p/2)}{\Gamma((p-1)/2)} \cdot \frac{1}{\|\mathbf{w}_1\|}\left(1 - \frac{t^2}{\|\mathbf{w}_1\|^2}\right)^{(p-3)/2} e^{t^2/2}$$

$$\times \sqrt{\frac{\|\mathbf{w}_1\|^2}{\|\mathbf{w}_1\|^2 - t^2}}\, \exp\left\{-\frac{1}{2}\frac{\|\mathbf{w}_1\|^2}{\|\mathbf{w}_1\|^2 - t^2}\left(k - \frac{t^2}{\|\mathbf{w}_1\|}\right)^2 + \frac{k^2}{2}\right\}$$

$$\times \frac{\sqrt{2}\,\Gamma((p-1)/2)}{\Gamma((p-2)/2)} \frac{1}{\|\mathbf{w}_3\|}\left(1 - \frac{l^2}{\|\mathbf{w}_3\|^2}\right)^{(p-4)/2} e^{l^2/2},$$

where $l = t^2(\|\mathbf{w}_1\| - k)^2/(\|\mathbf{w}_1\|^2 - t^2)$. Note that $\|\mathbf{w}_3\| = \|\mathbf{w}_2\|\,|\sin\theta|$. After simplification, we can obtain (2.10).

## APPENDIX B

### Proofs for subsection 3.1.

Proof of Lemma 3.1. First use Taylor expansion to get

$$|\sin\theta|^{-1} = 1 + O_p(\cos\theta^2).$$

Similarly,

$$\exp\left\{t^2 + \frac{p-4}{2}\log\left(1 - \frac{t^2\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\|\mathbf{x}_1\|^2\|\mathbf{x}_2\|^2\sin^2\theta}\right)\right\}$$

$$= \exp\left\{t^2 - t^2\frac{p}{2}\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\|\mathbf{x}_1\|^2\|\mathbf{x}_2\|^2} + O_p(p^{-1}) + O_p(\cos^2\theta)\right\}$$

$$= 1 + t^2\left(1 - \frac{p}{2}\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\|\mathbf{x}_1\|^2\|\mathbf{x}_2\|^2}\right)(1 + o_p(1)) + O_p(p^{-1}) + O_p(\cos^2\theta).$$

Further expansion of the leading term in the preceding expression gives

$$1 + t^2\left(1 - \frac{p}{2}\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\|\mathbf{x}_1\|^2\|\mathbf{x}_2\|^2}\right) = 1 + t^2\left(1 - \frac{1}{2}\left(\frac{p}{\|\mathbf{x}_1\|^2} + \frac{p}{\|\mathbf{x}_2\|^2}\right) + \frac{p}{\|\mathbf{x}_1\|\,\|\mathbf{x}_2\|}\cos\theta\right)$$

$$= 1 + t^2\left(1 - \frac{1}{2}\left(\frac{p}{\|\mathbf{x}_1\|^2} + \frac{p}{\|\mathbf{x}_2\|^2}\right) + \cos\theta\right)$$

$$+ o_p(\cos\theta).$$

This leads to Lemma 3.1. □

PROOF OF LEMMA 3.4.

$$pE\cos\theta = pE\frac{\mathbf{x}_1'\mathbf{x}_2}{\|\mathbf{x}_1\|\,\|\mathbf{x}_2\|}$$

$$= \left\|E\frac{\mathbf{x}}{\|\mathbf{x}\|/\sqrt{p}}\right\|^2$$

$$= \left\|E\mathbf{x}\left(\frac{1}{\|\mathbf{x}\|/\sqrt{p}} - 1\right)\right\|^2$$

$$= \sup_{\|b\|=1} E(b'\mathbf{x})\left(\frac{1}{\|\mathbf{x}\|/\sqrt{p}} - 1\right)$$

$$\le \left[\operatorname{var}\left(\frac{\sqrt{p}}{\|\mathbf{x}\|}\right)\right]^{1/2}.$$

This derives (3.16). Next, consider

$$E\cos^2\theta = E\frac{\mathbf{x}_1'\mathbf{x}_2\mathbf{x}_2'\mathbf{x}_1}{\|\mathbf{x}_1\|^2\|\mathbf{x}_2\|^2} = \operatorname{trace}\left[\left(E\frac{\mathbf{x}\mathbf{x}'}{\|\mathbf{x}\|^2}\right)^2\right].$$

Now take $b_1, \dots, b_p$ as the eigenvectors of the matrix $E\mathbf{x}\mathbf{x}'/\|\mathbf{x}\|^2$. We see that the last term in the above expression is equal to $\Sigma(E(b_i'\mathbf{x})^2/\|\mathbf{x}\|^2)^2$, which is no greater than

$$\sum E(b_i'\mathbf{x})^4 E\frac{1}{\|\mathbf{x}\|^4} \le \Lambda_p E\frac{1}{\|\mathbf{x}\|^4}.$$

This shows (3.17). □

PROOF OF LEMMA 3.5. Conditions (3.6) and (3.19) imply that $\operatorname{var}(\sqrt{p}/\|\mathbf{x}\|)$ converges to zero. Hence (3.22) follows from Lemma 3.4.

Next, due to (3.17), we need only show one-sided convergence:

$$pE\cos^2\theta \le p\sum\left(E\frac{(b_i'\mathbf{x})^2}{\|\mathbf{x}\|^2}\right)^2$$

$$= p\sum\left(E\frac{(b_i'\mathbf{x})^2}{\|\mathbf{x}\|^2}1\left(\left|\frac{\|\mathbf{x}\|^2}{p} - 1\right| > c\right)\right.$$

$$\left.+E\frac{(b_i'\mathbf{x})^2}{\|\mathbf{x}\|^2}1\left(\left|\frac{\|\mathbf{x}\|^2}{p} - 1\right| < c\right)\right)^2.$$

The first term inside the summation is no greater than $P\{1(|\|\mathbf{x}\|^2/p - 1| > c)\}$, which is of order $o(p^{-1})$ by assumption. The second term inside the summation is bounded by $1/[p(1-c)]$. Now we can obtain easily (3.23).

Consider (3.24). We have

$$pE\cos\theta\left(1 - \frac{p}{\|\mathbf{x}_1\|^2}\right) = E\frac{\sqrt{p}\,\mathbf{x}_1'}{\|\mathbf{x}_1\|}\left(1 - \frac{p}{\|\mathbf{x}_1\|^2}\right)E\frac{\mathbf{x}_2}{\|\mathbf{x}_2\|/\sqrt{p}}$$

$$= E\frac{\sqrt{p}\,\mathbf{x}_1'}{\|\mathbf{x}_1\|}\left(1 - \frac{p}{\|\mathbf{x}_1\|^2}\right)E\left(\frac{\sqrt{p}}{\|\mathbf{x}_2\|} - 1\right)\mathbf{x}_2.$$

The norm of the second expectation term converges to zero due to (3.22). We can bound the norm of first expectation term by

$$\sup_{\|b\|=1} E\frac{\sqrt{p}\,\mathbf{x}'b}{\|\mathbf{x}\|}\left(1 - \frac{p}{\|\mathbf{x}\|^2}\right).$$

Break the expectation into two parts, **I** + **II**, with

$$\mathbf{I} = E\frac{\sqrt{p}\,\mathbf{x}'b}{\|\mathbf{x}\|}\left(1 - \frac{p}{\|\mathbf{x}\|^2}\right)1\left(\frac{\|\mathbf{x}\|^2}{p} < c\right),$$

$$\mathbf{II} = E\frac{\sqrt{p}\,\mathbf{x}'b}{\|\mathbf{x}\|}\left(1 - \frac{p}{\|\mathbf{x}\|^2}\right)1\left(\frac{\|\mathbf{x}\|^2}{p} > c\right).$$

The **II** term is no greater than

$$c^{-1/2}(1 + 2/c)E|b'\mathbf{x}| \le c^{-1/2}(1 + 2/c).$$

The **I** term is no greater than

$$2p^{3/2}E\frac{\mathbf{x}'b}{\|\mathbf{x}\|^3}1\left(\frac{\|\mathbf{x}\|^2}{p} < c\right),$$

which in turn is bounded by

$$p^{3/2}E\frac{1}{\|\mathbf{x}\|^2}1\left(\frac{\|\mathbf{x}\|^2}{p} < c\right).$$

We have verified (3.24).

## APPENDIX C

**Proofs for subsection 3.2.** Proofs of Theorems 3.1 and 3.2 follow from the three lemmas given below.

LEMMA C.1. *Assume* (3.6) *and* (3.8). *Then* (3.3) *holds for* $t \neq 0$. *In addition, if the following additional condition also holds*:

$$E\frac{\sqrt{p}}{\|\mathbf{x}\|}1(B_1(c)) = o(1),$$

*then* (3.3) *holds for* $t = 0$.

LEMMA C.2. *Assume* (3.6), (3.8), (3.26) *and* (3.27). *Then* (3.4) *holds.*

LEMMA C.3. *Assume* (3.6), (3.8), (3.28) *and* (3.29). *Then* (3.5) *holds.*

PROOF OF LEMMA C.1. Take the derivative of $Q(\mathbf{x}, t)$ with respect to $\mathbf{x}$ and verify that the maximum of $Q(\mathbf{x}, t)$ is achieved at $\|\mathbf{x}\|^2 = t^2(p-2)$. The maximum value converges to $t^{-1}e^{(t^2-1)/2}$. This completes the proof of the first part. The second part is obvious. □

PROOF OF LEMMA C.2. Observe that $Q(\mathbf{x}_1, \mathbf{x}_2|t)$ is bounded by $p(\|\mathbf{x}_1\|\,\|\mathbf{x}_2\|\,|\sin\theta|)^{-1}$. Therefore it suffices to find a finite upper bound for

$$Ep(\|\mathbf{x}_1\|\,\|\mathbf{x}_2\|\,|\sin\theta|)^{-1}$$

We now use the Cauchy–Schwarz inequality to complete the proof. □

PROOF OF LEMMA C.3. For any small positive $c$, define

$$BAD = \left\{\mathbf{x}_1, \mathbf{x}_2: \left|\frac{\|\mathbf{x}_1\|^2}{p} - 1\right| > c, \text{ or } \left|\frac{\|\mathbf{x}_2\|^2}{p} - 1\right| > c, \text{ or } |\cos\theta| > c\right\}.$$

Then

$$\begin{aligned} e^{-t^2}E|\mathbf{x}_1'\mathbf{x}_2Q(\mathbf{x}_1, \mathbf{x}_2|t)|1(BAD) &\le pE\frac{1}{|\sin\theta|}1(BAD) \\ &\le pE\frac{1}{|\sin\theta|}1(B_2(c)) + p(1-c^2)^{-1}P\{BAD\}. \end{aligned}$$

This is seen to converge to zero by our assumptions. Similarly, define $L(\mathbf{x}, \mathbf{x}_2|t)$ to be the leading term of $\mathbf{x}_1'\mathbf{x}_2 Q_p(\mathbf{x}_1, \mathbf{x}_2|t)$, given in Corollary 3.1. We have

$$\begin{aligned} E|L(\mathbf{x}_1, \mathbf{x}_2|t)|e^{-t^2} &\le pE\left(1 + 2t^2 + \frac{1}{2}pt^2\left(\|\mathbf{x}_1\|^{-2} + \|\mathbf{x}_2\|^{-2}\right)\right)1(BAD) \\ &\le (1 + 2t^2)pP(BAD) + t^2 pE\frac{p}{\|\mathbf{x}\|^2}1(B_1(c)) \\ &\quad + t^2(1-c)^{-2}pE1(BAD), \end{aligned}$$

which converges to zero. This completes the proof. □

## APPENDIX D

**Proof of (4.4).** The distribution of the angle between $\mathbf{e}_1$ and $\mathbf{e}_2$ is the same as the distribution of the angle between a $p$-dimensional standard normal random vector $(z_1, \dots, z_p)$ and the first coordinate. Therefore the probability that the tangent of this angle has absolute value less than $t$ is equal to

$$P\left\{\frac{z_2^2 + \cdots + z_p^2}{z_1^2} < t^2\right\},$$

which equals

$$\frac{2\Gamma(p/2)}{\sqrt{p-1}\sqrt{\pi}\,\Gamma((p-1)/2)} \int_{\sqrt{p-1}/t}^{\infty} \left(1 + \frac{t^2}{p-1}\right)^{-p/2} dt.$$

Delete the constant "1" in the integrand and change variable to get a bound

$$\frac{2\Gamma(p/2)}{\sqrt{p-1}\sqrt{\pi}\,\Gamma((p-1)/2)} \frac{t^{p-1}}{p-1}.$$

Now use Stirling's formula and the relationship between sine and tangent to get (4.4). □

**Acknowledgment.** Ker-Chau Li is grateful for the support and hospitality of the Center for Mathematics and its Applications, Australian National University, where he was visiting when this work began.

## REFERENCES

BRILLINGER, D. R. (1977). The identification of a particular nonlinear time series system. *Biometrika* **64** 509–515.

BRILLINGER, D. R. (1983). A generalized linear model with "Gaussian" regressor variables. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 97–114. Wadsworth, Belmont, CA.

CARROLL, R. J. and LI, K. C. (1992). Measurement error regression with unknown link: dimension reduction and data visualization. *J. Amer. Statist. Assoc.* **87** 1040–1050.

CHEN, H. (1991). Estimation of a projection-pursuit type regression model. *Ann. Statist.* **19** 142–157.

CLEVELAND, W. S. ed. (1988). *The Collected Works of John W. Tukey, Volume V. Graphics: 1965–1985.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

CLEVELAND, W. S. and MACGILL, M. E. (1988). *Dynamic Graphics for Statistics.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

COOK, R. D. and WEISBERG, S. (1991). Discussion of "Sliced inverse regression for dimension reduction" by K. C. Li. *J. Amer. Statist. Assoc.* **86** 328–332.

DIACONIS. P and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815.

DONOHO, D. and JOHNSTONE, I. (1989). Projection-based approximation and a duality with kernel methods. *Ann. Statist.* **17** 58–106.

DUAN, N. and LI, K. C. (1991). Slicing regression: a link free regression method. *Ann. Statist.* **19** 505–530.

FRIEDMAN, J. (1987). Exploratory projection pursuit. *J. Amer. Statist. Assoc.* **82** 249–266.

FRIEDMAN, J. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

HAERDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995.

HALL, P. (1989a). On projection pursuit regression. *Ann. Statist.* **17** 573–588.

HALL, P. (1989b). On polynomial-based projection indices for exploratory projection pursuit. *Ann. Statist.* **17** 589–605.

HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061.

HUBER, P. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–525.

LI, K. C. (1990). Data visualization with SIR: a transformation based projection pursuit method. *UCLA Statistical Series* **24**.

LI, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86** 316–342.

LI, K. C. (1992a). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87** 1025–1039.

LI, K. C. (1992b). Uncertainty analysis for mathematical models with SIR. In *Probability and Statistics* (Z. Jiang, S. Yan, P. Cheng and R. Wu, eds.) 138–162. World Scientific, Singapore.

LI, K. C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052.

TIERNEY, L. (1990). *LISP-STAT: An Object-oriented Environment for Statistical Computing and Dynamic Graphics.* Wiley, New York.

WEGMAN, E. J. and DEPRIEST, D. J. (1986). *Statistical Image Processing and Graphics.* Dekker, New York.

WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 2601, AUSTRALIA

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024